Plant Molecular Biology **54:** 911–929, 2004.

Functional genomics of cell elongation in developing cotton fibers

A. Bulak Arpat[1], Mark Waugh[2], John P. Sullivan[2,4], Michael Gonzales[2], David Frisch[3,5], Dorrie Main[3], Todd Wood[3,6], Anna Leslie[1,7], Rod A. Wing[3,8] and Thea A. Wilkins[1,*]
[1]*Department of Agronomy and Range Science, University of California, One Shields Ave., Davis, CA 95616, USA (*author for correspondence; e-mail tawilkins@ucdavis.edu);* [2]*National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, NM 87505,USA;* [3]*Clemson University Genomics Institute, Biosystems Research Complex, 51 New Cherry St., Clemson, SC 29634, USA;* [4]*Computer Science Department, University of Massachusetts, Boston, Massachusetts, USA;* [5]*Genome Center of Wisconsin, 445 Henry Mall, Genetics Bldg. B19, University of Wisconsin, Madison, WI 53706, USA;* [6]*Center for Origins Research and Education, Bryan College, P.O. Box 7000, Dayton, TN 37321, USA;* [7]*College of Agriculture and Environmental Sciences Genomics Facility, University of California, One Shields Ave., Davis, CA 95616, USA;* [8]*Arizona Genomics Institute, Department of Plant Sciences, University of Arizona, 303 Forbes Bldg., Tucson, AZ 85721, USA*

Received 23 March 2004; accepted in revised form 24 May 2004

Key words: cell expansion, ESTs, oligonucleotide microarrays, primary cell wall, secondary cell wall

Abstract

Cotton fibers are single-celled seed trichomes of major economic importance. Factors that regulate the rate and duration of cell expansion control fiber morphology and important agronomic traits. For genetic characterization of rapid cell elongation in cotton fibers, ~ 14,000 unique genes were assembled from 46,603 expressed sequence tags (ESTs) from developmentally staged fiber cDNAs of a cultivated diploid species (*Gossypium arboreum* L.). Conservatively, the fiber transcriptome represents 35–40% of the genes in the cotton genome. In *silico* expression analysis revealed that rapidly elongating fiber cells exhibit significant metabolic activity, with the bulk of gene transcripts, represented by three major functional groups – cell wall structure and biogenesis, the cytoskeleton and energy/carbohydrate metabolism. Oligonucleotide microarrays revealed dynamic changes in gene expression between primary and secondary cell wall biogenesis showing that fiber genes in the dbEST are highly stage-specific for cell expansion – a conclusion supported by the absence of known secondary cell wall-specific genes from our fiber dbEST. During the developmental switch from primary to secondary cell wall syntheses, 2553 ''expansion-associated'' fiber genes are significantly down regulated. Genes (81) significantly up-regulated during secondary cell wall synthesis are involved in cell wall biogenesis and energy/carbohydrate metabolism, which is consistent with the stage of cellulose synthesis during secondary cell wall modification in developing fibers. This work provides the first in-depth view of the genetic complexity of the transcriptome of an expanding cell, and lays the groundwork for studying fundamental biological processes in plant biology with applications in agricultural biotechnology.

Abbreviations: dpa, days post-anthesis; EST, expressed sequence tag; GO, gene ontology; HMW, high molecular weight; NR, non-redundant; oligoNT, oligonucleotide; PCW, primary cell wall; PEPCase, phosphoenolpyruvate carboxylase; qPCR, quantitative real time reverse transcription-polymerase chain reaction; SCW, secondary cell wall; XTH, xyloglucan endotransglucosylase/hydrolases

Introduction

Plant trichomes found on vegetative and reproductive organs throughout the plant kingdom exhibit considerable diversity in terms of size and morphology, distribution, and origin (Werker, 2000). A number of biologically important roles have been ascribed to trichomes, including protection against biotic and abiotic factors, water absorption, secretion, alluring mechanisms, and seed dispersal. While the majority of plant trichomes are multicellular, cotton (*Gossypium* spp.) produces unicellular seed trichomes commonly called ''fibers'', which are of considerable economic importance. Cotton, a prized agricultural commodity for more than 8000 years, is the world's leading natural fiber and a mainstay of global economies. A multi-billion dollar industry, the production, marketing, consumption and trade of cotton-based products further stimulate the economy, making cotton the number one value-added crop. However, yield and fiber quality have declined over the last decade (Meredith, 2000) – a downward trend that has been attributed in part to erosion in genetic diversity of cotton varieties (Meredith, 2000).

Apart from their economic importance, cotton fibers provide an excellent single-celled model for studying fundamental biological processes in plants. Indeed, developing cotton fibers have been instrumental in providing novel insight into the mechanism of cellulose biosynthesis during secondary cell wall modification (Pear *et al.,* 1996; Delmer, 1999; Potikha *et al.,* 1999; Kurek *et al.,* 2002; Peng *et al.,* 2002). The near-synchronous growth of 21,000+ terminally differentiated fiber cells per ovule is characterized by four major discrete developmental stages – differentiation, expansion/primary cell wall (PCW) synthesis, secondary cell wall (SCW) synthesis and maturity (Basra and Malik, 1984; Wilkins and Jernstedt, 1999). Fibers, which are among the longest cells in the plant kingdom – reaching up to 5.0 cm in length in some genotypes, exhibit highly exaggerated growth rates well above the average of most plant cell types. As all plant cells undergo cell expansion to some extent at some stage of development, developing cotton fibers offer a unique opportunity to study cell expansion at the level of a single cell.

Fiber expansion and elongation, which commences on the day of anthesis (0 days post-anthesis [dpa]) and continues for a period of $\sim$ 21 days, occurs via a diffuse-growth mechanism that directs polarized growth of developing fiber cells (Tiwari and Wilkins, 1995; Wilkins and Jernstedt, 1999). Fiber length is dictated by the rate and duration of cell expansion, which is in turn, governed by developmental programs that coordinately regulate cell turgor, the driving force of cell expansion, and cell wall loosening (Smart *et al.,* 1998; Wilkins and Jernstedt, 1999; Ruan *et al.,* 2001). At present, the relatively few fiber genes characterized to date provide only a snapshot view of the molecular mechanisms underlying fiber growth (reviewed in Wilkins and Jernstedt, 1999), indicating that developmental regulation of fiber genes closely parallels the rate of expansion (Smart *et al.,* 1998). Expansion-associated fiber genes include those involved in osmoregulation and carbohydrate metabolism such as the vacuolar and plasma membrane H^+-ATPases (Wilkins 1993; Wilkins *et al.,* 1994; Hasenfratz *et al.,* 1995; Smart *et al.,* 1998; Ratajczak and Wilkins, 2000), phosphoenolpyruvate carboxylase (PEPCase) (Vojdani *et al.,* 1997), sucrose synthase, and sucrose and K^+ transporters (Nolte *et al.,* 1995; Ruan *et al.,* 2001, 2003), as well as cell wall modifying enzymes like expansin (Shimizu *et al.,* 1997; Orford and Timmis, 1998). Coincident with termination of fiber elongation at $\sim$ 21 dpa is an increase in fiber strength (Hsieh, 1999), presumably due to cross-linking of cellulosic and non-cellulosic matrices (Carpita and Gibeau, 1993), that is also accompanied by a major loss ($\sim$ 36%) of HMW noncellulosic polymers in the primary cell wall (Shimizu *et al.,* 1997). A similar loss of xyloglucans is reported in association with termination of auxin-induced expansion in pea hypocotyls (Hayashi *et al.,* 1984). The transition from primary to secondary cell wall synthesis, which occurs during the latter stage of expansion between $\sim$ 16 and 21 dpa, is distinguished by the re-orientation of microtubules and cellulose microfibrils to steeply pitched helical arrays (Seagull, 1992) in anticipation of secondary cell wall synthesis to produce a thick cell wall consisting of >94% cellulose. Expression of *GhCesA* genes encoding the catalytic subunit of cellulose synthase, first detected as the cell enters this

transition phase, dramatically increases in parallel with the rate of cellulose synthesis to peak levels at 24 dpa (Meinert and Delmer, 1977; Pear *et al.,* 1996). *GhCesA1 and GhCesA2* therefore serve as convenient stage-specific markers for secondary cell wall synthesis (Wilkins and Jernstedt, 1999).

One of our long term interests is elucidating fiber gene function in a developmental context as a means for manipulating output traits in the genetic improvement of fiber properties (Wilkins *et al.,* 2000). Mapping and comparative genomics of orthologous fiber gene sequences from cultivated allotetraploid species (*G. hirsutum* L. and *G. barbadense* L., $2n = 4x = 52$, AD genome) and diploid progenitor species ($2n = 26$, A or D genome) revealed that genome organization, gene sequences, as well as spatial and temporal expression patterns have been evolutionarily conserved (Cedroni *et al.,* 2003; Senchina *et al*., 2003; Rong *et al*., 2004). At last count, however, fewer than 50 fiber genes have been isolated and partially characterized using traditional molecular approaches. As the first step towards defining the cotton fiber transcriptome using genomic approaches, more than 46,000 ESTs from rapidly elongating fibers were generated from a cultivated diploid progenitor species, *G. arboreum* L. to maximize gene discovery and minimize redundancy due to polyploidy. The genetic complexity in elongating cotton fibers is very high, consisting of $\sim$ 14,000 fiber genes, the most abundant of which belong to functional categories that reflect a very metabolically active cell-type that is certainly consistent with the accentuated growth of developing fibers. To determine the stage-specific expression of fiber genes and identify candidate genes for genetic mapping and functional analysis, the dynamics of gene expression that accompanies the developmental switch from primary to secondary cell wall synthesis was obtained from expression signatures derived from oligonucleotide microarrays fabricated from non-redundant fiber ESTs. Expression profiles generated from comparison of 10 vs. 24 dpa fiber transcripts revealed a subset of more than 2500 stage-specific "expansion-associated" genes that are down regulated coincident with the termination of fiber elongation, as well as a group of 81 novel genes newly identified that are preferentially expressed during secondary cell wall synthesis.

Materials and methods

Plant material and RNA isolation

Developing fibers were harvested from greenhouse-grown cotton as described elsewhere (Smart *et al.,* 1998) to control for biological variability. PolyA RNA was purified from quality-controlled total RNA (Wan and Wilkins, 1994; Wilkins and Smart, 1996) using Promega's PolyAttract kit to construct a fiber cDNA library from *Gossypium arboreum* L. cv. AKA 8401. For microarrays, 10 and 24 dpa fibers were harvested from plants (*G. hirsutum* L. cv. TM-1) grown in a randomized complete block design to create three biological pools. LiCl-precipitated RNA prepared using the hot borate method (Wilkins and Smart, 1996) was suspended in 10 mM Tris/1 mM EDTA buffer, purified using RNeasy mini-spin columns (Qiagen), and ethanol-precipitated. The concentration of total RNA was determined spectrophotometrically and quality-controlled by agarose gel electrophoresis. RNA was isolated independently at least two times from each biological pool to create technical replicates.

Generation of cotton fiber ESTs

A high-quality directional cotton (*G. arboreum*) fiber cDNA library (1.2×10^6 pfu/ml) containing $< 0.5\%$ non-recombinant phage was constructed using Stratagene's λZAP Express cDNA cloning kit. Following mass-excision of phagemids, kanamycin-resistant bacterial colonies (92,160) were arrayed in 384-well microtiter plates containing LB-glycerol freezing media and stored at –80 °C. Automated DNA sequencing of $>$ 50,000 cDNAs from the 5′- and 3′-termini using universal T3 and T7 primers was performed using the Big Dye Terminator sequencing kit and Applied Biosystems (ABI) 377 or 3700 automated sequencers with $\geq$ 70% success rate. ESTs averaged 757 nucleotides (NT) in length with an average high quality sequence length of 416 NT after removal of vector sequences. The minimum length for ESTs released to GenBank (http://www.ncbi.nlm.nih.gov/dbEST) was $\geq$ 100 high quality NT defined by Phred scores $\geq$ 20.

The *Ga* (*G. arboreum*) Cotton Fiber dbEST (http://cfgc.ucdavis.edu) consists of four discrete data sets, (1) Ga_Ea (Ea) cDNAs (12,767)

randomly sequenced from the 5′-terminus, (2) Ga_Eb (Eb) sequences (13,613) obtained from the 5′-terminus following one round of normalization to remove redundant Ea sequences, (3) Ga_Ed (Ed) sequences (14,915) obtained from the 5′ and 3′ termini following a second round of normalization to remove redundant Ea and Eb sequences, and (4) Ga_Ec (Ec) sequences (3026) – a subset of Ea cDNAs sequenced from the 3′ terminus. Normalization of the fiber cDNA library was performed by sequential hybridization of high-density filter arrays to remove the most redundant (Ea and Eb) gene sequences. Fiber cDNA clones (92,160) were spotted in duplicate onto six high-density nylon membranes in 4×4 arrays as described (Maier *et al.,* 1994) and hybridized with radiolabeled probes (Lijavetzky *et al.,* 1999) generated from heterogeneous gene pools (20–25 cDNAs/pool) containing equal amounts of DNA from purified PCR products of unrelated gene sequences. Scanned images of autoradiographs were divided into six fields per image to generate output files containing total signal and background intensities for hybridized spots using ImaGene 4.2 software (BioDiscovery). Software programs written for data quantification, analysis and automated identification of corresponding plate address for each gene were used to create files for robotic rearraying of the cDNA library minus the redundant sequences before resuming random sequencing. Normalization probes were generated from a representative of 75 Ga_Ea gene clusters ($\geq$ 6 ESTs/cluster) and 95 Ga_Eb gene clusters ($\geq$ 7 ESTs/cluster) for the first and second rounds of normalization, respectively (Supplementary Tables 1 and 2; www.kluweronline.com/issn/0167-4412).

Raw EST sequences files were imported from an FTP site for processing, annotation and analysis using XGI™ (http://www.ncgr.org/xgi), an automated EST clustering and analysis pipeline (http://www.ncgr.org/xgi/). Only high-quality processed ESTs that passed through filters in the vector screener and quality control stages of the pipeline were clustered into consensus sequences and annotated using BLASTX against NCBIs nonredundant protein database, BLIMPS searching against the Blocks+ (Henikoff *et al.,* 1999) motif database and a suite of seven separate algorithms in InterProScan (Zdopnov and Apweiler, 2001) against the InterPro (Apweiler *et al.,* 2001) protein database. The resultant annotated sequences are available for searching through the UCD non-redundant (NR) Fiber EST consensus sequences v2.0 (Dec. 2002, http://cfgc.ucdavis.edu), 90.6% of which have BLASTX E values $\leq 10^{-10}$. Functional categories were assigned gene ontology (GO) annotations (Ashburner *et al.,* 2000) where possible using gene ontology annotation from curated GO annotations in InterPro as well as an automated method of GO mapping based on high-scoring hits to Swiss Prot entries.

Cotton oligonucleotide microarrays

Oligonucleotides (oligoNT [70-mers]) were synthesized by Operon Technologies against 12,227 NR fiber ESTs, excluding consensus sequences $\leq$ 100 nucleotides in length, or showing $\leq$ 85% similarity to other genes. Cotton fiber microarrays were fabricated by spotting oligoNTs (40 μM) in 1× Array-It Spotting Buffer Plus in duplicate on Telechem superaldehyde slides in 23 × 23 subarrays using the OmniGrid arrayer (Genomic Solutions) equipped with 16 (4×4) MicroQuill pins (Majer Precision Engineering). Experimental controls (71) included internal, positive and negative controls, transgene and vector controls, calibration spike-in controls, ratio spike-in controls, blank and buffer controls interspersed among the cotton oligoNTs and replicated $\geq$ 2 times. Additional controls included cotton sequences deposited in GenBank, but not found in our fiber dbEST. Post-printing processing of slides to chemically cross-link oligomers was performed according to slide manufacturer's instructions.

Hybridization probes were prepared using the aminoallyl labeling method as described (Hughes *et al.,* 2001). Total fiber RNA (20 μg) spiked with 2 μl of test or reference mRNA mix (Lucidea Universal Scorecard, Amersham Pharmacia) was reverse transcribed in the presence of aminoallyl-dUTP (Sigma). Following conjugation of Cy3- or Cy5-NHS esters (Amersham Pharmacia) to reverse-transcribed cDNA, unincorporated dye was removed from probes using QIAquick PCR Purification columns (Qiagen). Purified probes were hybridized at 42 °C for 16–20 h in humidified hybridization chambers (Telechem) as described elsewhere (Hedge *et al.,* 2000). Slides were scanned (10 mm resolution) using an Affymetrix Array

Scanner 428 from a total of 8 hybridizations, including 4 dye-swap treatments, producing 16 replicates for each fiber oligoNT. Self-hybridization controls were also performed. Signal intensities were quantified using ImaGene 4.2 software (BioDiscovery). Visually flagged spots and spots with a background corrected intensity smaller than the average plus two standard deviations of corrected intensity for blank spots ($N = 274$) were filtered. Normalization and analysis of microarray data were performed using GeneSpring 6.0 (Silicon Genetics). Normalization of the array dataset was based on intensity-dependent Lowess curve fitting ($f = 0.2$) and median of background subtracted intensities from control RNA set spiked into the query RNA samples at a 1:1 ratio. Statistical analysis of microarray data was performed using GeneSpring 6.0 cross-gene error model based on replicates for 10 vs. 24 dpa hybridizations and based on deviation from one for self-hybridizations. Significantly up- or down-regulated genes were filtered for expression ratios greater or smaller than 2 and 0.5, respectively, and for *t*-test *P*-value < 0.05. For multiple testing correction, Benjamini and Hochberg false discovery rate method was used.

Real time PCR

Expression analysis was performed to confirm microarray results using two-step quantitative real time RT-PCR (qPCR). A known amount of DNase-treated total cotton fiber RNA, spiked with non-plant RNA synthesized from a cloned human phosphomannomutase gene as an internal reference, was reverse transcribed using Invitrogen's Superscript II RTase kit. RT-PCR reactions were tracked on an ABI 7000 instrument (Applied Biosystems) using the Quantitect SYBR Green Master Mix (Qiagen) Each sample was PCR-amplified using the same amount of cDNA template in triplicate reactions in at least two independent experiments. Gene-specific qPCR primer-pairs for the spiked control and 16 fiber genes designed with Primer Express software (Applied Biosystems) are provided in Supplementary Table 3. Following an initial step in the thermal cycler for 15 min at 95 °C, PCR amplification proceeded for 40 cycles of 15 s at 95 °C and 1 min at 60 °C, and completed by melting curve analysis to confirm specificity of PCR products. The baseline and threshold values were adjusted according to manufacturer's instructions. Similar results were obtained from relative quantification of transcript abundance determined independently by the standard curve method described in Applied Biosystems User Bulletin 2 (Applied Biosystems, 2001).

Results

A genomic approach, based on ESTs and expression profiling, was used to characterize rapid cell elongation in developing cotton fibers. Three approaches were employed to maximize gene discovery: (1) Construction of a high-quality fiber cDNA library from a cultivated diploid species to minimize redundancy due to polyploidy, (2) deep sampling of the cDNA library and (3) normalization of the cDNA library to identify rare gene transcripts by removing highly redundant gene sequences.

*7–10 dpa cotton (*Gossypium arboreum *L.) fiber ESTs*

A high-quality cotton fiber cDNA library was constructed from the A-genome diploid species *G. arboreum* L. To obtain a stage-specific fiber library, 7–10 dpa developing fibers were harvested during rapid elongation, but well before detection of secondary cell wall (e.g. *GhCesA*) gene transcripts, which signal the onset of secondary cell wall synthesis (Pear *et al.,* 1996). A random sampling of 100 fiber cDNAs showed an average PCR amplicon size of 1.7 kbp, although cDNAs ≥ 3.0 kbp released by restriction digestion were not uncommon. Random sequencing of the arrayed cDNA library yielded 46,603 *G. arboreum* (Ga) cotton fiber ESTs, generated in four discrete data sets (Ea, Eb, Ec and Ed) before and after normalization (see Materials and methods). To show that fiber gene discovery in the diploid species was cost effective and highly successful in terms of minimizing redundancy relative to cultivated allotetraploids, an equivalent number of diploid Ea fibers ESTs were compared to *G. hirsutum* L. cv. Maxxa 5–6 dpa fiber ESTs downloaded from GenBank. Using BLASTN algorithm, Ga (Ea) fiber EST gene clusters yielded 61% novel gene sequences, which is slightly more

than two times the number of novel Gh genes, clearly indicating the higher redundancy in the allopolyploid species has a negative impact on the rate of gene discovery.

Discovery of novel genes in developing Ga cotton fibers was maintained by normalizing the arrayed library to remove redundant sequences by hybridization of high-density filters arrays. In the first round of normalization, redundant (Ea) sequences (Supplementary Table 1) were identified. Random sequencing of the re-arrayed library (minus redundant Ea clones) yielded Eb ESTs. A list of redundant Eb sequences (Supplementary Table 2), which is produced by clustering of Eb ESTs, was combined with redundant Ea sequences and used to generate normalization probes to remove redundant sequences before resuming sequencing to produce Ed ESTs. The rate of gene discovery was maintained at 50% following each normalization; and both Eb and Ed normalized data sets showed a disproportionate enrichment of singletons relative to the number of redundant gene clusters than would be expected by random sequencing. Clustering of Ea and Eb sequences revealed an average of $\sim$ 1:1 Ea:Eb ratio, whereas normalized gene clusters corresponding to normalization probes (Supplementary Tables 1 and 2) consistently produced an Ea:Eb transcript ratio of 5:1, indicating $\sim$ 80% of the most redundant sequences were eliminated from the re-arrayed library. Similar ratios were obtained in the Ed data set.

Transcriptome of rapidly elongating cotton fiber cells

The genetic complexity of the 7–10 dpa cotton fiber transcriptome is very high, consisting of 13,947 NR consensus sequences (9648 singletons and 4299 gene clusters) in the GA Cotton Fiber dbEST. BLASTX results returned 'no significant similarities' for 40.6% of cotton NR consensus sequences with an E-value greater than 10^{-5}, indicating a number of novel genes, at least some of which are likely unique to cotton.

The nature and scope of biological processes engaged during polar cell elongation was determined by *in silico* expression analysis of Ga_Ea (non-normalized) fiber EST gene clusters, 41.5% of which are singletons (Supplementary Figure 1; published online at http://www.kluweronline.com/issn/0167-4412). The 15% most highly expressed fiber genes are presented in Table 1, along with gene annotation and closest homolog exhibiting the lowest BLASTX E-value. With one notable exception of the most abundant gene cluster encoding α-tubulin, even the most abundantly expressed genes represent $<1\%$ of the Ea gene transcripts (Table 1). The major functional categories, which account for $\sim$ 76% of Ea transcripts include, Cytoskeleton (18.96%), Transport (17.96%), Cell Structure/Biogenesis/Cell Wall (15.74%), Protein Metabolism (12.64%), and Energy/Carbohydrate Metabolism (10.86%). Four major Ga_Ea clusters representing fiber genes that are highly expressed relative to all other genes encode two α-tubulin subunits, an aquaporin, and E6, a novel cotton protein (John and Crow, 1992). There are also a few genes (uncategorized) that are highly expressed, relatively speaking, that encode proteins of unknown function – three of which show significant homology to *Arabidopsis* genes. There are numerous examples of members of multigene families in our cotton fiber dbEST being differentially expressed. For example, *in silico* Northern analysis shows three α-tubulin genes with a transcript abundance of 78, 55 and 12 Ea ESTs, and 13 and 15 mRNA transcripts for two β-tubulin genes, respectively (Table 1).

Of the annotated consensus sequences, the 6682 or 47.9% assigned to functional categories using (GO) gene ontology (Ashburner *et al.,* 2000) that define the cotton fiber transcriptome are shown in Figure 1. Almost two-thirds of binned genes belong to three major functional categories – energy/carbohydrate metabolism, cellular structure, organization and biogenesis, and protein metabolism; a distribution that is consistent with very metabolically active cell types. *In silico* expression analysis of Ea ESTs also suggests that the GO energy/carbohydrate metabolism category consists of a large class of relatively moderately expressed genes.

Developmental regulation of gene expression during termination of fiber elongation

To assess developmental changes in the fiber transcriptome during the programmed switch from primary and secondary cell wall synthesis in 10 vs. 24 dpa fibers, respectively, expression profiling was performed using cotton oligonucleotide microarrays containing 12,227 fiber genes

Table 1. Consensus sequences of the top 15% most highly expressed genes in the *Gossypium arboreum* L. cv. AKA8401 Cotton Fiber dbEST (http://cfgc.ucdavis.edu/)[1].

Consensus Seq ID[2]	Putative gene function	Functional category[3]	log E-value	GenBank accession	# GA_Ea ESTs
132_04514	alpha-tubulin	Cytoskeleton	0	X67162	78
065_04510	alpha-tubulin	Cytoskeleton	0	U12589	55
049_04505	E6	CSO[4] /Cell wall	−117	U30507	45
050_04508	aquaporin PIP2-2	Transport	−146	AF141900	39
070_04512	elongation factor-1 alpha	Protein metabolism	−81	AB019427	31
088_04513	cytosolic ascorbate peroxidase	CSO	−111	AF159630	26
029_04480	polyubiquitin	Protein metabolism	−167	L05361	26
040_04500	lipid transfer protein 3 precursor	Transport	−60	AF228333	26
032_04493	glyceraldehyde-3-phosphate dehydrogenase C subunit	ECM[5]	−158	AC016829	26
026_04468	catechol O-methyltransferase	Secondary metabolism	−100	AF064694	23
039_04498	cyclophilin	ERD[6]	−84	AJ245940	22
030_04483	vacuolar acid invertase	ECM	0	AY048579	21
032_04490	arabinogalactan-like protein	CSO/Cell wall	−57	NM_125442	20
034_04495	myo-inositol-1-phosphate synthase	Lipid metabolism	0	AB032073	20
059_04509	adenosylhomocysteinase	ECM	0	NM_117468	19
031_04488	expansin	CSO/Cell wall	−135	AF043284	18
024_04456	flavonoid 3′,5′-hydroxylase	Secondary metabolism	0	Z22544	18
042_04503	thiol protease	Protein metabolism	−140	AF182079	18
027_04472	unknown protein	Uncategorized	−68	AY063786	18
021_04442	aquaporin PIP2-2	Transport	−139	AF141900	18
020_04434	Superoxide dismutase	CSO	−79	X73139	17
031_04486	unknown protein	Uncategorized	−123	NM_120332	16
022_04443	cytosolic phosphoglycerate kinase 1	ECM	0	AB018410	16
039_04499	translationally controlled tumor protein	Uncategorized	−64	AF091455	16
031_04487	S-adenosyl-L-homocysteine hydrolase	ECM	0	Z26881	16
024_04455	glycine-rich RNA-binding protein	Nucleic acid metabolism	−37	AJ245939	16
042_04502	beta-tubulin 1	Cytoskeleton	0	X70184	15
019_04429	lipid transfer protein	Transport	−35	X96716	15
049_04506	unknown protein	Uncategorized	−155	AF428325	14
025_04463	elongation factor-1 alpha	Protein metabolism	0	D63582	14
031_04485	heat shock protein 70	CSO	−165	AJ249331	14
016_04394	proline-rich protein	CSO/Cell wall	−9	J02746	14
021_04439	Cyclophilin	ERD	−82	AJ245940	14
018_04421	cotton fiber annexin	Transport	−170	U89609	13
014_04369	aquaporin PIP1-3	Transport	−144	AF141899	13
041_04501	methionine synthase	Protein metabolism	0	AF082893	13
049_04507	lipid transfer protein	Transport	−35	AC069471	13
022_04445	lipid transfer protein 3	Transport	−62	AF228333	13
019_04430	beta-tubulin 6	Cytoskeleton	0	L10633	13
015_04390	beta subunit of K + channels	Transport	−132	AJ000999	12
039_04497	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	Protein metabolism	0	Y14797	12
023_04453	DFR-like protein	Secondary metabolism	−119	AF092912	12
020_04435	alpha-tubulin	Cytoskeleton	−143	U12589	12
016_04395	fatty acid condensing enzyme CUT1	Lipid metabolism	0	AY070727	12

[1] Ga_Ea ESTs of 12,767 randomly sequenced cDNA clones.
[2] Abbreviated consensus sequence IDs. Complete IDs have the prefix "CON_".
[3] Based on Gene Ontology (GO) annotation (Ashburner *et al.,* 2000).
[4] CSO, cellular structure and organization.
[5] ECM, energy/carbohydrate metabolism.
[6] ERD, environmental response and defense.

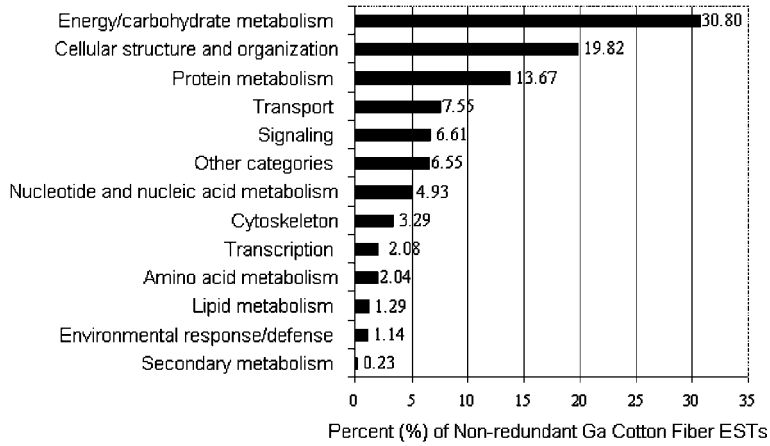

Figure 1. Functional groups of the cotton (*G. arboreum* L.) fiber transcriptome. UCD non-redundant fiber EST v2.0 consensus sequences (http://cfgc.ucdavis.edu) were assigned to functional categories using gene ontology (28). Not shown are uncategorized annotated genes (11.5%) and novel genes (40.6%) showing no significant similarity to genes in NCBI's non-redundant database.

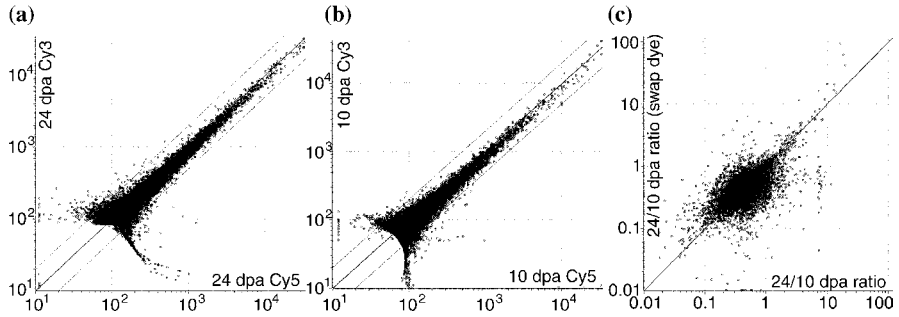

Figure 2. Experimental and biological variability and reproducibility of cotton fiber 70-mer oligonucleotide microarrays. (a) Single pool of total RNA from 24 dpa fibers self-hybridized with Cy3- and Cy5-labeld probes to assess experimental variability. (b) Total RNA from two independent pools from 10 dpa fibers labeled with Cy3 and Cy5 to assess biological variability. (c) Comparison of expression ratios obtained from swap-dye experiments. Signal intensities were normalized as described in Materials and methods. Lines indicate fold-changes (2:1, 1:1, and 1:2 in a and b, and 1:1 in c).

(excluding controls) spotted in duplicate. The high correlation coefficients obtained from control hybridizations indicated low biological and experimental variability (Figure 2a and b). Although fiber genes with a low absolute intensity did not correlate well, no outliers were identified when multi-layer filtering was applied (expression ratio ≤ 0.5 or ≥ 2 and *t*-test *P*-value < 0.05). To keep the likelihood of false positives $< 5\%$, a 2-fold expression ratio threshold was used for expression analysis of normalized data averaged from 8 hybridizations, including four dye-swap hybridizations. A high correlation ($r = 0.85$) between normalized expression ratios from dye-swap experiments using the same RNA preparations indicated microarray data was highly reproducible (Figure 2c). Following normalization, 12,139 fiber genes were available for analysis. The robustness of microarray data was confirmed by *in silico* expression, qPCR and the biological relevance of expression patterns in the context of fiber growth and development (Wilkins and Jernstedt, 1999).

About 21.7% of fiber genes expressed in 10 dpa fibers were developmentally regulated in 24 dpa fibers, the vast majority (21% of fiber genes) of which are down-regulated (Table 2; Supplementary Table 4). Fold-changes in expression of down-regulated genes varied up to 17.2-fold. Keeping in mind that $\sim 64\%$ of the fiber transcriptome in 10 dpa fibers are assigned to three major functional groups (Figure 1), almost two-thirds (66%) of the down-regulated genes belong

Table 2. Genes that are significantly up-regulated and down-regulated in 24 dpa cotton fibers relative to 10 dpa. Only the first 84 genes with the highest fold-change are reported for down-regulated genes, including control genes.

EST ID	Annotation	Fold change	FunctionalCategory[1]
Up-regulated genes			
FbL2A	fiber late 2A	37.8	Uncategorized
GhCesA-1	cellulase synthase A2	27.3	CSO
GA_Ed0053A09r	sucrose synthase isoform 3	15.3	ECM
GA_Ed0096C01f	nucleoid DNA-binding protein cnd41	12.5	Protein metabolism
GA_Eb0017L04f	no significant similarities	9.9	Uncategorized
GA_Ed0085G09f	calreticulin	9.8	Uncategorized
GhCesA-2	cellulase synthase A1	7.7	CSO
GA_Ea0014J22r	glucan 1	7.4	Uncategorized
GA_Ed0055A01r	unknown protein	6.6	Uncategorized
GA_Ed0097C08f	putative RING zinc finger protein	6.6	ECM
GA_Ea0001H11f	histone H2A	6.5	Nucleic acid metabolism
GA_Eb0006P15f	FS18A	6.4	Lipid metabolism
GA_Ea0019B23r	peroxidase	6.2	CSO
GA_Eb0014E15f	no significant similarities	6.2	Uncategorized
GA_Ed0065B07r	peptide transporter	5.9	Transport
GA_Ed0028C01r	no significant similarities	5.9	Uncategorized
GA_Eb0043C15f	unconventional myosin	5.8	Cytoskeleton
GA_Ea0005A15f	no significant similarities	5.7	Uncategorized
GA_Eb0032N04f	no significant similarities	5.6	Uncategorized
GA_Ed0075A10r	beta-1,3-glucanase-like protein	5.6	ECM
GA_Eb0003M12f	no significant similarities	5.2	Uncategorized
GA_Ea0008F01f	sterol glucosyltransferase	5.1	Uncategorized
GA_Ea0034L17r	unknown protein	5.1	Uncategorized
GA_Ed0028C01f	Regulatory-like protein	5.1	CSO
GA_Eb0005C05f	beta-1,3-glucanase-like protein	4.7	ECM
GA_Eb0014L17f	unknown protein	4.6	Uncategorized
GA_Eb0029L02f	no significant similarities	4.6	Uncategorized
GA_Ed0085C03f	ETAG-A3	4.5	ECM
GA_Ed0093B02f	Yippee-like protein	4.4	Uncategorized
GA_Ed0008A12f	unknown protein	4.3	Uncategorized
GA_Ed0101H11r	no significant similarities	4.2	Uncategorized
GA_Eb0002E23f	unknown protein	4.1	Uncategorized
GA_Eb0039K23f	peptide transporter	4.1	Uncategorized
GA_Ea0029D03f	Kelch repeat containing F-box protein family	4.1	Uncategorized
GA_Ed0083C06f	no significant similarities	4.0	Uncategorized
GA_Ed0105F11f	peroxidase	3.9	Uncategorized
GA_Ed0055G04r	unknown protein	3.8	ECM
GA_Ed0053A09f	sucrose synthase	3.7	ECM
GA_Ed0087G05f	no significant similarities	3.6	Uncategorized
GA_Eb0023C01f	unknown protein	3.6	ECM
GA_Ed0064C09f	unknown protein	3.5	Uncategorized
GA_Eb0037E17f	peroxidase	3.3	CSO
GA_Ea0026P20f	major latex-like protein	3.2	Uncategorized
GA_Ed0020A08r	no significant similarities	3.1	Uncategorized
GA_Ed0033A11f	proline-rich cell wall protein	3.1	CSO
GA_Eb0036I14f	cinnamyl alcohol dehydrogenase	3.0	ECM
GA_Ea0026I05f	no significant similarities	3.0	Uncategorized
GA_Eb0023B09f	steroid sulfotransferase-like protein	3.0	Uncategorized
GA_Ed0057D03r	no significant similarities	2.9	Uncategorized
GA_Ed0097C12r	unknown protein	2.7	CSO
GA_Ea0017O14f	dTDP-glucose 4-6-dehydratases-like protein	2.6	ECM
GA_Ea0032G16f	no significant similarities	2.5	Uncategorized
GA_Ed0036F03f	unknown protein	2.5	ECM

Table 2. Continued.

EST ID	Annotation	Fold change	Functional category[1]
GA_Eb0012G01f	unknown protein	2.4	Uncategorized
GA_Ed0102G03r	putative serine/threonine protein kinase	2.4	Other categories
GA_Ea0031M15f	auxin-repressed protein like-protein	2.4	Uncategorized
GA_Eb0022C02f	no significant similarities	2.4	Uncategorized
GA_Eb0033G08f	unknown protein	2.4	Uncategorized
GA_Ed0068E01f	shaggy-like kinase kappa	2.4	Transcription
GA_Ed0077C05f	unknown protein	2.3	CSO
GA_Ed0032A02r	no significant similarities	2.3	Uncategorized
GA_Ed0064B09r	membrane protein	2.3	CSO
GA_Ea0015L22f	beta tubulin 6	2.3	Cytoskeleton
GA_Eb0010F18f	nitrogen fixation like protein	2.3	Uncategorized
GA_Eb0027F23f	plasma membrane H + -ATPase	2.3	ECM
GA_Ed0052D12f	S-adenosyl-L-methionine synthetase	2.3	Protein metabolism
GA_Eb0020O06f	beta-galactosidase	2.3	ECM
GA_Ed0104E08r	carboxylesterase family	2.2	ECM
GA_Ea0030N06f	unknown protein	2.2	Uncategorized
GA_Ed0001A10r	GATA-binding transcription factor homolog 3	2.2	Uncategorized
GA_Ea0021G22f	beta tubulin 1	2.2	Cytoskeleton
GA_Ed0047G08f	unknown protein	2.2	Uncategorized
GA_Ed0035A06r	no significant similarities	2.1	Uncategorized
GA_Ea0024P08f	endo-1,4-beta-glucanase	2.1	ECM
GA_Eb0014G03f	unknown protein	2.1	Uncategorized
GA_Ea0029O20r	beta tubulin 1	2.1	Cytoskeleton
GA_Ea0029F03f	early nodulin 8 precursor	2.1	Uncategorized
GA_Eb0025C07f	unknown protein	2.1	Uncategorized
GA_Ea0027B03r	protein kinase, putative	2.1	Uncategorized
GA_Eb0019M24f	MutT domain protein-like	2.1	Uncategorized
GA_Ed0069H08f	unknown protein	2.0	Uncategorized
GA_Ed0107H10r	putative serine/threonine protein kinase	2.0	Other categories
GA_Ea0017B03f	no significant similarities	2.0	Uncategorized
GA_Eb0039K16f	no significant similarities	2.0	Uncategorized
Down-regulated genes[2]			
GA_Ea0021F16f	no significant similarities	17.2	Uncategorized
GA_Ed0107A01r	pumilio/Mpt5 family RNA-binding protein	15.9	Nucleic acid metabolism
GA_Ea0005I09f	lipid transfer protein 3 precursor	15.4	Transport
GA_Ea0021F16r	no significant similarities	12.9	Uncategorized
GA_Ed0098D11f	putative beta-galactosidase	12.8	ECM
GA_Ea0031A24f	RAB7A	12.6	Signaling
GA_Ea0029B01f	unknown protein	12.6	CSO
GA_Ea0017C17f	GDSL-motif lipase/hydrolase-like protein	12.0	Uncategorized
GA_Eb0033N23f	putative lipase/hydrolase	11.7	Lipid metabolism
GA_Ed0067F12r	no significant similarities	11.6	Uncategorized
GA_Ed0079H11r	multi-copper oxidase-related protein	11.4	Other categories
GA_Ea0014K19f	omega-3 fatty acid desaturase	11.1	Lipid metabolism
GA_Ea0012P05f	alpha-D-xylosidase	10.9	ECM
GA_Eb0023E15f	hypothetical protein	10.7	Uncategorized
GA_Ea0020F02f	ATP-dependent transmembrane transporter	10.4	Uncategorized
GA_Ed0068C05f	microsomal omega-3 acid desaturase	10.4	Lipid metabolism
GA_Ea0001K03f	unknown protein	10.2	Uncategorized
GA_Ed0064H01r	family II lipase EXL3	10.1	Other categories
GA_Ea0026C23f	pectinesterase (pectin methylesterase)	9.9	Other categories
GA_Ed0036G06f	fatty acid elongase - like protein	9.9	Lipid metabolism
GA_Ea0026O06f	putative ripening-related protein	9.4	Uncategorized
GA_Eb0023K18f	putative cinnamyl alcohol dehydrogenase	9.3	ECM
GA_Ed0043C09r	Gip1-like protein	9.1	ERD
GA_Eb0001L16f	glutathione S-transferase	8.9	Transcription

Table 2. Continued.

EST ID	Annotation	Fold change	Functional category[1]
GA_Eb0006G11f	unknown protein	8.5	Uncategorized
GA_Ea0032B01f	phi-1	8.4	CSO
GA_Eb0035E08f	tonoplast intrinsic protein bob TIP26-1	8.2	Transport
GA_Ea0028D14f	brassinosteroid biosynthetic protein LKB	8.1	Uncategorized
GA_Ea0031M24f	xyloglucan endotransglycosylase	8.0	ECM
GA_Eb0001C06f	putative cytochrome P450	8.0	ECM
GA_Ed0001H12r	unknown protein	7.9	Uncategorized
GA_Ea0028J16f	unknown protein	7.9	Uncategorized
GA_Ea0008H08f	ACE	7.9	ECM
GA_Ed0001A09f	unknown protein	7.8	Uncategorized
GA_Ea0012M05f	unknown protein	7.7	Uncategorized
GA_Ed0054C11r	no significant similarities	7.5	Uncategorized
GA_Ea0029C08f	sulfate transporter	7.2	Transport
GA_Ea0003G18f	pectinesterase	7.0	Uncategorized
GA_Ea0006P11f	glutathione S-transferase	6.9	Transcription
GA_Ea0018H11f	glutamine synthetase	6.8	Protein metabolism
GA_Eb0010A18f	putative glutamine synthetase	6.5	CSO
GA_Eb0018A24f	no significant similarities	6.5	Uncategorized
GA_Eb0018N24f	no significant similarities	6.4	Uncategorized
GA_Ed0034F03f	unknown protein	6.4	Uncategorized
GA_Eb0018D19f	no significant similarities	6.4	Uncategorized
GA_Eb0020G22f	no significant similarities	6.4	Uncategorized
GA_Ed0036B11f	P-glycoprotein	6.3	Uncategorized
GA_Ed0077G10f	GTP-binding protein	6.3	Signaling
GA_Ea0010D21f	hydroxymethylglutaryl coenzyme A synthase	6.2	Uncategorized
GA_Ed0098G08r	no significant similarities	6.2	Uncategorized
GA_Ea0007O12f	endo-beta-1,4-glucanase	6.1	ECM
GA_Ea0014E16f	unknown protein	6.1	CSO
GA_Ea0023C03f	MAP/ERK kinase 1	6.1	Uncategorized
GA_Eb0038J16f	no significant similarities	6.1	Uncategorized
GA_Ed0028F10f	seed coat BURP domain protein 1	6.0	Signaling
GA_Ed0077G09f	no significant similarities	6.0	Uncategorized
GA_Ea0024C12r	unknown protein	6.0	Uncategorized
GA_Ea0013F05f	unknown protein	6.0	Uncategorized
GA_Ea0008H12f	putative vacuolar acid invertase	6.0	ECM
GA_Ed0070H01f	phi-1	5.9	CSO
GA_Eb0024L14f	putative beta-galactosidase	5.8	ECM
GA_Eb0014P10f	cytosol aminopeptidase	5.8	Protein metabolism
GA_Ed0059A01r	lipid transfer protein	5.7	Uncategorized
GA_Ea0010F22f	unknown protein	5.7	Uncategorized
GA_Ed0026E12r	no significant similarities	5.7	Uncategorized
GA_Eb0012M15f	unknown protein	5.7	Uncategorized
GA_Ea0013K05f	unknown protein	5.7	Uncategorized
GA_Eb0032L22f	no significant similarities	5.7	Uncategorized
GA_Eb0026I04f	unknown protein	5.6	Uncategorized
GA_Ea0019L01f	nodulin-26	5.6	CSO
GA_Ed0015A02f	unknown protein	5.6	Uncategorized
GA_Eb0003B05f	unknown protein	5.6	Uncategorized
GA_Eb0015K21f	plastocyanin-like domain containing protein	5.6	Other categories
GA_Ea0024J19f	lipid-transfer protein-like protein	5.6	Transport
GA_Ea0012J18f	no significant similarities	5.6	Uncategorized
GA_Ea0012I03f	multi-copper oxidase-related protein	5.6	Other categories
GA_Ea0020O10f	no significant similarities	5.5	Uncategorized
GA_Ed0022B01r	GDSL-motif lipase	5.5	Uncategorized
GA_Ea0012I15f	polygalacuronase isoenzyme 1 beta subunit	5.5	ECM
GA_Ed0036A09f	pathogenesis-related protein	5.5	ERD

Table 2. Continued.

EST ID	Annotation	Fold change	Functional category[1]
GA_Eb0011B08f	unknown protein	5.5	Uncategorized
GA_Ed0099C09f	no significant similarities	5.5	Uncategorized
GA_Eb0025N10f	pathogenesis-related group 5 protein	5.5	CSO
GA_Ea0032I20f	lipid transfer protein	5.4	Transport

[1]Refer to footnotes of Table 1 for abbreviations of functional categories.
[2]For complete list of down-regulated genes, see Supplementary Table 4.

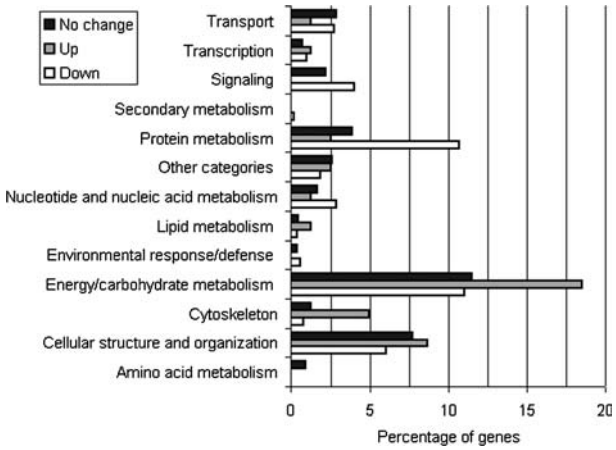

Figure 3. Functional categories of significantly up- and down-regulated genes in 24 dpa cotton (*G. arboreum* L.) fibers (Table 2; Supplementary Table 4) during cellulose synthesis and secondary cell wall biogenesis. Bars indicate consensus sequences within each functional category and expression group (up- or down-regulation, or no-change) expressed as a percent of the total number of consensus sequences within each expression group. Only consensus sequences assigned to a functional category are shown in the analysis.

to the same functional categories (Figure 3); energy/carbohydrate metabolism (11.1%), protein metabolism (10.7%), and cell structure, organization and biogenesis (6.0%). Uncategorized down-regulated fiber genes (57.8%) include proteins of unknown function as well as those for which no significant BLASTX results were returned. The number and functional categories of down-regulated genes, as well as timing in terms of development, are consistent with the termination of fiber elongation at the transcriptional level by 24 dpa. This includes the down-regulation of 26 transcription factors and 101 signaling molecules (Figure 3).

Very little is known about the fiber transcriptome in 24 dpa fibers (Wilkins and Jernstedt, 1999), and the temporal separation of primary and secondary cell wall synthesis in developing cotton fibers readily allows for independent study of these processes in 10 vs. 24 dpa fibers. Fiber genes specific to secondary cell wall synthesis (*GhCesA-1*, *GhCesA-2* and *FbL2A,* John and Crow, 1992; Pear *et al.,* 1996), and not present in the fiber dbEST were included as secondary cell wall controls on microarrays. *FbL2A* and *GhCesA-1* showed strong induced expression of 37.8- and 27.3-fold, respectively, in 24 dpa fibers, while *GhCesA-2* expression increased 7.7-fold (Table 2). A total of 81 genes were identified as up-regulated in 24 dpa fibers, or 0.7% of all differentially expressed genes. Excluding controls, the most up-regulated fiber gene (15.3-fold) is a member of the sucrose synthase gene family. Although many (42.9%) of the up-regulated fiber genes encode novel proteins of unknown function (no BLASTX results), slightly more than 75% of all up-regulated genes fall into three major categories – energy/metabolism (18.5%), cell structure, organization and biogenesis (8.6%), and cytoskeleton (4.9%) consistent with massive cellulose synthesis and secondary cell wall biogenesis

Table 3. Differential expression of fiber genes during primary (10 dpa) vs. secondary (24 dpa) cell wall biogenesis.

Expression categories in 10 dpa[1]	Number of genes differentially expressed in 24 dpa fibers			
	Up-regulated genes		Down-regulated genes	
	Moderate[2]	High[3]	Moderate[4]	High[5]
Low (0)	36	23	1213	88
Moderate (1–11)	14	8	1107	111
High (≥12)	0	0	25	8

[1] Categorized by 10 dpa transcript abundance (based on number of Ea ESTs per cluster).
[2] 2- to 4-fold increase.
[3] >4-fold increase.
[4] 2- to 4-fold decrease.

(Figure 3). Similar to the number of down-regulated genes, 58% of up-regulated genes have no GO assignments. Microarray expression data was analyzed in the context of Ea transcript abundance from *in silico* expression results to search for general trends (Table 3). For the most part, moderately expressed genes in 10 dpa fibers displayed a modest 2- to 4-fold increase in expression, while none of the highly expressed genes (Table 1) showed an increase in expression in 24 dpa fibers. By contrast, a subset of 23 up-regulated genes strongly induced during secondary cell wall synthesis in 24 dpa fibers, identified as singletons in our fiber dbEST, represent very low expressing genes or rare transcripts (normalized Eb or Ed ESTs only) in 10 dpa fibers. 75% of genes highly expressed in 10 dpa (Table 1) are down-regulated in 24 dpa fibers (Table 3). Most of the genes highly down-regulated in 24 dpa fibers showed moderate abundance levels in 10 dpa fibers.

While *in silico* northern analysis revealed that members of multigene families are differentially expressed, transcription profiling using microarrays showed that individual family members are also developmentally regulated. As an example, four α-tubulin and two β-tubulin genes are down-regulated $\leq$ 2.7-fold in 24 dpa fibers (Supplementary Table 4), including major α- and β-tubulin isoforms (Table 1). Intriguingly, three very low expressing β-tubulin genes were selectively up-regulated more than 2-fold in 24 dpa fibers (Table 2). No significant increase in expression was observed for members of the α-tubulin gene family. The differential and developmental regulation of a subset of the tubulin gene family was confirmed in a separate study by qPCR (Wilkins and Arpat, in preparation). The varying range in expression detected, both in terms of relative abundance and developmental changes in 10 vs. 24 dpa fibers, indicates that microarrays were capable of discriminating differences in expression among members of multigene families. Microarray results were corroborated by qPCR for 16 fiber genes, both up- and down-regulated and grouped as high, moderate and low expressing genes. Using gene-specific primers (Supplementary Table 3), qPCR expression profiles in 10 vs. 24 dpa fibers produced similar results as shown by the comparison of log ratios showing a strong correlation ($R^2 = 0.86$) in both methods (Supplementary Figure 2).

Discussion

Cotton functional genomics, driven by the EST gene discovery project and microarray expression profiling described here, has significantly advanced our understanding of the complexity of biological and cellular processes required for cotton fiber growth and development. Clearly, rate and duration of expansion, the type of growth mechanism involved (diffuse or tip-growth), and direction of expansion (cell polarity), together, play a crucial role in determining cell size and morphology, and hence, agronomic traits. The goal of this study was to define and characterize the transcriptome of elongating cotton fibers, and to assess gross changes in the transcriptome in response to developmental programs terminating cell expansion and signaling the switch from primary to secondary cell wall synthesis. As part of our

long-term interest in understanding the molecular mechanisms that control the size and shape of fibers, we elected to target fiber elongation as this stage governs important agronomic properties that dictate market value. Moreover, cell expansion is understandably one of the most fundamental processes in plant biology (Cosgrove, 2000) as all plant cells undergo expansion to varying degrees at some point during growth and development. The high genetic complexity of the fiber transcriptome reported here clearly indicates that rapid cell elongation in developing cotton fibers is a very metabolically active process that is subject to developmental cues that, in turn, govern fiber morphology and agronomic properties.

Transcriptome of rapidly elongating cotton fiber cells

Cultivated cotton species include Asiatic diploids *G. arboreum* L. and *G. herbaceum* L. ($2n = 26$, A genome) and allotetraploids *G. hirsutum* L. and *G. barbadense* L. ($2n = 4x = 52$, AD genome). The allotetraploids arose from a polyploidization event that took place $\sim$ 1.5–2 mya between an Old World A-genome species as the maternal parent, and a New World D-genome progenitor species (Galau and Wilkins, 1989; Wendel, 1989). Interestingly, fiber gene function is very highly conserved in the cotton genomes of wild D-genome diploid species, as well as cultivated diploid and tetraploid species, both in terms of nucleotide sequence and expression patterns, despite the millions of years of evolutionary history that separate the species (Cedroni *et al.,* 2003; Senchina *et al.,* 2003).

The fiber transcriptome in rapidly elongating cotton fibers consists of 13,947 unique (non-redundant) gene sequences assembled from over 46,000 ESTs. Based on an estimated 35,000–40,000 genes in diploid cotton, the fiber transcriptome, as it is currently defined, represents as much as 35–40% of the cotton genome. Given that the vast majority of fiber genes ($>10{,}000$) are expressed during both primary and secondary stages of cell wall development, and that the transcriptional activity and genetic complexity of other developmental stages is considerably lower (reviewed in Ryser, 1999), the genetic complexity of elongating fibers accounts for a significant fraction of the fiber transcriptome.

Based on our estimate of $\sim$ 14,000 genes in the fiber transcriptome of a cultivated progenitor species, and evidence of homoeologous genes from the A_T and D_T genomes in allotetraploid species (Senchina *et al.,* 2003; Rong *et al.,* 2004), the fiber transcriptome increases to an estimated 28,000+ genes, making redundancy an issue in gene discovery projects. To show that this is the case, comparison of a similar number of fiber ESTs sequenced randomly from diploid and tetraploid fiber cDNAs revealed that all genes identified in the tetraploid were present in our diploid fiber dbEST, and that discovery of novel genes was enhanced at least 2-fold, a significant number of which (40.6%) encode novel proteins. However, the discovery of more isoforms than may otherwise be expected is in keeping with a paleopolyploid origin for diploid cotton from an ancient polyploidization event between species bearing six or seven chromosomes (Wilkins *et al.,* 1994).

The high genetic complexity of the fiber transcriptome during turgor-driven cell expansion and elongation reflects the metabolic activity required to sustain the prolonged exaggerated growth of developing fibers, and provides a novel view of cell expansion at the level of a single cell. A considerable fraction (65%) of the most abundantly expressed gene transcripts (Table 1) fall into three main functional categories – cytoskeleton, cell wall, and metabolism. In addition, the bulk of moderately expressed genes are also metabolism-related. Similarly, metabolism-related genes are the most abundant gene transcripts among 569 glandular and eglandular trichome ESTs generated from subtracted libraries of *Medicago sativa* (Hayes and Skinner, 2001). A number of the most abundant fiber gene transcripts characterized at the molecular level are developmentally regulated during fiber elongation, in which transcripts increase dramatically during rapid growth, before declining coincident with the termination of expansion (Shimizu *et al.,* 1997; Smart *et al.,* 1998; Ruan *et al.,* 2001). Expression profiles that compare transcript abundance during and after fiber elongation revealed that $\sim$ 20% of the fiber transcriptome, including large numbers of metabolism related genes and cell wall proteins (Figure 3), is down-regulated coincident with termination of cell expansion. This subset of developmentally regulated genes is therefore highly specific to cell expansion, and provides candidate genes for the genetic modification of fiber

properties that target the timing, rate and duration of cell expansion and elongation.

Primary and secondary cell wall biosynthesis

Developmental programs regulate the temporal synthesis of fiber primary (PCW) and secondary (SCW) cell walls, which differ significantly in structure and composition. While the thin PCW (0.2–0.4 μm) deposited during fiber elongation contains $<30\%$ cellulose, the thick SCW (8–10 μm) is composed of $>94\%$ cellulose (Meinert and Delmer, 1977). In addition, the degree of polymerization of cellulose microfibrils also varies, being <5000 in PCW and $\sim 14{,}000$ in SCW (Marx-Figini, 1966). Fiber genes (2553) abundantly expressed in 10 dpa elongating fiber cells are moderately or highly down-regulated during SCW synthesis in 24 dpa fibers (Table 3). Although known secondary cell wall stage-specific genes (*GhCesA1, GhCesA2* and *FbL2A*) are absent in our elongating fiber dbEST, we discovered 81 novel fiber genes classified as very moderately abundant or rare transcripts in elongating fibers that are significantly up-regulated during SCW synthesis in 24 dpa fibers (Table 3), providing the very first glimpse of the genetic composition of the fiber transcriptome during secondary cell wall synthesis. Differential gene expression reported here for the first time clearly suggests putative stage-specific roles for isoforms in PCW synthesis or SCW synthesis.

PCW synthesis in elongating cotton fibers involves $\sim 15\%$ or approximately 2100 cell wall-related genes in the fiber transcriptome, including genes associated with cell expansion, such as expansins, xyloglucan endotransglucosylase/hydrolases, and endoglucanases (Cosgrove, 2000) as well as structural proteins. These genes represent a major portion of those genes down-regulated (Figure 3) coincident with the termination of fiber elongation and cessation of PCW biosynthesis. Expansins, which play a crucial role in cell wall loosening during turgor-driven cell expansion (Cosgrove *et al.,* 2002), are classified into two subgroups, α and β, based on structural differences. Based on transcript abundance, the fiber transcriptome includes one major α-expansin (Table 1), several minor isoforms, and at least one β-expansin in elongating fibers. While most α-expansin genes show varying degrees of down-regulation in expression during the PCW-SCW switch, only the major α-expansin abundantly expressed in elongating fibers (Table 1) is significantly down-regulated. The presence of major and minor isoforms of α-expansins in our dbEST, and their differential regulation, suggests that these isoforms might have specialized roles during cell elongation and termination of cell elongation. In other studies, regulation of cell wall extensibility is controlled, in part, by differential expression of expansin genes (Vogler *et al.,* 2003).

Xyloglucan endotransglucosylase/hydrolases (XTHs) play a central role in metabolism of xyloglucans, an important polysaccharide component of the primary cell wall in dicotyledonous plants. XTHs may be involved in wall restructuring by reversibly or irreversibly loosening existing wall material via endotransglucosylase activity, thereby enabling cell expansion. In *Arabidopsis*, XTHs are classified into three discrete groups (Rose *et al.,* 2002), in which unique expression profiles likely reflect unique physiological functions. In elongating cotton fibers, seven XTHs are found in our fiber dbEST (http://cfgc.ucdavis.edu), four of which share similarity with *AtXTH* groups 1 and 2, and are moderately expressed at 10 dpa based on the number of Ga_Ea transcripts. Consistent with the general trend of "expansion-associated" fiber genes (Table 3), cotton *GaXTH* genes are down-regulated by as much as 8-fold in 24 dpa fibers (Table 2, GA_Ea0031M24f; Supplementary Table 4, GA_Eb0043H09f, GA_Ea0012E09f). Interestingly, one cotton *GaXTH*, which shows significant similarity to *AtXTH28* (*XTR2*/*EXGT-A2*), a member of group 3 XTHs, is up-regulated 4.5-fold during the SCW synthesis phase. Several lines of evidence from Arabidopsis mutants defective in cell elongation suggest that *AtXTH28* plays a role in cell wall maintenance rather than cell wall extension. In contrast to expression profiles of five *AtXTH* genes in *acaulis* mutants, expression of *AtXTH28* was not altered (Akamatsu *et al.,* 1999). Similarly, *AtXTH28* expression was also not altered in an *angustifolia (an)* mutant defective in polar cell elongation (Kim *et al.,* 2002). Thus, *AtXTH28* may function in the regulated degradation of xyloglucans following cell elongation (Edwards *et al.,* 1986; Akamatsu *et al.,* 1999). Indeed, xyloglucan content in the primary cell walls of developing cotton fibers parallels the cell expansion rate, and gradually decreases by

more than 3-fold by the termination of fiber expansion (Shimizu *et al.,* 1997). During SCW synthesis, the increase observed in expression of some cell wall structural genes such as proline-rich cell wall protein (Table 2, GA_Ed0033A11f) could be explained as 'filling-in' the gaps left by the former xyloglucan network (Carpita and Gibeau, 1993). Despite the important role ascribed to extensins in stabilizing PCW structure (Carpita and Gibeau, 1993), attempts to isolate fiber extensin genes have not met with success (John and Crow, 1992). However, the role of extensins in 'fixing' PCW structure is likely mediated by the action of hydroxyproline residues (Iraki *et al.,* 1989; Carpita and Gibeau, 1993), and in the case of developing cotton fibers, increased expression of newly identified proline-rich proteins in 24 dpa fibers suggest these proteins likely perform this function.

Endo-β-1,4-glucanases (EGases), encoded by a multi-gene family in higher plants, potentially function in cell wall extension and cellulose biosynthesis (Darley *et al.,* 2001; Molhoj *et al.,* 2002). Although evidence for a direct role for EGases in cell wall extension is still missing, a strawberry EGase (*FaEG3*) that contains a putative cellulose-binding domain (CBD) has been associated with cell wall loosening (Trainotti *et al.,* 1999). A cotton fiber EGase (Table 2, GA_Ea0024P08f) showing high sequence similarity to the strawberry *FaEG3*, and moderately expressed in 10 dpa fibers, is down regulated 6.1-fold during termination of fiber elongation, suggesting a role for this EGase isoform in cell wall loosening during fiber elongation. In contrast, another EGase (Table 2, GA_Ea0007O12f) with similar transcript abundance in 10 dpa fibers, is up-regulated 2.1-fold in 24 dpa fibers. This fiber EGase is highly similar to the *Arabidopsis KORRIGAN* (*KOR*) gene. Recent studies indicate that fiber Korrigan (Kor) is involved in cellulose biosynthesis (Molhoj *et al.,* 2002; Peng *et al.,* 2002). Kor protein abundance increases during SCW synthesis in 24 dpa cotton fibers, and Kor activity appears to be required for *in vitro* synthesis of cellulose (Peng *et al.,* 2002). These results suggest that different EGases perform specialized roles in cell wall extension and cellulose biosynthesis during PCW and SCW synthesis in developing cotton fibers.

One of the most abundant cell wall proteins expressed during fiber elongation is the novel E6 asparagine- and glutamic acid-rich polypeptide of unknown function (John and Crow, 1992; Table 1) that appears to be cotton-specific, as no E6 homolog has been identified in the Arabidopsis genome (Arpat and Wilkins, unpublished data). Interestingly, however, antisense suppression of E6 gene expression in transgenic cotton did not discernibly alter fiber properties (John, 1999), although this does not preclude a feedback mechanism by which other structural proteins function to maintain structural integrity to compensate for the loss of E6.

Synthesis of cellulose and related polysaccharides in 10 dpa elongating cotton fibers involves expression of as many as 12 *GaCesA* and *GaCesA-like* (*GaCsl*) isoforms. At the amino acid level, six non-redundant consensus sequences share 91–98% identity with members of the *AtCesA* family (*AtCesA1*, *AtCesA2*, *AtCes3*, and *AtCesA10*). In Arabidopsis, *CesA1* and *CesA3* are required for PCW synthesis (Arioli *et al.,* 1998; Scheible *et al.,* 2001; Burn *et al.,* 2002), and *AtCesA2* is believed to have a role in PCW deposition (Burn *et al.,* 2002). Consistent with a function in PCW, all six cotton homologs show either no change in gene expression, or are down-regulated in 24 dpa fibers, although none of the cotton genes meet criteria for significant up- or down-regulated genes. A similar case can be presented for five non-redundant consensus sequences with similarity to *AtCslE1* and *AtCslD3* genes. *AtCslD3* (*KOJAK*) is likely involved in the biosynthesis of β-glucan-containing polysaccharides required during root hair elongation (Favery *et al.,* 2000). While microarray results indicate that all putative PCW *GaCsl* genes are down-regulated during SCW synthesis, Ga_Ed0108F05r, which shares 90% amino acid identity with AtCslD3, is down-regulated 3-fold and is the only significantly differentially expressed *GaCsl* gene (Supplementary Table 4). The correlation between putative gene function and gene regulation of these genes suggests that other members of CesA/Csl family such as CesA10 and CslE1, may have a role in PCW synthesis. In contrast to PCW synthesis, the peak rate in cellulose synthesis in 24 dpa fibers is accompanied by strong induced expression of *GhCesA1* and *GhCesA2* genes, as well as up-regulation of many energy and carbohydrate metabolism related genes, including the differential expression of two sucrose synthase (*SuSy*) genes (Table 2), consistent with cellulose biosynthesis and SCW

modification. Intriguingly, ~ 50% of SuSy is membrane-bound at this stage and is purportedly a component of the cellulose synthase complex (Amor *et al.,* 1995). The up-regulation of plasma membrane-related genes, and genes involved in sterol biosynthesis and modification, coupled with cytoskeleton dynamics (Seagull, 1992) are likely reflective of dynamic changes to the plant cell wall-plasma membrane continuum during the biogenesis and modification of the fiber SCW. Recent biochemical studies suggest that sitosterol-β-glucoside, synthesized by UDP-Glc:sterol glucosyltransferase (SGT), is the primer required for cellulose biosynthesis. Expression profiles of the three or more putative *SGT* genes in our cotton dbEST are differentially expressed during the switch from PCW to SCW. The 5-fold up-regulation of one fiber *SGT* (Table 2, GA_Ea0008F01f) is indeed consistent with a role for SGT in cellulose synthesis in 24 dpa fibers.

Model of fiber development

The high genetic complexity of the fiber transcriptome reflects the metabolic activity required to sustain rapid turgor-driven expansion in developing cotton fibers. Besides a large contingent of fiber genes that perform basic functions during development, the switch in developmental programs from primary to secondary cell wall synthesis and the termination of cell expansion is accompanied by dynamic changes in gene expression. About 2500 highly and moderately expressed fiber genes down-regulated in terminal stages of fiber elongation function selectively or preferentially during cell expansion. The regulation of these genes is in keeping with our developmental model in which expression of key genes is developmentally regulated in parallel with the rate of expansion (Smart *et al.,* 1998; Wilkins and Jernstedt, 1999). Transcript abundance and expression signatures identified major and minor isoforms, such that isoform-specific roles in synthesis of PCW and SCW can be projected for closely related members of a multi-gene family based on differential and developmental differences in expression. Moreover, the ability to distinguish stage-specific expression of major isoforms of functionally important genes may well account for the bulk of the genetic variability associated with major QTLs for fiber quality (Kohel *et al.,* 2001; Zhang *et al.,* 2003).

The next major challenge will be to determine fiber gene function using reverse genetic approaches, especially for the significant number of fiber genes of unknown function in long-term applications in agricultural biotechnology for the genetic improvement of cotton (Wilkins *et al.,* 2000).

Acknowledgements

The authors gratefully acknowledge the contributions of Drs. Bill Beavis, Callum Bell and Geoffrey Routh to the project. This work was supported by the NSF Plant Genome Program (DBI-9872630) to T.A.W.

References

Akamatsu, T., Hanzawa, Y., Ohtake, Y., Takahashi, T., Nishitani, K. and Komeda, Y. 1999. Expression of endoxyloglucan transferase genes in *acaulis* mutants of *Arabidopsis*. Plant Physiol. 121: 715–721.

Amor, Y., Haigler, C.H., Johnson, S., Wainscott, M. and Delmer, D.P. 1995. A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. Proc. Natl. Acad. Sci. USA 2: 9353–9357.

Applied Biosystems. 2001. ABI Prism 7700 Sequence Detection System User Bulletin 2. The Perkin-Elmer Corporation, P/N 4303859 Rev B, Stock No. 777802-002. http://docs.appliedbiosystems.com/pebiodocs/04303859.pdf

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D.R. *et al.* 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. 29: 37–40.

Arioli, T., Peng., L., Betzner, A.S., Burn, J., Wittke, W., Herth, W., Camilleri, C., Hofte, H., Plazinski, J., Birch, R., Cork, A., Glover, J., Redmond, J. and Williamson, R.E. 1998. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science 279: 717–720.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* 2000. Gene ontology: Tool for the unification of biology. Nat. Genet. 25: 25–29.

Basra, A.S. and Malik, C.P. 1984. Development of cotton fibers. Int. Rev. Cytol. 89: 65–113.

Bouyer, D., Kirik, V. and Hülskamp, M. 2001. Cell polarity in *Arabidopsis* trichomes. Cell Dev. Biol. 12: 353–356.

Burn, E.J., Hocart, C.H., Birch, R.J., Cork, A. and Williamson, R.E. 2002. Functional analysis of the cellulose synthase genes *CesA1*, *CesA2*, and *CesA3* in *Arabidopsis*. Plant Physiol. 129: 797–807.

Carpita, N.C. and Gibeau, D.M. 1993. Structural models of primary cell walls in flowering plants: Consistency of

molecular structure with the physical properties of the walls during growth. Plant J. 3: 1–30.

Cedroni, M.L., Cronn, R.C., Adams, K.L., Wilkins, T.A. and Wendel, J.F. 2003. Evolution and expression of *MYB* genes in diploid and polyploid cotton. Plant Mol. Biol. 51: 313–325.

Cosgrove, D.J. 2000. Expansive growth of plant cell walls. Plant Physiol. Biochem. 38: 109–124.

Cosgrove, D.J., Li, L.C., Cho, H.-T., Hoffmann-Benning, S., Moore, R.C. and Blecker, D. 2002. The growing world of expansins. Plant Cell Physiol. 43: 1436–1444.

Darley, C.P., Forrester, A.M. and McQueen-Mason, S.J. 2001. The molecular basis of plant cell wall extension. Plant Mol. Biol. 47: 179–195.

Delmer, D.P. 1999. Chapter 4. Cellulose biosynthesis in developing cotton fibers. In: A.M. Basra (Ed.), Cotton Fibers. Hawthorne Press, New York, pp. 85–106.

Edwards, M., Dea, I.C.M., Bulpin, P.V. and Reid, J.S.G. 1986. Purification and properties of a novel, xyloglucan-specific endo-(14)-β-D-glucanase from germinated nasturtium seeds (*Tropaeolum majus* L.). J. Biol. Chem. 261: 9489–9494.

Favery, B., Ryan, E., Foreman, J., Linstead, P., Boudonck, K., Steer, M., Shaw, P. and Dolan, L. 2001. *KOJAK* encodes a cellulose synthase-like protein required for root hair cell morphogenesis in *Arabidopsis*. Genes Dev. 15: 79–89.

Galau, G.A. and Wilkins, T.A. 1989. Alloplasmic male sterility in AD allotetraploid *Gossypium hirsutum* upon replacement of its resident A cytoplasm with that of D species *G. harknessii*. Theor. Appl. Genet. 78: 23–30.

Hasenfratz, M.-P., Tsou C.-L. and Wilkins, T.A. 1995. Expression of two related vacuolar H^+-ATPase 16 kD proteolipid genes is differentially regulated in a tissue-specific manner. Plant Physiol. 108: 1395–1404.

Hayashi, T., Wong, Y.-S. and Maclachlan, G.A. 1984. Pea xyloglucan and cellulose. 2. Hydrolysis by pea endo-1,4-beta-glucanases. Plant Physiol. 75: 605–610.

Hayes, D.B. and Skinner, D.Z. 2001. Development of an expressed sequence tag (EST) library for *Medicago sativa*. Plant Sci. 161: 517–526.

Hedge, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Lee, N. and Quackenbush, J. 2000. A concise guide to cDNA microarray analysis. Biotechniques 29: 548–557.

Henikoff, S., Henikoff, J.G. and Pietrokovski, S. 1999. Blocks+: A non-redundant database of protein alignment blocks derived from multiple compilations. Bioinformatics 15: 471–479.

Hsieh, Y.-L. 1999. Structural development of of cotton fibers and linkages to fiber quality. In: A.S. Basra (Ed.), Cotton Fibers, Hawthorne Press Inc., New York, pp. 137–144.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H.Y., He, Y.D.D., Stephaniants, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H. and Lindsey, P.S. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nature Biotech. 19: 342–347.

Iraki, N.M., Bressan, R.A., Hasegawa, P.M. and Carpita, N.C. 1989. Alteration of the physical and chemical structure of the primary cell wall of growth limited plant cells adapted to osmatic stress. Plant Physiol. 91: 39–47.

John, M.E. 1999. Genetic engineering strategies for cotton fiber modification. In: A.S. Basra (Ed), Cotton Fibers, Hawthorne Press, New York, pp. 271–292.

John, M.E. and Crow, L.J. 1992. Gene expression in cotton (*Gossypium hirsutum* L.) fiber: Cloning of the mRNAs. Proc. Natl. Acad. Sci. USA 89: 5769–5773.

Kim, G.T., Shoda, K., Tsuge, T., Cho, K.H., Uchimiya, H., Yokoyama, R., Nishitani, K. and Tsukaya, H. 2002. The *ANGUSTIFOLIA* gene of *Arabidopsis,* a plant *CtBP* gene, regulates leaf-cell expansion, the arrangement of cortical microtubules in leaf cells and expression of a gene involved in cell-wall formation. EMBO J. 21: 1267–1279.

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

Kurek, I., Kawagoe, Y., Jacob-Wilk, D., Doblin, M. and Delmer, D. 2002. Dimerization of cotton fiber cellulose synthase catalytic subunit occurs via oxidation of the zinc-bing domains. Proc. Natl. Acad. Sci. USA 99:11109–11114.

Lijavetzky, D., Muzzi, G., Wicker, T., Keller, B., Wing, R. and Dubcovsky, J. 1999. Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome 42: 1176–1182.

Maier, E., Meier-Ewert, S., Ahmadi, A.R., Curtis, J. and Lehrach, H. 1994. Application of robotic technology to automated sequence fingerprint analysis by oligonucleotide hybridization. J. Biotech. 35: 191–203.

Marx-Figini, M. 1966. Comparison of the biosynthesis of cellulose *in vitro* and *in vivo* in cotton bolls. Nature 210: 747–755.

Meinert, M.C. and Delmer, D.P. 1977. Changes in biochemical composition of cell wall of cotton fiber during development. Plant Physiol. 59: 1088–1097.

Meredith, Jr. W.R. 2000. Continued progress for breeding for yield in the USA? In: U. Kechagia (Ed.), Proceedings of the World Cotton Research Conference II, Athens, Greece, pp. 97–101.

Molhoj, M., Pagant, S. and Hofte, H. 2002. Toward understanding of the role of membrane-bound endo-β-1,4-glucanases in cellulose biosynthesis. Plant Cell Phyisol. 43: 1399–1406.

Nolte, K.D., Hendrix, D.L., Radin, J.W. and Koch, K.E. 1995. Sucrose synthase localization during initiation of seed development and trichome differentiation in cotton ovules. Plant Physiol. 109: 1285–1293.

Orford, S.J. and Timmis, J.N. 1998. Specific expression of an expansin gene during elongation of cotton fibres. BBA Genet. Struct. Exp. 1398: 342–346.

Pear, J.R., Kawagoe. Y., Schreckengost, W.E., Delmer, D.P. and Stalker, D.M. 1996. Higher plants contain homologs of the bacterial *celA* gene encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. USA 93: 12637–12642.

Peng, L., Kawagoe, Y., Hogan, P. and Delmer, D. 2002. Sitosterol-β-glucoside as primer for cellulose synthesis in plants. Science 295: 147–150.

Potikha, T.S., Collins, C., Johnson, D.I., Delmer, D.P. and Levine, A. 1999. The involvement of hydrogen peroxide in the differentiation of secondary walls in cotton fibers. Plant Physiol. 119: 849–858.

Ratajczak, R. and Wilkins, T.A. 2000. Energizing the tonoplast. Chapter 7. In: D.G. Robinson and J.C. Rogers (Eds.),

Vacuolar Compartments, Sheffield Academic Press Ltd., England, pp. 133–173.

Rong, J.K., Abbey, C., Bowers, J.E., Brubaker, C.L., Chang, C., Chee, P.-W., Delmonte, T.A., Ding, X.L., Garza, J.J., Marler, B.S., Park, C.H., Pierce, G.J., Rainey, K.M., Rastogi, V.K., Schultze, S.R., Trolinder, N.L., Wendel, J.F., Wilkins, T.A., Williams-Coplin, T.D., Wing, R.A., Wright, R.J., Zhao, X.P., Zhu, L.H. and Paterson, A.H. 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166: 389–417.

Rose, J.K.C., Braam, J., Fry, S.C. and Nishitani, K. 2002. The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature. Plant Cell Physiol. 43: 1421–1435.

Ruan, Y.L., Llewellyn, D.J. and Furbank, R.T. 2001. Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. Plant Cell 13: 47–60.

Ryser, U. 1999. Chapter 1. Cotton fiber initiation and histodifferentiation. In: A.M. Basra (Ed.), Cotton Fibers, Hawthorne Press, New York, pp. 1–45.

Scheible, W.R., Eshed, R., Richmond, T., Delmer, D. and Somerville, C. 2001. Modifications of cellulose synthase confer resistance to isoxalen and thiazolidinone herbicides in *Arabidopsis Ixrl* mutants. Proc. Natl. Acad. Sci. USA 98: 10079–10084.

Seagull, R.W. 1992. A quantitative electron microscopic study of changes in microtubule arrays and wall microfibril orientation during *in vitro* cotton fiber development. J. Cell Sci. 101: 561–577.

Senchina, D.S., Alvarez, I., Cronn, R.C., Liu, B., Rong, J., Noyes, R.D., Paterson, A.H., Wing, R.A., Wilkins, T.A. and Wendel J.F. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol. Biol. Evol. 20: 633–643.

Shimizu, Y., Aotsuka, S., Hasegawa, O., Kawada, T., Sakuno, T.F. and Hayashi, T. 1997. Changes in levels of mRNAs for cell wall-related enzymes in growing cotton fiber cells. Plant Cell Physiol. 38: 375–378.

Smart, L.B., Vojdani, F., Maeshima, M. and Wilkins, T.A. 1998. Genes involved in Osmoregulation during turgor-driven cell expansion of developing cotton fibers are differentially regulated. Plant Physiol. 116: 1539–1549.

Tiwari, S.C. and Wilkins, T.A. 1995. Cotton (*Gossypium hirsutum* L.) seed trichomes expand via diffuse growing mechanism. Can. J. Bot. 73: 746–757.

Trainotti, L., Spolaore, S., Pavanello, A., Baldan B. and Casadoro, G. 1999. A novel E-type endo-β-1,4-glucanase with a putative cellulose-binding domain is highly expressed in ripening strawberry fruits. Plant Mol. Biol. 40: 323–332.

Vogler, H., Caderas, D., Mandel, T. and Kuhlemeier C. 2003. Domains of expansin gene expression define growth regions in the shoot apex of tomato. Plant Mol. Biol. 53: 267–272.

Vojdani, F., Kim, W. and Wilkins, T.A. 1997. Phosphoenolpyruvate carboxylase cDNAs from developing cotton (*Gossypium hirsutum*) fibers (accession no. AF008939 and AF008940) (PGR 97-135). Plant Physiol. 115: 315.

Wan, C.-Y. and Wilkins, T.A. 1994. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). Anal. Biochem. 223: 7–12.

Wendel, J.F. 1989. New world tetraploid cottons contain old-world cytoplasm. Proc. Natl. Acad. Sci. USA 86: 4132–4136.

Werker, E. 2000. Trichome diversity and development. Adv. Bot. Res. 31: 1–35.

Wilkins, T.A. 1993. Vacuolar H^+-ATPase 69-kilodalton catalytic subunit cDNA from developing cotton (*Gossypium hirsutum*) ovules. Plant Physiol. 102: 679–680.

Wilkins, T.A. and Jernstedt, J.A. 1999. Chapter 9. Molecular genetics of developing cotton fibers. In: A.M. Basra (Ed.), Cotton Fibers, Hawthorne Press, New York, pp. 231–267.

Wilkins, T.A. and Smart, L.B. 1996. Chapter 2. Isolation of RNA from plants. In: P.A. Kreig (Ed.), A Laboratory Guide to RNA: Isolation, analysis and synthesis, Wiley & Sons Inc., New Jersey, pp. 21–41.

Wilkins, T.A., Rajasekaran, K. and Anderson, D.M. 2000. Cotton Biotechnology. Crit. Rev. Plant Sci. 15: 511–550.

Wilkins, T.A., Wan, C.Y. and Lu, C.C. 1994. Ancient origin of the vacuolar H^+-ATPase 69-kilodalton catalytic subunit superfamily. Theor. Appl. Genet. 89: 514–524.

Zdobnov, E.M. and Apweiler, R. 2001. InterProScan – An integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847–848.

Zhang, T., Yuan, Y., Yu, J., Guo, W. and Kohel, R.J. 2003. Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theor. Appl. Genet. 106: 262–268.