

Transcriptome Analysis of Ten Days Post Anthesis Elongating Fiber in the Upland Cotton (*Gossypium hirsutum*) Chromosome Substitution Line CS-B25

Chuan-Yu Hsu¹, Mark A. Arick II¹, Qing Miao^{2,3}, Sukumar Saha⁴, Johnie N. Jenkins⁴, Mirzakamol S. Ayubov⁵, Ibrokhim Y. Abdurakhmonov⁵, Daniel G. Peterson¹, Din-Pow Ma^{2*}

¹Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, Mississippi State, MS, USA

²Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS, USA

³Department of Pharmacology, Weill Cornell Medical College, Cornell University, New York, NY, USA

⁴USDA-ARS, Crop Science Research Laboratory, Mississippi State, MS, USA

⁵Center of Genomics and Bioinformatics, Academy Sciences of Uzbekistan, Tashkent, Uzbekistan

Email: *dma@bch.msstate.edu

How to cite this paper: Hsu, C.-Y., Arick II, M.A., Miao, Q., Saha, S., Jenkins, J.N., Ayubov, M.S., Abdurakhmonov, I.Y., Peterson, D.G. and Ma, D.-P. (2018) Transcriptome Analysis of Ten Days Post Anthesis Elongating Fiber in the Upland Cotton (*Gossypium hirsutum*) Chromosome Substitution Line CS-B25. *American Journal of Plant Sciences*, 9, 1334-1361.

<https://doi.org/10.4236/ajps.2018.96098>

Received: March 27, 2018

Accepted: May 27, 2018

Published: May 30, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

A chromosome substitution line, CS-B25, was developed by the substitution of chromosome pair 25 of *Gossypium hirsutum* TM-1 with the homologous pair of chromosome 25 from *G. barbadense*, a double haploid Pima 3-79 line. CS-B25 has improved fiber traits compared to its parent TM-1. To explore the molecule mechanisms underlying improved fiber traits, deep sequencing of total RNA was used to compare gene expression in fibers of CS-B25 and TM-1 at 10 days post anthesis (10-DPA). A total of 1872 differentially expressed genes (DEGs) were detected between the two lines, with 1175 up-regulated and 697 down-regulated in CS-B25. Gene Ontology (GO) enrichment analysis of the expression data by Generally Applicable Gene-set Enrichment (GAGE) and ReviGO indicated that the most prevalent Biological Process GO terms associated with DEGs included DNA-templated transcription, response to oxidative stress, and cellulose biosynthesis. Enriched Molecular Function GO terms included structural constituents of cytoskeleton, peroxidase activity, cellulose synthase (UDP-forming) activity, and transcription regulatory region sequence-specific DNA binding factors. GAGE was also used to find enriched KEGG pathways, and the highly represented pathways were Biosynthesis of Amino Acids, Starch and Sucrose Metabolism, Phenylpropanoid Biosynthesis, Protein Processing in Endoplasmic Reticulum, and Plant Hormone Signal Transduction. Many of the identified DEGs are involved in cy-

toskeleton and cell wall metabolism. The results of gene expression data have provided new insight into the molecular mechanisms of fiber development during the fiber elongation stage and would offer novel candidate genes that may be utilized in cotton fiber quality improvement.

Keywords

Chromosome Substitution Lines, Differentially Expressed Genes, Fiber Development, Quantitative Real-Time PCR, RNA Sequencing, Upland Cotton

1. Introduction

Cotton fibers are seed trichomes and differentiated from the epidermal cells of a developing seed. The development of cotton fiber consists of four partially overlapping stages: initiation, elongation (primary wall synthesis), secondary wall thickening, and maturation [1]. Fiber cells initiate from 3 days before anthesis to 3 days post anthesis (DPA). After initiation, fiber cells immediately enter into the elongation stage. Fiber elongation reaches peaks at 6 - 12 DPA, but can last up to 22 - 26 DPA. The secondary wall synthesis begins at 16 - 19 DPA, reaches a maximal rate at 25 DPA, and lasts up to 40 DPA. Fiber maturation, the last stage of fiber development, occurs at 40 - 50 DPA [1]. Fiber cell initiation and elongation are the two most important developmental stages with regard to fiber quality [2].

Cotton provides the largest renewable natural fiber for the textile industry and is also an important source of food, feed, fuel and other products [3]. Over 90% of the world cotton production is from Upland cotton, *Gossypium hirsutum* L. [4]. Upland cotton is widely adapted and has a high percentage of lint fiber and a relatively high yield. Pima cotton (*G. barbadense* L.), the second most widely-grown cotton species, generally has a lower lint percentage and lower yield, but its fiber qualities are superior to Upland cotton in terms of length, strength, and fineness. The Pima fiber offers superior spinning and manufacturing performance and better textile products, and thus normally commands a 30% to 50% price advantage over fiber from high-yielding Upland cottons. Pima cotton, however, requires long growing seasons to produce profitable yields of high quality fiber. One of the major goals of cotton breeders is to develop Upland cotton with Pima-like elite fiber quality combined with high yield and adaptability. Combining the adaptability and yield of Upland cotton with the fiber quality of *G. barbadense* into an Upland cotton cultivar would maximize quality and productivity in an increasingly competitive market. However, crosses between *G. hirsutum* and *G. barbadense* at the whole genome level have not been highly successful due to technical and biological challenges associated with the conventional methods of interspecific introgression. An alternative approach to interspecific introgression is to use alien chromosome substitution lines where individual chromosomes or chromosome arms of *G. barbadense* are substituted into

G. hirsutum. Seventeen interspecific chromosome substitution lines (CS-B lines) of Upland cotton (TM-1 background) containing whole chromosomes or chromosome arms of *G. barbadense* (Pima 3-79) chromosomes have been developed and released to the public [5]. These lines are genetically identical except that each differs by the replacement of a specific homologous pair of chromosomes from Pima 3-79 (*G. barbadense*) into the Upland background. CS-B25, a line in which the chromosome 25 pair of TM-1 has been replaced with the homologous chromosome 25 pair from *G. barbadense*, exhibits several improved fiber quality traits including increased fiber length, increased fiber strength, and lower micronaire [6] [7] [8] [9] [10]. Results from crosses of CS-B25 with selected USA improved cultivars also demonstrated that CS-B25 had several beneficial alleles with additive gene effects for improving fiber strength and decreasing micronaire in Upland cultivars [9] [10]. Additionally, CS-B25 has some beneficial alleles for improving agronomic traits with dominance genetic effects for lint percentage, seed cotton, and lint yield [11].

One of the serious impediments in cotton genetic improvement is the paucity of information of genes and their association with important fiber traits. The CS-B lines, being isogenic lines with the recurrent parent TM-1 except for the substituted chromosome pair from the alien species of *G. barbadense*, provide excellent analytical tools to dissect genetically complex fiber traits at the molecular level. The CS-B22sh line (chromosome 22 short arm substitution from Pima 3 - 79 into TM-1) was first used as a tool to identify differentially expressed transcripts in cotton [12]. The authors reported that 36 genes were differentially expressed in 15-DPA fiber in CS-B22sh, including genes that encoded tubulins, actin, cellulose synthase, glycosyl hydrolase, glycoside hydrolase, glycoside pyruvate decarboxylase and genes involved in signal transduction and nucleic acid, protein and lipid metabolism. Another report showed that CS-B22sh had positive additive effect on decreasing micronaire and a positive homozygous dominance effect on fiber strength [9].

In this study a genome wide transcriptome analysis was used to identify differentially expressed genes (DEGs) in 10-DPA fibers in CS-B25 as compared to TM-1. These DEGs and KEGG pathways analysis have provided insight into the mechanisms of fiber development during the fiber elongation stage and offered novel candidate genes that can be utilized in the genetic improvement of fiber quality.

2. Materials and Methods

2.1. Plant Materials and Growth Conditions

G. hirsutum L. cv. TM-1 and *G. hirsutum* CS-B25 plants were grown in the field at the R. R. Foil Plant Science Research Center at Mississippi State University (geographic coordinate: latitude N33.47 and longitude W88.77). TM-1 (Texas Marker-1) is a homozygous inbred line, serving as the genetic and cytogenetic standard for Upland cotton, *G. hirsutum*. TM-1 was developed by selection from

Deltapine 14 at Texas A&M AgriLife Research, College Station, Texas. CS-B25 line was developed using a special cytogenetically modified backcross-inbred breeding method [5]. The method used a specific hypoaneuploid type of *G. hirsutum*, primary monosomic for chromosome 25, as the recurrent parent which was previously backcrossed for more than 30 generations to inbred TM-1 and was thus quasi-isogenic hypoaneuploid type for chromosome 25 to TM-1. The recurrent hypoaneuploid for chromosome 25 monosomic plant was crossed with Pima 3-79, a double haploid *G. barbadense* line, as the male parent in developing the CS-B25 line. The cytologically identified monosomic F1 hybrid was used in a program with continuous backcrossing with TM-1 as the recurrent parent until the BC5F1 hypoaneuploid F1 plant was recovered. Upon identification, a hypoaneuploid BC5F1 segregate was selfed to identify a disomic, true-breeding disomic substitution. The substitution line was named as CS-B25 according to the expectedly substituted alien chromosome for 25 from the *G. barbadense* 3-79. The CS-B25 line is isogenic to the parent TM-1 for 25 chromosome pairs except the pair of chromosome 25 from the *G. barbadense* 3-79. Flowers were tagged on the day of anthesis (0 DPA). Cotton bolls at 10-DPA were harvested. Fibers were carefully dissected from cotton bolls, immediately frozen in liquid nitrogen, and then stored at -80°C for RNA extraction.

2.2. Extraction and Purification of Fiber RNA

Total RNA was isolated from 10-DPA fibers of TM-1 and CS-B25 lines using a modified hot borate method [13]. Genomic DNA contamination in the RNA samples was removed by DNase I (Promega, Madison, Wisconsin, USA). After DNase treatment, other potential contaminants were removed from the RNA sample with a Qiagen RNasey Mini Kit (Qiagen, Valencia, California, USA). The quality of the total RNA sample was pre-checked on a 1% - 1.2% (w/v) agarose gel by electrophoresis, and the RNA concentration was estimated with a NanoDrop 1000 Spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, USA) and a Qubit® Fluorometer by using the Qubit RNA BR Assay Kit (Invitrogen, Carlsbad, California, USA). The quality of the total RNA was finally determined using both agarose gel electrophoresis and a Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, California, USA).

2.3. RNA Library Preparation, Illumina Sequencing and Data Collection

The isolated total RNA samples from 10-DPA fibers (1.5 μg from each sample) were used to construct strand-specific cDNA libraries using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina, San Diego, California, USA) according to the manufacturer's instructions. Three biological replicate samples of the two cotton lines CS-B25 and TM-1 were used to generate 6 RNA-Seq libraries which were pooled and sequenced on one lane of an Illumina HiSeq 2000 sequencer (pair end sequencing, 50 bp per end) (Illumina, San Diego, California,

USA). The raw reads were mapped to the *G. hirsutum* (AD1) NAU-NBI genome [14] using Tophat 2 [15]. Next HTSeq [16] was used to estimate gene expression from the Tophat alignments. The mRNA sequencing reads were deposited into NCBI Sequence Read Archive (SRA) with accession numbers SRR6015224-6 and SRR6015221-3 for TM-1 and CS-B25, respectively.

2.4. Identification and Functional Analysis of Differentially Expressed Genes

EdgeR [17], a Bioconductor (<http://www.bioconductor.org/>) software package for differential gene expression analysis, was used to identify DEGs in CS-B25 10-DPA fibers via comparison with TM-1 10-DPA fibers. Generalized linear models were used to test for differential gene expression between CS-B25 and TM-1. Genes with an FDR adjusted p-value ≤ 0.05 were considered differentially expressed. GAGE [18], a Bioconductor software package for differential gene set analysis, was used to identify differentially expressed GO terms [19] and KEGG pathways [20]. GO gene sets were compiled from the functional annotation available from CottonGen [21]. REVIGO [22] was used to summarize and visualize the enriched GO terms. Gene expression data was mapped to Arabidopsis KEGG pathways using gene homologies reported on CottonGen.

2.5. Quantitative Real-Time PCR

Twelve DEGs including 8 up-regulated and 4 down-regulated in CS-B25 were selected for validation using qRT-PCR. One microgram of fiber 10-DPA total RNA was used as template for first-strand cDNA synthesis using the random hexamer (Promega, Madison, Wisconsin, USA) and M-MLV reverse transcriptase (Promega, Madison, Wisconsin, USA) according to the manufacturer's instructions. The synthesized cDNA was then used as template for qRT-PCR reactions using PowerUp SYBR Green Master Mix and a QuantStudio 5 Real-Time PCR System (Applied Biosystems, Waltham, Massachusetts, USA). Cotton 18S rRNA was used as a reference gene to normalize the expression levels of RT-PCR products [23]. All reactions were performed with three replicates. The $2^{-\Delta\Delta Ct}$ method [$\Delta Ct = Ct$ (differentially expressed gene) – Ct (18S rRNA), $\Delta\Delta Ct = \Delta Ct$ (CS-B25) – ΔCt (TM-1), $2^{-\Delta\Delta Ct} = \text{Relative Expression}$] was used to calculate relative expression of differentially expressed genes [24]. All primers were designed by using Primer3 plus and the sequences of the primers are listed in **Table S1**.

2.6. Analysis of SSR DNA Markers Associated with Chromosome 25

Seven chromosome 25-specific nuclear SSR markers associated with improved fiber traits of fineness (micronaire), length, and strength were selected from the previous QTL mapping of fiber quality loci (see **Table S2**). These QTLs identify genetic relationship among the selected improved *G. hirsutum* and *G. barbadense* accessions including the recurrent parent TM-1 and the donor parent Pima 3-79 of the CS-B25 line. The seven chromosome-specific SSR markers were

used to construct a dendrogram based on the average coefficient of similarity using JMP Genomics program (SASTM). The method of Egamberdiev *et al.* [25] was used to verify SSRs via PCR amplification and capillary electrophoresis.

3. Results

3.1. Transcriptome Analysis of 10 DPA Fiber from CS-B25 and TM-1

Approximately 300 million total sequence reads of 50 bp were generated for this study. Analysis of sequencing data of CS-B25 and TM-1 from the six RNA libraries revealed that more than 80% of reads were mapped to the TM-1 reference genome. **Table 1** shows the total reads statistics of the six libraries. Of the 70,478 genes annotated in the *G. hirsutum* reference, 57,366 had at least one sequence read alignment. A total of 1870 genes (**Table S3**) were found to be differentially expressed in 10-DPA fiber when comparing CS-B25 to the TM-1 control at the false discovery rate (FDR) cutoff of 5% (adjusted $p \leq 0.05$) and an absolute value of $\log_2\text{ratio} \geq 1$. Among them, 1173 and 697 genes were up-regulated and down-regulated in CS-B25, respectively.

3.2. Gene Function Annotation by GO Enrichment Analysis

GO enrichment analysis by GAGE and ReviGO indicated that Biological Process GO terms related to DNA-templated transcription, response to oxidative stress, and cellulose biosynthetic processes were enriched (**Figure 1(a)**). In addition, the Molecular Function GO terms including structural constituent of cytoskeleton, peroxidase activity, cellulose synthase (UDP-forming) activity, and transcription regulatory region sequence-specific DNA binding were predominant (**Figure 2(a)**). The 57,366 *G. hirsutum* genes with expression data mapped to 14,610 homologous Arabidopsis genes were used to find enriched KEGG pathways. Most enriched pathways were found to be related to biosynthesis of secondary metabolites, biosynthesis of amino acids, starch and sucrose metabolism, amino sugar and nucleotide sugar metabolism, and plant hormone signal transduction (**Table 2**). The genes differentially expressed in CS-B25 10-DPA fibers were functionally annotated, and representatives of these differentially expressed genes (DEGs) are listed in **Table 3** and **Table S3**.

Table 1. Total reads statistics of RNA-Seq data

RNA-Seq library ID	TM1-1	TM1-2	TM1-3	CS-B25-1	CS-B25-2	CS-B25-3
Total reads	48,362,375	44,800,820	106,254,636	46,940,992	45,866,594	48,500,739
Total unique reads	14,097,884	12,148,341	16,029,773	16,444,626	15,765,544	16,855,253
Overall %	99.2	99.2	99.2	99.3	99.3	99.1
Concordant	85.8	85.2	80.5	87.8	87.5	87.7

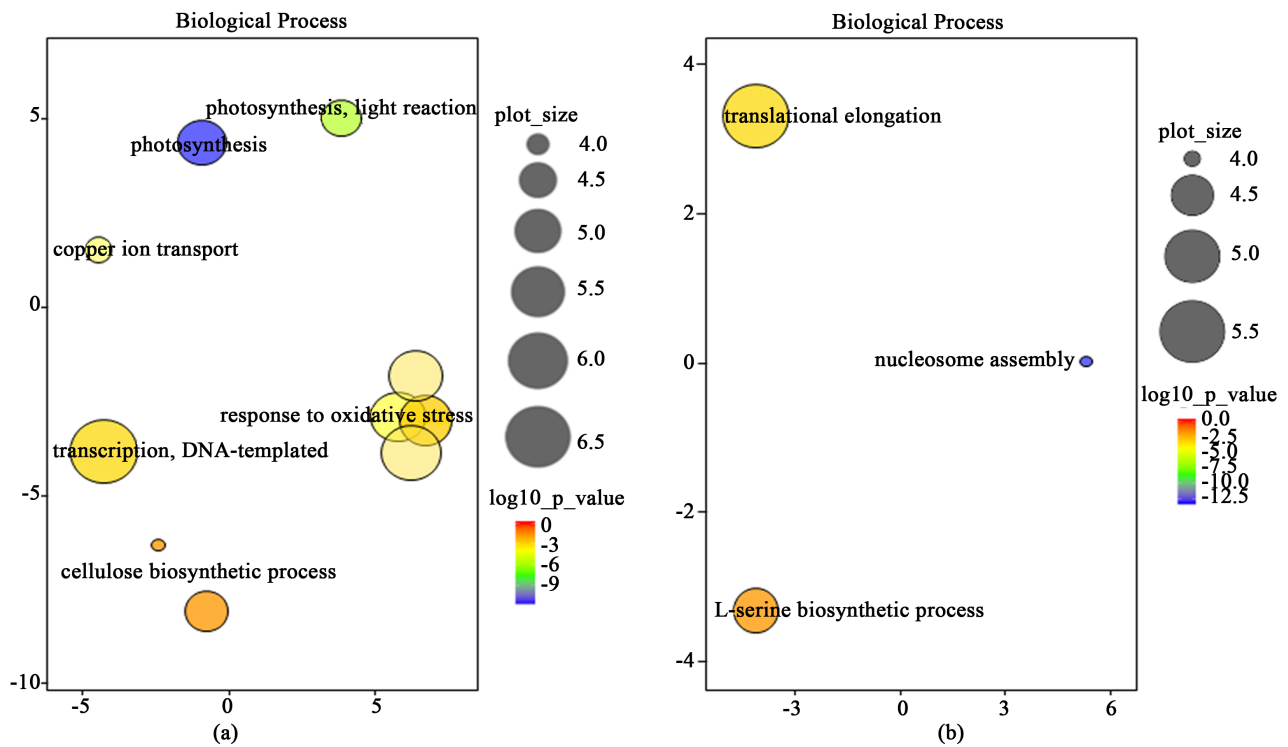


Figure 1. A ReViGo scatter plot of GO analyses in biological process from (a) up-regulated and (b) down-regulated DEGs in the CS-B25 10-DPA fibers. Related GO terms are clustered and presented with bubbles with similar color shades. Bubble colors represent p-values and bubble sizes indicate the relative frequency of the GO terms.

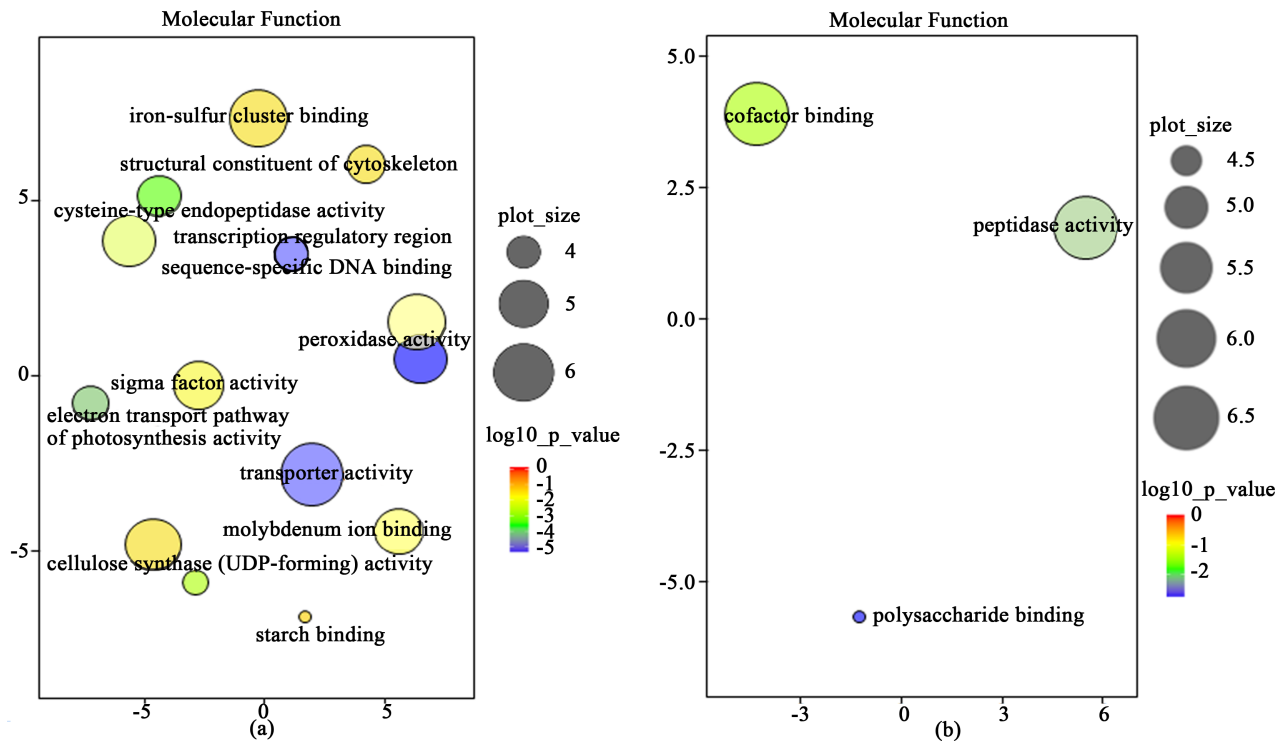


Figure 2. A ReViGo scatter plot of GO analyses in molecular function from (a) up-regulated and (b) down-regulated DEGs in the CS-B25 10-DPA fibers. Related GO terms are clustered and presented with bubbles with similar color shades. Bubble colors represent p-values and bubble sizes indicate the relative frequency of the GO terms.

Table 2. KEGG pathway analysis of KOs.

Pathway	KOs	Pathway ID
Metabolic pathways	113	ko01100
Biosynthesis of secondary metabolites	71	ko01110
Biosynthesis of antibiotics	27	ko01130
Microbial metabolism in diverse environments	19	ko01120
Biosynthesis of amino acids	17	ko01230
Ribosome	15	ko03010
Starch and sucrose metabolism	14	ko00500
Plant hormone signal transduction	13	ko04075
Amino sugar and nucleotide sugar metabolism	11	ko00520
Carbon metabolism	11	ko01200
Plant-pathogen interaction	10	ko04626
Photosynthesis	10	ko00195

Table 3. Annotation of genes differentially expressed in CS-B25 10-DPA fibers.

Gene ID	Tair 10 ID	Description	FC* (CS-B25/TM-1)
Hormone signal transduction			
Gh_A08G1512	AT4G16770	1-aminocyclopropane-1-carboxylate oxidase-like	8.59
Gh_A06G1558	AT5G04530	3-ketoacyl-CoA synthase 19	0.045
Gh_D10G2187	AT5G04530	3-ketoacyl-CoA synthase 21-like	3.21
Gh_A04G0224	AT1G52820	gibberellin 20 oxidase	11
Gh_A04G0228	AT1G52820	gibberellin 3-beta-dioxygenase 4	10.4
Gh_A06G0045	AT2G18420	gibberellin-regulated family protein	4.97
Gh_A01G2049	AT1G78440	gibberellin 2-oxidase 1	0.058
Gh_D08G2151	AT3G51770	ethylene-overproduction 1-like TRR-containing protein	17.77
Gh_A07G0377	AT5G07580	Ethylene-responsive transcription factor ERF106	8.76
Gh_A04G0096	AT5G55340	long-chain alcohol O-fatty-acyltransferase 5	277.51
Gh_A04G0106	AT1G50640	ethylene responsive element binding factor 3	0.008
Gh_D06G0023	AT1G75540	zinc finger transcription factor Constans	23.45
Gh_A08G0537	AT2G44840	ethylene responsive 13	0.055
Gh_A08G0780	AT5G57740	RING E3 ligase XBAT32	0.0009
Gh_A02G0011	AT5G01990	auxin influx carrier family protein	215.15
Gh_A03G1429	AT5G54510	auxin-responsive GH3 family protein	73.05
Gh_D02G2389	AT4G14550	auxin-responsive protein IAA14-like	40.88
Gh_A06G0351	AT1G77690	auxin influx transport protein	7.41

Continued

Gh_D02G2219	AT4G34770	SAUR-like auxin-responsive protein	0.047
Gh_D06G2010	AT3G15540	auxin-induced protein AUX22-like	0.1
Gh_A03G1429	AT5G54510	indole-3-acetic acid-amido synthetase GH3.6-like	73.05
Gh_A10G2215	AT5G60850	zinc finger transcription factor DOF5	74.64
Gh_A13G0392	AT2G14960	auxin-responsive GH3 family protein	0.12
Gh_D09G0013	AT2G34740	protein phosphatase 2C family protein	70.17
Gh_A02G0999	AT5G59220	highly ABA-induced PP2C gene 1	0.096
Cell wall biosynthesis/modification			
Gh_A08G0457	AT2G04780	fasciclin-like arabinogalactan protein 7	3338.37
Gh_D08G0544	AT2G04780	fasciclin-like arabinogalactan protein 7	210.99
Gh_D10G2231	AT1G35710	LRR receptor-like serine/threonine protein kinase	293.42
Gh_D12G0597	AT5G44700	LRR receptor-like serine/threonine protein kinase	21.09
Gh_D02G0280	AT5G46330	LRR receptor-like serine/threonine protein kinase	23.75
Gh_D08G0544	AT2G04780	Ascorbate peroxidase 2	210.99
Gh_A04G0652	AT3G09640	Ascorbate peroxidase 2	49.27
Gh_A09G1963	AT4G23990	cellulose synthase-like protein G3 isoform X1	117.67
Gh_D09G2166	AT4G23990	cellulose synthase-like protein G3	155.29
Gh_D02G0079	AT2G43820	UDP-glycosyltransferase 74-F2-like	120.09
Gh_D13G0415	AT4G34131	UDP-glucosyltransferase 73B3	0.0037
Gh_Sca004909G03	AT2G43820	UDP-glucosyltransferase 74F2	0.084
Gh_D11G0846	AT2G45220	pectin methylesterase 2	183.15
Gh_A11G0727	AT2G45220	pectin methylesterase 2	208.35
Gh_A08G0741	AT5G56350	pyruvate kinase, cytosolic isozyme	369.33
Gh_D02G0381	AT2G48130	non-specific lipid transfer protein-like protein	232.45
Gh_A02G0317	AT3G22600	lipid transfer protein-like protein	219.08
Gh_A01G1193	AT2G37870	non-specific lipid transfer protein-like	45.76
Gh_D07G2095	AT2G03090	expansin A15	0.039
Gh_Sca012883G01	AT5G23860	tubulin beta-8	30.51
Gh_D09G1351	AT2G36710	pectin lyase-like superfamily protein	155.96
Gh_D08G1862	AT3G14300	pectinesterase family protein	87.69
Gh_A09G1349	AT2G36710	pectin lyase-like superfamily protein	25.6
Gh_A07G1458	AT3G07010	pectin lyase-like superfamily protein	0.061
Gh_D08G2205	AT3G05610	pectin methylesterase inhibitor superfamily	0.017
Gh_A06G0439	AT3G12500	chitinase	13.73

Continued

Gh_A12G2714	AT5G24090	chitinase A	0.092
Gh_D12G2669	AT5G24090	chitinase A	0.033
Transcriptional regulation of cellular metabolic process			
Gh_A08G0993	AT1G63910	myb domain protein 103	172.34
Gh_A13G1873	AT5G12870	transcription factor MYB46	155.87
Gh_D11G2061	AT3G49690	myb domain protein 84	83.18
Gh_D13G1452	AT2G32460	myb domain protein 101	45.57
Gh_D08G1266	AT1G63910	myb domain protein 103	143.47
Gh_A09G2017	AT3G24310	myb-related protein 305-like	40.58
Gh_D11G1132	AT2G47190	myb domain protein 2	40.54
Gh_D05G1913	AT3G18100	myb domain protein 4r1	20.74
Gh_A02G1357	AT5G16770	myb domain protein 9	0.067
Gh_D05G2967	AT1G14600	myb family transcription factor	30.75
Gh_A08G0907	AT2G46680	homeobox protein 7	97.47
Gh_D02G1259	AT5G39760	homeobox protein 23	58.09
Gh_A04G0732	AT5G02030	POX (plant homeobox) family protein	31.45
Gh_A08G0909	AT3G61890	homeobox protein 12	13.57
Gh_A03G0158	AT5G41410	POX (plant homeobox) family protein	9.73
Gh_D06G1066	AT5G25200	KNOTTED1-like homeobox gene 3	10.89
Gh_A11G2219	AT1G14440	homeobox protein 31	10.13
Gh_D05G1962	AT1G46480	WUSCHEL related homeobox 4	28.28
Gh_A05G2722	AT1G23380	KNOTTED1-like homeobox gene 6	0.065
Gh_A03G0887	AT1G69490	NAC transcription factor 29	60.7
Gh_A08G1691	AT4G28530	NAC domain containing protein 74	50.5
Gh_A05G1339	AT4G28500	NAC domain containing protein 73	9.99
Gh_A08G1086	AT4G01550	NAC domain containing protein 69	0.0032
Gh_A11G1172	AT2G30590	WRKY transcription factor 21	16.78
Gh_A02G1042	AT4G23550	WRKY transcription factor 29	3.27
Phenylpropanoid biosynthesis			
Gh_A02G1258	AT2G02400	cinnamoyl-CoA reductase 1	4.94
Gh_D01G1091	AT4G24510	shikimate O-hydroxycinnamoyltransferase	33.16
Gh_A01G1688	AT1G68540	cinnamoyl-CoA reductase 2	49.79
Gh_D11G1234	AT3G06350	dehydroquinate dehydratase/shikimate dehydrogenase	5.78

*FC represents fold change

3.3. DEGs Involved in Hormone Synthesis and Signaling

Three phytohormones of gibberellin, ethylene, and auxin have been shown to play important roles in cotton fiber development [26]. Several genes participating in gibberellin synthesis and in response to gibberellin stimulus were up-regulated in CS-B25 10-DPA fiber (**Table 3** and **Table S3**). These included genes for gibberellin 20 oxidase (Gh_A04G0224, 11 fold upregulation), gibberellin 3-beta-dioxygenase 4 (Gh_A04G0228, 10.41 fold upregulation) and a gibberellin-regulated family protein (Gh_A06G0045, 4.97 fold upregulation). Genes involved in ethylene synthesis and signaling were also found to be up-regulated. These include genes for 1-aminocyclopropane-1-carboxylate oxidase (ACCO) (Gh_A08G1512, 8.59 fold and Gh_D05G1663, 4.91 fold), a key enzyme involved in ethylene biosynthesis [27] [28], ethylene-overproduction 1-like tetratricopeptide repeat (TRR)-containing protein (Gh_D08G2151, 17.77 fold upregulation) which is involved in chromatin remodeling and transcription regulation, ethylene responsive transcription factor ERF106 (Gh_A07G0377, up-regulated 8.76 fold), and a membrane bound long-chain-alcohol O-fatty-acyltransferase 5 (Gh_A04G0096, up-regulated 277.5 fold) involved in synthesis of very long-chain-fatty acids which had been shown to promote fiber elongation by stimulating an increase in ACCO transcript and ethylene production [28] (**Table 3** and **Table S3**). In addition, genes involved in auxin synthesis and transport including auxin-responsible protein IAA14 (Gh_D02G2389, 40.9 fold), auxin transporter-like protein 3 (Gh_A06G0351, 7.41 fold), and auxin-responsive protein IAA14-like (Gh_A02G0341, 3.87 fold), auxin influx carrier family protein (Gh_A02G0011, 215 fold), and indole-3-acetic acid-amido synthetase GH3.6-like (auxin-responsive GH3 family protein) (Gh_A03G1429, 73.05 fold) were up-regulated. Genes encoding zinc finger transcription factors including *CONSTANS* (Gh_D06G0023, 23.44 fold, involved in ethylene biosynthesis) and *DOF5* (Gh_A10G2215, 74.6 fold, affecting auxin homeostasis) [29] were also up-regulated. RING E3 ligase XBAT32 (Gh_A08G0780) was down-regulated 1,113 fold in CS-B25 compared to TM-1 (**Table 3** and **Table S3**). XBAT32 reduces ethylene biosynthesis by decreasing the abundance of aminocyclopropane-1-carboxylic acid synthases [30], the enzymes participating in the first step of ethylene biosynthesis [31]. The gene Gh_D09G0013 encoding a protein phosphatase 2C family protein, a negative regulator of abscisic acid signaling [32], was up-regulated by 70.17 fold, suggesting that down regulation of ABA might contribute to a longer period of fiber elongation [33] and resulting in longer fibers in CS-B25.

3.4. DEGs Involved in Cell Wall Synthesis

Many genes involved in primary cell wall (PCW) biosynthesis were up-regulated in CS-B25 10-DPA fiber cells (**Table 3** and **Table S3**). These include genes coding for fasciclin-like arabinogalactan proteins 7 (Gh_A08G0457, 3338.37 fold and Gh_D08G0544, 211 fold) [34], non-specific lipid transfer proteins (LTPs)

(Gh_A02G0317, 219.1 fold, similar to At3G22600; Gh_D02G0381, 232.44 fold, similar to At2G48130 protein; Gh_A01G1193, 45.76 fold; Gh_A08G0129, 18.73 fold) [35], cellulose synthase-like protein G3 isoform X1 (Gh_A09G1963, 117.67 fold and Gh_D09G2166, 155.29 fold), UDP-glycosyltransferase 74F2-like (Gh_D02G0079, 120.09 fold), pectin methylesterase 2 (PME) (Gh_A11G0727, 208.35 fold and Gh_D11G0846, 183.15 fold) and pectin methylesterase 15 (Gh_D09G1351, 155.96 fold and Gh_A09G1349, 25.63 fold) [36]. Other genes related to the cytoskeleton, including tubulins beta 8 (Gh_Sca012883G01) was also up-regulated 30.5 fold. It had been suggested that specific tubulins might affect microtubule organization and microfibril deposition in elongating fiber cells [37]. An ABIL-1 gene encoding a subunit of the WAVE complex was highly expressed (1,262 fold upregulation). The WAVE complex is involved in regulation of actin and microtubule organization [38]. Genes for LRR receptor-like serine/threonine-protein kinases (Gh_D10G2231, 294 fold; Gh_D12G0597, 21.11 fold; Gh_D02G0280, 23.75 fold) were highly or moderately up-regulated, and they are known to participate in signal transduction pathways [39], cellulose deposition [40], and maintenance of cell wall integrity [41].

3.5. Transcription Factors Involved in Fiber Development

Many MYB-type transcription factors, found to regulate plant cell wall biosynthesis [42] [43], were also up-regulated in CS-B25 (Table 3 and Table S3). Genes for MYB transcription factors including MYB domain protein 103 (Gh_A08G0993, 172.34 fold; Gh_D08G1266, 143.47 fold; and Gh_A03G0166, 122.3 fold), MYB46 (Gh_D13G2261, 103.67 fold; Gh_A13G1873, 155.86 fold; Gh_D09G1082, 9.76 fold), MYB domain protein 305 (Gh_A09G2017, 40.58 fold) [44], MYB domain protein 101 (Gh_D13G1452, 45.57 fold), MYB domain protein 2 (Gh_D11G1132, 40.54 fold), MYB family transcription factor (Gh_D05G2967, 30.75), MYB44 (myb domain protein 4r1) (Gh_D05G1913, 20.74 fold), and MYB domain protein 84 (Gh_D11G2061, 83.18 fold) were up-regulated. Several homeobox leucine zipper proteins, including homeobox protein 7 (Gh_A08G0907, 97.47 fold), homeobox protein 23 (Gh_D02G1259, 58.08 fold), homeobox protein 12 (Gh_A08G0909, 13.57 fold) were up-regulated in CS-B25 as well. NAC transcription factors that activate MYBs, including NAC transcription factor 98 (Gh_D13G1561, 20.82 fold), NAC domain-containing protein 21/22-like (Gh_A08G1691, 50.56 fold), NAC transcription factor 100 (Gh_A11G0633, 7.36 fold), NAC-domain containing protein 73-like (Gh_A05G1339, 9.98 fold), and NAC transcription factor 29-like (Gh_A03G0887, 60.55 fold), a negative regulator of ABA signaling with function in transition from active cell division to cell expansion, were also up-regulated. NAC domain-containing protein 67-like (Gh_A08G1086, 0.0032), however, was down-regulated. NAC transcription factors and NAC domain-containing proteins are a very large family in plants [45], and the SND protein classified as NAC domain protein 1 was reported in regulating secondary wall synthesis by targeting the MYB46 transcription factor [42]. WRKY transcription factors 21

(Gh_A11G1172, 16.78 fold) and 29 (Gh_A02G1042, 3.27 fold) were also up-regulated in CS-B25.

3.6. DEGs Involved in Phenylpropanoid Synthesis

DEGs involved in phenylpropanoid biosynthesis were slightly up-regulated in CS-B25 (**Table 3** and **Table S3**). Genes for cinnamoyl-CoA reductase 2 (Gh_A02G1258, 4.96 fold), caffeic acid 3-O-methyltransferase (Gh_A12G2227, 8.54 fold and Gh_A08G0929, 3.86 fold) (lignin synthesis), and bifunctional 3-dihydroquinate dehydratase/shikimate dehydrogenase (chloroplast-like isoform X1) (Gh_D11G1234, 5.78 fold) (lignin synthesis) were up-regulated. These expression data indicated that lignin/secondary cell wall synthesis also occurred in the fiber elongation step but the difference between CS-B25 and TM-1 was less dramatic.

3.7. DEGs with Function in Oxidoreductase Activity

Genes involved in oxidoreductase activity were up-regulated in CS-B25 10-DPA fiber (**Table S3**), which include genes for alcohol dehydrogenases (Gh_A04G1326, 878.37 fold), cytochrome P450 CYP736A12-like (Gh_A09G1848, 82.98 fold), cytochrome P450 84-A1 like (Gh_A11G3018, 58.28 fold), and cytochrome P450 72A1 (Gh_D02G0287, 25.78 fold). Genes for cytosolic pyruvate kinase (Gh_A08G0741, 369.33 fold) [46] and ascorbate peroxidase (Gh_A04G0652, 49.27 fold) [47] involved in fiber elongation were also highly up-regulated in CS-B25 (**Table S3**).

In comparison to the numbers of up-regulated genes, fewer genes were down-regulated in CS-B25 10-DPA fiber (**Table S3**). Genes involved in translational elongation, L-serine biosynthetic processes, and nucleosome assembly in the Biological Process GO group are down-regulated (**Figure 1(b)**). Genes in Molecular Function including cofactor binding, polysaccharide binding, and peptidase activity are down-regulated (**Figure 2(b)**). Genes for ribosomal proteins including the 60S ribosomal protein L11 (Gh_D10G0336, 0.0696 fold), 60S ribosomal protein L5-like (Gh_D02G0103, 0.032 fold), 40S ribosomal protein S5-like (Gh_A04G0473, 0.0083 fold), and ribosomal protein L46 (Gh_A08G1142, 0.1118 fold) were down-regulated. Cyclin-J18 isoform X2 (Gh_A12G1420, 0.0145 fold), heme-binding-like protein (Gh_A04G0508, 0.0093 fold), 14-3-3-like protein (Gh_A01G0110, 0.027 fold), RAD-like transcription factor (Gh_A13G0293, 0.0042 fold), hypothetical protein (Gh_A10G1995, 0.0013 fold), Peptidase M20/M25/M40 family protein (Gh_A04G1433, 0.00235 fold), B3-domain-containing protein Os01g0234100-like (Gh_A01G0064, 0.0068 fold), E3 ubiquitin-protein ligase XBAT32 (E3-Ring finger, Gh_A08G0780, 0.000895 fold), and ethylene-responsive transcription factor 3-like (ERF3, Gh_A04G0106, 0.0081 fold) were down-regulated.

3.8. Validation of Differentially Expressed Genes by qRT-PCR

To test whether the gene expression results from RNA-seq were reliable, RNA

transcripts of 16 DEGs detected in CS-B25 were selected for qRT-PCR analysis. All the 16 transcripts were successfully detected by qRT-PCR and 12 transcripts showed agreement with RNA-seq results (see **Figure 3**). Both RNA-seq and qRT-PCR detected that the expression levels of genes for lipid transfer protein and cellulose synthase were comparable, and CS-B25 had the two RNA transcripts that were more than 70-fold higher than those in TM-1. Based on the qRT-PCR data, the four genes encoding B3 domain-containing protein Os01g0234100-like, E3 ubiquitin-protein ligase XBAT32, ethylene-responsive transcription factor 3-like, and 40S ribosomal protein S5-like had lower transcript levels in CS-B25, with fold change from 0.1 to 0.24 when compared with those in TM-1. The RNA-seq data, however, showed that TM-1 had higher transcripts levels of the four genes.

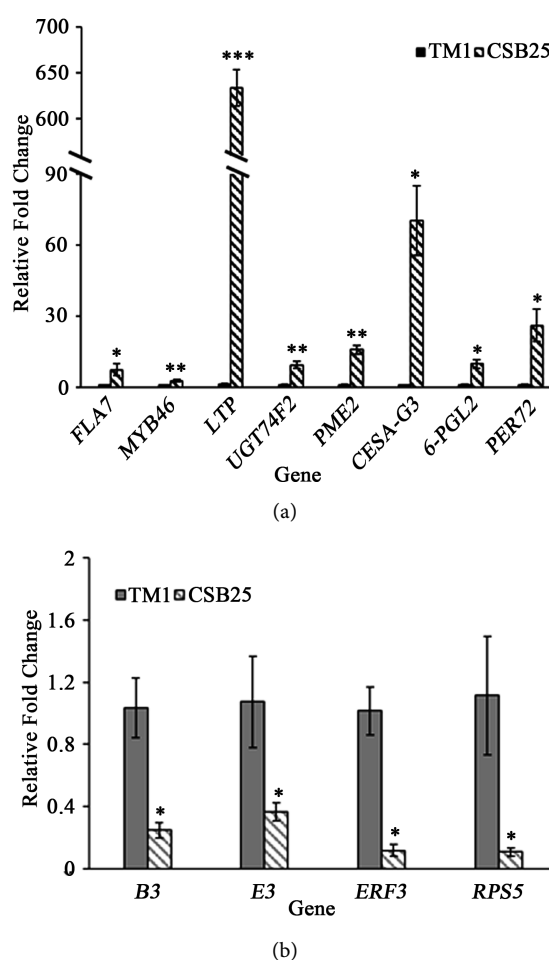


Figure 3. Quantitative RT-PCR analysis of differentially expressed genes in 10 DPA cotton fibers from CS-B25 and TM1 lines. Compared to TM1, eight selected up-regulated genes (a) and four selected down-regulated genes (b) identified in CS-B25 line from RNA-Seq analysis were validated with quantitative RT-PCR assays. At least three biological replicates were used in the assays, cotton 18S rRNA was used as a reference gene to normalize the expression level. The significant difference in the expression of each transcript between CS-B25 and TM1 was analyzed by t-test (* $p \leq 0.05$; ** $p \leq 0.005$; and *** $p \leq 0.005$). The error bars represent standard error of the mean.

3.9. Dendrogram Based on the Chromosome 25-Specific Simple Sequence Repeat (SSR) Markers

Results from the dendrogram analysis divided all of the accessions into two broad groups of *G. hirsutum* and *G. barbadense* based on the average coefficient of similarity (IBD value indicating identical alleles), suggesting that the chromosome 25 is distinctly different at the molecular level and fiber traits between the two species accessions including the recurrent parent TM-1 and the donor parent Pima 3-79 of the CS-B25 line (Figure 4). Pima 3-79, the donor parent, was clustered with other improved *G. barbadense* accessions in the dendrogram based on the chromosome 25-specific nuclear SSR markers associated with improved fiber traits from previous reports. This observation suggests that the substituted alien species chromosome 25 pair in CS-B25 would have the potential to carry genes of improved fiber traits including those associated with micronaire, fiber length and fiber strength. The results from field experiments demonstrated that the substituted chromosome 25 pair from *G. barbadense* in CS-B25 had significant effects on improved micronaire (Figure 5), which is one of important fiber quality traits used in determining fiber price.

4. Discussion

Results from GO enrichment analysis by GAGE and ReviGO indicated that many genes were up-regulated in CS-B25 10-DPA fiber with the highest fold change (FC) in expression of the gene coding for a fasciclin-like arabinogalactan protein 7 (AGP, Gh_A08G0457, 3338.37 fold upregulation) in CS-B25 compared to TM-1 (Table 3 and Table S3). Using over-expression and RNAi approaches,

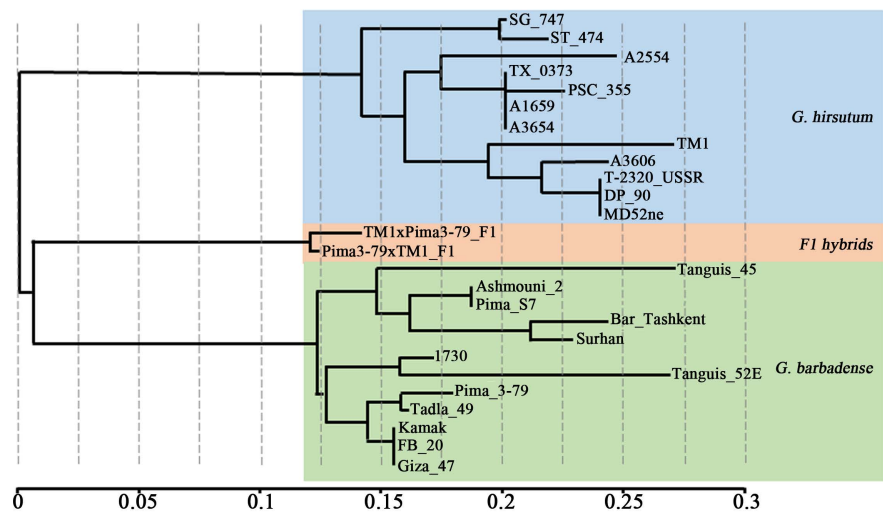


Figure 4. The dendrogram analysis of cotton fiber traits-associated SSR markers. The dendrogram of a set of improved *Gossypium hirsutum* and *G. barbadense* cotton lines constructed using JMP Genomics program on the basis of association of fiber traits with seven chromosome 25 specific SSR markers (Table S2). The results suggested that chromosome 25 is distinctly different between *G. hirsutum* and *G. barbadense* in fiber traits and molecular markers. The scale is presented as the branch length distance.

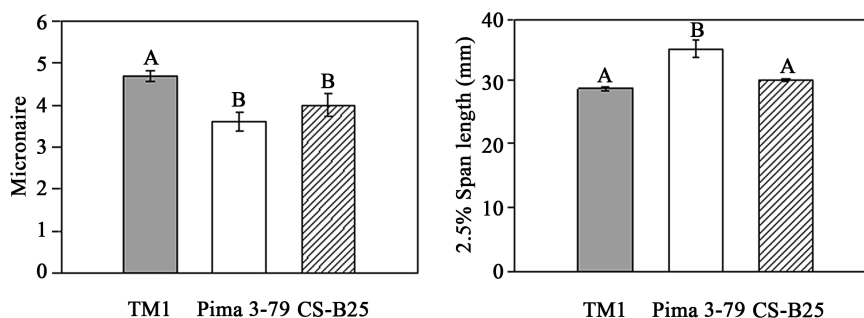


Figure 5. Comparison of micronaire and 2.5% span length of CS-B25 fiber with TM-1 and Pima 3-79 lines. Some of the differentially expressed transcripts in CS-B25 might be associated with improved fiber quality traits. CS-B25 and TM-1 do not have significant differences in Fiber 2.5% Span length, although CS-B25 had average values higher than TM-1. The letters marked above bars indicate statistically significant differences ($p \leq 0.005$ in the micronaire analysis and $p \leq 0.0005$ in the analysis of 2.5% span length). The error bars represent standard deviation of the mean.

Huang *et al.* [34] reported that AGPs are involved in fiber initiation and elongation. They reported that overexpression of a fasciclin-like arabinogalactan protein (GhFLA1) in *G. hirsutum* promoted fiber elongation and caused an increase in fiber length. In contrast, the suppression of *GhFLA1* expression slowed down fiber initiation and elongation and produced shorter mature fiber. In addition, expression levels of *GhFLAs* and the genes related to primary cell wall biosynthesis were notably enhanced in the *GhFLA1* overexpression transgenic fibers, whereas the transcripts of these genes were dramatically reduced in the fibers of *GhFLA1* RNAi cotton plants. Using immunostaining methods, Huang *et al.* [34] further showed that both AGP composition and primary cell wall composition were changed in the transgenic fibers. The CS-B25 had several differentially expressed AGP-like transcripts upregulated (Table S3), suggesting a major role of these genes at the fiber elongation stage. Previous studies revealed that the CS-B25 had several improved fiber quality traits including fiber length, strength, and micronaire [6] [7] [8] [9] [10]. These results were further confirmed from the top crosses with the selected USA improved cultivars that CS-B25 had several beneficial alleles with additive gene effects for improving fiber strength and micronaire in Upland cultivars [9] [10]. Our results suggest that the improvement of fiber traits in CS-B 25 line was in part associated with very high upregulation of *GhFLA* genes in fiber cell due to the introgression of the substituted chromosome of Pima 3-79 into the CS-B25 line. The dendrogram results based on the SSR markers also separated all lines into two broad groups of *G. hirsutum* and *G. barbadense*. It will be interesting in future to investigate the relation of differentially expressed transcripts of fasciclin-like arabinogalactan proteins among the lines of these two species with reference to the improved fiber traits including elongation. This is important considering that fiber elongation is one of the most important economic traits for determining the fiber price in global textile market.

A large scale analysis of fiber gene expression in the elongation step had been previously conducted via cDNA array [48] [49] and deep RNA sequencing [50] [51]. Ji *et al.* [48] used subtractive PCR and cDNA macroarrays to identify 172 genes that were up-regulated in 10-DPA elongating fibers. They found that genes encoding a V-ATPase catalytic subunit, arabinogalactan proteins (AGPs), a kinesin-like calmodulin binding protein, cytochrome P450-like protein I, and proline-rich protein 5 were up-regulated. They also found that the mRNAs for proteins involved in signal transduction, lipid metabolism (such as LTP and ketoacyl CoA synthase), and cell wall-loosening enzymes (xyloglucan endotransglycosylase XET and expansins) were abundant. Using microarray analysis, Shi *et al.* [49] found that the ethylene biosynthesis pathway was up-regulated during fiber elongation, indicating that ethylene plays a role in promoting fiber elongation. Via high-throughput RNA sequencing, Wang *et al.* [50] conducted a global analysis of transcriptomes of two cotton lines, a wild type and a fuzzless/lintless mutant, at the early fiber development. They identified many genes differentially expressed in 2 - 8 DPA ovules, which included genes for a long-chain fatty acid elongation enzyme, 3-ketoacyl-CoA synthase, an ACC oxidase, NAC domain protein NAC2, P450 monooxygenase, WRKY transcription factor 45, MYB19, pectin esterase, chitinase, lipid transfer protein, and tubulin β -1 chain. Transcriptome analysis of short fiber mutant ligo lintless-1 had shown that DEGs involved in very-long-chain fatty acid biosynthesis and cell wall metabolism were differentially expressed in fiber bearing ovules at 3 and 8 DPA [51]. In this study, the identified DEGs in 10-DPA CS-B25 fiber were very similar to those found by Ji *et al.* [48], Wang *et al.* [50], and Liang *et al.* [51]. We found that many genes involved in primary cell wall synthesis were up-regulated in CS-B25 during the fiber elongation stage, which included fasciclin-like arabinogalactan proteins [34], non-specific lipid transfer proteins, cellulose synthase, UDP-glycosyltransferase, pectin methylesterases, xyloglucan endotransglucosylase (XET), glycine-rich cell wall structural proteins, cytosolic pyruvate kinase [46], and ascorbate peroxidase [47]. Pectin methylesterases (PME), a multigene family, catalyze the demethylesterification of galacturonic acid units in pectins, which are the major components of plant cell wall. PME have been shown to play an important role in plant cell wall extension during stem elongation, pollen germination and pollen tube growth, and root development [52]. XETs are generally considered to be important enzymes in cell wall loosening, but they have been also shown to function in secondary cell wall genesis during the expansion process [53] [54]. Both PME and XET were found to be up-regulated in a cotton MD52ne line which had high fiber strength [55]. Several cotton non-specific lipid transfer protein genes, including *pFS6*, *LTP3*, *LTP12*, were highly expressed in fiber cells [56] [57] [58]. LTP proteins have been proposed to participate in lipid transport for primary cell wall synthesis in fiber cells [59]. Recently, the functional role of cytosolic pyruvate kinase and ascorbate peroxidase in fiber elongation was elucidated. Zhang and Liu [46] showed that a cotton *GhPK6* gene encoding a cytosolic pyruvate kinase was preferentially expressed in fibers,

and its transcript level increased during the fiber elongation stage from 5 to 20 DPA and reached the maximum at the late elongation stage (20 DPA). Pyruvate kinase catalyzes the conversion of PEP to pyruvate in glycolysis, and pyruvate then undergoes oxidative phosphorylation in TCA cycle to generate the by-product, H_2O_2 , a form of reactive oxygen species (ROS) which was also found to have elevated levels during fiber elongation [46]. The GhPK6 activity was negatively correlated with the rate of fiber elongation [46]. An ascorbate peroxidase, a reactive oxygen species (ROS) scavenging enzyme encoded by the *GhAPX1* gene, was shown to be up-regulated in 10-DPA fibers [47]. Both the cytosolic pyruvate kinase and ascorbate peroxidase were involved in fiber elongation regulation in a ROS-mediated manner by modulating ROS homeostasis. A tubulin protein of tubulin beta 8 (Gh_Sca012883G01) and an ABI-1 like protein (Gh_A08G1131) in the assembly of cytoskeleton were also up-regulated (Table 3 and Table S3). It had been suggested that specific tubulins might affect microtubule organization and microfibril deposition in elongating fiber cells [37]. Three LRR receptor-like serine/threonine-protein kinases (Gh_D10G2231, Gh_D12G0597, and Gh_D02G0280) were highly up-regulated in CS-B25 (Table 3 and Table S3). It had been reported that these membrane-associated kinases were involved in signal transduction pathways [39], participating in cellulose deposition [40] and maintenance of cell wall integrity [41]. An LRR receptor-like kinase GhRLK1 was reported to be involved in secondary cell wall formation during fiber development [60]. Seventy-nine genes encoding *Catharanthus roseus* receptor-like kinases (CrRLK1Ls) had been identified in *G. hirsutum* TM-1 line. Some of them were found to be up-regulated in cotton fiber of chromosomal segment introgression lines (CSILs) with excellent fiber traits during the elongation stages from 10 to 20 DPA, but down-regulated at the early (5-DPA) and later (25-DPA) stages of fiber elongation. These observations revealed that RLKs positively affected fiber development [61].

The results of transcriptome analysis of 10 DPA fibers of the CS-B25 and TM-1 lines indicated that the synthesis and signaling pathways of GA, ethylene and auxin increased in CS-B25 when compared with TM-1. Many publications have indicated that ethylene modulates the elongation of cotton fiber cells. In this study, genes encoding the enzyme for ethylene synthesis, 1-aminocyclopropane-1-carboxylate oxidase (ACCO) and ethylene responsive transcription factors are up-regulated. Genes coding for ethylene overproduction 1-like protein (in chromatin remodeling and transcription regulator) and NAC transcription factors were also up-regulated. The gene encoding a RING E3 ligase XBAT32 (Gh_A08G0780), which targets the ACS proteins, was down-regulated in the CS-B25 fiber. This down regulation resulted in a slight increase of the ACS level. Auxin has been shown to promote fiber initiation and elongation and play an important role in fiber development [26]. The transcript levels of auxin-responsive GH3 protein, auxin-responsive IAA14 protein, and auxin influx carrier family protein in CS-B25 10 DPA fiber were up-regulated, suggesting that they may have important functions in the fiber elongation.

Several genes encoding MYB transcription factors were found to be highly up-regulated in the CS-B25 elongating fibers. A total of 524 non-redundant MYB genes had been identified in the upland cotton genome [62]. Among them, the *G. hirsutum* MYB genes *MYB2*, *MYB3*, *MYB5*, *MYB7*, *MYB20*, *MYB25*, *MYB42*, *MYB43*, *MYB46*, *MYB52*, *MYB54*, *MYB69*, *MYB85*, *MYB103*, and *MYB109* were shown to be highly expressed in fiber [42] [62], suggesting that they play important roles in fiber development. *GhMYB109* was reported to participate in fiber development, and the suppression of its expression resulted in the reduction of fiber length [63]. *GhMYB25* was shown to regulate trichome formation and act as a key factor in early cotton fiber development [64]. We had previously reported that a cotton MYB gene *GhMYB7* was differentially expressed in developing fibers, and *GhMYB7* was involved in transcriptional regulation of a *LTP3* gene encoding lipid transfer protein [65]. The *GhMYB7* was later shown to activate genes involved in secondary cell wall thickening and biosynthesis [66]. Besides the MYB transcription factors, homeobox leucine zipper (HD-Zip) transcription factors are also up-regulated in CS-B25. HD-Zip proteins were shown to play important roles in cotton fiber differentiation [67] [68]. Many NAC transcription factors were also highly expressed in CS-B25 10-DPA fibers. A total of 143 and 145 NAC genes have been identified in the genomes of the two diploid cottons *G. arboreum* and *G. raimondii*, respectively [69]. Among them, more than 30 had been shown to have higher expression levels in 10 - 20 DPA fiber. NAC genes had been shown to control secondary cell wall thickening during fiber development [69]. Our transcript profile observations during fiber development are similar to those of Al-Ghazi *et al.* [70]. Similar results had also been observed in some of Upland cotton chromosome introgression lines (CSILs) carrying different *G. barbadense* chromosomal segments in the recurrent parent TM-1 [71] [72]. These CSIL lines with longer and stronger fiber had increased expression in genes involved in cell wall biosynthesis, hormone signal transduction, and in genes encoding transcription factors during the fiber elongation stage. Both CS-B25 and chromosome segment introgression lines (CSILs) were useful resources for transcriptome profiling to study molecular mechanism of positive fiber traits.

Genome wide analysis has identified 116 and 102 WRKY genes in *G. raimondii* and *G. hirsutum*, respectively [73]. We have only found that two WRKY transcription factors, WRKY 21 and WRKY 29, were moderately up-regulated in CS-B25. This observation is consistent with the results of Dou *et al.* [73], in that many *GhWRKY* transcription factors had high expression levels at 0, 20, and 25 DPA fiber and they are involved in fiber initiation and secondary wall synthesis. Using the homologous alignment method, 143 genes for NAC-transcription factors in the *G. arboreum* genome were identified [69]. Many of them were highly expressed in 15 DPA fiber cells and might be involved in the fiber secondary wall thickening.

The genetic effects in the CS-B25 line could be due to any of the following reasons: 1) the genes on the substituted chromosome, 2) the genes on the re-

remaining chromosomes of its recurrent parent, and 3) an interaction between the gene(s) on the substituted chromosome and the remaining chromosomes of its recurrent parent [8]. A comparative analysis of CS-B25 with TM-1 is expected to show the differential transcripts from the direct effect of alien species substituted chromosome or an interaction effect of the substituted chromosome and the remaining chromosomes of CS-B25. Results showed that the majority genes up- and down-regulated in 10-DPA fiber in CS-B25 were not located on Pima 3-79 chromosome 25, which is mapped to Gh_D06G in the *G. hirsutum* (AD1) NAU-NBI genome used in this study (Table S3) [74]. This observation suggested that chromosome interactions between the substituted *G. barbadense* chromosome 25 pair and other non-substituted chromosomes of TM-1 might be involved in potential gene regulation. This is important considering that molecular markers (including EST markers) are used to construct linkage map of important QTLs with the expectation that these markers are directly associated with the QTL of interest in the same chromosomal region and can be used in marker assisted selection using genomic DNA to expedite the breeding program. However, our results showed that some of the ESTs may not have direct association with the chromosome but can cause a chromosomal effect on the genes. For example, the substituted *G. barbadense* chromosome 25 may contain potential regulatory elements that interact with gene located on other chromosomes in CS-B25 to cause an effect. Our results showed that several transcription factor like elements were associated with differentially expressed transcripts in CS-B25 line. It would be interesting in future investigation to identify the roles of some inter-chromosomal interactions in the activation or repression of gene transcription. The role of long-range chromosomal interactions involved in gene regulation has been reviewed in details by different research groups [75] [76]. The CS-B25 line can be used as a tool for the study of cotton chromosomal interactions and their role in transcriptional regulation because of the unique genetic background of CS-B25. The Chromosome Conformation Capture (3C), a high throughput technique developed by Dekker *et al.* [77], has been widely used to study chromosomal conformation/interaction. Using the 3C-based technology would allow us to study chromosomal interactions in CS-B25 and reveal how the interactions might affect up or down regulation of gene expression.

The results from the dendrogram and fiber QTL analyses suggested that chromosome 25 from *G. barbadense* is associated with potential valuable fiber traits (Figure 4). A previous report [8] also documented that CS-B25 had important association with several improved fiber traits, including decreasing micronaire, increasing fiber length and strength compared to TM-1 and some USA commercial cultivars, suggesting that CS-B25 would be an important genetic tool to discover many useful genes associated with fiber development. Cotton fiber, an elongated and thickened single cell of the seed epidermis in which the fiber cell wall polymers help to determine primarily the cotton fiber quality. Cotton fiber cell undergoes several phase of changes since its initiation: 1) extreme elongation (to >2.5 cm) via primary wall synthesis till about 10 days; 2) then

transitional wall thickening and primary wall remodeling; and 3) secondary wall thickening via deposition of nearly pure cellulose; and 4) final maturation and cell death processes before boll opening to reveal the fluffy fiber within the cotton bolls [78]. Our results showed that CS-B25 had several up or down regulated differentially expressed transcripts in 10 DPA fiber associated with potential cytoskeleton or cell wall genes (Table 3). It is important to note that most of the improved cotton fiber quality traits are primarily controlled by the cytoskeleton or cell wall related genes during fiber development [77]. These results suggested that future investigation with these differentially expressed transcripts, especially very high upregulated expression of fasciclin-like arabinogalactan protein 7 genes in CS-B25 10 DPA fiber cell, will have great potential to elucidate genes associated with fiber quality traits at the molecular level.

5. Conclusion

This study showed that the CS-B25 line is a valuable genetic resource to discern many novel differentially expressed genes due to the effect of the substituted alien species chromosome 25 or its interaction with other chromosomes in CS-B25. The CS-B25 line had several valuable fiber quality traits including micronaire and fiber length associated with the substituted chromosome 25 from *G. barbadense*. We discovered that several genes involved in hormone signal transduction, cell wall synthesis, and transcription factors in regulating cell wall synthesis are up-regulated in 10 DPA fiber of the CS-B25 line. These differentially expressed genes could be used as potential resources for fiber quality improvement. This research will facilitate our understanding of molecular mechanisms of fiber cell elongation and will allow us to develop new approaches for cotton fiber improvement.

Acknowledgements

The authors gratefully acknowledge that this research was made possible through the collaborative efforts between USDA/ARS and Mississippi State University/Mississippi Agriculture and Forestry Experimental Station (MAFES). We thank Uzbek government in providing training opportunities for Mr. Mirazakamol A. Ayubov in the USA laboratories at USDA and Mississippi State University, Mississippi State, Mississippi.

References

- [1] Basra, A.S. and Malik, C. (1984) Development of the Cotton Fiber. *International Review of Cytology*, **89**, 65-113. [https://doi.org/10.1016/S0074-7696\(08\)61300-5](https://doi.org/10.1016/S0074-7696(08)61300-5)
- [2] DeLanghe, E.A.L. (1986) Lint Development. In: Mauney, J.R. and Stewart, J.M., Ed., *Cotton Physiology*, The Cotton Foundation, Memphis, 325-349.
- [3] <http://cottonaustralia.com.au/cotton-library/fact-sheets/cotton-fact-file-cotton-properties-and-products>
- [4] Tyagi, P., Gore, M.A., Bowman, D.T., Campbell, B.T., Udall, J.A. and Kuraparthy, V. (2014) Genetic Diversity and Population Structure in the US Upland Cotton

- (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics*, **127**, 283-295.
<https://doi.org/10.1007/s00122-013-2217-3>
- [5] Stelly, D.M., Saha, S., Raska, D.A., Jenkins, J.N., McCarty, J.C. and Gutierrez, O.A. (2005) Registration of 17 Upland (*Gossypium hirsutum*) Cotton Germplasm Lines Disomic for Different *G. barbadense* Chromosome or Arm Substitutions. *Crop Science*, **45**, 2663-2665. <https://doi.org/10.2135/cropsci2004.0642>
- [6] Saha, S., Wu, J., Jenkins, J.N., McCarty, J.C., Stelly, D.M., Percy, R.G., Raska, D.A. and Gutierrez, O.A. (2004) Effect of Chromosome Substitutions from *Gossypium barbadense* L. 3-79 into *G. hirsutum* L. TM-1 on Agronomic and Fiber Traits. *The Journal of Cotton Science*, **8**, 162-169.
- [7] Saha, S., Jenkins, J.N., Wu, J., McCarty, J.C., Gutierrez, O.A., Percy, R.G., Cantrell, R.G. and Stelly, D.M. (2006) Effects of Chromosome-Specific Introgression in Upland Cotton on Fiber and Agronomic Traits. *Genetics*, **172**, 1927-1938.
<https://doi.org/10.1534/genetics.105.053371>
- [8] Wu, J., Jenkins, J.N., McCarty, J.C., Saha, S. and Stelly, D.M. (2006) An Additive Dominance Model to Determine Chromosomal Effects in Chromosome Substitution Lines and Other Germplasms. *Theoretical and Applied Genetics*, **112**, 391-399.
<https://doi.org/10.1007/s00122-005-0042-z>
- [9] Jenkins, J.N., McCarty, J.C., Wu, J., Saha, S., Gutierrez, O.A., Hayes, R. and Stelly, D.M. (2007) Genetic Effects of Thirteen *Gossypium barbadense* L. Chromosome Substitution Lines in Top Crosses with Upland Cotton Cultivars: II. Fiber Quality Traits. *Crop Science*, **47**, 561-570. <https://doi.org/10.2135/cropsci2006.06.0396>
- [10] Wu, J., Jenkins, J.N., McCarty, J.C. and Saha, S. (2010) Genetic Effects of Individual Chromosomes in Cotton Cultivars Detected by Using Chromosome Substitution Lines as Genetic Probes. *Genetica*, **138**, 1171-1179.
<https://doi.org/10.1007/s10709-010-9507-3>
- [11] Saha, S., Wu, J., Jenkins, J.N., McCarty, J., Hayes, R. and Stelly, D.M. (2010) Genetic Dissection of Chromosome Substitution Lines Discovered Novel Alleles in *Gossypium barbadense* L. with Potential for Improving Agronomic Traits Including Yield. *Theoretical and Applied Genetics*, **120**, 1193-1205.
<https://doi.org/10.1007/s00122-009-1247-3>
- [12] Wu, Z., Soliman, K.M., Bolton, J.J., Saha, S. and Jenkins, J.N. (2008) Identification of Differentially Expressed Genes Associated with Cotton Fiber Development in a Chromosomal Substitution Line (CS-B22sh). *Functional & Integrative Genomics*, **8**, 165-174. <https://doi.org/10.1007/s10142-007-0064-5>
- [13] Wan, C.Y. and Wilkins, T.A. (1994) A Modified Hot Borate Method Significantly Enhances the Yield of High-Quality RNA from Cotton (*Gossypium hirsutum* L.). *Analytical Biochemistry*, **223**, 7-12. <https://doi.org/10.1006/abio.1994.1538>
- [14] Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Sasaki, C.A., Scheffler, B.E., Stelly, D.M., et al. (2015) Sequencing of Allotetraploid Cotton (*Gossypium hirsutum* L. acc. TM-1) Provides a Resource for Fiber Improvement. *Nature Biotechnology*, **33**, 531-537. <https://doi.org/10.1038/nbt.3207>
- [15] Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions. *Genome Biology*, **14**, R36.
<https://doi.org/10.1186/gb-2013-14-4-r36>
- [16] Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq—A Python Framework to Work with High-Throughput Sequencing Data. *Bioinformatics*, **31**, 166-169.
<https://doi.org/10.1093/bioinformatics/btu638>

- [17] Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics*, **26**, 139-140. <https://doi.org/10.1093/bioinformatics/btp616>
- [18] Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D. and Woolf, J.P. (2009) GAGE: Generally Applicable Gene Set Enrichment for Pathway Analysis. *BMC Bioinformatics*, **10**, 161. <https://doi.org/10.1186/1471-2105-10-161>
- [19] Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. (2000) Gene Ontology: Tool for the Unification of Biology. *Nature Genetics*, **25**, 25-29. <https://doi.org/10.1038/75556>
- [20] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **27**, 29-34. <https://doi.org/10.1093/nar/27.1.29>
- [21] Yu, J., Jung, S., Cheng, C.-H., Ficklin, S.P., Lee, T., Zheng, P., Jones, D., Percy, R.G. and Main, D. (2013) CottonGen: A Genomics, Genetics and Breeding Database for Cotton Research. *Nucleic Acids Research*, **42**, D1229-D1236. <https://doi.org/10.1093/nar/gkt1064>
- [22] Supek, F., Bosnjak, M., Skunca, N. and Smuc, T. (2011) REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, **6**, e21800. <https://doi.org/10.1371/journal.pone.0021800>
- [23] Sun, Q., Jiang, H., Zhu, X., Wang, W., He, X., Shi, Y., Yuan, Y., Du, X. and Cai, Y. (2013) Analysis of Sea-Island and Upland Cotton in Response to *Verticillium dahlia* Infection by RNA Sequencing. *BMC Genomics*, **14**, 852. <https://doi.org/10.1186/1471-2164-14-852>
- [24] Livak, K.J. and Schmittgen, T.D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods*, **25**, 402-408. <https://doi.org/10.1006/meth.2001.1262>
- [25] Egamberdiev, S., Saha, S., Salakhutdinov, I., Jenkins, J.N., Deng, D., Khurshut, E. and Abdurakhmonov, I. (2016) Comparative Assessment of Genetic Diversity in Cytoplasmic and Nuclear Genome of Upland Cotton. *Genetica*, **144**, 289-306. <https://doi.org/10.1007/s10709-016-9898-x>
- [26] Liao, W., Zhang, J., Xu, N. and Peng, M. (2010) The Role of Phytohormones in Cotton Fiber Development. *Russian Journal of Plant Physiology*, **57**, 462-468. <https://doi.org/10.1134/S1021443710040023>
- [27] Ruduś, I., Sasiak, M. and kepczyński, J. (2013) Regulation of Ethylene Biosynthesis at the Level of 1-Aminocyclopropane-1-Carboxylate Oxidase (ACO) Gene. *Acta Physiologiae Plantarum*, **35**, 295-307. <https://doi.org/10.1007/s11738-012-1096-6>
- [28] Qin, Y.-M., Hu, C.-Y., Pang, Y., Kastaniotis, A.J., Hiltunen, J.K. and Zhu, Y.-X. (2007) Saturated Very-Long Fatty Acids Promote Cotton Fiber and Arabidopsis Cell Elongation by Activating Ethylene Biosynthesis. *Plant Cell*, **19**, 3692-3704. <https://doi.org/10.1105/tpc.107.054437>
- [29] Kim, H.-S., Kim, S.J., Abbasi, N., Bressan, R.A., Yun, D.-J., Yoo, S.-D., Kwon, S.-Y. and Choi, S.-B. (2010) The DOF Transcription Factor Dof5.1 Influences Leaf Axial Patterning by Promoting Revoluta Transcription in Arabidopsis. *The Plant Journal*, **64**, 524-535. <https://doi.org/10.1111/j.1365-313X.2010.04346.x>
- [30] Prasad, M.E., Schofield, A., Lyzenga, W., Liu, H. and Stone, S. (2010) Arabidopsis RING E3 Ligase XBAT32 Regulates Lateral Root Production through Its Role in Ethylene Biosynthesis. *Plant Physiology*, **153**, 1587-1596.

- <https://doi.org/10.1104/pp.110.156976>
- [31] Pech, J.C., Latché, A. and Bouzayen, M. (2010) Ethylene Biosynthesis. In: Davies, P.J., Ed., *Plant Hormones: Biosynthesis, Signal Transduction, Action*, 3rd Edition, Springer, Dordrecht, 115-136. https://doi.org/10.1007/978-1-4020-2686-7_6
- [32] Gosti, F., Beaudoin, N., Serizet, C., Webb, A.A.R., Vartanian, N. and Giraudat, J. (1999) ABI1 Protein Phosphatase 2C Is a Negative Regulator of Abscisic Acid Signaling. *Plant Cell*, **11**, 1897-1909. <https://doi.org/10.1105/tpc.11.10.1897>
- [33] Gilbert, M.K., Bland, J.M., Shockey, J.M., Cao, H., Hinchliffe, D.J., Fang, D.D. and Naoumkina, M. (2013) A Transcript Profiling Approach Reveals an Abscisic Acid-Specific Glycosyltransferase (UGT73C14) Induced in Developing Fiber of Ligon Lintless-2 Mutant of Cotton (*Gossypium hirsutum* L.). *PLoS ONE*, **8**, e75268. <https://doi.org/10.1371/journal.pone.0075268>
- [34] Huang, G.-Q., Gong, S.-Y., Xu, W.-L., Li, W., Li, P., Zhang, C.-J., Li, D.-D., Zheng, Y., Li, F.-G. and Li, X.-B. (2013) A Fasciclin-Like Arabinogalactan Protein, GhFLA1, Is Involved in Fiber Initiation and Elongation of Cotton. *Plant Physiology*, **161**, 1278-1290. <https://doi.org/10.1104/pp.112.203760>
- [35] Deng, T., Yao, H., Wang, J., Wang, J., Xue, H. and Zuo, K. (2016) GhLTPG1, a Cotton GPI-Anchored Lipid Transfer Protein, Regulates the Transport of Phosphatidylinositol Monophosphates and Cotton Fiber Elongation. *Scientific Reports*, **6**, Article No. 26829.
- [36] Liu, Q., Talbot, M. and Llewellyn, D.J. (2013) Pectin Methyltransferase and Pectin Remodeling Differ in the Fibre Walls of Two *Gossypium* Species with Different Fibre Properties. *PLoS ONE*, **8**, e65131. <https://doi.org/10.1371/journal.pone.0065131>
- [37] Whittaker, D.J. and Triplett, B.A. (1999) Gene Specific Changes in α -Tubulin Transcript Accumulation in Developing Cotton Fibers. *Plant Physiology*, **121**, 181-188. <https://doi.org/10.1104/pp.121.1.181>
- [38] Szymanski, D.B. (2005) Breaking the WAVE Complex: The Point of Arabidopsis Trichomes. *Current Opinion in Plant Biology*, **8**, 103-112. <https://doi.org/10.1016/j.pbi.2004.11.004>
- [39] Afzal, A.J., Wood, A.J. and Lightfoot, D.A. (2008) Plant Receptor-Like Serine Threonine Kinases: Roles in Signaling and Plant Defense. *Molecular Plant-Microbe Interactions*, **21**, 507-517. <https://doi.org/10.1094/MPMI-21-5-0507>
- [40] Xu, S.-L., Rahman, A., Baskin, T.I. and Kieber, J.J. (2008) Two Leucine-Rich Repeat Receptor Kinases Mediate Signaling, Linking Cell Wall Biosynthesis and ACC Synthase in Arabidopsis. *Plant Cell*, **20**, 3065-3079. <https://doi.org/10.1105/tpc.108.063354>
- [41] Hamann, T. (2015) The Plant Cell Wall Integrity Maintenance Mechanism—A Case Study of a Cell Wall Plasma Membrane Signaling Network. *Phytochemistry*, **112**, 100-109. <https://doi.org/10.1016/j.phytochem.2014.09.019>
- [42] Zhong, Z., Lee, C., Zhou, J., McCarthy, R.L. and Ye, Z.-H. (2008) A Battery of Transcription Factors Involved in the Regulation of Secondary Cell Wall Biosynthesis in Arabidopsis. *Plant Cell*, **20**, 2763-2782. <https://doi.org/10.1105/tpc.108.061325>
- [43] Ko, J.-H., Kim, W.-C., Kim, J.-Y., Ahn, S.-J. and Han, K.-H. (2012) MYB46-Mediated Transcriptional Regulation of Secondary Wall Synthesis. *Molecular Plant*, **5**, 961-963. <https://doi.org/10.1093/mp/sss076>
- [44] Zhong, R. and Ye, Z.H. (2012) MYB46 and MYB83 Bind to the SMRE Sites and Directly Activate a Suite of Transcription Factors and Secondary Wall Biosynthetic Genes. *Plant and Cell Physiology*, **53**, 368-380. <https://doi.org/10.1093/pccp/pcr185>

- [45] Shang, H., Li, W., Zou, C. and Yuan, Y. (2013) Analyses of the NAC Transcription Factor Gene Family in *Gossypium raimondii* Ulbr: Chromosomal Location, Structure, Phylogeny, and Expression Patterns. *Journal of Integrative Plant Biology*, **55**, 663-676. <https://doi.org/10.1111/jipb.12085>
- [46] Zhang, B. and Liu, J.Y. (2016) Cotton Cytosolic Pyruvate Kinase GhPK6 Participates in Fast Fiber Elongation in a ROS-Mediated Manner. *Planta*, **244**, 915-926. <https://doi.org/10.1007/s00425-016-2557-8>
- [47] Guo, K., Du, X., Tu, L., Tang, W., Wang, P., Wang, M., Liu, Z. and Zhang, X. (2016) Fibre Elongation Requires Normal Redox Homeostasis Modulated by Cytosolic Ascorbate Peroxidase in Cotton (*Gossypium hirsutum*). *Journal of Experimental Botany*, **67**, 3289-3301. <https://doi.org/10.1093/jxb/erw146>
- [48] 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 Analysis of Genes Preferentially Expressed during Early Cotton Fiber Development by Subtractive PCR and cDNA Array. *Nucleic Acids Research*, **31**, 2534-2543. <https://doi.org/10.1093/nar/gkg358>
- [49] Shi, Y.H., Zhu, S.W., Mao, X.Z., Feng, J.X., Qin, Y.M., Zhang, L., *et al.* (2006) Transcriptome Profiling, Molecular Biological, and Physiological Studies Reveal a Major Role for Ethylene in Cotton Fiber Elongation. *Plant Cell*, **18**, 651-664. <https://doi.org/10.1105/tpc.105.040303>
- [50] Wang, Q.Q., Liu, F., Chen, X.S., Ma, X.J., Zeng, H.Q. and Yang, Z.M. (2010) Transcriptome Profiling of Early Developing Cotton Fiber by Deep-Sequencing Reveals Significantly Differential Expression of Genes in a Fuzzless/Lintless Mutant. *Genomics*, **96**, 369-376. <https://doi.org/10.1016/j.ygeno.2010.08.009>
- [51] Liang, W., Fang, L., Xiang, D., Hu, Y., Feng, H., Chang, L. and Zhang, T. (2015) Transcriptome Analysis of Short Fiber Mutant Ligon Lintless-1 (Li1) Reveals Critical Genes and Key Pathways in Cotton Fiber Elongation and Leaf Development. *PLoS ONE*, **10**, e0143503. <https://doi.org/10.1371/journal.pone.0143503>
- [52] Micheli, F. (2001) Pectin Methylsterases: Cell Wall Enzymes with Important Roles in Plant Physiology. *Trends in Plant Science*, **6**, 414-419. [https://doi.org/10.1016/S1360-1385\(01\)02045-3](https://doi.org/10.1016/S1360-1385(01)02045-3)
- [53] Bourquin, V., Nishikubo, N., Abe, H., Brumer, H., Denman, S., Eklund, M., Christiernin, M., Teeri, T.T., Sundberg, B. and Mellerowicz, E.J. (2002) Xyloglucan Endotransglycosylases Have a Function during the Formation of Secondary Cell Walls of Vascular Tissues. *Plant Cell*, **14**, 3073-3088. <https://doi.org/10.1105/tpc.007773>
- [54] Lee, J., Burns, T.H., Light, G., Sun, Y., Fokar, M., Kasukabe, Y., *et al.* (2010) Xyloglucan Endotransglycosylase/Hydrolase Genes in Cotton and Their Role in Fiber Elongation. *Planta*, **232**, 1191-1205. <https://doi.org/10.1007/s00425-010-1246-2>
- [55] Islam, M.S., Fang, D.D., Thyssen, G.N., Delhom, C.D., Liu, Y. and Kim, H.J. (2016) Comparative Fiber Property and Transcriptome Analyses Reveal Key Genes Potentially Related to High Fiber Strength in Cotton (*Gossypium hirsutum* L.) Line MD52ne. *BMC Plant Biology*, **16**, 36. <https://doi.org/10.1186/s12870-016-0727-2>
- [56] Orford, S.J. and Timmis, J.N. (2000) Expression of a Lipid Transfer Protein Gene Family during Cotton Fibre Development. *Biochimica et Biophysica Acta*, **1483**, 275-284. [https://doi.org/10.1016/S1388-1981\(99\)00194-8](https://doi.org/10.1016/S1388-1981(99)00194-8)
- [57] Hsu, C.Y., Creech, R.G., Jenkins, J.N. and Ma, D.P. (1999) Analysis of Promoter Activity of Cotton Lipid Transfer Protein Gene *LTP6* in Transgenic Tobacco Plants. *Plant Science*, **143**, 63-70. [https://doi.org/10.1016/S0168-9452\(99\)00026-6](https://doi.org/10.1016/S0168-9452(99)00026-6)
- [58] Liu, H.C., Creech, R.G., Jenkins, J.N. and Ma, D.P. (2000) Cloning and Promoter Analysis of the Cotton Lipid Transfer Protein Gene *Ltp3*. *Biochimica et Biophysica*

- Acta*, **1487**, 106-111. [https://doi.org/10.1016/S1388-1981\(00\)00072-X](https://doi.org/10.1016/S1388-1981(00)00072-X)
- [59] Ma, D.P., Liu, H.C., Tan, H., Creech, R.G., Jenkins, J.N. and Chang, Y.F. (1997) Cloning and Characterization of a Lipid Transfer Protein Gene Specifically Expressed in Fiber Cells. *Biochimica et Biophysica Acta*, **1344**, 111-114. [https://doi.org/10.1016/S0005-2760\(96\)00166-X](https://doi.org/10.1016/S0005-2760(96)00166-X)
- [60] Li, Y.-L., Sun, J. and Xia, G.-X. (2005) Cloning and Characterization of a Gene for an LRR Receptor-Like Protein Kinase Associated with Cotton Fiber Development. *Molecular Genetics and Genomics*, **273**, 217-224. <https://doi.org/10.1007/s00438-005-1115-z>
- [61] Niu, E., Cai, C., Zheng, Y., Shang, X. and Fang, L. (2016) Genome-Wide Analysis of CrRLK1L Gene Family in *Gossypium* and Identification of Candidate CrRLK1L Genes Related to Fiber Development. *Molecular Genetics and Genomics*, **291**, 1137-1154. <https://doi.org/10.1007/s00438-016-1169-0>
- [62] Salih, H., Gong, W., He, S., Sun, G., Sun, J. and Du, X. (2016) Genome-Wide Characterization and Expression Analysis of MYB Transcription Factors in *Gossypium hirsutum*. *BMC Genetics*, **17**, 129. <https://doi.org/10.1186/s12863-016-0436-8>
- [63] Pu, L., Li, Q., Fan, X., Yang, W. and Xue, Y. (2008) The R2R3 MYB Transcription Factor GhMYB109 Is Required for Cotton Fiber Development. *Genetics*, **180**, 811-820. <https://doi.org/10.1534/genetics.108.093070>
- [64] Walford, S.A., Wu, Y., Ilewellyn, D.J. and Dennis, E.S. (2011) GhMYB25-Like: A Key Factor in Early Cotton Fibre Development. *The Plant Journal*, **65**, 785-797. <https://doi.org/10.1111/j.1365-3113X.2010.04464.x>
- [65] Hsu, C.Y., Jenkins, J.N., Saha, S. and Ma, D.P. (2005) Transcriptional Regulation of the Lipid Transfer Protein Gene *LTP3* in Cotton Fibers by a Novel MYB Protein. *Plant Science*, **168**, 167-181. <https://doi.org/10.1016/j.plantsci.2004.07.033>
- [66] Huang, J., Chen, F., Wu, S., Li, J. and Xu, W. (2016) Cotton GhMYB7 Is Predominantly Expressed in Developing Fibers and Regulates Secondary Cell Wall Biosynthesis in Transgenic Arabidopsis. *Science China Life Sciences*, **59**, 194-205. <https://doi.org/10.1007/s11427-015-4991-4>
- [67] Walford, S.A., Wu, Y., Ilewellyn, D.J. and Dennis, E.S. (2012) Epidermal Cell Differentiation in Cotton Mediated by the Homeodomain Leucine Zipper Gene, GhHD-1. *The Plant Journal*, **71**, 464-478.
- [68] Zahur, M., Asif, M.A., Zeeshan, N., Mehmood, S., Malik, M.F. and Asif, A.R. (2013) Homeobox Leucine Zipper Proteins and Cotton Improvement. *Advances in Bioscience and Biotechnology*, **4**, 15-20. <https://doi.org/10.4236/abb.2013.410A3003>
- [69] Shang, H., Wang, Z., Zou, C., Zhang, Z., Li, W., Li, J., Shi, Y., Gong, W., Chen, T., Liu, A., Gong, J., Ge, Q. and Yuan, Y. (2016) Comprehensive Analysis of NAC Transcription Factors in Diploid *Gossypium*: Sequence Conservation and Expression Analysis Uncover Their Roles during Fiber Development. *Science China Life Sciences*, **59**, 142-153. <https://doi.org/10.1007/s11427-016-5001-1>
- [70] Al-Ghazi, Y., Bourot, S., Arioli, T., Dennis, E.S. and Ilewellyn, D.J. (2009) Transcript Profiling during Fiber Development Identifies Pathways in Secondary Metabolism and Cell Wall Structure That May Contribute to Cotton Fiber Quality. *Plant and Cell Physiology*, **50**, 1364-1381. <https://doi.org/10.1093/pcp/pcp084>
- [71] Fang, L., Tian, R., Chen, J., Wang, S., Li, X., Wang, P. and Zhang, T. (2014) Transcriptomic Analysis of Fiber Strength in Upland Cotton Chromosome Introgression Lines Carrying Different *Gossypium barbadense* Chromosomal Segments. *PLoS ONE*, **9**, e94642. <https://doi.org/10.1371/journal.pone.0094642>
- [72] Fang, L., Tian, R., Li, X., Chen, J., Wang, S., Wang, P. and Zhang, T. (2014) Cotton

- Fiber Elongation Network Revealed by Expression Profiling of Longer Fiber Lines Introgressed with Different *Gossypium barbadense* Chromosome Segments. *BMC Genomics*, **15**, 838. <https://doi.org/10.1186/1471-2164-15-838>
- [73] Dou, L., Zhang, X., Pang, C., Song, M., Wei, H., Fan, S. and Yu, S. (2014) Genome-Wide Analysis of the WRKY Gene Family in Cotton. *Molecular Genetics and Genomics*, **289**, 1103-1121. <https://doi.org/10.1007/s00438-014-0872-y>
- [74] Yu, J.Z., Ulloa, M., Hoffman, S.M., Kohel, R.J., Pepper, A.E., Fang, D.D., Percy, R.G. and Burke, J.J. (2014) Mapping Genomic Loci for Cotton Plant Architecture, Yield Components, and Fiber Properties in an Interspecific (*Gossypium hirsutum* L. x *G. barbadense* L.) RIL Population. *Molecular Genetics and Genomics*, **289**, 1347-1367. <https://doi.org/10.1007/s00438-014-0930-5>
- [75] Bartkuhn, M. and Renkawitz, R. (2008) Long Range Chromatin Interactions Involved in Gene Regulation. *Biochimica et Biophysica Acta*, **1783**, 2161-2166. <https://doi.org/10.1016/j.bbamcr.2008.07.011>
- [76] Miele, A. and Dekker, J. (2008) Long-Range Chromosomal Interactions and Gene Regulation. *Molecular BioSystems*, **4**, 1046-1057. <https://doi.org/10.1039/b803580f>
- [77] Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) Capturing Chromosome Conformation. *Science*, **295**, 1306-1311. <https://doi.org/10.1126/science.1067799>
- [78] Haigler, C.H., Betancur, L., Stiff, M.R. and Tuttle, J.R. (2012) Cotton Fiber: A Powerful Single-Cell Model for Cell Wall and Cellulose Research. *Front Plant Science*, **3**, 104. <https://doi.org/10.3389/fpls.2012.00104>
- [79] Mei, H., Zhu, X., Guo, W., Cai, C. and Zhang, T. (2014) Exploitation of Chinese Upland Cotton Cultivar Germplasm Resources to Mine Favorable QTL Alleles Using Association Mapping. In: Abdurakhmonov, I.Y., Ed., *World Cotton Germplasm Resources*, Intech, Rijeka, 55-85.
- [80] Zhang, Z., Li, J.W., Muhammad, J., Cai, J., Jia, F., Shi, Y.Z., *et al.* (2015) High Resolution Consensus Mapping of Quantitative Trait Loci for Fiber Strength, Length and Micronaire on Chromosome 25 of the Upland Cotton (*Gossypium hirsutum* L.). *PLoS ONE*, **10**, e0135430. <https://doi.org/10.1371/journal.pone.0135430>
- [81] Abdurakhmonov, I.Y., Saha, S., Jenkins, J.N., Buriev, Z.T., Shermatov, S.E., Scheffler, B.E., *et al.* (2009) Linkage Disequilibrium Based Association Mapping of Fiber Quality Traits in *G. hirsutum* L. Variety Germplasm. *Genetica*, **136**, 401-417. <https://doi.org/10.1007/s10709-008-9337-8>
- [82] Sun, F.D., Zhang, J.H., Wang, S.F., Gong, W.K., Shi, Y.Z., Liu, A.Y., Li, J.W., Gong, J.W., Shang, H.H. and Yuan, Y.L. (2012) QTL Mapping for Fiber Quality Traits across Multiple Generations and Environments in Upland Cotton. *Molecular Breeding*, **30**, 569-582. <https://doi.org/10.1007/s11032-011-9645-z>
- [83] Liang, Q., Hu, C., Hua, H., Li, Z. and Hua, J. (2013) Construction of a Linkage Map and QTL Mapping for Fiber Quality Traits in Upland Cotton (*Gossypium hirsutum* L.). *Chinese Science Bulletin*, **58**, 3233-3243. <https://doi.org/10.1007/s11434-013-5807-1>
- [84] Qin, H., Chen, M., Yi, X., Bie, S., Zhang, C., Zhang, Y., *et al.* (2015) Identification of Associated SSR Markers for Yield Component and Fiber Quality Traits Based on Frame Map and Upland Cotton Collections. *PLoS ONE*, **10**, e0118073. <https://doi.org/10.1371/journal.pone.0118073>

Supplement

Table S1. Primers used for quantitative RT-PCR.

Gene ID	Primer Name	Sequence (5'-3')	Tm (°C)
Gh_A08G0457	FLA7-F	GTTGCAATGAGCACAAAGCCCGACACCG	65.8
	FLA7-R	GTACGAAAATCGTAATGCCTTCAACCG	58.2
Gh_D02G0381	LTP-F	TTACCACCCCATTCAAATGATACCC	57.2
	LTP-R	TTAGGAGATGGTGGTCCAAGTTGAAG	58.3
Gh_A11G0727	PME2-F	CTTCGAAGCGGTCCGGAAGTGGAAAGG	64.7
	PME2-R	CACTGCCTGATGGTTAGTAGGACCAG	60.4
Gh_D08G1266	MYB46-F	CTGATATCAGAAGAGGGAGATTTACTC	54.2
	MYB46-R	AGTACTTGGTGGAGGAGCAGTGCTTG	62.3
Gh_D09G2166	CESA-G3-F	TCAACAACATTCACGACGGGTGACG	62.9
	CESA-G3-R	GTCGCAGTCAAGCATTAGAATATATGG	55.7
Gh_D02G0079	UGT74F2-F	GATGAAGGCAGATTCTTTGAAGTTAGG	55.9
	UGT74F2-R	TATGTAATTCACAGCGCACGCCTGCG	63.3
Gh_D06G0375	6-PGL2-F	GGAAAACGTAGTCCCTAAGCGACACC	60.9
	6-PGL2-R	GACAGTCGGTTGATCGGGAAAACCTG	61.3
Gh_A11G0400	PER72-F	CTTCCAAGACAGTTGGAGGAGGATAC	58.4
	PER72-R	AGCTGAGTTCCTGTTTGGATTGACACC	60.7
Gh_A08G0780	E3-F	GATTCCTTGTCTCATGCCAAGATCC	57.8
	E3-R	TAACATTGGTTTTGCTCCAGGGAGC	59.0
Gh_A01G0064	B3-F	ACAGAGGATAAGCTCTCTCAAGCAGG	59.6
	B3-R	TGCTCGTTCATTACAGCCAATTTGG	59.9
Gh_A04G0106	ERF3-F	CTCGAATACTACCGCTTTTCCGATCG	59.0
	ERF3-R	GATGATATTTCCGGTCCGACACAACC	59.4
Gh_A04G0473	RPS5-F	GCCTCACTGCAAGACCCAACCAAGTC	63.6
	RPS5-R	GCACTGAGCCTTTCTAAAACGCTTAAC	58.4
	18S rRNA-F	GGAATTGACGGAAGGGCACCACCAGGC	78.0
	18S rRNA-R	GGACATCTAAGGGCATCACAGACCTG	68.0

Table S2. Association of seven chromosome 25 specific SSR markers with fiber traits.

SSR marker	Fiber trait	References
BNL3436	Fineness	Mei <i>et al.</i> 2014 [79]
BNL3806	Length, strength	Zhang <i>et al.</i> 2015 [80]
	Length	Abdurakhmonov <i>et al.</i> 2009 [81]
DPL0067	Micronaire, elongation, uniformity	Sun <i>et al.</i> 2012 [82]
	Length	Zhang <i>et al.</i> 2015 [80]
GH220	Elongation, micronaire	Sun <i>et al.</i> 2012 [82]
	Elongation	Liang <i>et al.</i> 2013 [83]
BNL0827	Strength	Qin <i>et al.</i> 2015 [84]
BNL1064	Uniformity	Sun <i>et al.</i> 2012 [82]
BNL1440	Strength, length, uniformity, micronaire	Sun <i>et al.</i> 2012 [82]

Table S3. List of differentially expressed genes (DEGs) in CS-B25 10 DPA fibers.

<https://gitlab.com/IGBB/OBL/ma.csb25-fiber/raw/master/Table%20S3.xlsx>