TITLE: **Transcriptome Profiling Identifies Genes Preferentially Expressed During Cotton Fiber Initiation Period**

DISCIPLINE: **Genomics and Biotechnology**

AUTHORS:
Wen-Qian Mei
Yong-Hui Shi
Yu-Xian Zhu (Corresponding author)
National Laboratory of Protein Engineering and Plant Genetic Engineering
College of Life Sciences
Peking University
Beijing 100871, China
Phone: 86-10-62751193
Fax: 86-10-62754427
E-mail: zhuyx@water.pku.edu.cn

Transcriptome Profiling Identifies Genes Preferentially Expressed During Cotton Fiber Initiation Period

ABSTRACT

Cotton fiber initiation is a complex biological process including fiber cell protrusion from ovule surface and early expansion, while the regulatory mechanisms are not well understood. Here we obtained 30,154 uniESTs by sequencing a 0 to 10 days post anthesis (DPA) cotton fiber cDNA library. Comparative analysis of transcriptomes by microarray showed that 2980 genes were up-regulated during fiber initiation period and 255 of them are closely related with ovule expansion. Significant expression increase in genes encoding putative transcription factors such as MYB, WRKY, bZIP, bHLH, YABBY,ERF and genes encoding predicted protein kinases including receptor-like kinase, casein kinase, MAP kinase kinase, CTR1-like and protein phosphotase 2C reveal the great importance of both transcriptional and post-translational regulation in fiber cell initiation. KOBAS analysis identified 19 metabolic pathways highly up-regulated in fiber initiation such as ion channels activity, protein and ATP synthesis and carbon fixation, while 18 pathways highly induced in fl ovule expansion more concern about cell cycle signaling and lipids metabolism.

KEY WORDS:

cotton fiber, initiation, transcriptome, transcription factor, protein kinase, metabolic pathway

Cotton fibers are single-celled thichomes differentiated from outer integuments of the ovule. As one of the major natural raw materials for textile industry, they have considerable economic importance. For studying fundamental biological processes, cotton fibers also provide an excellent single-celled model (Wilkins et al., 2004). Fiber development consists of four overlapping stages: fiber initiation, fiber elongation, secondary cell wall deposition and maturation (Ji et al., 2003). The fiber initiation stage occurs from -3 to 3 day post-anthesis (DPA) and is characterized by the enlargement and protrusion of epidermal cells from the ovular surface. Only about 30% of epidermal cells will eventually differentiate into fiber cells. During the fiber elongation period (5-25 DPA.) cells demonstrate vigorous polarized expansion with peak growth rates of >2 mm/day until the fiber reaches its final dimensions. During secondary cell wall deposition (20-45 DPA) cellulose biosynthesis predominates until the fiber contains ~90% cellulose. In the final stage of maturation (45-50 DPA) fibers undergo dehydration and become mature cotton fibers. Taken together, these stages include the biomechanical processes of rapid cell division, differentiation and expansion, secondary cell wall synthesis, suggesting that a large number of genes are involved in the regulation of fiber development.

Several studies have reported genes that are highly or preferentially expressed in cotton fibers and play important roles on different fiber development stage. GhACT1, which is expressed predominantly in fiber cells, participates in fiber elongation process (Li et al., 2005). Antisense suppression of a sucrose synthase gene disrupted fiber elongation, indicating the involvement of osmosis regulation during the elongation stage (Ruan et al., 2003). GaMYB2 (*Gossypium arboretum*), a MYB family transcription factor related to

AtMYBGL1, complements the Arabidopsis glabrous1 (gl1) mutant. Moreover, ectopic expression of GaMYB2 induces a single trichome from the epidermis of Arabidopsis seed (Wang et al., 2004). Using microarray and in vitro culture system analyses has indicated that ethylene is involved in fiber cell elongation (Shi et al., 2006). Antisense-mediated suppression of GhDET2 which encodes a steroid 5d-reductase in brassinosteroid (BR) biosynthesis, inhibited both fiber initiation and fiber elongation (Luo et al., 2007). Collectively these data suggest critical roles for phytohormones in fiber cell development

Microarray is a powerful tool to reveal transcriptional regulation of genes and pathways during fiber development, especially when a large number of cotton ESTs were obtained from public databases such as NCBI and EMBL. Recently, Approximately 185,000 *Gossypium* EST sequences from more than 30 libraries have been amassed and analyzed through international collaboration. 51,107 uniESTs were identified and 33,665 of them were partial or full-length non-repeated coding regions (Udall et al., 2006). Our previous work to compare the transcriptomes of cotton fiber with fuzzless-linless (fl) mutant ovules based on 12,233 uniESTs found that 778 cDNAs Preferentially Expressed in Developing Cotton Fibers (Shi et al., 2006). Another cDNA array analysis of fiber cell elongation and secondary cell wall synthesis also identified several upregulated metabolic pathways during cell elongation and cellulose synthesis (Gou et al., 2007). Proteomic study also found that proteins in carbon partitioning pathway for cellulose synthesis such as fructose kinase (FEK), UDPG-pyrophosphorylase and UDPG-dehydrogenase were highly accumulated in 10 DPA wild-type ovules but not in fl mutant ovules (Xu et al., 2007). An investigation of cotton immature ovules transcriptomes of *Gossypium hirsutum* and naked seed mutant (N1N1)

showed that cotton homologs to transcription factor MIXTA, MYB5, GL2 and 8 genes on auxin, BR, GA and ethylene pathways were induced during fiber initiation but repressed in the mutant, indicating the similar regulation mechanism between cotton fiber initiation and leaf trichome development (Yang et al., 2006).

So far, more molecular and genomic studies shed light on mechanisms responsible for fiber cell elongation than fiber cell initiation. To survey the events crucial for fiber initiation process, we obtained 30,154 uniESTs from 0 to10 DPA ovules of a tetraploid species (*G. hirsutum*). A large number of genes and metabolic pathways upregulated during fiber initiation and ovule expansion processes were identified using a custom-designed cDNA microarray. Analysis of these genes suggest the important role for transcription factors, protein phosphorylation, ion channels during the early stage of fiber development.

RESULTS

30,154 uniESTs in total were obtained by sequencing 110,000 cDNAs from 0 to 10 DPA cotton ovules. A high-quality cDNA library was constructed using RNA samples harvested from 0 to 10 DPA tetraploid wild-type upland cotton ovules with their fiber attached (Lu, 2002). Previously, by random sequencing this cDNA library, 29,992 high quality cotton ESTs were obtained and clustered into 12,223 uniESTs (Shi et al., 2006). In this work, after further sequencing, the total number of cotton ESTs was reached to 110,303 and clustered into 30,154 uniESTs. All the uniESTs were used for PCR amplification and subsequent microarray construction.

Clustering analysis of all expressed genes in fiber initiation process. All the uniESTs were successfully amplified by PCR and printed on amino silane slides to construct

a cDNA microarray. In order to identify genes specifically or preferentially expressed in cotton fiber initiation stage, we hybridize the microarray with RNA samples from wild type ovules harvested at 0, 1, 2, 3 DPA and fl mutant ovules at 3 DPA, Using a -3 DPA wild type ovules sample as a reference. From STM (scanning electron microscope) photos we also see that cotton fiber emerges initially as a bulge at 0 DPA and then elongates by tip growth whereas at -3 DPA and in fl mutant, there is no fiber bulge at all (Fig. 1). After hybridization, A total of 19293 genes passed multiple testing (with false discovery rate [FDR]–corrected P values <0.001 in at least one of the above growth stages) on the MAANOVA program. Expression pattern of these genes were clustered and divided into three primary groups. Genes in group (a) have increased expression levels from 0 DPA to 3 DPA, indicating that their expression has a close connection with fiber initiation. Genes of group (b) showed a decreased expression pattern along with fiber and ovule development. The expression levels of group (c) genes are higher at 0 DPA and than decrease from 1 DPA to 3 DPA. We consider these genes may involved in fiber cell protrusion from ovule surface (Fig. 2A). Notably, some transcription factors, protein kinases and protein phosphatases were detected in both group (a) and group (c) (Fig. 2B,C). The expression levels of transcription factors which belong to MYB, zinc finger, bZIP, WD40, ERF gene families respectively and protein kinases such as receptor-like, NAK-like, ethylene CTR1-like and casein kinase increased significantly during fiber initiation (from 0-3 DPA) whereas none of them were induced in 3DPA fiber less (fl) mutant. Other transcription factors including family members of bHLH, YABBY, WRKY; MAP kinase kinase, another receptor-like kinase and protein phosphatase 2C (PP2C) have higher levels at 0 DPA and then decreased along the early stage

development. But in 3DPA fl mutant, no decreased expressions were seen, suggesting their roles maybe more concern about fiber cell protrusion.

The quality of the microarray data was assessed by correlation coefficients. The r values calculated from different samples were used as measures of biological reproducibility, and the r values obtained from swap-dye experiments of individual biological samples were used as measures of technical reproducibility (Table 1). All of them are higher than 0.85 and 77% are higher than 0.90, indicating excellent reproducibility of our experiment.

Genes dramatically induced during fiber initiation and ovule expansion. A total of 2980 genes showed significant increase of expression in 3DPA wild type cotton ovules while 613 genes were highly expressed in 3DPA fl mutant comparing with -3DPA ovules (FDR-adjusted P-value<0.01). There are 255 genes appeared in both of the two groups and they are considered ovule expansion-related genes (Fig. 3). The most highly expressed genes (>10 times) during wild type fiber initiation and ovule expansion were listed in Table 2. Cell wall proteins such as alpha-expansins and proline-rich proteins have particularly high levels of expression during the initiation stage. 28% genes in the table showed "no hits" in their annotation which indicated they are new genes and their functions need further investigation.

Compare with wild type ovules, 10 times higher expressed genes during fl mutant ovule expansion are obviously less (Table 3).

Several metabolic pathways are significantly upregulated during fiber initiation and ovule expansion. All the ESTs with expression in the chip were analyzed using KOBAS (for KEGG Orthology Based Annotation System) (Mao et al., 2005) to identify the metabolic

pathways upregulated during fiber initiation and ovule expansion. Metabolic pathways significantly upregulated ($P < 0.05$) during the wild-type fiber initiation and fl ovule expansion are 19 and 18 respectively (Table 4). Both of them have methionine metabolism pathway but other pathways are quit different. In wild-type fiber initiation process, upregulated metabolic pathways are mainly about protein and ATP synthesis, carbon fixation and ion channels activity, while during fl ovules expansion, cell cycle signaling and lipids metabolism are show great changes.

DISCUSSION

Although a global collection of 185,000 cotton ESTs were reported in April 2006, only 38% of the ESTs (69,853 ESTs) were from tetraploid upland cotton *Gossypium hirsutum* which is widely used for industrial cotton lint production (Udall et al., 2006). In this work, we obtained 110,303 ESTs derived from 0 DPA to 10 DPA cotton (*Gossypium hirsutum*) ovules covering fiber initiation and active elongation stage. ESTs assembling process produced 30,154 unigenes and were used in cDNA microarray procedure.

The majority published data of genome-wide expression analysis using microarray are focus on fiber elongation and second wall synthesis stage (Arpat et al., 2004; Shi et al., 2006; Gou et al., 2007). The molecular events of fiber initiation are poorly understood. So we fabricated a cotton cDNA microarray and hybridized it with RNA samples prepared throughout the period of fiber initiation (-3 DPA and 0-3 DPA). We also used a genetic mutant that produced no fiber cells (fl) to confirm the fiber specificity of the ESTs. After data clustering and processing, we found the expression of some transcription factors and protein phosphorylation regulators (kinases and phosphatases) showed close relations with fiber

differentiation and initiation. The role of cotton MYB transcription factors in cotton fiber development has also been documented (Loguercio et al., 1999; Suo et al., 2003; Wang et al., 2004). TTG2/AtWRKY44 (TRANSPARENT TESTA GLABRA2), the first WRKY gene to be functionally characterized, controls trichome development in leaves and the production of mucilage and tannin in Arabidopsis seed coats (Johnson et al., 2002). The trichomes in the ttg2 mutants are unbranched and reduced in number compared with the wild-type. A basic helix-loop-helix proteins (GL3 or EGL3) also play a role in determining epidermal trichome cell patterning in Arabidopsis leaves (Ramsay and Glover, 2005). YABBY transcription factors involved in abaxial cell fate in the carpel and ovule development in Arabidopsis (Fang et al., 2007). Our result indicated that homologs of these transcription factors in cotton may have the similar function during fiber cell initiation. Protein phosphorylation regulates protein function in post-translational level, its role was of great importance especially in signal transduction pathways. Our findings confirmed that protein phosphorylation was also involved in fiber cell development. Although we didn't find the ethylene synthesis pathway upregulated during fiber initiation, the increased expression of ethylene responsive protein kinase CTR1, MAPKK and ethylene-responsive element binding protein (ERF3) suggested that ethylene signaling may involve in this process.

MATERIALS AND METHODS

Plant materials. Upland cotton (*Gossypium hirsutum* cv Xuzhou 142) and the fl mutant, originally discovered in the same cotton field in China (Zhang and Pan, 1992), were grown in soil mixture in a fully automated green house. Bolls were tagged and the flowering

day was taken as 0 DPA. Ovules were excised from bolls and collected from 3 days post anthesis (-3 DPA) to 3 DPA, then frozen and stored in liquid nitrogen immediately.

Scanning Electron Microscope (SEM). SEM was performed using a modified protocol (Jinsuk et al., 2006). In brief, ovules from -1 to 2 DPA were dissected from bolls of wild-type and the fl mutant. The ovules were fixed in a solution containing 3% each of formaldehyde and glutaraldehyde in 0.1 M sodium cacodylate buffer (pH=7.4) and rinsed in 0.2 M sodium cacodylate buffer (pH = 7.4) three times. The ovules were washed in an ethanol series and dehydrate in 100% ethanol for SEM analysis. The specimens were prepared by critical-point, dried with CO2 consecutively, and mounted by conductive gold paint and sputter coating. The samples were then scanned and analyzed using S-800 (HITACHI) SEM. Images were scanned and stored as TIFF files.

RNA isolation, cDNA library construction, and sequencing. Total RNA samples were prepared from 0 to 10 DPA ovules using a modified hot borate method (Lu, 2002). cDNA was synthesized and cloned into the EcoRI-XhoI sites of the ZAP Express vector using a cDNA synthesis kit, and the ligationmixture was packaged using a ZAP-cDNA Gigapack Gold III cloning kit (Stratagene). Bacterial colonies containing fiber cDNAs, with average insert size of 1.6 to 1.7 kb, were obtained after in vivo excision. A total of 110,303 clones were subjected to single-pass sequencing reactions from the 5' end. Vectors and sequences shorter than 300 bp or containing >1.5% of imprecise nucleotides were removed. The remaining ESTs were compared with the GenBank database using BLASTX (http://www.ncbi.nlm.nih.gov/BLAST/), with 10^{-2} as the cutoff e-value, and named after the

homologous sequences in GenBank. ESTs with e-values higher than 10^{-2} were designated as unknown. EST assembly was performed to obtain uniESTs using Stackpack2.1 software (Christoffels et al., 2001). Clusters and singletons with alternatively spliced isoforms were reassembled manually.

cDNA microarray design and hybridization. Total of 30154 cotton uniEST clones were selected and their inserts were amplified by PCR, using M13 forward and reverse primers. Aliquots (1 µL) of the PCR reactions were analyzed in a 1% agarose gel to verify the success of the PCR. The remaining cDNA was prepared for arraying. Eight sequences derived from intergenic regions of the yeast genome showing no significant homology with any existing cotton sequences in GenBank were used as external controls. These sequences were PCR amplified and cloned into plasmid pSP64Poly(A) (Promega) to produce poly(A)-RNA after in vitro transcription. PCR products representing control sequences were spotted 12 times onto the microarray in different subgrids. RNAs from the 0, 1, 2, 3 DPA wild type as well as 3 DPA fl mutant ovules were compared to the -3 DPA wild type ovule RNAs for fiber initiation gene expression analysis. Probes preparation and slide hybridization were carried out as described previously (Shi et al., 2006).

Image acquisition, data processing and clustering. All microarrays were scanned with a ScanArray Express scanner using ScanArray 2.0 software (Packard Bioscience). We quantified signal intensities of individual spots from the 16-bit TIFF images using GenePix Pro 4.0 (Axon Instruments). The microarray slides were hybridized with RNA prepared from three biological replicate samples for each DPA. As a measure of technical replication, one swap-dye experiment was performed on each of the biological samples so

that a total of six data points were available for every EST on the microarrays. The linear normalization method was used for data analysis, based on the expression levels of 40 cotton housekeeping genes in combination with the yeast external controls. Normalized data was log transformed and loaded into MAANOVA under R environment for multiple testing, by fitting a mixed effects ANOVA model (Wu et al., 2003). Microarray spots with FDR-corrected P values <0.001 in the F-test were regarded as differentially expressed genes. Hierarchical clustering with the average linkage method was employed on all expressed genes in fiber initiation process defined as 0/-3 (0 DPA versus -3 DPA RNA samples, etc.), 1/-3, 2/-3, 3/-3 and 3fl /-3wt. We visualized the cluster data by the Treeview program (Eisenet al., 1998).

Identifying fiber initiation and ovule expansion preferential pathways using KOBAS. We used the software KOBAS (Mao et al., 2005) to identify biochemical pathways involved in cotton fiber initiation and ovule expansion process by setting BLAST similarity cutoff e-values $<1e^{-5}$, rank <10, and sequence identity >30%. We also manually reviewed all identified pathways for quality control. The P value of a particular pathway was set following a hyper geometric distribution. FDR correction was then applied to control the overall Type I error rate of multiple hypotheses testing using GeneTS (2.8.0) in the R (2.2.0) statistics software package (Storey, 2002; Wichert et al., 2004). Pathways with FDR-corrected P values <0.05 were considered statistically significant.

REFERENCES

Arpat, A.B., M. Waugh, J.P. Sullivan, M. Gonzales, D. Frisch, D. Main, T. Wood, A. Leslie, R.A. Wing, and T.A. Wilkins. 2004. Functional genomics of cell elongation in developing cotton fibers. Plant Mol. Biol. 54: 911–929.

Gasser, C., R. Brown, T. Gallagher, D. Kelley, M. Monfared, and C. Passerini. 2007. Organ formation and gene expression patterning in ovule development. 18th Int. Conf. Arabidopsis Res.

Gou, J.Y., L.J. Wang, S.P. Chen, W.L. Hu, and X.Y. Chen. 2007. Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. Cell Res. 17: 422-434.

Ji, S.J., Y.C. Lu, J.X. Feng, G. Wei, J. Li, Y.H. Shi, Q. Fu, D. Liu, J.C. Luo, and Y.X. Zhu. 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.

Johnson, C.S., B. Kolevski, and D.R. Smyth. 2002. TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. Plant Cell 14: 1359–1375.

Li, X.B., X.P. Fan, X.L. Wang, L. Cai, and W.C. Yang. 2005. The cotton ACTIN1 gene is functionally expressed in fibers and participates in fiber elongation. Plant Cell 17: 859–875.

Luo, M., Y.H. Xiao, X.B. Li, X.F. Lu, W. Deng, D.M. Li, L. Hou, M.Y. Hu, Y. Li, and Y. Pei. 2007. GhDET2, a steroid 5a-reductase, plays an important role in cotton fiber cell initiation and elongation. Plant J. online early article.

Ramsay, N.A., and B.J. Glover. 2005. MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci. 10: 63–70.

Ruan, Y.L., D.J. Llewellyn, and R.T. Furbank. 2003. Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. Plant Cell 15: 952–964.

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

Udall, J.A., J.M. Swanson, and K. Haller, et al. 2006. A global assembly of cotton ESTs. Genome Res. 16: 441–450.

Wang, S., J.-W. Wang, N. Yu, C.-H. Li, B. Luo, J.-Y. Gou, L.-J. Wang, and X.-Y. Chen. 2004. Control of plant trichome development by a cotton fiber MYB gene. Plant Cell 16: 2323–2334.

Xu, Y., H.B. Li, and Y.X. Zhu. 2007. Molecular biological and biochemical studies reveal new pathways important for cotton fiber development. J. Int. Plant Biol.49: 69-74.

Yang, S.S., F. Cheung, J.J. Lee, M. Ha, N.E. Wei, S.H. Sze, D.M. Stelly, P. Thaxton, B. Triplett, C.D. Town, and J.Z. Chen. 2006. Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. Plant J. 47: 761-775.

Table 1. Correlation coefficients obtained from microarray hybridization experiments with total RNA samples prepared from wild-type or mutant cotton ovules harvested at different growth stages.

	technical coefficient			biological coefficient		
Microarray samples	**1-1´**	**2-2´**	**3-3´**	**1-2**	**1-3**	**2-3**
0wt/-3wt	0.9278	0.9353	0.9458	0.9051	0.8596	0.8982
1wt/-3wt	0.9202	0.9368	0.9569	0.9138	0.8676	0.871
2wt/-3wt	0.944	0.9488	0.9209	0.9465	0.8858	0.9283
3wt/-3wt	0.9009	0.9201	0.9069	0.8821	0.9155	0.9062
3fl/-3wt	0.9508	0.9269	0.9142	0.9417	0.9109	0.8794

Table 2. Highly expressed genes during wild-type cotton fiber initiation and ovule expansion period.

ARRAY ID	3wt/-3wt	Annotation
UFL084E11	77.9	alpha-expansin 1 [Gossypium hirsutum]
UFL062H08	63.2	alpha-expansin precursor [Gossypium hirsutum]
UFL016C02	47.8	No Hits
UFL033G12	38.6	No Hits
UFL079D10	37.8	No Hits
UFL038D01	34.6	alpha-expansin precursor [Gossypium hirsutum]
CM123E12	31.7	No Hits
UFL136C02	30.1	E6-4 [Gossypium hirsutum]
UFL190E05	29.8	hypothetical protein AN8715.2 [Aspergillus nidulans FGSC A4]
UFL188H05	27.8	No Hits
UFL087B02	26.9	E6
UFL129H12	26.6	protein kinase [Gossypium hirsutum]
UFL209A12	26.3	No Hits
UFL207G02	25.9	No Hits
UFL064A05	25.4	No Hits
UFL010F11	25.0	receptor protein kinase-like protein [Arabidopsis thaliana]
UFL098A02	24.5	No Hits
UFL016F01	22.0	unknown protein [Arabidopsis thaliana]
UFL015G01	21.0	No Hits
UFL078E01	20.7	mKIAA4001 protein [Mus musculus]
CM005F03	19.7	alpha-expansin precursor [Gossypium hirsutum]
UFL104C04	17.9	Sterol desaturase [Medicago truncatula]
UFL101F09	17.7	long chain fatty acid elongation enzyme [Gossypium hirsutum]
UFL063B12	17.2	unknown protein [Arabidopsis thaliana]
CM032F06	17.2	proline-rich protein-1 [Gossypium hirsutum]
UFL046A05	17.0	unknown protein [Arabidopsis thaliana]
UFL031D02	16.7	P9L; DNA binding [Arabidopsis thaliana]
UFL155C09	16.5	unknown protein [Arabidopsis thaliana]
UFL080E09	16.1	putative lipid transfer protein GPI-anchored [Cicer arietinum]
UFL047B11	15.8	PS60 [Nicotiana tabacum]
UFL048B10	15.4	E-class P450, group I [Medicago truncatula]
UFL110H10	14.4	dehydration-induced protein RD22-like protein 1 [Gossypium arboretum]
CM079H06	14.4	At1g76160/T23E18_10 [Arabidopsis thaliana]
CM059D06	14.3	5´ start site is putative; putative
CM103G04	14.0	DNA helicase, putative [Plasmodium falciparum 3D7]
CM054F10	13.7	proline-rich protein-1 [Gossypium hirsutum]
UFL053B07	13.7	polyphenol oxidase [Populus balsamifera subsp. trichocarpa x Populus deltoides]
UFL147C09	13.5	acyltransferase [Arabidopsis thaliana]

UFL008B01	13.1	delta-COP [Zea mays]
CM097A04	12.9	signal recognition particle 54 kDa subunit precursor [Pisum sativum]
UFL171H04	12.8	dehydration-induced protein RD22-like protein [Gossypium hirsutum]
UFL167A10	12.6	Alcohol acyl-transferase [Prunus mume]
CM049F02	12.6	fiber protein E6 (clone CKE6-4A) – upland cotton
UFL205F05	12.5	acyl-transferase [Arabidopsis thaliana]
CM033A03	12.3	expansin [Gossypium hirsutum]
UFL009E06	12.3	RNA-directed DNA polymerase (Reverse transcriptase) [Medicago truncatula]
UFL099F10	11.6	CER1 (ECERIFERUM 1) [Arabidopsis thaliana]
CM110C12	11.5	sulfotransferase family [Arabidopsis thaliana]
UFL119G05	11.4	hydrolase, hydrolyzing O-glycosyl compounds [Arabidopsis thaliana]
UFL144A01	11.3	kinesin-like calmodulin binding protein [Gossypium hirsutum]
UFL188H02	11.3	No Hits
UFL130G06	11.2	No Hits
CM005H11	11.1	proline-rich protein-1 [Gossypium hirsutum]
UFL048H06	11.0	No Hits
CM085A01	10.7	auxin-induced (indole-3-acetic acid induced) protein family [Arabidopsis thaliana]
UFL060B11	10.6	No Hits
CM036H12	10.6	GDSL-motif lipase/hydrolase protein [Arabidopsis thaliana]
CM097A09	10.5	expressed protein [Arabidopsis thaliana]
UFL193F07	10.4	No Hits
UFL161C05	10.3	No Hits
CM025A04	10.2	proline-rich protein-1 [Gossypium hirsutum]
UFL177D11	10.0	No Hits

Table 3. Highly expressed genes during fl mutant ovule expansion period.

ARRAY ID	3fl/-3wt	Annotation
UFL064A05	36.2	No Hits
UFL020B11	31.6	SBT1 [Lycopersicon esculentum]
UFL008D04	17.8	No Hits
CM114D09	12.8	invertase/pectin methylesterase inhibitor – related [Arabidopsis thaliana]
UFL073F12	10.5	permease [Arabidopsis thaliana]

Table 4. KOBAS analysis of highly expressed metabolic pathways in fiber initiation and ovule expansion.

Pathway (+3wt/-3wt)	up-regulated (836)	Overall (6046)	P-value	Q-value
Gap junction	45	111	2.21E-12	3.92E-10
Methionine metabolism	27	52	6.51E-11	5.76E-09
Antigen processing and presentation	30	89	1.32E-06	7.81E-05
Ribosome	115	556	1.95E-06	8.64E-05
Pores ion channels	16	39	2.59E-05	0.0009
Epithelial cell signaling in Heliobacter pylori infection	15	49	0.0018	0.0542
Selenoamino acid metabolism	20	79	0.0044	0.1119
Glycosaminoglycan degradation	11	36	0.0074	0.1567
Citrate cycle (TCA cycle)	20	83	0.0080	0.1567
Other lipid metabolism	4	7	0.0090	0.1589
ATP synthesis	19	80	0.0112	0.1802
Linoleic acid metabolism	8	24	0.0123	0.1813
Glycosphinogolipid biosynthesis-ganglioseries	10	34	0.0138	0.1884
N-Glycan degradation	13	50	0.0159	0.2006
Leukocyte transendothelial migration	7	21	0.0188	0.2224
Glycine, serine and threonine metabolism	20	92	0.0243	0.2532
Carbon fixation	31	157	0.0234	0.2532
Ascorbate and aldarate metabolism	36	194	0.0371	0.3644
Streptomycin biosynthesis	10	40	0.0416	0.3876

Pathway (+3fl/-3wt)	up-regulated (298)	Overall (6046)	P-value	Q-value
Gap junction	33	111	0	0
Methionine metabolism	13	52	8.72E-07	6.23E-05
Cell cycle – yeasts	12	78	0.0004	0.0182
Thiamine metabolism	6	24	0.0009	0.0221
Metabolism of xenobiotics by cytochrome P450	7	33	0.0009	0.0221
Phenylalanine metabolism	13	94	0.0006	0.0221
Cell cycle – mammals	12	91	0.0016	0.0320
Nitrogen metabolism	11	86	0.0031	0.0555
Stilbene, coumarine and lignin biosynthesis	20	211	0.0035	0.0562
Bile acid biosynthesis	8	54	0.0046	0.0651
Starch and sucrose metabolism	21	241	0.0075	0.0975
Ether lipid metabolism	4	18	0.0102	0.1218
Oxidative phosphorylation	13	130	0.0113	0.1247
Fatty acid metabolism	10	90	0.0126	0.1287
1- and 2-Methylnaphthalene degradation	5	30	0.0145	0.1384
Antigen processing and presentation	9	89	0.0304	0.2716

Glycerolipid metabolism	**11**	**121**	**0.0356**	**0.2991**
Galactose metabolism	**9**	**94**	**0.0412**	**0.3270**

FIGURE CAPTIONS

Figure 1. SEM photos of wild-type fiber initiation phenotypes (-1wt, 0wt, 1wt, 2wt, upper panel) and fl ovule expansion phenotypes (-1wt, 0wt, 1wt, 2wt, lower panel).

Figure 2. TreeView Representation of gene expression pattern in fiber cell initiation

A. Hierarchical clustering of 19293 ESTs that showed FDR-corrected P values <0.001 in at least one of the growth stages. The signals are shown in a red-green color scale, where red represents higher expression and green represents lower expression. The numbers represent the DPA of ovule harvest of the hybridizing RNA. An RNA sample from -3 DPA wild-type ovules was used as the reference for each hybridization. (a): genes induced from 0 DPA and maintained at relatively high levels throughout the experimental period. (b): genes of which expression decreased during the whole early development stage. (c): genes induced in 0 DPA wild-type ovules and 3 DPA fl ovules, but repressed drastically around 3 DPA wild-type ovules.

B. transcription factors and protein kinases highly induced in 0-3 DPA wild-type but not in 3 DPA fl ovules.

C. transcription factors and protein kinases or phosphotase repressed significantly after 0 DPA but not in 3 DPA fl ovules.

Figure 3. Total number of genes up-regulated in fiber initiation and ovule expansion period (Q-value < 0.01).

Fig. 1

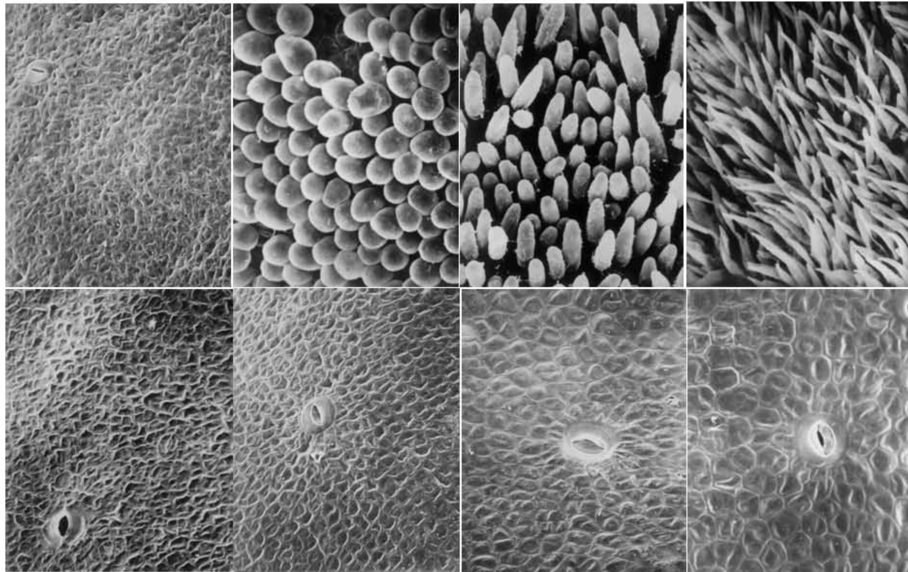
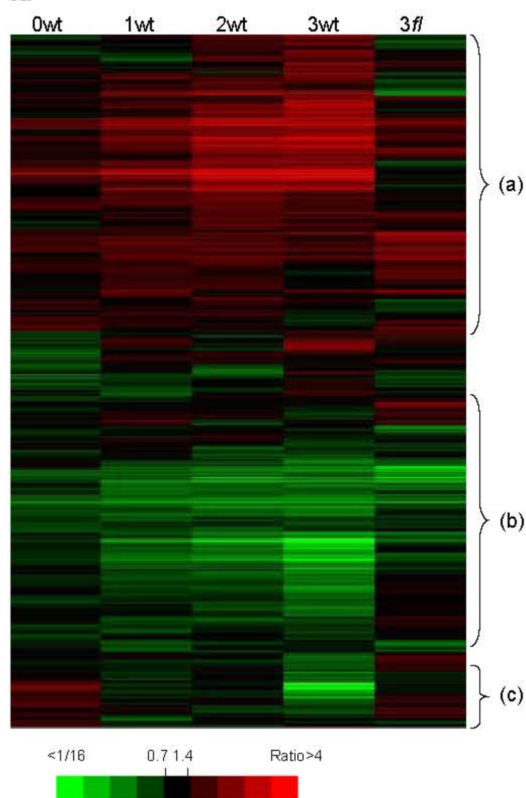


Fig.2

A.



B.

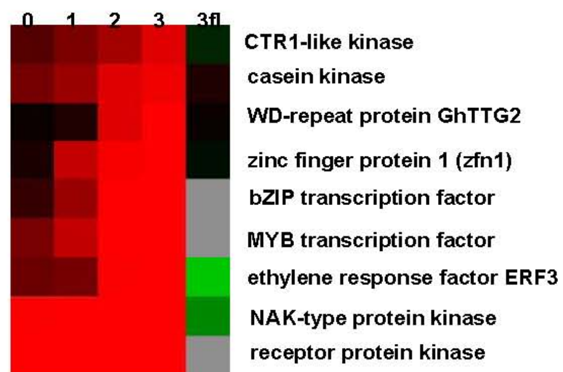


Fig.2 continued

C.

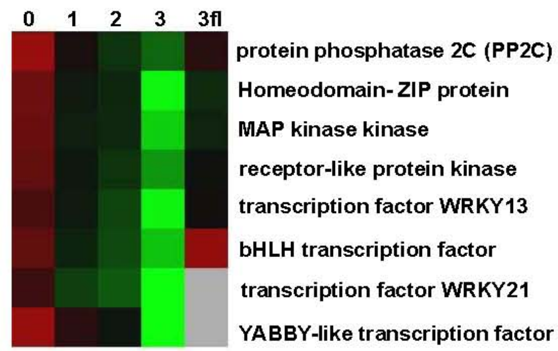


Fig.3