Planta (2008) 227:319–329
DOI 10.1007/s00425-007-0619-7

ORIGINAL ARTICLE

A majority of cotton genes are expressed in single-celled fiber

Ran Hovav · Joshua A. Udall · Einat Hovav · Ryan Rapp · Lex Flagel · Jonathan F. Wendel

Received: 24 July 2007 / Accepted: 24 August 2007 / Published online: 12 September 2007
© Springer-Verlag 2007

Abstract Multicellular eukaryotes contain a diversity of cell types, presumably differing from one another in the suite of genes expressed during development. At present, little is known about the proportion of the genome transcribed in most cell types, nor the degree to which global patterns of expression change during cellular differentiation. To address these questions in a model plant system, we studied the unique and highly exaggerated single-celled, epidermal seed trichomes ("cotton") of cultivated cotton (*Gossypium hirsutum*). By taking advantage of advances in expression profiling and microarray technology, we evaluated the transcriptome of cotton fibers across a developmental time-course, from a few days post-anthesis through primary and secondary wall synthesis stages. Comparisons of gene expression in populations of developing cotton fiber cells to genetically complex reference samples derived from 6 different cotton organs demonstrated that a remarkably high proportion of the cotton genome is transcribed, with 75–94% of the total genome transcribed at each stage. Compared to the reference samples, more than half of all genes were up-regulated during at least one stage of fiber development. These genes were clustered into seven groups of expression profiles that provided new insight into biological processes governing fiber development. Genes implicated in vesicle coating and trafficking were found to be overexpressed throughout all stages of fiber development studied, indicating their important role in maintaining rapid growth of this unique plant cell.

Keywords Cotton · *Gossypium hirsutum* · Fiber · Single-cell · Microarray · Vesicle coating proteins

Abbreviations

DPA Days Post Anthesis
FDR False Discovery Rates
SNARE Soluble NSF attachment protein receptor

Electronic supplementary material The online version of this article (doi:10.1007/s00425-007-0619-7) contains supplementary material, which is available to authorized users.

R. Hovav · J. A. Udall · E. Hovav · R. Rapp · L. Flagel · J. F. Wendel (✉)
Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA
e-mail: jfw@mail.adp.iastate.edu

R. Hovav
e-mail: ran@iastate.edu

J. A. Udall
Department of Plant and Animal Science, Brigham Young University, Provo, UT 84602, USA

Introduction

Higher eukaryotes contain a multitude of cell types at maturity, each initiating its developmental program from undifferentiated progenitor cells (Honys and Twell 2004; Galbraith and Birnbaum 2006). In plants, it has been estimated that there are approximately 40 different cell types, each with their own function, structure, and location (Demura et al. 2002). Cellular differentiation is accompanied by myriad changes in transcription and translation, metabolism, and synthesis of intracellular structures. Insights into these complex processes require understanding the dynamics of transcription during cellular differentiation, growth, and maturation. At present, relatively little is known about the portion and proportion of the total transcript pool that is transcribed in most cell

types or the degree to which global patterns of expression change during growth and maturation. In prokaryotes, whose entire lifecycle is completed in a single cell, the entire transcribed component of the genome is expressed at one or more stages, with the exception of genes induced by specific environmental stresses or challenges. In multicellular eukaryotes, however, this need not be the a priori prediction, because the number of genes is larger and the multiplicity of differentiated cell types.

Advances in technologies for harvesting specific cell types and in amplifying messenger RNA pools for expression profiling have stimulated studies of the transcriptome at the cellular level in plants (Galbraith and Birnbaum 2006). Explorations of changes during differentiation under natural or in vitro systems have been facilitated by isolating cells using sorting procedures or by laser microdissection and capture (Birnbaum et al. 2003; Leonhardt et al. 2004; Casson et al. 2005). Transcription profiling of single cells using microarrays has made it possible to evaluate the involvement of thousands of genes in biological processes, providing a powerful tool for analyzing cell differentiation and development. Recent applications include wood-forming cambial meristem tissue (Schrader et al. 2004), the analysis of the *Arabidopsis thaliana* pollen transcriptome (Pina et al. 2005), the quiescent-center cells of developing roots (Nawy et al. 2005), epidermis cells or vascular tissue of maize (Nakazono et al. 2003), and rapidly expanding cotton fiber initial cells (Wu et al. 2007). These experiments have a common obstacle of sample preparation and cell-type isolation that could impact interpretations of gene expression. Measuring gene expression in a single, abundant cell type will not have as much experimental induced error.

Here, we studied global patterns of gene expression during development of the single-celled epidermal trichomes of cotton seeds (*Gossypium hirsutum*). Seed trichomes, colloquially termed “cotton fiber”, comprise the world’s most important textile fiber and a vital component of the agricultural economy of over 50 nations. They also represent one of the most distinct single cell types in the plant kingdom. In some domesticated varieties, cotton fibers may attain a final length of 6 cm, or about one-third the height of an entire *Arabidopsis* plant (Kim and Triplett 2001). A single cotton ovary contains ~500,000 elongating cells representing a single cell type. On the day of anthesis (flower opening), approximately one in four epidermal ovular cells has already been destined to become a cotton fiber, initially appearing as a spherical protrusion and subsequently elongating through stages of primary wall synthesis, secondary wall synthesis, maturation, and cell death. Here, we used cotton oligonucleotide microarrays containing 13,178 probes (Udall et al. 2007) to evaluate the transcriptome of cotton fibers across a developmental time-course, from two days post-anthesis (DPA) through the stages of primary and secondary wall synthesis. We show that the transcriptome of the single cell type cotton fiber is extremely complex, with most genes expressed at one or more stages during development, and that more than half of all genes are “up-regulated” when compared to a reference sample. We also present new insight into the biological processes leading to cotton fiber cell elongation and development.

Materials and methods

Plant material and RNA preparation

Gossypium hirsutum cv. TM1 plants were grown in four separate replicates of four to eight plants in the Horticulture Greenhouse at Iowa State University under supplemental lighting (16 h days). For each replicate, ovules were excised, immediately frozen in liquid nitrogen, and stored in –80°C. A heterogeneous mix of tissues was constructed by combining RNAs from, leaves, stems, petals, anthers, calyx, and bracts into a single reference sample. Roots were excluded from the reference tissue due to potentially high homology between roots hairs and cotton fiber cells. Other tissues were selected to ensure a diverse and complex transcriptomic representation, including various developmental stages and different plant organs. At each developmental time-point, fibers were isolated from ovules using a liquid nitrogen/glass bead shearing approach, as described (Taliercio and Boykin 2007). Initially, ovules were visually inspected for cell damage and the fibers were visually monitored under microscope to avoid contaminating epidermis cells. Subsequent RNA extractions were performed using a hot borate method (Wilkins and Smart 1996). RNA quality was confirmed on a BioAnalyzer (Agilent, Palo Alto, CA). Equimolar amounts of RNA (A260) from each separate replicate were pooled into a single sample for the six nonfiber samples destined to become part of the transcriptionally complex reference sample.

RNA amplification and labeling

For microarray analyses, an indirect labeling procedure of amplified aminoallyl a-RNA was used. Amplified-RNA (aRNA) was created for the reference RNA sample and for each fiber growth stage using the TargetAmpTM 1-Round Aminoallyl-aRNA Amplification kit (Epicentre Biotechnologies, Madison, WI, USA). About 0.5 μg of total RNA was used as starting material for one round of aRNA amplification resulting in 20–60 μg of aRNA.

Two dyes, Cy3 and Cy5, were coupled to 8 µg aliquots of aRNA using the Post-Labeling Aminoallyl-aRNA Cy-Dye reactive dyes (Amersham Biosciences, Pittsburgh, PA, USA). The Cy3- and Cy5-labeled aRNA probes were purified using the Qiagen RNA easy Mini kit (Qiagen, Germantown MD, USA). Labeled aRNA products were analyzed for purity and yield (260 nm) using the NanoDrop spectrophotometer and for incorporation of Cy3 (550 nm) and Cy5 (650 nm) dyes.

Microarray hybridization and image analysis

For microarray hybridization, 300 ng of Cy3- and Cy5-labeled aRNA was used per slide. Slides were pre-hybridized using the Pronto!TM Plus system protocol (Promega Corporation, Madison WI, USA) with minor modifications as described below. Slides from each replicate were immersed in 200 ml of Pronto Universal Pre-Soak solution containing 2 ml of liquid sodium borohydride for 20 min at 42°C. Slides were transferred to fresh containers with Wash Solution 2 at room temperature for 2 min and were then immersed in 200 ml of hybridization buffer (5× SSC; 0.1× SDS; BSA 0.1 mg/ml). Slides were incubated with fresh Wash Solution 2 at room temperature for 2 min, and washed two additional times with Wash Solution 3 at room temperature for 2 min each. Finally slides were immersed in nuclease-free water and dried by centrifugation at 1,600*g* for 3 min. All hybridizations and post-hybridization washes were performed as described in the Pronto!TM Plus system protocol.

Microarray images were captured using an arrayWoRx® Biochip Reader (Applied Precision, Issaquah, WA, USA) using a light exposure period of 0.5 s for each channel (Cy5 and Cy3) at ~10 µm resolution. GenPix® Pro (v 5.1, Molecular Devices, Sunnyvale, CA, USA) was used to manually align each block of feature positions to the microarray hybridization images. The signal intensity values were quantified as the background-adjusted median of pixels within the area of each feature's spot circle.

Experimental design and statistical analysis

For each biological replication of each of the five fiber developmental time-points, hybridizations were performed using each fiber developmental stage paired against the same reference sample. With four biological replications, five time-points, and two dye swaps, we generated a total of 20 microarrays. Statistical analyses were performed using R and SAS statistical software (code available upon request).

Background-corrected signal intensity data were log (base 10) transformed and median-centered. For purposes of estimating whether a gene was expressed (see below), we transformed each normalized median value by adding the lowest value in each channel to the actual log/normalized intensity values, thereby forcing all intensity values to be positive. Two approaches were utilized to diagnose "presence" or "absence" of gene expression: (1) *t*-tests, to determine if mean intensity values were significantly different from zero, and (2) comparisons to ten noncotton control genes (SpotReport® Oligo Array Validation System, Stratagene, CA, USA). For the latter, log-intensities formed a bimodal distribution that showed two genes with high spot intensities values (data not shown); these two genes were omitted from subsequent analysis.

Calculated ratios of the genes inferred as expressed in fibers relative to the mixed reference samples (PM) used this formula:

$$\mathrm{PM} = \frac{\text{Percentage of fiber genes inferred as expressed}}{\text{Percentage of reference sample genes inferred as expressed}} \times 100$$

Log (base 10) transformed, median-normalized values of the 13,178 genes were examined for expression differences between each fiber developmental stage and the reference sample. We considered a standard mixed linear model for the data for each gene as

$$y_{ijkl} = \mu + \delta_i + \tau_j + s_k + r_l + e_{ijkl}$$

where y_{ijkl} denotes the normalized log-scale signal intensity (averaged over duplicate spots) for dye *i*, treatment *j*, slide *k*, and replication *l*; μ denotes an intercept parameter; δ_i denotes the effect of dye *i*; τ_j denotes the effect of treatment *j*; s_k denotes the random effect of slide *k*; r_l denotes the random effect of replication *l*; and e_{ijkl} denotes a random error term that is intended to capture all other sources of variability. Tests for differential expression between fiber and reference samples were conducted using this model. The 13,178 *p*-values from each comparison were converted to *q*-values using the method of Storey and Tibshirani (2003). These *q*-values were used to identify the number of differentially expressed gene for a given comparison when controlling the false discovery rate (FDR) at various levels.

Cluster analysis of fiber gene expression was performed using a k-medoids clustering (Kaufman and Rousseeuw 1990) on the 5,430 genes inferred to be up-regulated in one or more stages of fiber development relative to the mixed sample. For these analyses intensity values were standardized so that each profile had a mean of 0 and a standard

deviation of 1. The number of clusters was determined using the gap statistic approach (Tibshirani et al. 2001).

We used Blast2GO (http://www.blast2go.de/) to identify biochemical pathways involved in cotton fiber development and to calculate the statistical significance of each pathway. Blast2GO includes the Gossip package (Bluthgen et al. 2005) for statistical assessment of annotation differences between two sets of sequences, using Fisher's exact test for each GO term. False discovery rate (FDR) controlled *p*-values were corrected for all differentially significant metabolic pathways.

Results and discussion

Cotton fiber transcriptome analysis

Simple mechanical detachment of fibers by treatment with liquid nitrogen and glass beads proved effective in isolating populations of fiber cells, from which RNA extraction yielded fiber-enriched RNAs from developmental time-points as early as 2 DPA, when fiber length is less than 1 mm (Fig. 1). RNAs isolated from five developmental time-points, 2, 7, 10, 20, and 25 DPA, were amplified and hybridized to cotton oligonucleotide microarrays containing 13,178 probes derived from deep EST sampling of diverse tissues and organs (Udall et al. 2007). Microscopic examination of fibers showed that at 20 DPA, fibers were undergoing transition to secondary wall synthesis (data not shown). Fibers extracted from this time-point represent the transition stage. Four biological replicate comparisons between the fiber developmental stages and a transcriptionally diverse reference sample derived from equimolar RNA mixtures of six different cotton organs showed that a high proportion of the cotton genome is transcribed. Two different *t*-test analyses demonstrated this point, the comparison of normalized mean hybridization signal intensity of each spot to zero, and their comparison to a set of eight negative control spots. These data revealed that 80.0–84.2% of the fiber genes are differentially expressed relative to zero and 53–67% are differentially expressed relative to the presumably negative control genes (Table 1). The foregoing estimates are somewhat arbitrary, in that it is difficult to define a transcription state as "on" or "off", and also because of errors caused by nonspecific hybridization and cross-hybridization. Accordingly, to gain further insight into the percentage of the transcriptome expressed at each developmental time-point in fibers, we calculated ratios of the percentage of genes inferred as expressed in fibers relative to the percentage of genes inferred as expressed in mixed reference samples. Our hypothesis was that inference errors caused by nonspecific hybridization and cross-hybridization would be similar for fiber and mixed reference samples. Using this approach, 75.4–94.8% of the genes are inferred as expressed relative to the reference samples (Table 1). Moreover, calculating the percentages of genes that were highly expressed (genes significantly overexpressed relative to the general experimental median value) showed that in some cases, like in 2 DPA, over than 100% of the genes inferred as expressed in the fiber relative to the reference sample (data not shown). Given that the total number of genes in the cotton genome may be approximately 40,000 (Rabinowicz et al. 2005), the 13,178 genes tested here include perhaps 30% of the total genic diversity. For this fraction, therefore, we conclude that the transcriptome of this single cell type is extraordinarily diverse, with most genes expressed at one or more stages of fiber development.

Fig. 1 Isolation of fiber cells from ovules. Cotton ovules were harvested 2 days after flowering (*left*) and fibers were removed using a liquid nitrogen/glass bead shearing method (modified from Taliercio and Boykin 2007). *Right* "naked" ovule after the procedure. Isolated fiber cells with glass beads are shown in the middle

Table 1 Gene expression in cotton fiber and a mixed reference sample

	N1	PT1	PM1	N2	PT2	PM2
2 days	11,094	84.2	93.1	8,818	66.9	94.8
7 days	10,278	80.0	86.2	7,774	59.0	83.6
10 days	10,814	82.1	90.7	7,006	53.2	75.4
20 days	11,078	84.0	93.0	7,644	58.0	82.2
25 days	10,716	81.3	89.9	8,312	63.1	89.4
Reference sample	11,913	90.4	–	9,294	–	

N1 denotes the number of genes expressed at levels significantly different from zero ($P < 0.01$), PT1 indicates the percentage of total genes expressed relative to the total number of genes on the chip, and PM1 is the ratio of $PT1_{fiber}/PT1_{reference}$. N2 denotes the number of genes expressed at levels significantly different from those of the control genes ($P < 0.01$); PT2 indicates the percentage of total genes expressed relative to the total number of genes on the chip, and PM2 is the ratio of $PT2_{fiber}/PT2_{reference}$

Dynamic changes across cotton fiber development and differentiation

We generated a cotton transcriptome database by comparing the fiber-cell transcript pool with that of the reference sample derived from mixed tissues. Compared with the reference samples, there were 6,802 up-regulated genes and 4,714 down-regulated genes across fiber developmental stages (Table 2). This shows that under common statistical thresholds for microarray experiments (q-values ≤ 0.05), 51.6% of the genes are up-regulated in fiber cells when compared to the mixed reference sample at one or more time-points during fiber development. These results indicate that about half the transcriptome can be shown to be up-regulated in the fiber, emphasizing the fact that most of the genome is not only "expressed" but is actually up-regulated relative to a complex mix of other cell types. Given that many other genes, perhaps thousands, are also expressed but are not up-regulated relative to this reference, a conservative conclusion is that greater than half of the genome is expressed at one or more stages during the life of this single cell.

Few other studies offer comparable analyses of the transcriptome of single cell types. In a study of the transcriptome of human oocytes (Kocabas et al. 2006), amplified human metaphase II oocyte mRNA was compared to a reference sample consisting of a mixture of total RNA from 10 different normal human tissues. Compared with reference samples, there were 5,331 transcripts significantly up-regulated and 7,074 transcripts significantly down-regulated. These results mirror our own, showing that a single cell in both animal and plant systems may express many thousands of genes that are differentially expressed relative to more complex admixtures. Other studies have evaluated the transcriptome of developing haploid pollen cells in *Arabidopsis*, in which the pollen tube transcriptome was reduced compared to vegetative tissues, with a high proportion of enriched or selectively expressed genes families important for pollen tube growth (Honys and Twell 2004; Pina et al. 2005). In one of these studies (Honys and Twell 2004), when the pollen transcriptome was compared to seven sporophytic tissues, it was shown that 61.9% of the genome is transcribed in at least one developmental stage. Here we studied gene expression during development of one cell type in the sporophyte, including the most metabolically active stages of rapid fiber cell elongation and primary and secondary wall synthesis. Interestingly, our results, describing another asymmetrically growing cell, yield comparable values.

To uncover patterns of co-expression during fiber development we conducted cluster analysis on the 6,802 genes that were up-regulated relative to the reference sample. K-medoids clustering (Kaufman and Rousseeuw 1990) was used to group the 6,802 gene expression patterns into seven clusters (Fig. 2b, Supplemental Table 1). Gene ontology (GO) was used to assign putative functions to cotton genes targeted by the oligonucleotide probes (Fig. 2c and Supplemental Table 2). Compared to the reference sample, clusters 1 and 2 represent genes up-regulated at 2 DPA, with decreased expression at later developmental time-points, hence representing the initial stage of cotton fiber development. Genes from these clusters belong to GO families representing biosynthesis of proteins and carbohydrates, ribonucleases, ribosomal RNA/proteins and cytoplasm organization, including the organization and constitution of the endoplasmatic reticulum (ER). These processes are in agreement with a new study (Taliercio and Boykin 2007) that identified an increase in genes related to ER construction and ER size in fiber initials (as early as 1 DPA) in comparison to a *fiberless* mutant and 10 DPA normal fibers. Genes connected to respiratory functions and energy production are also overrepresented, consistent with a hypothesis that fibers are exceptionally active metabolically at this stage.

The foregoing results are compatible with another microarray analysis showing energy and cell components as important processes at the elongation phase of fibers (Arpat et al. 2004). In that experiment, two time points (10 and 24 DPA) were studied, representing primary and secondary wall synthesis stages. Our results show that energy and metabolism processes are more reflective of the initial stage rather than the elongation stage. Cluster 3 contains a quite

Table 2 Number of differentially expressed, up- and down-regulated genes in cotton fiber in comparison to reference samples

	$q < 0.1$		$q < 0.05$		$q < 0.02$		$q < 0.01$	
	N	Total (%)	N	Total (%)	N	Total (%)	N	Total (%)
Differentially expressed genes	11,431	86.74	10,533	79.93	9,447	71.69	8,670	65.79
Up-regulated genes	7,557	57.35	6,802	51.62	5,978	45.36	5,430	41.21
Down-regulated genes	5,274	40.02	4,714	35.77	4,107	31.17	3,718	28.21

Median-normalized values of the 13,178 genes were examined for expression differences between each fiber developmental stage and the reference sample. Numbers of differentially expressed genes represent sums across developmental stages, using different threshold values (q) for controlling False Discovery Rates (FDR)

large group of around 600 genes exclusively up-regulated at 7 DPA; genes overrepresented in this cluster encode components of the ubiquitin/proteasome protein catabolism complex involved in signaling pathways such as hormone-regulated processes, cell cycle control, photomorphogenesis, and senescence (Schaller 2004). This is the first time this group of the ubiquitin/proteasome complex has been observed in developing cotton at this stage. Genes included in clusters 4 and 5 are overexpressed during the time period of most rapid cotton fiber cell elongation (between 10 DPA and 20 DPA) (Applequist et al. 2001). These clusters are enriched for genes involved in water channel activity, microtubule and cytoskeleton formation, and regulation of the cell cycle, processes previously described as

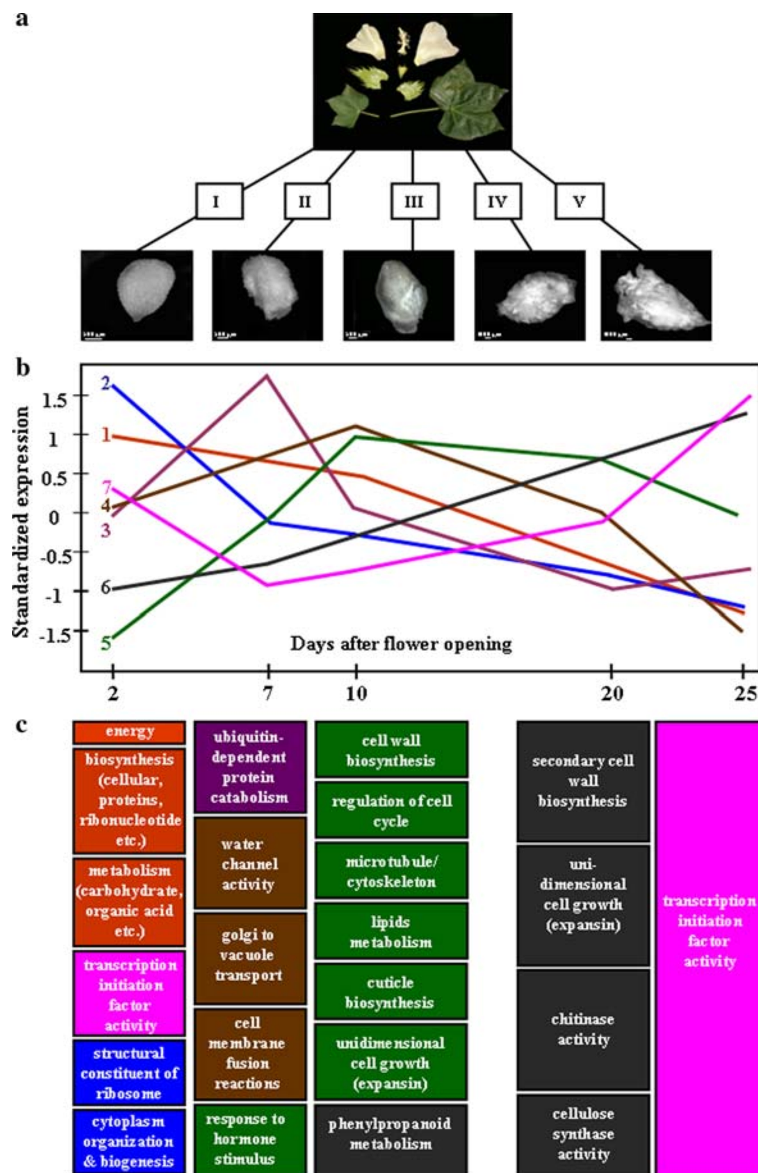

Fig. 2 Cotton fiber-cell gene expression during development. (**a**) Microarray analyses were performed comparing a heterogeneous mix of leaves, stems, petals, anthers, calyx, and bracts to isolated fiber cells from five developmental stages (2, 7, 10, 20 and 25 days post-anthesis). (**b**) Cluster analysis 6,802 of genes that were up-regulated relative to the reference sample at one or more time-points during fiber development revealed seven statistically significant clusters (1–7) representing different gene expression patterns. Each line on the graph represents the arithmetic mean among genes in each cluster. (**c**) Putative function of differentially expressed gene classes involved in cotton fiber development. Gene family and biological process are shown under each developmental stage for those classes exhibiting the greatest difference in comparison to the previous stage (except for 2 DPA, where the most significantly overrepresented gene classes in compare to 7 DPA are shown). Each family or biological process is shown using the color matching its representative cluster. The list in the figure is partial, including only the most significant processes (see complete list of genes in Supplemental Table 1, and list of over representative families in Supplemental Tables 2 and 3; see also http://cottonevolution.info/)

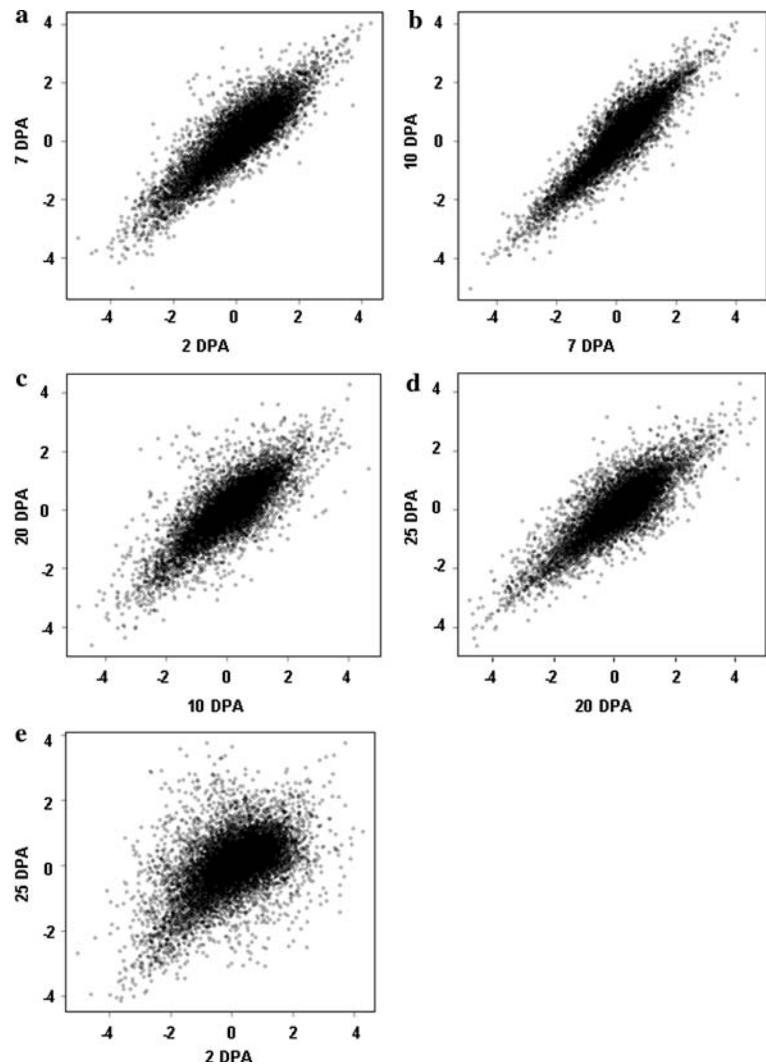

Fig. 3 Correlation of gene expression between adjacent time-points during cotton fiber development. For each spot on each microarray the difference in log-normalized values was calculated between the fiber and reference samples. Values presented are means of four biological replications. (**a**) Correlation between 2 DPA and 7 DPA. (**b**) Correlation between 7 DPA and 10 DPA. (**c**) Correlation between 10 DPA and 20 DPA. (**d**) Correlation between 20 DPA and 25 DPA (**e**) Correlation between values calculated for the two most temporally distant developmental stages (2, 25 DPA)

participating in cotton fiber elongation (Smart et al. 1998; Whittaker and Triplett 1999; Taliercio et al. 2005; Liu et al. 2006). Other gene families expressed in these clusters include components of cell wall biosynthesis, lipid metabolism, and cuticle biosynthesis, indicating the essential role of these cellular components during rapid elongation. Cluster 6 contains genes that gradually increase in expression during fiber development, with an over-representation of genes characteristic of primary and secondary cell wall biosynthesis, chitinase activity, and cellulose synthase activity, processes known to be enhanced during the later stages of fiber development (Arpat et al. 2004; Zhang et al. 2004). Genes belonging to cluster 7 show increased expression in two different stages, 2 and 25 DPA, and as a class include an over-representation of signal transduction pathway genes that may play a role in initiation and maturation processes, respectively. An example of expression patterns of eight well-characterized fiber development related genes are presented in Supplemental Fig. 1.

Gene expression analyses between adjacent time-points during fiber development

To better appreciate patterns of change in global expression during fiber development, we tracked differences in expression between the fiber and reference sample treatments for all genes at all developmental stages (Fig. 3). As expected, gene expression in fiber from adjacent time-points during development was more similar than between ends of the developmental spectrum; for example, genes up-regulated (or down-regulated) at 2 DPA relative to the reference sample also tended to be up-regulated (down-regulated) at 7 DPA (Fig. 3a). Also, the correlation between the two most temporally distant developmental stages (2 DPA vs. 25 DPA) is much lower (Fig. 3e), consistent with our hypothesis and indicating a gradual but extensive alteration in the cotton fiber transcriptome during development. Of course, some individual gene expression profiles changed abruptly during development, in accordance with the

classical model of modular development of the cotton fiber (Wilkins and Jernstedt 1999), and as evidenced by points located out of the main group in Fig. 3. On an overall basis, however, the transcriptome is shown here to change more gradually. More interesting, perhaps, is the large transcriptional difference between the loop end-points, underscoring the cumulative effects of this gradual change in the transcriptome during development of this single cell type.

Differential GO annotations between consecutive developmental stages are presented in Supplemental Table 3 and Fig. 2. The results of these analyses parallel to those of the cluster analyses, but provide additional perspective on important biological processes. Genes up-regulated at 2 DPA compared to 7 DPA, for example, are biased toward cellular biosynthesis, ribosome structure, and transcription. Up-regulated genes at 7 DPA (compared to 2 DPA) include those implicated in hormone responses, such as gibberellic acid and ethylene-mediated signaling. By using microarray analysis on wild-type cotton ovules and fiberless mutant, Shi et al. (2006) has previously described the importance of hormone response in cotton fiber elongation. Another gene family overexpressed at 7 DPA is involved with cell membrane fusion reactions (SNARE binding, diacylglycerol O-acyltransferase activity, golgi to vacuole transport), suggesting that processes involved in increasing membrane production are required for the rapid unidirectional cell elongation that characterizes the cotton fiber. Additional biological processes are highlighted during fiber elongation, such as phenylpropanoid biosynthesis (which stops before 20 DPA) and expansin metabolism (Fig. 2). The expansin gene family is known as a key player in cell loosening and is considered to be important during cotton fiber elongation (Ruan et al. 2001).

The foregoing analyses demonstrate that global patterns of gene expression are highly dynamic during development of this single type of epidermal plant cell, with most of the genome being transcribed at one or more stages of cell growth. As we did not sample earlier stages of fiber initiation nor the later stages of maturation and cell death, both the conclusion of high transcriptome representation and that of clustered expression likely underestimate the true levels of both phenomena.

Biological process during fiber development

To gain insight into the biological processes associated with different stages of fiber development studied (from 2 DPA to 25 DPA), we inspected the lists of genes that were overexpressed at all developmental stages in fiber with comparison to the reference tissues. Our initial goal was to elucidate processes controlling fiber development that were not stage-specific. This procedure led to the recognition of 1,111 genes that were up-regulated in fiber, compared to the reference sample, throughout the developmental period studied (2–25 DPA). Assessment of annotation differences (Fisher's exact test) showed that genes in this group are biased toward processes including intracellular transport, protein polymerization, organic anion transporter activity, and vesicle coating (Table 3).

The most remarkable group of genes detected in this fashion are those encoding vesicle coating proteins, in which 17 of 19 genes on the microarray were up-regulated in fiber. The importance of vesicle coating and trafficking to cell polarization and unilateral growth has been widely reviewed (Aroeti et al. 1998; Neumann et al. 2003; Macara and Spang 2006; Campanoni and Blatt 2007). The basic mechanism for vesicle transport membrane trafficking is similar between animals and plants, involving similar regulatory and structural proteins (Sanderfoot and Raikhel 1999; Pratelli et al. 2004). In animals, it has been shown

Table 3 Classes of statistically overrepresented and under-represented biological processes throughout fiber development, relative to the reference sample. *N*—number of genes in each group

GO	Name	FDR	*N* in test	*N* in reference	Over/Under
GO:0046907	Intracellular transport	1.92E-08	47	156	Over
GO:0030135	Coated vesicle	2.73E-08	17	19	Over
GO:0051258	Protein polymerization	4.19E-04	8	10	Over
GO:0008514	Organic anion transporter activity	2.15E-06	9	17	Over
GO:0003677	DNA binding	7.38E-06	40	1062	Under
GO:0009725	Response to hormone stimulus	7.18E-05	25	590	Under
GO:0009651	Response to salt stress	0.00268	2	134	Under
GO:0009723	Response to ethylene stimulus	0.003	5	192	Under
GO:0009733	Response to auxin stimulus	0.00356	2	130	Under
GO:0009607	Response to biotic stimulus	0.04643	8	192	Under
GO:0009737	Response to abscisic acid stimulus	0.0468	14	281	Under
GO:0009751	Response to salicylic acid stimulus	0.04787	5	143	Under

Table 4 List of Oligo Identification Numbers (ID), cotton gene target and gene product names of 17 genes belonging to GO: 0030135 (coated vesicle) (Table 3)

Oligo ID	Oligo gene target	Blast results from Arabidopsis nr database
625	Cotton12_01187_01	Sintaxin
1197	Cotton12_06960_01	(Q93ZN7) T25N20_16 (COPII domain)
1475	Cotton12_00008_11	(Q6ZGX8) Putative clathrin coat assembly protein
3573	Cotton12_19390_01	Nonclathrin coat protein zeta1-COP
4054	Cotton12_17795_01	(Q5JMS0) Putative syntaxin 6
4760	Cotton12_03874_01	Expressed protein (transporter activity)
6206	Cotton12_03609_01	(Q8LJR4) Syntaxin
6565	Cotton12_40610_01	(Q7X9R1) Clathrin coat assembly protein
9844	Cotton12_25700_01	(Q6ZDG9) Putative SEC23
10112	Cotton12_08998_01	(Q9LW87) Coatomer protein complex
10262	Cotton12_11199_01	(Q39834) Clathrin heavy chain
10525	Cotton12_12489_01	(Q6JJ39) Putative adapitin protein
10601	Cotton12_04867_01	(Q9FXB1) Putative clathrin-associated adaptor
10696	Cotton12_01124_01	(Q9SB50) Clathrin coat assembly like protein
11782	Cotton12_11076_01	(Q8S0N4) Vesicle transport SNARE protein-like
12082	AW587452	Coatomer delta subunit (Delta-coat protein)
9396	Cotton12_38560_01	(Q93ZN7) At1g05520 (COPII domain)

that vesicle transport and localization play a key role in the distal end of the single-celled neuron elongation by fusion of intracellular, specific and dynamic vesicles specialized for plasmalemmal expansion in the growth cone (Pfenninger and Friedman 1993; Hirling et al. 2000; Steiner et al. 2004). In plants, most studies of the role of vesicle coating and transport regulation on cell tip growth have been on root hairs and pollen tubes (reviewed by Campanoni and Blatt 2007). It is known that membrane trafficking is linked to ion gradients and is fundamental to tip growth, particularly in supplying lipid and protein to the new plasma membrane and cell wall. It also has a complementary role for endocytosis in retrieving excess membrane and in recycling various protein fractions. Recent studies even suggest that proteins in the coated vesicle, like SNARE super-family (vesicle soluble NSF attachment protein receptor), are essential not only as housekeeping or "greasing" factors but for cell signaling as well. In this regard, coated vesicle trafficking is considered to be an active regulator of ion channel turnover and activity through its localization in the membrane during polar cell growth (Leyman et al. 1999; Pratelli et al. 2004; Sutter et al. 2006).

Cotton fibers, like plant root hairs and pollen tubes and like animal neurons, are among the most rapidly elongating cells in nature. Unlike the tip-growing cells of pollen tubes and root hairs, cotton fibers elongate by a diffuse growing mechanism (Seagull 1990). Still, to support its remarkable growth rate cotton fibers must possess an appropriate turnover of cytoskeleton, cytoplasmic structures, and organelles, and an accelerated traffic of membrane vesicles that deliver membrane and cell wall material to the points of deposition. Taliercio and Boykin (2007) demonstrated the importance of terms associated with "membrane bound organelles" and "intrinsic to membrane" in initials (1 DPA) and elongating (10 DPA) fibers. Here, we show that these processes are important through secondary wall biosynthesis and maturation. A list of genes encoding proteins participating in vesicle coating and transporting is presented in Table 4. Some of the gene products are known to be associated with cell growth and elongation. Syntaxin, for example, is a common golgi-localized protein that is required for ER-golgi traffic, and has been shown to regulate trafficking of chitin synthase III to polarized growth sites in yeast (Holthuis et al. 1998), and golgi reassembly following cell division in mammalian cells (Rabouille et al. 1998). In *Arabidopsis*, mutants reported to display defects in pollen development are members of the Syp2 and Syp4 subfamilies of syntaxin-like proteins. These two subfamilies are thought to contribute to vesicle traffic between the vacuole and trans-golgi network (Sanderfoot et al. 2001). Clathrin heavy chain is another gene product that has been shown to be localized to rapidly growing pollen tubes (Blackbourn and Jackson 1996). Vesicle transport SNARE protein is another example, mediating the final stages of vesicle fusion throughout the endomembrane system and at the plasma membrane, and is also considered to play a regulatory role (Pratelli et al. 2004). Interestingly, to date, no clear connection has yet been made between plant cell tip growth and SNARE proteins (Campanoni and Blatt 2007) and its particular function in regulating cotton fiber growth, as much as other vesicle coating members, is a subject for further investigations.

Acknowledgments R.H., J.A.U., and J.F.W. conceptualized the experiment. R.H. managed the bench experiments and the data analysis. EH conducted part of the bench experiments. J.A.U., L.F., and R.R. contributed to data analysis. All authors assisted in drafting the manuscript. The authors thank Ryan Percifield for technical assistance, Alan Gingle for database management, Prof. Candace Haigler for help in secondary cell microscopic observations and the US National Science Foundation Plant Genome Program for financial support.

References

Applequist WL, Cronn R, Wendel JF (2001) Comparative development of fiber in wild and cultivated cotton. Evol Dev 3:3–17

Aroeti B, Okhrimenko H, Reich V, Orzech E (1998) Polarized trafficking of plasma membrane proteins: emerging roles for coats, SNAREs, GTPases and their link to the cytoskeleton. Biochim Biophys Acta 1376:57–90

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

Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN (2003) A gene expression map of the *Arabidopsis* root. Science 302:1956–1960

Blackbourn HD, Jackson AP (1996) Plant clathrin heavy chain: sequence analysis and restricted localization in growing pollen tubes. J Cell Sci 109:777–786

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

Campanoni P, Blatt MR (2007) Membrane trafficking and polar growth in root hairs and pollen tubes. J Exp Bot 58:65–74

Casson S, Spencer M, Walker K, Lindsey K (2005) Laser capture microdissection for the analysis of gene expression during embryogenesis of *Arabidopsis*. Plant J 42:111–123

Demura T, Tashiro G, Horiguchi G, Kishimoto N, Kubo M, Matsuoka N, Minami A, Nagata-Hiwatashi M, Nakamura K, Okamura Y, Sassa N, Suzuki S, Yazaki J, Kikuchi S, Fukuda H (2002) Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. Proc Natl Acad Sci U S A 99:15794–15799

Galbraith DW, Birnbaum K (2006) Global studies of cell type-specific gene expression in plants. Annu Rev Plant Biol 57:451–475

Hirling H, Steiner P, Chaperon C, Marsault R, Regazzi R, Catsicas S (2000) Syntaxin 13 is a developmentally regulated SNARE involved in neurite outgrowth and endosomal trafficking. Eur J Neurosci 12:1913–1923

Holthuis JCM, Nichols BJ, Pelham HRB (1998) The syntaxin Tlg1p mediates trafficking of chitin synthase III to polarized growth sites in yeast. Mol Biol Cell 9:3383–3397

Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. Genome Biol DOI:10.1186/gb-2004-5-11-r85

Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis. Wiley, New York

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

Kocabas AM, Crosby J, Ross PJ, Otu HH, Beyhan Z, Can H, Tam WL, Rosa GJM, Halgren RG, Lim B, Fernandez E, Cibelli JB (2006) The transcriptome of human oocytes. Proc Natl Acad Sci U S A 103:14027–14032

Leonhardt N, Kwak JM, .Robert N, Waner D, Leonhardt G, Schroeder JI (2004) Microarray expression analyses of *Arabidopsis* guard cells and isolation of a recessive abscisic acid hypersensitive protein phosphatase 2C mutant. Plant Cell 16:596–615

Leyman B, Geelen D, Quintero FJ, Blatt MR (1999) A tobacco syntaxin with a role in hormonal control of guard cell ion channels. Science 283:537–540

Liu D, Zhang X, Tu L, Zhu L, Guo X (2006) Isolation by suppression-subtractive hybridization of genes preferentially expressed during early and late fiber development stages in cotton. Mol Biol 40:741–749

Macara IG, Spang A (2006) Closing the GAP between polarity and vesicle transport. Cell 125:419–421

Nakazono M, Qiu F, Borsuk LA, Schnable PS (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. Plant Cell 15:583–596

Nawy T, Lee JY, Colinas J, Wang JY, Thongrod SC, Malamy JE, Birnbaum K, Benfey PN (2005) Transcriptional profile of the *Arabidopsis* root quiescent center. Plant Cell 17:1908–1925

Neumann U, Brandizzi F, Hawes C (2003) Protein transport in plant cells: in and out of the golgi. Ann Bot 92:167–180

Pfenninger KH, Friedman LB (1993) Sites of plasmalemmal expansion in growth cones. Brain Res Dev Brain Res 71:181–192

Pina C, Pinto F, Feijo JA, Becker JD (2005) Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol 138:744–756

Pratelli R, Sutter JU, Blatt MR (2004) A new catch in the SNARE. Trends Plant Sci 9:187–195

Rabinowicz PD, Citek R, Budiman MA, Nunberg A, Bedell JA, Lakey N, O'Shaughnessy AL, Nascimento LU, McCombie WR, Martienssen RA (2005) Differential methylation of genes and repeats in land plants. Genome Res 15:1431–1440

Rabouille C, Kondo H, Newman R, Hui N, Freemont P, Warren G (1998) Syntaxin 5 is a common component of the NSF- and p97-mediated reassembly pathways of golgi cisternae from mitotic golgi fragments in vitro. Cell 92:603–610

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

Sanderfoot AA, Raikhel NV (1999) The specificity of vesicle trafficking: coat proteins and SNAREs. Plant Cell 11:629–642

Sanderfoot AA, Pilgrim M, Adam L, Raikhel NV (2001) Disruption of individual members of *Arabidopsis* syntaxin gene families indicates each has essential functions. Plant Cell 13:659–666

Schaller AA (2004) cut above the rest: the regulatory function of plant proteases. Planta 220:183–197

Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M Sandberg G (2004) A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. Plant Cell 16:2278–2292

Seagull RW (1990) Tip growth and transition to secondary cell wall synthesis indeveloping cotton hairs. In: Heath IB (ed) Tip growth in plant and fungal cells. Academic Press, San Diego, pp 261–284

Shi Y, Zhu S, Mao X, Feng J, Qin Y, Zhang L, Cheng J, Wei L, Wang Z, Zhu Y (2006) Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell 18:651–664

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

Steiner P, Kulangara K, Sarria JCF, Glauser L, Regazzi R, Hirling H (2004) Reticulon 1-C/neuroendocrine-specific protein-C interacts with SNARE proteins. J Neurochem 89:569–580

Storey JD, Tibshirani R (2003) SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In: Parmigiani G, Garrett ES, Irizarry RA, Zeger SL (eds) The analysis of gene expression data: methods and software. Springer, New York, pp 272–290

Sutter JU, Campanoni P, Tyrrell M, Blatt MR (2006) Selective mobility and sensitivity to SNAREs is exhibited by the *Arabidopsis* KAT1 k+ channel at the plasma membrane. Plant Cell 18:935–954

Taliercio EW, Hendrix B, Stewart JM (2005) DNA content and expression of genes related to cell cycling in developing *Gossypium hirsutum* (Malvaceae) fibers. Am J Bot 92:1942–1947

Taliercio EW, Boykin D (2007) Analysis of gene expression in cotton fiber initials. BMC Plant Biol DOI:10.1186/1471-2229-7-22

Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Ser B Stat Methodol 63:411–423

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

Whittaker DJ, Triplett BA (1999) Gene-specific changes in alpha—tubulin transcript accumulation in developing cotton fibers. Plant Physiol 121:181–188

Wilkins TA, Jernstedt JA (1999) Molecular genetics of developing cotton fibers. In: Basra AS (ed) Cotton Fibers. Haworth Press, New York, pp 231–267

Wilkins TA, Smart LB (1996) Isolation of RNA from plant tissue. In: Kreig PA (ed) A laboratory guide to RNA isolation, analysis, and synthesis. Wiley-Liss, New York, pp 21–41

Wu Y, Liewellyn DJ, White R, Ruggiero K, Al-Ghazi Y, Dennis ES (2007) Laser capture microdissection and cDNA microarrays used to generate gene expression profiles of rapidly expanding fiber initials cells on surface of cotton ovules. Planta DOI 10.1007/s00425-007-0580-5

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