

Citation for this Article: Haigler CH, Singh B, Wang G, Zhang D (2009) Genomics of cotton fiber secondary wall deposition and cellulose biogenesis. In: Paterson AH (ed.) *Genetics and Genomics of Cotton*, Plant Genetics and Genomics: Crops and Models 3, Springer Science + Business Media, LLC: New York, pp. 385-417, DOI 10.1007/978-0-387-70810-2\_16

## Genomics of Cotton Fiber Secondary Wall Deposition and Cellulose Biogenesis

Candace H. Haigler<sup>1</sup>, Bir Singh<sup>2</sup>, Guirong Wang<sup>3</sup>, and Deshui Zhang<sup>4</sup>

<sup>1</sup> North Carolina State University, Department of Crop Science and Department of Plant Biology, candace\_haigler@ncsu.edu

<sup>2</sup> North Carolina State University, Department of Crop Science, bsingh4@ncsu.edu

<sup>3</sup> Data Systems Group, guirong2004@yahoo.com

<sup>4</sup> Ventria Bioscience, dzhang@ventria.com

### Abstract

The deposition of > 90% cellulose in the cotton fiber secondary wall makes this unique cell powerful for understanding cellulose biogenesis, a process with great importance in nature and industry. This chapter provides an overview of cellulose biogenesis, summarizes how cotton fiber has previously facilitated unique insights in this field, and explains how cellulose is important in terms of cotton fiber physical properties. The nature of the cotton fiber secondary wall transcriptome is discussed, including comparisons to primary-wall-stage fiber and the Arabidopsis proteome. Microarray data, including validation by quantitative reverse transcription PCR, are described to show that transcriptomes for secondary wall deposition in cotton fiber and xylem are similar. The functional context of selected genes that are up-regulated for secondary wall deposition is discussed.

### 1 Introduction

Cellulose is an essential renewable material in nature and industry, and we need to know more about how plants carry out a nanoscale manufacturing process to produce it. Cellulose is defined chemically as polymeric  $\beta$ -1,4-glucan, but it occurs naturally as partly crystalline fibrils within plant cell walls. Therefore, the production of cellulose is best referred to as a “biogenetic” rather than simply a “biosynthetic” process. Cellulose fibrils are typically 3 – 6 nm in diameter and several  $\mu$ m long; for example, a secondary wall cellulose molecule with 10,000 glucose units would be 4.75  $\mu$ m long. The term “microfibril” historically was used to denote the primary unit of cellulose biogenesis in higher plants, but given their diameter and the importance of nanoscale phenomena in their synthesis and interactions with other cell wall molecules, “nanofibril” is more appropriate (Haigler 2006). Due to their high stiffness and strength, cellulose nanofibril orientation is a primary control point for the plant morphogenetic processes that depends on directional constraint of cell expansion. Thin, extensible, primary walls that typically contain 10 – 20% cellulose surround all plant cells. After cell expansion is largely complete, some cells deposit cellulose-rich secondary walls that are at least 4  $\mu$ m thick. These confer extra

strength to cells such as fibers and water-conducting tracheary elements (Haigler 1985; Niklas 1992; Stone 2001).

Cellulose quantity and physical properties are central to virtually every industrial use of fiber and chemical cellulose. Wood is composed mainly of vascular fibers and water conducting cells with secondary walls, and the cellulose component rather than the lignin is the useful component for many products. Cotton fiber, which is essentially a cellulosic secondary cell wall when mature, is a major resource for the textile and chemical industries. Since secondary walls derive most of their extra strength through modified cellulose properties (Hsieh 1999; Benedict et al. 1999), there is a particular need to understand the regulation of secondary wall cellulose biogenesis as a basis for devising strategies to stabilize and/or change the quantity and quality of cellulose. This need has been emphasized recently by the interest in making biofuels from renewable plant biomass, with the cellulose component being the most abundant and useful for saccharification.

For numerous reasons, cotton fiber provides a valuable experimental tool and genomics system for understanding the control of cellulose biogenesis in plants in general (Kim and Triplett 2001). **(1)** The normal cotton fiber secondary wall has the highest percentage of cellulose known in plants. Residual weight of crystalline cellulose after solubilization of other components equaled 90% of original fiber weight (Haigler et al. 2007). However, since both the primary wall and non-crystalline cellulose would have been removed by the acetic/nitric solvent, >90% cellulose existed in the secondary wall *per se*. In contrast, typical secondary walls in xylem/wood contain 40-50% cellulose, although bending-induced tension wood has special G-fibers with a secondary wall layer that is richer in cellulose (Pilate et al. 2004). **(2)** Cotton fibers, after the initiation and earliest elongation stages, are easily separated from the seed in large quantity using minimally stressful procedures, so gene expression in the context of a single cell is easily analyzed. **(3)** The prolonged (45 – 60 day) developmental program of *Gossypium hirsutum* includes distinct phases of primary wall deposition (for fiber elongation) and secondary wall deposition (for fiber thickening), as well as an intervening transitional stage. **(4)** Experimentation is made more efficient by year-round fiber development in greenhouses or growth chambers or on cultured ovules. **(5)** The genus *Gossypium* has a well-documented and fascinating evolutionary history that includes recent polyploidization. Both diploid and allotetraploid species were domesticated to create cultivated crops by several different groups of indigenous people beginning ~6000 years ago, and breeding has continued until the present day (Fryxell 1979; Endrizzi et al. 1985; Adams and Wendel 2004). **(6)** Genomic tools for cotton are rapidly emerging, as described further elsewhere in this volume and below.

This chapter will emphasize genomic research in cotton and other plants that provides additional evidence and/or extends concepts described previously. From a pre-genomic perspective, the fundamental biology of cotton fiber cellulose was reviewed recently including: (a) the role and changing properties of cellulose during fiber development; (b) control of cellulose biogenesis by genetic, cellular, and biochemical processes; and (c) limitations on cellulose accumulation imposed by abiotic stresses (Haigler 2007). In the brief summary below, primary citations are often not repeated if they can be found there or in other reviews on cellulose

biogenesis (Delmer 1999; Haigler et al. 2001; Scheible and Pauly 2004; Saxena and Brown 2005; Somerville 2006). In addition, a recent multi-author book provides a good resource for understanding diverse aspects of cellulose biogenesis (Brown and Saxena 2007), and other reviews give more comprehensive treatment of several aspects of cell wall biosynthesis discussed here (e.g. Farrokhi et al. 2006).

## 2 Overview of Cellulose Biogenesis

Cellulose biogenesis is under complex, multi-level, control including regulation through genetics, cell biology, biophysics, biochemistry, hormones, and environmental interaction. This speaks to both the nanoscale cellular manufacturing of crystalline fibrils and the advantage to the plant of regulating carefully the strong, irreversible, carbon sink of secondary wall cellulose deposition. Cellulose biogenesis bridges both sides of the plasma membrane, from cortical cytoskeleton to cell wall space. Plant cellulose fibrils are synthesized by a large organized complex of proteins, which has been detected as a “rosette” of 6, ~8 nm diameter, particles in the inner face of the fractured plasma membrane. The rosette marks the existence of protein transmembrane domains within an orderly aggregation of ~24-36 cellulose synthase (CESA) enzymes, and probably other proteins that participate in cellulose synthesis. Microfibril-terminal-complex (TC) or “rosette” are traditional names for this complex, but “cellulose synthase complex (CSC)” has been suggested as an alternative (Diotallevi and Mulder 2007). However, CSC to denote “cellulose synthesis complex” may be preferred to allow the possibility that other proteins besides CESAs are part of the complex; this terminology will be used further in this article.

The close proximity of synthetic enzymes facilitates self-assembly and co-crystallization of newly polymerized  $\beta$ -1,4-glucan chains. At least the core part of the CSC is transported to the plasma membrane by fusion of Golgi vesicles, and CSC integrity in the plasma membrane may be under oxidative and proteolytic regulation. Supporting a long-standing hypothesis, there are now both *in vivo* and biophysical modeling evidence that CSCs move at about  $6 \times 10^{-9}$  m/s through the fluid plasma membrane as cellulose nanofibrils crystallize behind them (Paradez et al. 2006; Diotallevi and Mulder, 2007). Microtubules and the cellulose nanofibrils themselves are seemingly interdependent in determining the localization and orientation of cellulose fibril biogenesis, at least in some cell types. A plasma membrane- and/or cytoskeleton-associated form of sucrose synthase participates in secondary wall cellulose biogenesis by cleaving sucrose to release UDP-glucose (UDP-glc) to the polymerizing enzymes. Metabolic control, interfacing with hormonal and environmental regulation, is required to regulate the extent of cellulose deposition in different cell types and in response to different growing conditions. In reference to cotton genes that are up-regulated for secondary wall deposition, we will discuss some proteins that participate in this process in more detail below. However, it is important to note that there is currently a poor understanding of exact mechanisms of protein action in most cases or how diverse participants carry out the integrated process of cellulose biogenesis.

### 3 Cotton Fiber Illuminates the Control of Cellulose Biogenesis

Although extensive cell wall thickening is characteristic of the seed hairs of all *Gossypium* species (Fryxell 1979), unless specifically noted otherwise, research data summarized here are for *Gossypium hirsutum*. Allotetraploid *G. hirsutum* (AADD genome) arose 1-2 mya by hybridization of two diploid genomes, AA (typified by *G. arboreum*) and DD (typified by *G. raimondii*) (Wendel and Cronn 2002). Domesticated *G. hirsutum* is now grown most commonly for fiber crop production worldwide, and *G. arboreum* is sometimes grown on a much smaller scale in Asia. However, *G. raimondii* produces short fiber that is not economically useful, although a similar DD genome contributed useful quantitative trait loci (QTL) for fiber traits to the allotetraploid (Jiang et al. 1998). Timing of developmental events in cotton fiber is referenced as days post anthesis (DPA) because fiber initiation occurs at or near anthesis (flower opening). Typical primary walls, ~ 0.5  $\mu\text{m}$  thick and containing 20-25% cellulose along with pectin xyloglucan, and protein, are synthesized during fiber elongation. In *G. hirsutum*, primary wall deposition proceeds alone until 15 - 19 DPA. Next, a transition phase with concurrent primary and secondary wall deposition begins within the range of 16 - 20 DPA. At this stage the wall begins to thicken rapidly, including synthesis of cellulose fibrils that are oriented at 45- 55° relative to the long fiber axis within the transiently-synthesized "winding" cell wall layer (Ryser 1985). Beginning within the range of 19 - 24 DPA, continued fiber wall thickening occurs via > 90% cellulose deposition until 45 - 50 DPA when the boll opens and the fiber dries (Meinert and Delmer 1977). The ranges of DPA for developmental stages shown above reflect variability within cultivars and across different environments.

Because of the distinct phase of secondary wall deposition within its prolonged development program, cotton fiber has a long history of revealing new information about the control of cellulose biogenesis and secondary wall deposition. For example, research on cotton fiber: **(1)** allowed the first plant *CESA* genes to be identified (Pear et al. 1996); **(2)** illuminated effects of microtubules and actin on the orientation of cellulose fibrils (Seagull 1993); **(3)** showed for the first time that multiple tubulin isotypes existed in a single cell, with some isotypes under developmental regulation (Whittaker and Triplett 1999); **(4)** showed changing rates of cellulose biogenesis and changing length of cellulose chains for primary and secondary wall deposition (Meinert and Delmer 1977; Timpa and Triplett 1993); **(5)** implicated UDP-glc as the substrate for cellulose biogenesis (Franz 1969; Carpita and Delmer 1981); **(6)** demonstrated a role for membrane-associated sucrose synthase acting degradatively in supplying UDP-glc for secondary wall cellulose biogenesis (Amor et al. 1995; Salnikov et al. 2003); **(7)** illuminated the cool temperature sensitivity of cellulose biogenesis, the associated block in intra-fiber sucrose synthesis, and the ability of additional sucrose supply to boost fiber filling under stressful conditions (Gipson 1986; Roberts et al. 1992; Martin and Haigler 2004; Haigler et al. 2007); **(8)** revealed a role for an oxidative burst, probably mediated by a small GTPase Rac, in initiating secondary wall deposition (Potikha et al. 1999); **(9)** showed that oxidation can promote *CESA* interaction via their N-terminal zinc-binding domain, and thereby probably affect CSC stability (Kurek et

al. 2002); and (10) showed that, unusually for a membrane protein, CESA proteins have a short half-life in the plasma membrane ( $\leq 30$  min), and that their turnover and degradation are likely mediated through activity of a metallothionein protein and proteolysis (Jacob-Wilk et al. 2006). Other tantalizing clues about cellulose biogenesis obtained from cotton fiber are in need of further exploration, e.g. lipid intermediate that may exist (Peng et al. 2002, although for critique see Schrick et al. 2004). These past results lead to the expectation that further research on cotton fiber will be useful not only for cotton improvement, but also for revealing novel mechanisms controlling secondary wall deposition.

#### 4 The Relationship of Cellulose to Fiber Physical Properties

A survey of the impact of cellulose on cotton fiber utility makes it clear why genomic analysis to identify underlying control mechanisms for cellulose properties is so important for our future potential to improve the usefulness of cotton fiber through breeding complemented with genetic engineering or marker assisted selection. In contrast to being a bulk commodity, future cotton crops may have special fiber and cellulose properties that are directed toward particular high value products. However, achieving this goal requires complete understanding of underlying mechanisms governing all aspects of fiber differentiation.

A key character of technologically useful cotton fiber is an intermediate thickness (maturity) of the secondary wall relative to the fiber diameter. Fiber must be neither immature (see below) nor overly mature because the lumen (former vacuolar space) must remain large enough to allow collapse into an ellipsoid/ribbon-like shape and development of spiral twists (convolutions) that aid spinning into yarn (Hutchinson et al. 1945). A fiber that is over filled with secondary wall remains cylindrical and unconvoluted upon drying (see diagram in Haigler et al. 2005). Modern cotton typically has 3.5 – 6.9 convolutions/mm fiber, with more twists increasing single fiber strength (Hsieh 1999). Convolute fiber is characteristic of and considered diagnostic of human domestication of archaeological cotton in the New World, first to make twined textiles and nets then woven cloth (Stephens 1970). The desirable degree of secondary wall thickening increases fiber mass (yield) along with single fiber strength, dyeing intensity, and water absorption. The effects of increased secondary wall thickness on single fiber strength, Micronaire, and Maturity Ratio are illustrated by a family of transgenic cotton plants over-expressing sucrose phosphate synthase (Haigler et al. 2007).

In addition, molecular properties of cellulose, including degree of polymerization (DP), fibril orientation, percent crystallinity, and crystallite size affect fiber quality characteristics of modern cultivars as summarized briefly here and reviewed thoroughly elsewhere (Lewin and Pierce 1998; Benedict et al. 1999; Hsieh 1999; Bradow and Davidonis 2000). Percent crystallinity, meaning the percentage of cellulose in crystalline vs. disordered state (e.g. on fibril surfaces), has been positively correlated with fiber bundle strength and single fiber strength in some studies (Hindeleh et al. 1980; Hsieh et al. 1997), but not others (Timpa and Ramey 1994). These relationships may be variety dependent, and comparing different studies is especially problematic (Hsieh et al. 1999). Percent crystallinity is also

maximized in correlation with more cellulose fibril bundling in the fiber secondary wall (Willison and Brown 1977) where cellulose can self-assemble without substantial interference from other molecules (Haigler 1991).

Cellulose crystallite size, meaning the average width of single crystals, varies with developmental stage. Cotton primary wall cellulose exists as cellulose IV due to poor order in the lateral dimension, indicating small fibril diameter and crystallite size (Chanzy et al. 1978). Cotton secondary walls contain more orderly cellulose I fibrils with higher crystallite size (3.5 - 5.5 nm) (Ryser 1999; Hu and Hsieh 1996). Crystallite size is mediated by the number of  $\beta$ -1,4 glucan chains synthesized in close proximity and by matrix molecules that can interfere with cellulose self-assembly (Haigler 1991). Cellulose crystallite size is positively correlated with single fiber breaking force (Hsieh et al. 1997), but negatively correlated with water and dye absorption. The latter negative effect arises because the crystalline core of the cellulose nanofibril cannot be penetrated by water or dyes (Rowland and Bertoniere 1985).

The cellulose degree of polymerization (DP<sub>w</sub>, weight averaged) in cotton fiber primary walls was lower (~4000) than in secondary walls (~11,000), with distribution around the mean occurring in each case (Timpa 1991; Timpa and Triplett 1993). The average cellulose DP in solubilized whole fiber and in acetic/nitric acid-resistant crystalline regions were positively correlated with bundle strength (Timpa and Ramey 1989, 1994; Benedict et al. 1994). The DP distribution has its own importance for fiber quality, greatly impacting flexibility and processing potential (Schneider et al. 1996), and it also affects manufacturing of cellulose derivatives (Heinze 1998).

More highly oriented cellulose fibrils increase fiber strength (Yatsu 1983), and a small fibril angle relative to the longitudinal fiber axis correlates with higher elastic modulus (implying less extensibility or extension to break) and strength. Fibril angle is also correlated with resilience, work, recovery, and permanent set (Rebenfield 1990). There is a developmental shift in the orientation of cellulose fibrils relative to the longitudinal fiber axis, with angles of ~70°-90°, 45°-55°, and 20° during the primary wall, transition, and secondary wall stages, respectively (Arthur 1990; Rebenfield 1990). The mechanisms mediating these changes and interrelating cellulose fibrils and microtubules that shift in parallel are subjects of ongoing research (Baskin 2001).

## **5 Analysis of the Cotton Fiber Secondary Wall Transcriptome**

### **5.1 Specialized cDNA Library Production and EST Sequencing**

In order to complement prior sequencing of genes expressed at the primary wall stage of fiber development (Arpat et al. 2004; <http://demeter.bio.bnl.gov/>), a set of *G. hirsutum* fiber secondary wall (G.h.fbr-sw ESTs) was generated at Michigan State University from a suppression subtraction hybridization (SSH) library (Diatchenko et al. 1996). By use of 20 DPA fiber of *G. hirsutum* cv. Deltapine90 as tester and 6 DPA fiber as driver, clones that were up-regulated at 20 DPA compared to 6 DPA

were expected to be enriched, and 9,240 high quality sequences were deposited into GenBank with accession numbers C0490611-C0499850 (Haigler et al. 2005). Both analysis of wall thickness in the fiber used to make the library and subsequent analysis of gene expression patterns (see below) indicated that 20 DPA was early in the primary- to secondary- wall transition period for fiber developing under relatively cool temperature in a Michigan greenhouse (J. Klug, personal communication) or subsequently in a tightly controlled 26/22°C greenhouse at North Carolina State University (U. Avci and C. Haigler, unpublished). During the transition, gene expression required for primary and secondary wall deposition is expected to be decreasing and increasing, respectively.

By reference to a comprehensive assembly of 150,436 cotton ESTs (Udall et al. 2006; [http://agcol.arizona.edu/cgi-bin/pave/adv\\_search.cgi?AssemblyID=Cotton12](http://agcol.arizona.edu/cgi-bin/pave/adv_search.cgi?AssemblyID=Cotton12)), we analyzed the new gene discovery rate arising from three strategies for making cotton fiber cDNA libraries (Table 1). (Note that the G.h.fbr-sw ESTs are called GH\_SCW in the cotton comprehensive gene assembly.)

**Table 1.** Analysis of individual cotton fiber cDNA libraries in relationship to the cotton comprehensive assembly (Udall et al. 2006).

Fiber EST data set	Genotype and fiber stages	Library ESTs in the assembly	Unigenes with library ESTs	Average ESTs/unigene <sup>a</sup>	Rate of gene discovery (%) <sup>b</sup>
GA_Ea,	<i>G. arboreum</i> cv AKA8401; 7-10 DPA fiber	28,080 (of 46,603 in GenBank)	11,816	2.38	42.08
GH_BNL	<i>G. hirsutum</i> cv Acala Maxxa; 5 DPA fiber	7,590 (of 8,910 in GenBank)	4,798	1.58	63.21
G.h.fbr-sw a.k.a. GH_SCW	<i>G. hirsutum</i> cv DP 90; 20 DPA fiber	7,372 (of 9,240 in GenBank)	3,253	2.27	44.13

<sup>a</sup> Calculated as (library ESTs included in the assembly/unigenes containing library ESTs). <sup>b</sup> Calculated inversely as (unigenes containing library ESTs/library ESTs included in the assembly) x 100.

The first generation cotton assembly contained 17,355 tentative unigenes that included ESTs from at least one of the three fiber-specific libraries. The *G. arboreum* GA\_Ea ESTs were generated from a standard cDNA library representing 7 – 10 DPA fiber, but during sequencing two successive rounds of filter colony hybridization were used to remove the most abundant clones already sequenced (Arpat et al. 2004). The *G. hirsutum* GH\_BNL ESTs were generated from a cDNA library normalized by use of reassociation kinetics hybridization (Bonaldo et al. 1996) using RNA isolated from 5 DPA fiber (B. Burr, personal communication and <http://demeter.bio.bnl.gov/>).

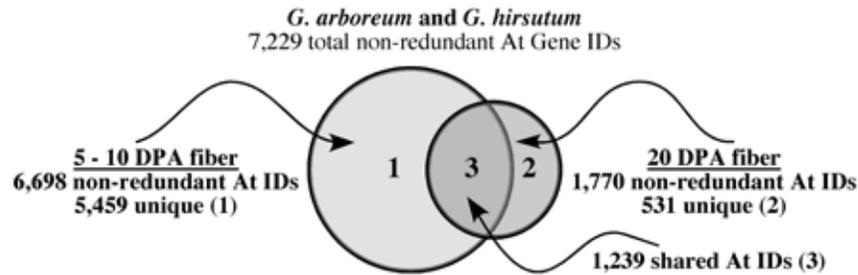
Normalization is essentially a self-subtraction process using the same cDNA population as both tester and driver in order to create similar representation of all transcripts regardless of their original abundance in cells.

As shown in Table 1, the GH\_BNL sequence set was the least redundant, although redundancy would have been inevitably increased through the much deeper GA\_Ea sequencing. In another analysis in which assembly unigenes with 1, 2, or 3 ESTs from any one library were considered together as low redundancy sequences, 75% of the unigenes including GH\_BNL ESTs had low redundancy compared to 52% and 50% of unigenes including GH\_SCW or GA\_Ea ESTs, respectively (data not shown). Therefore, in terms of new gene discovery, sequencing of the GH\_BNL normalized fiber library was most cost effective, as has also been observed in other species (Zhang et al. 2004a). The SSH cDNA library construction method that combined subtraction and normalization did not aim to enhance the rate of gene discovery *per se*; rather it helped to enrich differentially expressed genes within the G.h.fbr-sw/GH\_SCW ESTs. These outcomes were confirmed by comparison to the Arabidopsis proteome (see below).

## 5.2 Cotton Fiber Gene Expression Compared to Arabidopsis

The proteome of the model plant *Arabidopsis thaliana* was used as a reference point for analyzing the changing complexity of gene expression during fiber development. Unigenes from the GA\_Ea, GH\_BNL, and GH\_SCW libraries, which had been subjected to uniform length and quality editing, were downloaded from the cotton comprehensive gene assembly (numbers in Table 1). Then each unigene set was compared by blastx to the Arabidopsis proteome with an E-10 cut-off value as used previously to indicate significantly similar plant sequences (Gutierrez et al. 2004). The number of unique *Arabidopsis thaliana* gene identifiers (At Gene IDs, found at TAIR, <http://www.arabidopsis.org>) matched by each unigene set were: (a) 5,682 (48.09% of unigenes) for the GA\_Ea library; (b) 3,149 (65.63% of unigenes) for the GH\_BNL library; and (c) 1,770 (54.44% of unigenes) for the GH\_SCW library. The inclusion of both A genome *G. arboreum* and AD genome *G. hirsutum* in such a conceptual analysis is justified by the relatively recent polyploidization event and by the fact that both species produce commercially useful fiber that develops similarly. Note that a large set of *G. hirsutum* sequences (11,692 unigenes) recently deposited into GenBank derived from ovules with attached fiber (see p. 654 in Shi et al. 2006), so these sequences were not included in the results shown here. However, when this sequence set was included in the same type of analysis of *G. hirsutum* sequences only, it does not change the conclusion regarding unique gene expression at the secondary wall stage of fiber development (see below).

Comparison of cotton fibers ESTs to the Arabidopsis proteome showed that a single cell is not a simple cell in terms of its set of expressed genes. In the cotton comprehensive gene assembly, unigenes with contributions from GA\_Ea, GH\_BNL, and G.h.fbr-sw/GH\_SCW fiber libraries were classified as including ESTs from the: (1) primary wall stage; (2) secondary wall stage; or (3) both stages of fiber development. Each group of translated fiber-associated unigenes was then compared by blastx to the Arabidopsis proteome (E-10 cut-off value), and each list was made

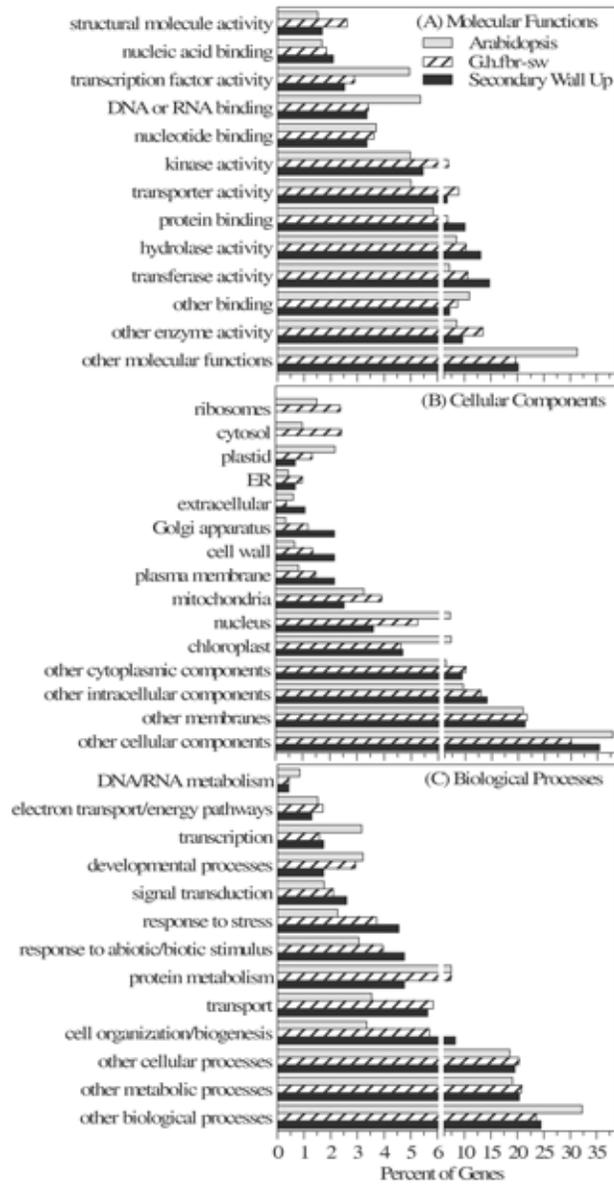


**Fig. 1.** Venn Diagram illustrating genetic complexity during primary and secondary wall deposition in cotton fiber development in reference to homology ( $\leq E-10$ ) of translated cotton unigenes to the Arabidopsis proteome. The 5 – 10 DPA group was based on cotton unigenes with contributions from GA\_Ea and/or GH\_BNL sequences, whereas the 20 DPA group derived from only GH\_SCW unigenes.

non-redundant in terms of At Gene IDs recognized. The outcomes were used to generate a Venn diagram illustrating current knowledge about fiber-specific gene expression complexity (Fig. 1). The 7,229 total non-redundant At Gene IDs represent 36.4% of 19,867 cotton fiber-associated unigenes analyzed. Interestingly, in Group 1 of primary wall genes, 1,016 of 6,698 At Gene IDs arose from the GH\_BNL ESTs, which shows that normalization-based sequencing from *G. hirsutum* was useful in complementing the large scale sequencing effort from *G. arboreum*.

Of 6,698 At Gene IDs matched by primary wall sequences and 1,770 matched by secondary wall sequences, 1,239 At Gene IDs were shared between the stages of development. At the onset of secondary wall deposition, cotton genes were expressed that were homologous to 531 unique At Gene IDs, representing 7.3% of the total. This reflects a minimum estimate of gene expression for secondary wall deposition because only 19.6% of the unigenes in the three fiber libraries arose from sequencing at the secondary wall stage of fiber development. Remembering that cotton genes without close Arabidopsis homologues are not represented in this analysis, these results demonstrate that genes equivalent to at least 25 – 30% of the proteome of the whole Arabidopsis plant are expressed during cotton fiber development.

Previously, G.h.fbr-sw sequences were assigned to MIPS major categories of protein function by analogy to homologous Arabidopsis proteins. Highly represented categories were consistent with control of cellular organization and transport processes, as well as metabolic processes, in transition stage cotton fiber. Cotton sequences related to these functions were also over-represented compared to the Arabidopsis proteome, whereas transcriptional functions were under-represented (Haigler et al. 2005). A similar analysis was done more recently for GO functional categories, along with comparison of categories for cotton genes up-regulated on the microarray to be discussed later (Fig. 2). In the Molecular Function category, high enzymatic activity as well as transport activity was indicated for transition-stage cotton fiber (hatched bars), whereas transcriptional activity was implicated to a lesser extent than for the entire Arabidopsis proteome (gray bars). The latter difference is consistent with the more restricted set of developmental changes occurring at one



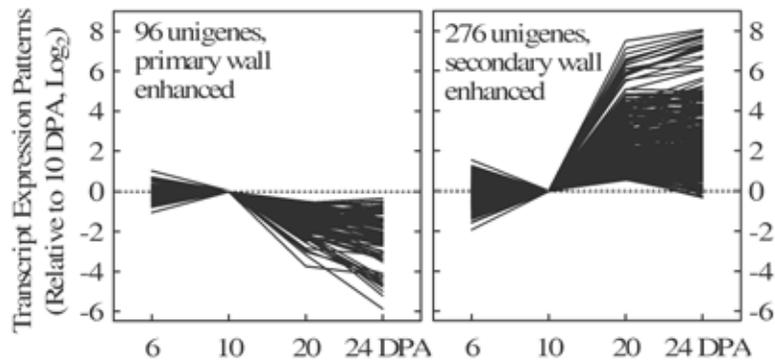
**Fig. 2.** GO Annotation Groups (A, B, C) for the G.h.fbr-sw sequence set (striped), for cotton genes up-regulated at 20 DPA on the microarray (black), and for the Arabidopsis proteome (grey). The scale is expanded for 0 – 6%. Cotton genes were grouped in analogy with their homologous Arabidopsis gene ( $\leq E-10$ ; <http://www.arabidopsis.org/tools/bulk/go/index.jsp>). Percentages are based on the entire Arabidopsis proteome or 1770 or 181 unique At Gene IDs corresponding to G.h.fbr-sw sequences or its subset that was up-regulated at 20 DPA, respectively.

stage in a single cell compared to a whole plant. However, regulatory activity via kinases and, potentially, protein binding is well represented within the G.h.fbr-sw sequences. Other analyses show that substantial numbers of the hydrolases and transferase are related to sugar- and/or cell wall metabolism (data now shown), which is consistent with dynamic changes in the fiber cell wall occurring at the transition stage. In the Cellular Components category, most G.h.fbr-sw sequences were in the “other” categories just as in Arabidopsis. However, compared to Arabidopsis, there was an enhancement of the percentage of G.h.fbr-sw sequences in the categories of ER, Golgi, cell wall and plasma membrane, which makes sense given the main activities of the secondary-wall-stage cotton fiber. In the Biological Processes category, G.h.fbr-sw sequences were again highly represented in the “other” categories, but there was over-representation compared to Arabidopsis in the categories of transport and cell organization and biogenesis.

### 5.3 A Discrete Transcriptome for Secondary Wall Deposition

Microarray analysis was done by Curtis Wilkerson and Jeff Landgraf at the Research Technology Support Facility, Michigan State University. Spotted onto slides in duplicate were 3,185 cDNA clones after PCR amplification; these represented 93.1% of 3,420 internally-assembled unigenes (Note that 167 fewer GH\_SCW unigenes were included in the cotton comprehensive gene assembly because of preliminary uniform editing of sequence length and quality between all sequence sets included.) RNA used to make probes was isolated from fiber at 6 and 10 DPA (elongation via primary wall deposition), 20 DPA (transition to secondary wall deposition under relatively cool conditions), and 24 DPA (onset of high rate secondary wall cellulose biogenesis). The experiment was designed to reveal genes that were up-regulated for secondary wall deposition. To exemplify the sensitivity of the array, for clones up-regulated at 20 vs.10 DPA, the minimum significant change ( $p \leq 0.050$  on combined duplicates of each cDNA spotted onto each slide) was  $\log_2 = 0.5$ , or 1.4-fold, but 84.8% of significant changes had magnitudes  $\geq \log_2 = 1$ , or 2-fold.

General trends in the microarray data are further described here in terms of clones that were significantly up-regulated at 20 DPA vs. 6 DPA and 10 DPA ( $p \leq 0.050$  on combined duplicates). This is a stringent requirement to enhance the chances that real changes relative to onset of secondary wall deposition are discussed. There were 276 clones that were up-regulated at the transition to secondary wall synthesis vs. 96 with enhanced expression during primary wall synthesis (Fig. 3). The existence of 2.9 times more clones in the up-regulated category reflects the initial bias of the spotted cDNAs toward the secondary wall stage of fiber development. Graphing the expression patterns of genes with changes at 20 DPA shows, as expected, that the discrete developmental stages of primary and secondary wall deposition in cotton fiber are underpinned by substantially different programs of gene transcription. The 276 up-regulated clones, when translated *in silico* and compared to the Arabidopsis proteome by BLAST, were homologous ( $E \leq 10^{-10}$ ) to 181 unique Arabidopsis proteins. (Accounting for the total of 276 up-regulated clones, 47 did not match Arabidopsis sequences at this cut-off level and 48 matched



**Fig. 3.** Graphical summary of microarray results for genes with significantly different expression at 20 DPA compared to 6 and 10 DPA ( $p \leq 0.05$ ). The expression pattern of each gene is shown relative to its own level at 10 DPA. The graphs show that developmental shift between primary and secondary wall deposition (around 20 DPA in this experiment) has a strong transcriptional basis. Rigorous methods were used to generate these results, including two copies of each cDNA spotted on each slide and 3 biological replications and two technical replications (dye swaps) for each data point, followed by quality control then statistical analysis of the results (see also R Development Core Team, 2005; Smyth 2004, 2005; Smyth and Speed 2003; Smyth et al. 2005). Briefly, RNA was labeled with the amino-allyl procedure (<http://www.tigr.org/tdb/microarray/protocolsTIGR.shtml>). Hybridization was performed using SlideHyb #1 buffer (Ambion). Slides were scanned using an Affymetrix 428 ArrayScanner and analyzed using the GenePix Pro 3.0 software (Axon). Array normalization and statistical analysis were performed using the “limma: Linear Models for Microarray Data” library module (version 2.2.0) of the R statistical package (version 2.2.0). Slide intensity data were normalized using the global loess method. The least squares method was used for the linear model fit utilizing the Benjamini and Hochberg method to control the false discovery rate.

23 At Gene IDs redundantly). Of the 181 unique Arabidopsis homologs of the up-regulated genes, only 23 were held in common with genes up-regulated at 24 DPA in *G. arboreum* within the GA\_Ea sequences derived from the primary wall stage of fiber development (Arpat et al. 2004). The GO categories of the up-regulated G.h.fbr-sw genes were analyzed according to GO annotations of the Arabidopsis homologs (Fig. 2, black bars). In the Molecular Functions group, categories that were over-represented compared to the entire G.h.fbr-sw sequence set were protein-binding, hydrolase, and transferase. In the Cellular Components group, over-represented categories included Golgi apparatus, cell wall, and plasma membrane. In the Biological Processes group, genes up-regulated at the onset of secondary wall deposition were mainly proportional to their overall representation in the G.h.fbr-sw sequences. These trends are consistent with emphasis on cell wall synthesis and remodeling during the primary-to-secondary wall transition in cotton fiber development.

It was important to validate microarray data by real-time reverse-transcription PCR (qPCR). To establish a basis for this between 6 – 30 DPA in cotton fiber, three possible endogenous controls, *EIF-5*, *Tua-4*, and *UBQ*, were tested (Fig. 4). The most variability of expression was observed for *UBQ*, and its  $\Delta R_n$  vs. Cycle (log graph, where  $\Delta R_n$  is the intensity of dye fluorescence as a function of cycle number) showed amplification before 10 cycles (data not shown). The difference between the controls is clarified by examining the expression of a cotton chitinase-like gene, *GhCTL1*, normalized to each one. *GhCTL1* was used for this purpose because of background data showing that it was secondary-wall-specific in several types of cotton cells, including fiber (Zhang et al. 2004b). Normalization to *UBQ*, with its early cycle amplification, often resulted in lower values for *GhCTL1* expression compared to the other two controls. The expression of *Tua-4* dropped at 30 DPA, so, when it was used as a normalizer, the apparent expression of *GhCTL1* increased compared to 24 DPA. On a linear scale, the enhancement in *GhCTL1* expression relative to 6 DPA was 4,493-fold at 24 DPA and 4,299-fold at 30 DPA when normalized to *EIF-5*. The corresponding values after normalization to *Tua-4* were 3,584-fold at 24 DPA and 6,916-fold at 30 DPA. Overall, normalization with *EIF-5* resulted in a *GhCTL1* expression pattern that was more consistent with results of promoter-GUS assay and Northern blotting; the latter showed approximately equal expression for *GhCTL1* during the later stages of secondary wall deposition (Zhang et al. 2004b).

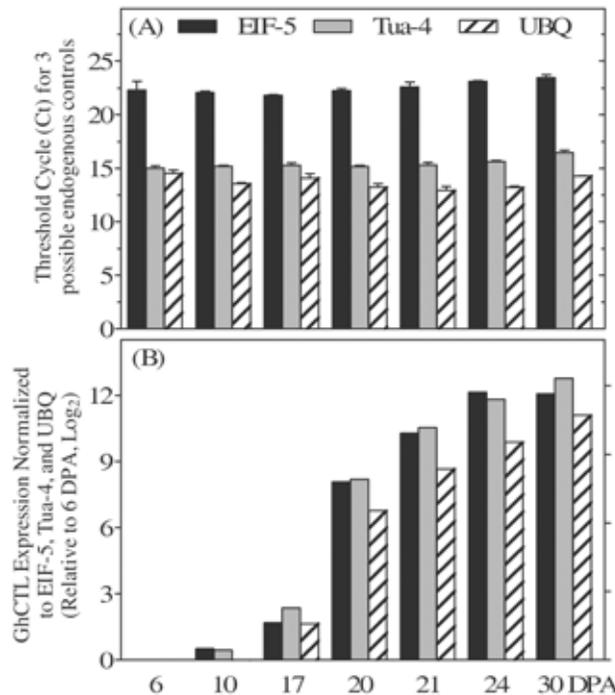
To verify the expression patterns for selected cotton sequences showing changes on the array, qPCR was performed (Table 2 and Fig. 5). Information from *Arabidopsis* was used to aid primer design since the cotton genome is not completely sequenced. This strategy was justified in light of the close relationship between the two species: *Arabidopsis* (Brassicales) and *Gossypium* (Malvales) are constituents of Rosids II, a large subclade within Rosids in the plant phylogenetic tree (Soltis et al. 2005). Cotton contigs/ESTs were analyzed via blastx to identify the set of closest *Arabidopsis* protein homologs. In order to minimize the chances of amplification of multiple related cotton genes by avoiding highly conserved domains, primers were designed within regions of the cotton sequence that showed divergence from the most closely related *Arabidopsis* sequences. Primer Express 3.0 software (Applied Biosystems) was used to design forward and reverse primers for an amplicon size between 50-150 base pairs. Primers were analyzed by BLAST to show that they only matched one known cotton unigene (or sometimes two closely related sequences).

Fig. 5 shows a comparison of gene expression profiles obtained by qPCR (histograms) and microarray analysis (in bold type over each histogram bar for the same DPA). Importantly, for several up- and down-regulated genes, the qPCR data demonstrate that 20 DPA was transitional between the gene expression levels at 10 DPA (high-rate elongation) vs. 24 DPA (high-rate secondary wall deposition). On 20 DPA, expression of genes putatively related to auxin response, which is most important during fiber elongation, and primary cell wall flexibility via expansin action were declining, whereas expression of a secondary-wall-specific isoform of Cesa was increasing (see also Guo et al. 2007). Corresponding to its higher dynamic range (Busch and Lohmann 2007), qPCR often revealed greater magnitudes of up/down-regulation than microarray analysis. In a large percentage of cases, the

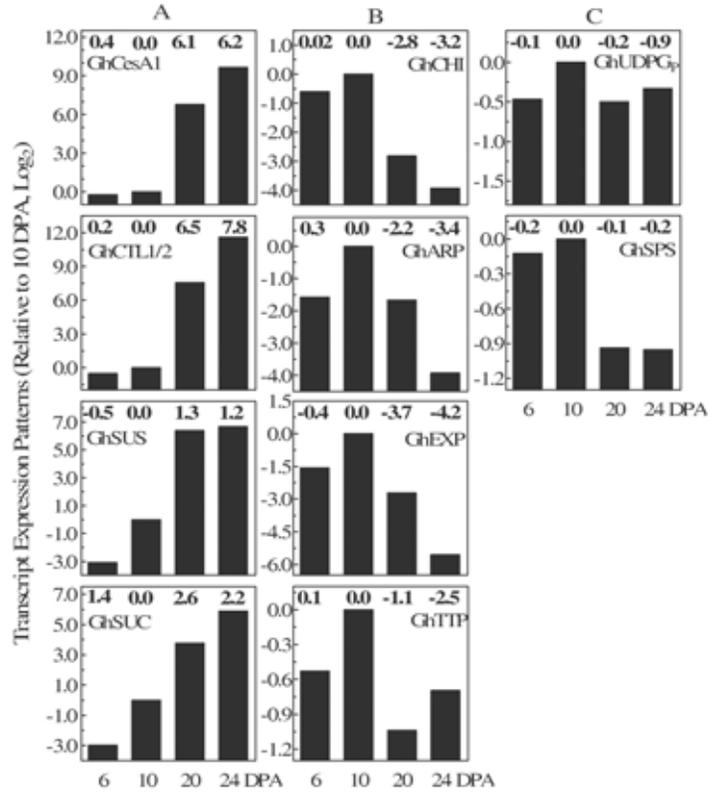
same pattern of gene expression was revealed by the two techniques. However, GhSPS showed apparently constant expression on the microarray, whereas the isoform assayed by qPCR was down-regulated at the onset of secondary wall deposition. Spotted cDNAs may hybridize with multiple closely related genes (Busch and Lohmann 2007), whereas, in general, primers for qPCR are more discriminatory. However, this cannot be fully evaluated for any organism with an unsequenced genome.

**Table 2.** Cotton cDNAs for which microarray gene expression patterns were checked by qPCR. E-values are for the cDNA spotted on the array, and longer assembled cotton unigenes with better homology to the Arabidopsis protein may exist.

GenBank Accession Number	Arabidopsis homolog	E- value	Abbreviated Name/ Annotation
Up-regulated on the microarray			
CO493183	At3g16920	1E-39	GhCTL1/2; chitinase-like,
CO498265	At4g18780, CesA8	6E-94	GhCesA1; cellulose synthase
CO497506	At3g43190 SUS4	1E-111	GhSUS; sucrose synthase
CO491535	At1g71880 SUC1	1E-10	GhSUC; sucrose transporter
Down-regulated on the microarray			
CO495014	At5g05270	1E-60	GhCHI; chalcone-flavanone- isomerase-like
CO496859	At2g39700 EXPA4	3E-57	GhEXP; expansin
CO492432	At1g19840	2E-12	GhARP; auxin-responsive-SAUR- protein-like
CO499674	At1g26690	1E-34	GhTTP; transmembrane trafficking protein-like
Unchanged on the microarray			
CO492429	At5g20280 SPS1F	3E-54	GhSPS; sucrose-phosphate synthase
CO498005	At5g17310	2E-49	GhUDPG <sub>p</sub> ; UDP-glucose pyrophosphorylase



**Fig. 4.** Comparison of endogenous controls for qPCR analysis of cotton fiber gene expression between 6 – 30 DPA. (A) Expression pattern of three potential endogenous controls, EIF-5 (eukaryotic translation initiation factor-5, Accession: CO492947), GhTua-4 (alpha-tubulin-4, Accession: AY345605), and GhUBQ, (ubiquitin, Accession: AY189972). (B) GhCTL expression normalized to each control. Forward (F) and Reverse (R) primer pairs (5' to 3') were: (a) for EIF-5, (F) GGTTGCCATTGTGCAAGGA, (R) CCGTAGGTGAGCGTTAATCAGA; (b) for GhTua-4, (F) GGCGCTGAGTCTGGTGAAG (R) CAGTAGCAAACGGATAACGGTACA; and (c) for GhUBQ, (F) AAGCTCGCAGTGCTCCAGTT (R) GAACTCTTTCTCAGCCTCTGA. Total RNA was isolated from  $\leq 100$  mg cotton fiber ground under liquid nitrogen (Spectrum Plant Total RNA Kit, Sigma), purified (On Column DNase Digestion Kit, Sigma), and reverse-transcribed ( $\sim 800$  ng/40  $\mu$ l TaqMan Reagents, Applied Biosystems). The 25  $\mu$ l PCR reactions (with 12.5  $\mu$ l SYBR green, 150 nM (F) and (R) primers, 10 ng cDNA) were run [50°C, 2 min; 95°C, 10 min; followed by 40 cycles of (95°C, 20 sec and 60°C, 1 min)] on a Real Time PCR Model 7300 (Applied Biosystems). Dissociation curves were analyzed to verify the absence of non-specific amplification and primer-dimer formation. SDS software was used for automatic analysis of threshold cycle (Ct) values. Each data point derived from three biological samples, each containing fiber from 2 – 10 bolls. For EIF5 and Tua4, the entire assay was repeated twice starting from RNA isolation, and similar results were observed in other experiments for UBQ. Bars = SE values.

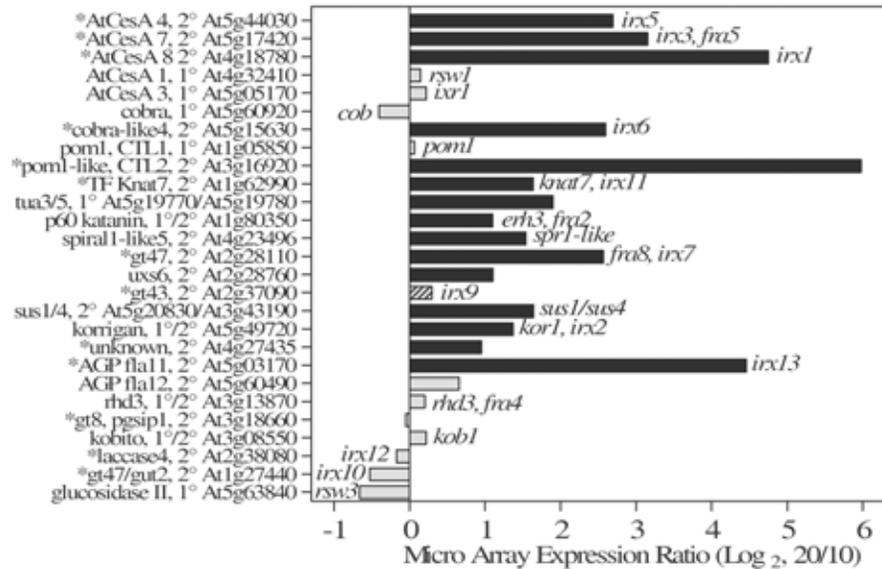


**Fig. 5.** Comparison between gene expression patterns relative to 10 DPA derived from microarray and qPCR analyses. Columns A, B, and C show genes that were up- or down-regulated or unchanged on the microarray, respectively, as shown by  $\log_2$  values in bold above each corresponding ratio in the qPCR histogram. qPCR data (from three biological replicates) were normalized to EIF-5. See Table 2 for accession numbers and Fig. 4 for basic methods. Forward (F) and Reverse (R) primer pairs (5' – 3') were: (a) for *GhCTLI/2*, (F) CCGACCAAGAACGACACGTT, (R) ACCTCGCCACAAACTTGAT; (b) for *GhCesA1*, (F) GCAGCAGACGATACAG AATTTCG, (R) CGTTGTTGATTGCGTCTGAAAC; (c) for *GhSUS*, (F) ACATGG CAGATTTACTCCGAGAA, (R) TCCATGCATGCAACATTTC; (d) for *GhSUC*, (F) AGCCATTCCGGTTGAGTTGA, (R) CCAGCCTTTGTATTGGGGTCTA; (e) for *GhCHI*, (F) AGCCTGGGAACGTGCTAGCT, (R) TTGGGAACCTTTGATCT CCTTACT; (f) for *GhEXP*, (F) TTTGACCAAACTGCCAATGG, (R) CAGAACT GGCAGTCAAATGCA; (g) for *GhARP*, (F) GCCAGCAACAACACCATCAA, (R) TGAGACTTGTGCCCACACAGA; (h) for *GhTTP*, (F) CCCAAAAGCAAGCCA TGTA, (R) TTCATATGGGAATAATTTTCACTCG; (i) for *GhSPS*, (F) GTCTCATGTTCTTGGTGAGCAAAT, (R) CACCAGATAGAAGAGCA GCAGAGT; (j) for *GhUDPG<sub>p</sub>*, (F) CAACTTTTCTCCTGGTTTTGCA, (R) TCAAGACAA-TCCCAAGCATCAT.

## 6 Cotton Fiber Secondary Wall Transcriptome vs. Xylem

We aimed to allow functional information derived in *Arabidopsis* to illuminate the possible significance of homologous gene expression in cotton fiber. The most abundant secondary walls in plant vegetative tissue occur in xylem, and xylem secondary walls in dicotyledonous plants contain 40 – 50% cellulose as well as substantial amounts of xylan and lignin. As supported by the gene expression data described below, secondary wall deposition in cotton fiber and in water-conducting xylem cells shares common elements. This similarity is reasonable since formation of xylem secondary walls in vascular tissue was a foundational event for the diversification of land plants (Bowman et al. 2007). Figure 6 shows a fusion of *Arabidopsis* gene names and mutant information with gene expression data (histogram bars) in isolated fiber for the cotton homolog of the *Arabidopsis* gene (see legend for further details). The homology relationships, and possible orthology, were confirmed by reciprocal BLAST of the longest cotton unigene vs. the *Arabidopsis* proteome and the identified *Arabidopsis* protein vs. the cotton comprehensive gene assembly.

Included in Fig. 6 are many genes known to be required for cell wall synthesis, many identified through mutant analysis and a few through bioinformatics alone. Mutant series that appear more than once in Fig. 6 are: (a) *irx*, or irregular/collapsing xylem resulting from weak secondary walls in water-conducting cells; (b) *fra*, derived from the fragile inflorescence stems of mutants with thin secondary walls in interfascicular fibers; and (c) *rsw*, derived from a root radial swelling phenotype. The *irx* and *fra* phenotypes have been particularly useful in revealing genes required for secondary wall deposition. Note that the *ixr1* mutant name is different from *irx*, denoting resistance to the cellulose-synthesis inhibitor isoxaben. Three bioinformatics studies also formed the basis for Fig. 6. One comparative transcriptome analysis focused on identifying a “core xylem gene set” already highlighted xylem-expressed genes with homologs in cotton fiber (Ko et al. 2006). Some of these genes that had G.h.fbr-sw homologs spotted on the array and about which we have some functional information are included in Fig. 6. Two analyses of genes co-expressed in *Arabidopsis* with secondary wall CESA genes identified a total of 52 unique sequences (Brown et al. 2005; Persson et al. 2005). These studies were recently synthesized to show a set of 19 commonly identified secondary-wall-associated genes, along with any collapsing xylem (*irx*) phenotypes observed and/or cellulose depletion documented by other means (Table 1 in Joshi and Mansfield 2007). ESTs representing cotton homologs of 13 of these 19 genes were sequenced from 20 DPA cotton fiber and included on the microarray (see gene names with asterisks in Fig. 6). (An EST from a 14<sup>th</sup> cotton homolog, for the unknown protein At1g09610, was in the G.h.fbr-sw sequences but did not amplify for the spotted cDNA array. The other 5 genes among the 19 genes dually found to be co-regulated with *Arabidopsis* secondary wall CESA genes were not the closest homolog of a known cotton fiber gene, data not shown.)



**Fig. 6.** Microarray results (20/10 DPA expression ratio) for cotton homologs of Arabidopsis cell-wall-associated genes. Black bars indicate significant changes ( $p \leq 0.05$ ), and the striped bar indicates that other data show up-regulation of this cotton gene in transition-stage fiber (Wu and Liu 2005). Arabidopsis gene names and At Gene IDs are shown on the y-axis, and names for corresponding Arabidopsis mutants are italicized beside the histogram bars. The designations 1° and/or 2° indicate whether the genes have been associated with primary and/or secondary wall synthesis in Arabidopsis (and/or other species) through bioinformatics, analysis of mutant phenotypes, or both. Regarding gene names, two Arabidopsis homologs are shown for tubulin and sucrose synthase because each of these pairs represents closely related isoforms with equal homology to the cotton cDNA spotted on the array.

## 7 Function of Genes Related to Secondary Wall Deposition

Twenty-one of the genes in Fig. 6 were associated with secondary wall deposition in xylem (see 2° notation), and 15 of the cotton fiber homologs showed up-regulation at the onset of secondary wall deposition on the spotted cDNA microarray. The functional context of these genes is briefly summarized here. For findings from species other than cotton, often only one recent publication is cited to provide an entry point into the relevant literature. Information from Arabidopsis is emphasized, but many parallels also exist between gene expression for cotton fiber secondary wall deposition and formation of woody stem tissue in trees and grain stalks (e.g. see Hertzberg et al. 2001). Acknowledging the ultimate need for confirmation of each gene expression pattern by qPCR, we will omit the word “putatively” in reference to gene expression patterns discussed further here. Also, some genes may have failed to show significant up-regulation on the spotted cDNA microarray, whereas changes would be detected by a gene-specific oligomer arrays or qPCR. Finally, increased

gene expression does not always correlate with translation or increased protein activity, but other evidence supports the similarity of secondary wall deposition in cotton fiber and xylem. In addition to similar results for cotton (Zhang et al. 2004b), when the promoter of secondary-wall-expressed *GhCESA4* was used in a reporter gene construct in transgenic Arabidopsis, GUS activity was observed in a variety of Arabidopsis cells with thickened walls, including the xylem (Kim and Triplett 2001).

### 7.1 Specific Isoforms Within Families Are Up-regulated

First, we describe three cases illustrating isoform-specific involvement within one protein family in secondary wall deposition in cotton fiber, as also occurs in xylem of other species. The *CESA* genes encode proteins included in the CSC that mediates the biogenesis of crystalline cellulose nanofibrils. These Family 2 processive glycosyltransferases can add UDP-glc to continue the elongation of a  $\beta$ -1,4-glucan chain. Arabidopsis has 10 related genes (*AtCESA1-10*), and three genes have non-redundant roles for primary wall synthesis (*AtCESA1*, -3, -6) or for secondary wall synthesis (*AtCESA4*, -7, -8). In addition, *AtCESA2* is required for normal cellulose synthesis in primary walls of some aerial Arabidopsis organs (Chu et al. 2007). Now, CESAs from many plant species have been shown to group into clades of “primary wall type” and “secondary wall type” proteins. Increasing evidence supports the existence of a multimeric CSC in which particular CESA isoforms occupy non-interchangeable sites (Taylor et al. 2003; Wang et al. 2006). In the fiber microarray data, the cotton homologs of *AtCESA4*, -7, -8 were strongly up-regulated at the transition to secondary wall deposition on 20 DPA. The expression of *GhCESA1* (Accession U58283), the cotton homolog of *AtCESA8*, was shown previously to be strongly up-regulated at the onset of fiber secondary wall synthesis (Pear et al. 1996) and the encoded protein also appeared at that time (Peng et al. 2002). In contrast, the cotton homologs of at least two Arabidopsis primary-wall-related CESAs, *AtCESA1* and *AtCESA3*, were still being transcribed at 20 DPA, but they were not up-regulated since fiber elongation via primary wall synthesis was slowing or stopped. Supporting these results, a cotton gene with 98% nucleotide identity to the Gh.fbr.sw EST that was homologous to *AtCESA3* was previously sequenced from 24 DPA fiber (*GhCESA3*, Accession AF150630). However, preliminary evidence suggested that its expression peaked during fiber elongation (Laosinchai 2002). Probably arising from cotton allotetraploidy and/or emphasis on cellulose synthesis in cotton fiber, *GhCESA4* (AF413210), a homolog of *GhCESA1* with 96% nucleotide identity to it, was reported to have secondary-wall specific expression in fiber (Kim et al. 2002). However, another *GhCESA1*-type sequence (AF139448, called at first *CelA2*) shows evidence of pseudogenization that originated in the ancestral diploid A-genome and persisted in allotetraploid *G. hirsutum* (Cronn et al. 1999). In analogy with other crop plants (reviewed in Farrokhi et al. 2006), *Gossypium* is likely to contain numerous distinct Cesa genes, with changes with possible functional consequences having occurred within species.

A second case of isoform-specific up-regulation for secondary wall deposition occurs within the *COBRA* gene family. COBRA proteins are of unknown function, but they localize to the external face of the plasma membrane and may function

coordinately with microtubules as well as binding to a cell wall polysaccharide. Loss of COBRA was associated with abnormal anisotropic cell expansion, cellulose deficiency, and poorly oriented cellulose fibrils in Arabidopsis roots. The failure of its cotton fiber homolog to up-regulate for secondary wall deposition can be explained as for *AtCESA1* and *AtCESA3*. In contrast, the cotton homolog of Arabidopsis *COBRA-LIKE4* was up-regulated for fiber secondary wall deposition, which is consistent with orthologous genes being required to achieve the normal degree of xylem secondary wall deposition in rice, corn, and Arabidopsis (Brady et al. 2007). A third case occurs within a family of chitinase-like genes. Mutation in *POM1/AtCTL1* was associated with defects in root development, incomplete primary walls, and pleiotropic stress phenotypes (Zhong et al. 2002) as well as cellulose deficiency in Arabidopsis (Piling and Hofte 2003). Subsequently, *AtCTL2* was revealed by bioinformatics to be associated with expression of Arabidopsis secondary-wall-*CESA* genes, and this gene is the closest match to cotton genes with secondary-wall-enhanced expression as shown by Northern blotting and promoter:GUS assay (Zhang et al. 2004b), as well as by microarray and qPCR analysis (Fig. 5, 6). The expression patterns of two cotton *CTL*-type genes on the array correspond to the primary- or secondary-wall roles of the two (apparent) Arabidopsis orthologs.

## 7.2 Transcriptional Control

The discussion now continues approximately in order of how up-regulated genes are likely to function in the transition to cotton fiber secondary wall deposition. Clearly there are transcription factors that mediate the large-scale gene transcriptional changes that occur during this developmental shift. One of these is homologous to the Arabidopsis homeodomain transcription factor, *KNAT7* (knotted1-like7). Typically, KNAT transcription factors regulate development via networks of protein interactions (Truernit et al. 2006). In Arabidopsis xylem, this transcription factor acts downstream of NAC domain transcription factors that are master regulators of the whole process of xylem secondary wall deposition, including cellulose, lignin, and xylan synthesis (Zhong et al. 2007). In cotton fiber, lignin synthesis does not occur, so it is tempting to speculate that *KNAT7* homologs have a specific role in stimulating secondary wall cellulose biogenesis. However, we cannot disallow a possible role in xylan synthesis at this time (see below).

## 7.3 Microtubule-related Genes

Among the first visible changes at the transition to secondary wall deposition in cotton fiber are changes in the angle of microtubules and cellulose fibrils. As secondary wall synthesis progresses, microtubules become more numerous, longer, closer to the plasma membrane, and oriented in a more steeply pitched helix that is mirrored by cellulose nanofibril orientation (Seagull 1993). The expression of tubulin was also up-regulated for secondary wall synthesis (Fig. 6), which correlates with increasing microtubule density, and, possibly, functional differences between tubulin isoforms. Also related to microtubule function, the expression is up-regulated of the

cotton homolog of the p60 subunit of *KATANIN*, which encodes a widely-expressed and broadly-important AAA-ATPase with microtubule-severing capacity. Among diverse phenotypes in plants, some of its mutant alleles decreased the quantity and organization of cellulose in secondary walls of Arabidopsis interfascicular fibers (Burk and Ye 2002). Microtubule severing likely aids in reconfiguring the microtubule array to support the deposition of more cellulose in a different orientation in secondary walls compared to primary walls. Probably for similar reasons, a cotton homolog of *SPIRALI-LIKE5* was up-regulated at 20 DPA. *SPIRAL1* (*SPR1*) has unknown biochemical function, but it localizes to microtubules and mediates directional growth of rapidly expanding cells in diverse plant organs. All *SPIRALI-LIKE* genes are at least partly functionally redundant with *SPIRAL1* in Arabidopsis, but double mutants lead to randomly oriented cortical microtubule arrays (Nakajima et al. 2006). Clearly, orderly microtubule arrays are important for cotton secondary wall deposition.

#### 7.4 Synthesis of Non-cellulosic Polysaccharides

The cotton homolog of At2g28110 (*fra8, irx7*), in glycosyltransferase family 47, was up-regulated on the array. Although mutations caused both cellulose and xylan deficiency in Arabidopsis xylem, this enzyme participates in the synthesis of a special 5-sugar terminator on the reducing end of glucuronoxylan (Pena et al. 2007). In addition, the cotton homolog of *UXS6*, a UDP-D-glucuronate decarboxylase yielding UDP-D-xylose, was up-regulated on the array. This nucleotide sugar can provide xylose for biosynthesis of xyloglucan and arabinoxylan, as well as undergo epimerization to UDP-L-arabinose for arabinoxylan and pectin biosynthesis. Previously some data implicated an increase in xylose during cotton fiber secondary wall formation (reviewed in Ryser 1985), but, at most, only minor amounts of any typical cell wall matrix molecules exist in this thick wall. However, the transitional “winding” layer, which is similar to the S1 layer in a wood fiber, may contain molecules not yet characterized in detail in cotton. The possible existence of a xylan-type molecule requires explanation of why the cotton homolog of At2g37090 (*irx9*) was not shown to be significantly up-regulated on the array. This gene encodes a Family 43 beta-glucuronyltransferase involved in glucuronoxylan backbone elongation (Pena et al. 2007). Microarray results are never completely accurate, and it is also possible that 20 DPA was too late to capture the up-regulation of the cotton homolog of At2g37090 as shown by other results. The G.h.fbr-sw cDNA spotted on the array is contained within *GhGlcAT1* (*G. hirsutum* glucuronic acid transferase 1, AY346330), which was shown by Northern blotting to have a sharp peak of expression in 15 DPA cotton fiber as well as being expressed in stems that contain xylem (Wu and Liu 2005). The histogram bar for the cotton At2g37090 homolog was striped in Fig. 6 to indicate these prior results. Therefore, cotton fiber does express genes closely related to those required for the synthesis of glucuronoxylan in xylem cells. In this case, analysis of gene expression drives the need to look in detail for cellular mechanisms not yet fully characterized in cotton fiber development.

## 7.5 Supply of Carbon to Cellulose Synthetic Enzymes

Sink strength strongly increases for secondary wall deposition, and the expression of sucrose synthase (*SUS* (called previously *SuSy*; E.C. 2.4.1.13;  $\text{suc} + \text{UDP} \leftrightarrow \text{UDP-glc} + \text{fru}$ ) increases to support the supply of more UDP-glc to CESA enzymes as shown previously by Northern blotting (Amor et al. 1995) and also in Fig. 5, 6. Increased *SUS* gene expression is correlated with increased *SUS* protein near the cortical microtubules and plasma membrane (the site of CSCs) as cotton fiber secondary wall deposition commences, as well as in the exoplasmic space where callose ( $\beta$ -1,3-glucan) is deposited and cellulose fibrils crystallize (Salnikov et al. 2003). Recently, maize *SUS1* was shown to have intrinsic membrane-binding domains, with membrane association promoted by high *SUS1* protein and sucrose concentration and lower pH (Hardin et al. 2006). In *Arabidopsis*, genetic analyses provided no evidence that a particular *SUS* isoform was specifically associated with cellulose synthesis; even a *sus1sus4* double mutant grew normally except when flooded (Bieniawska et al. 2007). However, regulation of the intracellular behavior of *SUS* transcripts or protein may modulate cellulose synthesis under normal conditions in all plants (Haigler et al. 2001). Furthermore, crop plants that have been domesticated to develop large storage organs have shown detrimental effects of down-regulation of particular *SUS* isoforms (Bieniawska et al. 2007). The proven role of *SUS* in cotton fiber initiation and elongation based on anti-sense suppression of cotton *SUS* (U73588) has been reviewed recently, including preliminary evidence that transgenic suppression of *SUS* reduces fiber secondary wall cellulose content (Ruan 2007). It will be interesting to explore the functional significance of the putative sucrose transporter, *GhSUC*, that is up-regulated along with secondary wall cellulose synthesis (Fig. 5), especially because another fiber sucrose transporter (*GhSUT1* with unspecified sequence, Ruan 2007) has maximal expression during elongation. Furthermore, current models suggest that there is less need for sucrose to be transported across membranes when plasmodesmata are open at the fiber foot during secondary wall deposition (Ruan 2007).

Sucrose degradation by *SUS* releases only half the carbon as UDP-glc to be used immediately by CESA enzymes. To recycle the coordinately-released fructose (after its phosphorylation) toward additional sucrose to support the extensive cellulose synthesis during cotton fiber secondary wall deposition, sucrose phosphate synthase (*SPS*; E.C. 2.4.1.14;  $\text{fru-6-P} + \text{UDP-glc} \rightarrow \text{suc-P} + \text{UDP}$ ) may play a role. Hindrance in the intra-fiber synthesis of sucrose is correlated with the limitation of cellulose synthesis rate by cool temperatures, which in turn can result in fiber immaturity in the field (Martin and Haigler 2004). Transgenic cotton with *SPS* activity elevated in leaves and, to a lesser extent, in fiber showed increased fiber secondary wall deposition in growth chamber experiments (Haigler et al. 2007). Nonetheless, one native cotton *SPS* gene assayed by qPCR (Fig. 5) showed down-regulation for secondary wall synthesis. Cotton is likely to contain several *SPS* isoforms (Lutfiyya et al. 2006), and additional data are needed to clarify whether or not one of them has a particular relationship to secondary wall cellulose synthesis.

In addition, it is important to remember that metabolism is often controlled post-transcriptionally. Even though the UDP-glc pyrophosphorylase gene assayed here did

not show increased expression for secondary wall deposition (Fig. 5), the size of the UDP-glc pool does increase at this time. There is likely metabolic rather than transcriptional control of UDPG<sub>p</sub> enzyme activity, and SPS activity is also highly regulated by phosphorylation and interaction with co-factors and other proteins (Haigler et al. 2001).

## 7.6 Genes of Unknown Function

The cotton homolog of the gene for KORRIGAN is also up-regulated on the array, which is consistent with a large increase in KORRIGAN protein at the onset of cotton fiber secondary wall deposition (Peng et al. 2002). In Arabidopsis, KORRIGAN, which has cellulase or 1,4-β-D-glucanase activity *in vitro* but an unknown substrate *in vivo*, is required for both primary and secondary wall synthesis. Severe phenotypes arise from its functional loss. KORRIGAN cycles between the endomembrane system and the plasma membrane in a microtubule-dependent manner, but available evidence suggests that it is not part of the CSC that contains CESA proteins (Robert et al. 2005; Szyjanowicz et al. 2004). Speculations about the role of KORRIGAN include cleaving a primer for cellulose synthesis, editing glucan chains prior to cellulose crystallization, cleaving glucan chains to release tensional stress of cellulose crystallization, or releasing CSCs from the cellulose fibril.

Other genes up-regulated on the array have even less known about their function. (1) The cotton homolog of At4g27435 (AAR07956; Fb34) was associated with an informative EST-SSR (expressed sequence tag simple sequence repeat) in a mapping population derived from a cross of *G. hirsutum* and *G. barbadense*, with the latter producing the highest quality cotton fiber (Han et al. 2004). The poplar homolog of this gene is up-regulated during wood secondary wall formation, and one Arabidopsis knock-out mutant showed about 20% less secondary xylem production under special growth conditions that induce that process in Arabidopsis (Ubeda-Tomas et al. 2007). (2) The up-regulated fasciclin-like protein FLA11 (*irx13*) belongs to a group of arabinogalactan proteins (AGPs); these secreted proteoglycans contain only 1-10% protein. Homologous genes to Arabidopsis FLA11 and FLA12 are strongly expressed in cellulose-rich tension wood (Lafarguette et al. 2004), and a gene with closest homology to Arabidopsis FLA12 was sequenced from 10 DPA cotton fiber and shown not to be present in ovules of wild-type cotton or a fiberless mutant (GhAGP1, AAO92753; Ji et al. 2003). The Gh.fbr.sw EST spotted on the array is nearly identical to *GhAGP1*, and further tests by qPCR would be needed to determine if true up-regulation of this particular AGP occurs for cotton fiber secondary wall deposition since the increase shown for the FLA12 cotton homolog at 20 DPA by the microarray was non-significant. Based on the presence of their putative cell adhesion domains (fasciclin domains) and immunolocalization in tension wood at the interface between the cell wall and plasma membrane [to which AGPs may be attached through glycosylphosphatidylinositol (GPI) anchors], it was proposed that the AGPs may act as adhesive proteins or participate in a signaling networks based on N-acetyl-glucosamine (Lafarguette et al. 2004).

## 7.7 Cell-wall Related Genes Not Shown to be Up-regulated

If not already mentioned, below is a brief summary of the functional context in other species of genes that were not up-regulated on the array. Additional experiments (including checking the gene expression pattern by qPCR) would be required to test whether or not these or related proteins have similar roles in cotton fiber. **(1)** In Arabidopsis, RHD3, which has GTP-binding motifs, is thought to be involved in endomembrane trafficking of molecules required for cell wall synthesis, thereby helping to control anisotropic cell expansion as well as secondary wall thickening (Hu et al. 2003; Yuen et al. 2005). **(2)** Down-regulation of *PGSIP1*, a glycosyltransferase family 8 gene, in Arabidopsis is associated with starch deficiency, and the protein likely primes starch synthesis. Starch is not present after early elongation in cotton fiber (Ryser 1985), but the expression pattern of the poplar homolog suggests a possible role in priming synthesis of another polysaccharide during wood secondary wall formation (Aspeborg et al. 2005). **(3)** The serine-rich protein, KOBITO, in Arabidopsis has a role in cell elongation, cellulose synthesis, vascular differentiation, and root meristem maintenance. However, multiple downstream phenotypes arose from alteration of an ABA- and Glc-responsive signaling network (Lertpiriyapong and Sung 2003; Brocard-Gifford et al. 2004), and the same regulatory network might not operate in cotton fiber. **(4)** For the multi-copper oxidase, *LACCASE4*, down-regulation of a close poplar homolog resulted in increased soluble phenolics in xylem parenchyma cells and misshapen secondary-walled conducting cells, which also had secondary walls that were poorly adhering internally or to the adjacent primary wall (Ranocha et al. 2002). These authors inferred that polymerization of non-lignin phenolics was important to maintain xylem secondary wall integrity. **(5)** For the glycosyltransferase family 47 gene (At1g27440), the poplar homolog is related to cell wall cross-linking in wood. Based on analysis of a tobacco homolog, it may be that this gene modifies pectin type II rhamnogalacturonan by addition of glucuronic acid, which in turn can affect borate cross-linking (Ubeda-Tomas et al. 2007). Neither phenolic or borate cross-linking are known to be important in secondary wall cotton fibers.

The secondary-wall-biased G.h.fbr-sw sequence set did not include cotton homologs of several genes that support primary wall cellulose synthesis (and probably secondary wall synthesis) at a fundamental level in Arabidopsis including: (a) *CYT1* and *KNF* that, along with *RSW3*, encode proteins that carry out N-glycan processing in the endoplasmic reticulum; (b) *FK*, *HYD1*, and *SMT1/CPH* that encode proteins involved in sterol biosynthesis (Schrack et al. 2004); and (c) *RSW10*, encoding a ribose 5-phosphate isomerase that is likely generates UDP-glc (see review in Somerville 2006 and Howles et al. 2006). Probably their absence from these experiments denotes their constitutive expression in cotton fiber.

## 8.0 Conclusions

We have illustrated the power of cotton fiber gene expression to reveal genes that are required across species for secondary wall deposition, particularly high-rate cellulose synthesis. We showed that the highly developed genomic resources for Arabidopsis

can aid functional understanding of cotton genes. The similarities of cotton fiber secondary wall gene expression with xylem reveal an opportunity to analyze the “pure system” of cotton fiber for as yet unknown participants in cellulose synthesis. In addition to up-regulated genes, there are surely constitutively expressed genes that also have important roles. Further research on secondary wall deposition in cotton fiber is likely to be beneficial to the improvement of numerous cellulose-rich fiber crops.

In addition, one of the most interesting challenges is to understand what makes cotton fiber different, e.g. why does it not synthesize large amounts of matrix molecules in its secondary wall? What is the role of proteins encoded by cotton-specific and fiber-specific genes? Are there any unique aspects of cellulose synthesis in the fiber? How are the developmental transitions in the fiber controlled and interfaced with environmental signals? How is cellulose synthesis down-regulated under stress?

Similar analyses to those reported here using cDNA arrays have recently been reported for *G. hirsutum* (Guo et al. 2007) and *G. barbadense* (Tu et al. 2007), and soon we will see the application of more comprehensive gene-specific oligomer arrays to understand changes in gene expression at the onset of secondary wall deposition. All such research quickly results in the need to assign gene function through experimentation, a process that is much slower than demonstrating changes in gene expression. Achieving functional understanding will require more integration of gene expression studies with, for example, analysis of protein-protein interactions, enzymatic activity, metabolite levels and flux, and cell structure via advanced imaging techniques. More rapid means of testing gene function in cotton, e.g. by more efficient and faster production of stably transformed cotton or by alternative means such as virus induced gene silencing, would certainly be useful.

Given that the cotton genome sequence is not yet available, molecular means of cotton improvement must rely on the existing large body of cotton ESTs and ability to analyze the expression and significance of genes they represent. Such information can help bridge cotton and model plants, including Arabidopsis and Poplar, for comparative genome analysis. It can also be directly used for genetic improvement of fiber quality and yield traits. For example, if fiber ESTs can be associated with fiber traits by QTL mapping, fine mapping of the QTL region can lead to development of molecular markers for marker-assisted breeding.

## Acknowledgments

Cotton Incorporated, Cary, NC supported the microarray and bioinformatics analyses, with additional support from North Carolina State University. Special thanks go to Curtis Wilkerson and Jeff Landgraf, Research Technology Support Facility/Bioinformatics at Michigan State University, for carrying out the microarray analysis. NSF Plant Genome Research Program grants, #DBI-0211797, R98RA1829, and #DBI-0110173, supported making and sequencing the G.h.fbr-sw SSH library.

## References

- Adams, K.L. and Wendel, J.F. (2004) Exploring the genomic mysteries of polyploidy in cotton. *Biol. J. Linn. Soc.* 82: 573-581.
- 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* 92: 9353-9357.
- Arpat, A., Waugh, M.P., Sullivan, J., Gonzales, M., Frisch, D., Main, D., Wood, T., Leslie, A., Wing, R. and Wilkins, T. (2004) Functional genomics of cell elongation in developing cotton fibers. *Plant Mol. Biol.* 54: 911-929.
- Arthur, J.C. (1990) Cotton. In: J.I. Kroschwitz (Ed.), *Polymers: Fibers and Textiles, a Compendium*, John Wiley and Sons, New York, pp. 118-141.
- Aspeborg, H., Schrader, J., Coutinho, P.M., Stam, M., Kallas, A., Djerbi, S., Nilsson, P., Denman, S., Amini, B., Sterky, F., Master, E., Sandberg, G., Mellerowicz, E., Sundberg, G., Henrissat, B. and Teeri, T.T. (2005) Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. *Plant Physiol.* 137: 983-997.
- Baskin, T.I. (2001) On the alignment of cellulose microfibrils by cortical microtubules: a review and a model. *Protoplasma* 215: 150-171.
- Benedict, C.R., Kohel, R.J. and Jividen, G.M. (1994) Crystalline cellulose and cotton fiber strength. *Crop Sci.* 34: 147-151.
- Benedict, C.R., Kohel, R.J. and Lewis, H.L. (1999) Cotton fiber quality, In: C.W. Smith, J.T. Cothren (Eds.), *Cotton: Origin, History, Technology, and Production*, John Wiley & Sons, New York, pp. 269 – 288.
- Bieniawska, Z., Paul Barratt, D.H., Garlick, A.P., Thole, V., Kruger, N.J., Martin, C., Zrenner, R. and Smith, A.M. (2007) Analysis of the sucrose synthase gene family in *Arabidopsis*. *Plant J.* 49: 810-828.
- Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.* 6:791-806.
- Bowman, J.L., Floyd, S.K. and Sakakibara, K. (2007) Green genes—comparative genomics of the green branch of life. *Cell* 129: 229-234.
- Bradow, J.M. and Davidonis, G.W. (2000) Review: Quantitation of fiber quality and the cotton production-processing interface: A physiologist's perspective. *J. Cotton Sci.* 4: 34-64.
- Brady, S.M., Song, S., Dhugga, K.S., Rafalski, J.A. and Benfey, P.N. (2007) Combining expression and comparative evolutionary analysis. The COBRA gene family. *Plant Physiol.* 143:172-87.
- Brocard-Gifford, I., Lynch, T. J., Garcia, E. M., Malhotra, B. and Finklestein, R. R. (2004) The *Arabidopsis thaliana* abscisic acid insensitive8 locus encodes a novel protein mediating abscisic acid and sugar responses essential for growth. *Plant Cell* 16: 406-421.
- Brown, D.M., Zeef, L. A.-H., Ellis, J., Goodacre, R. and Turner, S.R. (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell*: 17: 2281-2295.

- Brown, R.M. Jr. and Saxena, I.M. (Eds.) (2007) *Cellulose: Molecular and Structural Biology*. Springer, Dordrecht, 379 pp.
- Burk, D.H. and Ye, Z.-H. (2002) Alteration of oriented deposition of cellulose microfibrils by mutation of a katanin-like microtubule severing protein. *Plant Cell* 14: 2145-2160.
- Busch, W. and Lohmann, J.U. (2007) Profiling a plant: expression analysis in *Arabidopsis*. *Curr. Opin. Plant Biol.* 10: 136-141.
- Carpita, N.C. and Delmer, D.P. (1981) Concentration and metabolic turnover of UDP-glucose in developing cotton fibers. *J. Biol. Chem.* 256: 308-315.
- Chanzy, H., Imada, K. and Vuong, R. (1978) Electron diffraction from the primary wall of cotton fibers. *Protoplasma* 94: 299-306.
- Chu, A., Chen, H., Zhang, Y., Zhang, Z., Zheng, N., Yin, B., Yan, H., Zhu, L., Zhao, X., Yuan, M., Zhang, X. and Xie, Q. (2007) Knockout of the *AtCESA2* gene affects microtubule orientation and causes abnormal cell expansion in *Arabidopsis*. *Plant Physiol.* 143: 213-224.
- Cronn, R.C., Small, R.L. and Wendel, J.F. (1999) Duplicated genes evolve independently after polyploidy formation in cotton. *Proc. Natl. Acad. Sci. USA* 96: 14406-14411.
- Delmer, D.P. (1999) Cellulose biosynthesis in developing cotton fibers. In: A.S. Basra (Ed.), *Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing*, The Haworth Press, New York, pp. 85-112.
- Diatchenko, L., Lau, Y.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K. Gurskaya, N., Sverdlov, E.D. and Siebert, P.D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* 93:6025-6030.
- Diotallevi, F. and Mulder, B. (2007) The cellulose synthase complex: a polymerization driven supramolecular motor. *Biophys. J.* 92: 2666-2673.
- Endrizzi, J.E., Turcotte, E.L. and Kohel, R.J. (1985) Genetics, cytology, and evolution of *Gossypium*. *Adv. Gen.* 23: 271-354.
- Farrokhi, N., Burton, R.A., Brownfield, L., Hrmova, M., Wilson, S.M., Bacic, A. and Fincher, G.B. (2006) Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to identification of key genes. *Plant Biotech. J.* 4: 145-167.
- Franz, G. (1969) Soluble nucleotides in growing cotton hair. *Phytochem.* 8: 737-741.
- Fryxell, P.A. (1979) *The Natural History of the Cotton Tribe (Malvaceae, Tribe Gossypieae)*. Texas A&M University Press, College Station, 245 pp.
- Gipson, J.R. (1986) Temperature effects on growth, development, and fiber properties. In: J.R. Mauney and J. McD. Stewart (Eds.), *Cotton Physiology*, The Cotton Foundation, Memphis, pp. 47-56.
- Guo, J.-Y., Wang, L.-J., Chen, S.-P., Hu, W.-L. and Chen, X.-Y. (2007) Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary wall synthesis. *Cell Research* 2007: 1-13.
- Gutierrez, R.A., Green, P.J., Keegstra, K. and Ohlrogge, J.B. (2004) Phylogenetic profiling of the *Arabidopsis thaliana* proteome: what proteins distinguish plants

- from other organisms. *Genome Biology* 5:R53 (<http://genomebiology.com/2004/5/8/R53>).
- Haigler, C.H. (1985) The functions and biogenesis of native cellulose, In: S.H. Zeronian and T.P. Nevell TP (Eds.), *Cellulose Chemistry and its Applications*, Ellis Horwood, Chichester, pp 30-83.
- Haigler, C.H. (1991) The relationship between polymerization and crystallization in cellulose biogenesis, In: C.H. Haigler and P. Weimer (Eds.), *Biosynthesis and Biodegradation of Cellulose*, Marcel Dekker, New York, pp 99-124.
- Haigler, C.H., Ivanova-Datcheva, M., Hogan, P.S. Salnikov, V.V., Hwang, S., Martin, L.K. and Delmer, D.P. (2001) Carbon partitioning to cellulose synthesis. *Plant Mol. Biol.* 47: 29-51.
- Haigler, C.H., Zhang, D. and Wilkerson, C.G. (2005) Biotechnological improvement of cotton fibre maturity. *Physiol. Plant.* 124: 285-294.
- Haigler, C.H. (2006) Establishing the cellular and biophysical context of cellulose synthesis. In: T. Hayashi (Ed.), *The Science and Lore of the Plant Cell Wall: Biosynthesis, Structure and Function*, BrownWalker Press: Boca Raton, pp. 97 – 105.
- Haigler, C.H. (2007) Substrate supply for cellulose synthesis and its stress sensitivity in the cotton fiber. In: R.M. Brown Jr, and I. Saxena (Eds.), *Cellulose: Molecular and Structural Biology*, Springer: New York, pp. 145 – 166.
- Haigler, C.H., Singh, B., Zhang, D., Hwang, S., Wu, C., Cai, W.X., Hozain, M., Kang, W., Kiedaisch, B., Strauss, R.E., Hequet, E.F., Wyatt, B.G., Jividen, G.M. and Holaday, A.S. (2007) Transgenic cotton over-producing spinach sucrose phosphate synthase showed enhanced leaf sucrose synthesis and improved fiber quality under controlled environmental conditions. *Plant Mol. Biol.* 63: 815-832.
- Han, Z.G., Guo, W.Z., Song, X.L. and Zhang, T.Z. (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. *Mol. Gen. Genom.* 272: 308-327.
- Hardin, S.C., Duncan, K.A. and Huber, S.C. (2006) Determination of structural requirements and probable regulatory effectors for membrane association of maize sucrose synthase. *Plant Physiol.* 141: 1106-1119.
- Heinze, T. (1998) New ionic polymers by cellulose functionalization. *Macromol. Chem. Phys.* 199: 2341-2364.
- Hertzberg, M., Aspeborg, M., Schrader, J., Andersson, A., Erlandsson, R., Blomqvist, K., Bhalerao, R., Uhlén, M., Teeri, T.T., Lundeberg, J., Sundberg, B., Nilsson, and P. Sandberg, G. (2001) A transcriptional roadmap to wood formation. *Proc. Natl. Acad. Sci. USA* 98: 14732-14737.
- Hindeleh, A.M., Johnson, D.J. and Monatgue, P.E. (1980) Computational methods for profile resolution and crystallite size evaluation in fibrous polymers, In: A.D. French and K.H. Gardener (Eds.), *Fiber Diffraction Methods: ACS Symposia no. 141*, p. 149-182.
- Howles, P.A., Birch, R.J., Collings, D.A., Gebbie, L.K., Hurley, U.A., Hocart, C.H., Arioli, T. and Williamson, R.E. (2006) A mutation in an Arabidopsis ribose 5-phosphate isomerase reduces cellulose synthesis and is rescued by exogenous uridine. *Plant J.* 48: 606-618.

- Hsieh, Y.-L. (1999) Structural development of cotton fibers and linkages to fiber quality. In: A.S. Basra (Ed.), *Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing*, The Haworth Press, New York, pp. 137-166.
- Hsieh, Y.-L., Hu, X.P. and Nguyen, A. (1997) Strength and crystalline structure of developing Acala cotton. *Text. Res. J.* 67: 529 – 536.
- Hu, X.P. and Hsieh, Y.-L. (1996) Crystalline structure of developing cotton fibers. *J. Polym. Sci.: Part B: Polym. Phys.* 34: 1451-1459.
- Hu, H.Y., Zhong, R., Morrison, W. and Ye, Z. (2003) The Arabidopsis RHD3 gene is required for cell wall biosynthesis and actin organization. *Planta* 217: 912-921
- Hutchinson, J.B., Stephens, S.G. and Dodds, K.S. (1945) The seed hairs of *Gossypium*. *Ann. Bot.* IX (36): 360-368.
- Jacob-Wilk, D., Kurek, I., Hogan, P. and Delmer, D.P. (2006) The cotton fiber zinc-binding domain of cellulose synthase A1 from *Gossypium hirsutum* displays rapid turnover in vitro and in vivo. *Proc. Natl. Acad. Sci. USA* 103: 12191-12196.
- Ji, S.J., Lu, Y.C., Feng, J.X., Wei, G., Li, J., Shi, Y.H., Fu, Q., Liu, D., Luo, J.C. and Zhu, Y.X. (2003) Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. *Nucleic Acids Res.* 31: 2534-2543
- Jiang, C., Wright, R.J., El-Zik, K.M. and Paterson, A.H. (1998) Polyploid formation created unique avenues for response to selection in *Gossypium*. *Proc. Natl. Acad. Sci. USA* 95(8): 4419-4424.
- Joshi, C.P. and Mansfield, S.D. (2007) The cellulose paradox—simple molecule, complex biosynthesis. *Curr. Opin. Plant Biol.* 10: 220 – 226.
- Ko, J.-H., Beers, E.P. and Han, K.H. (2006) Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 276: 517-531.
- Kim, H.J. and Triplett, B.A. (2001) Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* 127: 1361-1366.
- Kim, H.J., Williams, M.Y. and Triplett, B.A. (2002) A novel expression assay system for fiber-specific promoters in developing cotton fibers. *Plant Mol. Biol. Rep* 20: 7-18.
- Kurek, I., Kawogoe, Y., Jacob-Wilk, D., Doblin, M. and Delmer, D. (2002) Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domain. *Proc. Natl. Acad. Sci.* 99: 11109-11114.
- Lafarguette, F., Leple, J.-C., Dejardin, A., Laurans, F., Costa, G., Lesage-Descauses, M.-C. and Pilate, G. (2004) Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytol.* 164: 107-121.
- Laosinchai, W. (2002) Molecular and biochemical studies of cellulose and callose synthase. Ph.D. Dissertation, The University of Texas at Austin, 207 pp.
- Lertpiriyapong, K. and Sung, Z. R. (2003) The elongation defective1 mutant of *Arabidopsis* is impaired in the gene encoding a serine-rich secreted protein. *Plant Mol. Biol.* 53: 581-595.

- Lewin, M. and Pearce, E.M. (Eds.) (1998) *Handbook of Fiber Chemistry*, 2<sup>nd</sup> edn, Marcel Dekker: NY, 724 pp.
- Lutfiyya, L.L., Xu, N., D'Ordine, R.L., Morrell, J.A., Miller, P.W. and Duff, S.M.G. (2006) Phylogenetic and expression analysis of sucrose phosphate synthase isozymes in plants. *J. Plant Physiol.* Doi:10.1016/j.jpiph.2006.04.014.
- Martin, L.K. and Haigler, C.H. (2004) Cool temperature hinders flux from glucose to sucrose during cellulose synthesis in secondary wall stage cotton fibers. *Cellulose* 11:339-349.
- Meinert, M.C. and Delmer, D.P. (1977) Changes in biochemical composition of the cell wall of the cotton fiber during development. *Plant Physiol.* 59: 1088-1097.
- Nakajima, K., Kawamura, T. and Hashimoto, T. (2006) Role of the *SPIRALI* gene family in anisotropic growth of Arabidopsis. *Plant Cell Physiol.* 47: 513-522
- Niklas, K.J. (1992) *Plant Biomechanics, An Engineering Approach to Plant Form and Function*, Univ Chicago Press, Chicago, 607 pp.
- Paradez, A.R., Someville, C.R. and Erhardt, D.W. (2006) Visualization of cellulose synthase demonstrates functional association with microtubules. *Science* 312: 1491-1495.
- Pear, J., Kawagoe, Y., Schreckengost, W., Delmer, D.P. and Stalker, D. (1996) Higher plants contain homologs of the CelA genes that encode the catalytic subunit of the bacterial cellulose synthases. *Proc. Natl. Acad. Sci. USA* 93: 12637-12642.
- Pena, M.J., Zhong, R., Zhou, G.-K., Richardson, E.A., O'Neill, M.A., Darvill, A.G., York, W.S. and Ye, Z.-H. (2007) Arabidopsis *irregular xylem8* and *irregular xylem 9*: Implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell* 19: 549 – 563.
- Peng, L., Kawagoe, Y., Hogan, P. and Delmer, D. (2002) Sitosterol- $\beta$ -glucoside as primer for cellulose synthesis in plants. *Science* 295: 147-150.
- Persson S., Wei, H., Milne, J., Page, G.P. and Somerville, C.R. (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl. Acad. Sci. USA* 102: 8633-8638.
- Pilate, G., Dejardin, A., Laurans, F. and Leple, J.-C. (2004) Tension wood as a model for functional genomics of wood formation. *New Phytol.* 164: 63-72.
- Piling, E. and Hofte, H. (2003) Feedback from the wall. *Curr. Opin. Plant Biol.* 6: 611-616.
- Potikha, T.S., Collins, C.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.
- R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, URL <http://www.R-project.org>.
- Ranocha, P., Chabannes, M., Chamayou, S., Danoun, S., Jauneau, A., Boudet, A.-M. and Goffner, D. (2002) Laccase down-regulation causes alterations in phenolic metabolism and cell wall structure in poplar. *Plant Physiol.* 129: 145-155
- Rebenfield, L. (1990) Fibers. In: J.I. Kroschwitz (Ed.), *Polymers: Fibers and Textiles, a Compendium*. John Wiley and Sons, New York, pp. 219-305.

- Robert, S., Bichet, A., Grandjean, O., Kierzkowski, D., Satiat-Jeunemaitre, B., Pelletier, S., Hauser, M.-T., Hofte, H. and Vernhettes, S. (2005) An Arabidopsis endo-1,4- $\beta$ -D-glucanase involved in cellulose synthesis undergoes regulated intracellular cycling. *Plant Cell* 17: 3378-3389.
- Roberts, E.M., Nunna, R.R., Huang, J.Y., Trolinder, N.L. and Haigler, C.H. (1992) Effects of cycling temperatures on fiber metabolism in cultured cotton ovules. *Plant Physiol.* 100: 979-986.
- Rowland, S.P. and Bertoniere, N.R. (1985) Chemical methods for studying supramolecular structure. In: T.P. Nevell and S.H. Zeronian (Eds.), *Cellulose Chemistry and its Applications*, Ellis Horwood, Chichester, pp. 112-137.
- Ruan, Y.-L. (2007) Rapid cell expansion and cellulose synthesis regulated by plasmodesmata and sugar: insights from the single-celled cotton fiber. *Func. Plant Biol.* 34: 1 – 10
- Ryser, U. (1985) Cell wall biosynthesis in differentiating cotton fiber. *Eur. J. Cell Biol.* 39: 236 – 256
- Ryser, U. (1999) Cotton fiber initiation and histodifferentiation. In: A.S. Basra (Ed.), *Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing*, The Haworth Press, New York, pp. 1 – 46.
- Salnikov, V., Grimson, M.J., Seagull, R.W. and Haigler, C.H. (2003) Localization of sucrose synthase and callose in freeze substituted, secondary wall stage, cotton fibers. *Protoplasma* 221: 175-184.
- Saxena, I.M. and Brown, R.M. (2005) Cellulose biosynthesis: Current views and evolving concepts. *Ann. Bot.* 96: 9-21.
- Scheible, W.-R. and Pauly, M. (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. *Curr. Opin. Plant. Biol.* 7: 1-11.
- Schneider, M., Schwart, J.P.J. and Sanderson, R.D. (1996) Biological variation in the degree of polymerization of cotton lint cellulose produced in South Africa Part II: Comparing Acala 1517 of the Middle Transvaal and the Northern Cape. *Text. Res. J.* 66: 428-435.
- Schrack, K., Fukioka, S., Takatsuto, S., Stierhof, Y.-D., Stransky, H., Yoshida, S. and Jurgens, G. (2004) A link between sterol biosynthesis, the cell wall, and cellulose in Arabidopsis. *Plant J.* 38: 227-243.
- Seagull, R.W. (1993) Cytoskeletal involvement in cotton fiber growth and development. *Micron* 24: 643-660.
- Shi, Y.-H., Zhu, S.-W., Mao, X.-Z., Feng, J.-X., Qin, Y.-M., Zhang, L., Cheng, J., Wei, L.-P., Wang, Z.-Y. and Zhu, Y.-X. (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18: 651-664.
- Smyth, G. K. and Speed, T. P. (2003). Normalization of cDNA microarray data. *Methods* 31: 265–273.
- Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3(1): Article 3.
- Smyth, G. K. (2005). Limma: linear models for microarray data. In: R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber (Eds.), *Bioinformatics and*

- Computational Biology Solutions using R and Bioconductor*, Springer, New York, pages 397–420.
- Smyth, G. K., Michaud, J. and Scott, H. (2005). The use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21: 2067–2075.
- Soltis, P.S., Endress, P.K., Chase, M.W. and Soltis, D.E. (2005) *Phylogeny & Evolution of Angiosperms*. Sinauer Associates, Inc.: Sunderland, p. 190.
- Somerville, C. (2006) Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* 22: 53-78.
- Stephens, S.G. (1970) The botanical identification of archaeological cotton. *Amer. Antiquity* 35 (3): 368-373.
- Stone, B. (2001) Cellulose: Structure and distribution, in *Encyclopedia of Life Sciences*, Nature Publishing Group, pp. 1 – 9.
- Szyjanowicz, P.M.J., McKinnon, I., Taylor, N.G., Gardiner, J., Jarvis, M.C. and Turner, S.R. (2004) The *irregular xylem 2* mutant is an allele of *korrigan* that affects the secondary cell wall of *Arabidopsis thaliana*. *Plant J.* 37: 730-740.
- Taylor, N. G., Howells, R. M., Huttly, A. K., Vickers, K. and Turner, S. R. (2003) Interactions among three distinct CesaA proteins essential for cellulose synthesis, *Proc. Natl. Acad. Sci. USA* 100: 1450-1455.
- Timpa, J.D. (1991) Application of universal calibration in gel permeation chromatography for molecular weight determination of plant cell wall polymers: Cotton fiber. *J. Agric. Food Chem.* 39: 270-275.
- Timpa, J.D. and Triplett, B.A. (1993) Analysis of cell-wall polymers during cotton fiber development. *Planta* 189: 101-108.
- Timpa, J.D. and Ramey, H.H. (1989) Molecular characterization of three cotton varieties. *Text. Res. J.* 59: 661-664.
- Timpa, J.D. and Ramey, H.H. (1994) Relationship between cotton fiber strength and cellulose molecular weight distribution: HVI calibration standards. *Text. Res. J.* 64: 557-562.
- Truernit, E., Siemering, K.R., Hodge, S., Vojislava, G. and Haseloff, J. (2006) A map of KNAT gene expression in the *Arabidopsis* root. *Plant Mol. Biol.* 60: 1-20.
- Tu, L.-L., Zhang, X.-L., Liang, S.-G., Liu, D.-Q., Zhu, L.-F., Zeng, F.-C., Liu, D.-Q., Zhu, L.-F., Zeng, F.-C., Nie, Y.-C., Guo, X.-P., Deng, F.-L., Tan, J.-F. and Xu, L. (2007) Gene expression analysis of sea-island cotton (*Gossypium barbadense* L.) during fiber development. *Plant Cell Rep* DOI 10.1007/s00299-007-0337-4
- Ubeda-Tomas, S., Edvardsson, E., Eland, C., Singh, S.K., Zadik, D., Aspeborg, H., Gorzsas, A., Teeri, T.T., Sundberg, B., Persson, P., Bennett, M. and Marchant, A. (2007) Genomic-assisted identification of genes involved in secondary growth in *Arabidopsis* utilizing transcript profiling of poplar wood-forming tissues. *Physiol. Plant.* 129: 415-428.
- Udall, J.A., Swanson, J.M., Haller, K., Rapp, R.A., Sparks, M.E., Hatfield, J., Yu, Y., Wu, Y., Dowd, C., Arpat, A.B., Sickler, B.A., Wilkins, T.A., Guo, J.Y., Chen, X.Y., Scheffler, J., Talierco, E., Turley, R., McFadden, H., Payton, P., Allen, R., Zhang, D., Haigler, C., Wilkerson, C., Suo, J., Schulze, S.R., Pierce,

- M.L., Essenberg, M., Kim, H., Llewellyn, D.J., Dennis, E.S., Kudrna, D., Wing, R., Paterson, A.H., Soderlund, C. and Wendel, J.F. (2006) A global assembly of cotton ESTs. *Gen. Res.* 16: 441-50.
- Wang, J., Howles, P.A., Cork, A.H., Birch, R.J. and Williamson, R.E. (2006) Chimeric proteins suggest that the catalytic and/or C-terminal domains give CesA1 and CesA3 access to their specific sites in the cellulose synthase of primary walls. *Plant Physiol.* 142: 685-695.
- Wendel, J.F. and Cronn, R.C. (2002) Polyploidy and evolutionary history of cotton. *Adv. Agron.* 78: 139-186.
- Whittaker, D.J. and Triplett, B.A. (1999) Gene-specific changes in alpha-tubulin transcript accumulation in developing cotton fibers. *Plant Physiol.* 121: 181-188.
- Willison, J.H.M. and Brown, R.M. (1977) An examination of the developing cotton fiber: wall and plasmalemma. *Protoplasma* 92: 21-41.
- Wu, Y.-T. and Liu, J.-Y. (2005) Molecular cloning and characterization of a cotton glucuronosyltransferase gene. *J. Plant Physiol.* 162: 573-582.
- Yatsu, L.Y. (1983) Morphological and physical effects of colchicine treatment on cotton (*Gossypium hirsutum* L.) fibers. *Text. Res. J.* 53: 515-519.
- Yuen, C.Y.L., Sedbrook, J.C., Perrin, R.M., Carroll, K.L. and Masson, P.H. (2005) Loss-of-function mutations of ROOT HAIR DEFECTIVE3 suppress root waving, skewing, and epidermal cell file rotation in Arabidopsis. *Plant Physiol.* 138: 701-714.
- Zhang, D., Choi, D.W., Wanamaker, S., Fenton, R.D., Chin, A., Malatrasi, M., Turuspekov, Y., Walia, H., Akhunov, E.D., Kianian, P., Otto, C., Simons, K., Deal, K.R., Echenique, V., Stamova, B., Ross, K., Butler, G.E., Strader, L., Verhey, S.D., Johnson, R., Altenbach, S., Kothari, K., Tanaka, C., Shah, M.M., Laudencia-Chingcuanco, D., Han, P., Miller, R.E., Crossman, C.C., Chao, S., Lazo, G.R., Klueva, N., Gustafson, J.P., Kianian, S.F., Dubcovsky, J., Walker-Simmons, M.K., Gill, K.S., Dvorak, J., Anderson, O.D., Sorrells, M.E., McGuire, P.E., Qualset, C.O., Nguyen, H.T. and Close, T.J. (2004a) Construction and evaluation of cDNA libraries for large-scale expressed sequence tag sequencing in wheat (*Triticum aestivum* L.). *Genetics* 168: 595-608.
- Zhang, D., Hrmova, M., Wan, C.-H., Wu, C., Balzen, J., Cai, W., Wang, J., Densmore, L.D., Fincher, G.B., Zhang, H. and Haigler, C.H. (2004b) Members of a new group of chitinase-like genes are expressed preferentially in cotton cells with secondary walls. *Plant Mol. Biol.* 54:353-372.
- Zhong R., Kayes, S. J., Schroeder, B. P. and Ye, Z.H. (2002) Mutation of a chitinase-like gene causes ectopic deposition of lignin, aberrant cell shapes, and overproduction of ethylene. *Plant Cell* 14: 165-179.
- Zhong, R., Richardson, E.A. and Ye, Z.-H. (2007) Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. *Planta* 225: 1603-1611