

Transcriptome Analysis Suggests That Chromosome Introgression Fragments from Sea Island Cotton (*Gossypium barbadense*) Increase Fiber Strength in Upland Cotton (*Gossypium hirsutum*)

Quanwei Lu,^{*,†,‡,1} Yuzhen Shi,^{†,1} Xianghui Xiao,^{†,‡} Pengtao Li,[†] Juwu Gong,[†] Wankui Gong,[†] Aiying Liu,[†] Haihong Shang,[†] Junwen Li,[†] Qun Ge,[†] Weiwu Song,[†] Shaoqi Li,[†] Zhen Zhang,[†] Md Harun or Rashid,[†] Renhai Peng,[‡] Youlu Yuan,^{†,2} and Jinling Huang^{*,2}

^{*}College of Agriculture, Shanxi Agricultural University, Taigu 030801, Shanxi, China, [†]State Key Laboratory of Cotton Biology, Key Laboratory of Biological and Genetic Breeding of Cotton, The Ministry of Agriculture, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China, and [‡]School of Biotechnology and Food Engineering, Anyang Institute of Technology, Anyang 455000, Henan, China

ORCID ID: 0000-0001-9970-0952 (Q.L.)

ABSTRACT As high-strength cotton fibers are critical components of high quality cotton, developing cotton cultivars with high-strength fibers as well as high yield is a top priority for cotton development. Recently, chromosome segment substitution lines (CSSLs) have been developed from high-yield Upland cotton (*Gossypium hirsutum*) crossed with high-quality Sea Island cotton (*G. barbadense*). Here, we constructed a CSSL population by crossing CCRI45, a high-yield Upland cotton cultivar, with Hai1, a Sea Island cotton cultivar with superior fiber quality. We then selected two CSSLs with significantly higher fiber strength than CCRI45 (MBI7747 and MBI7561), and one CSSL with lower fiber strength than CCRI45 (MBI7285), for further analysis. We sequenced all four transcriptomes at four different time points postanthesis, and clustered the 44,678 identified genes by function. We identified 2200 common differentially-expressed genes (DEGs): those that were found in both high quality CSSLs (MBI7747 and MBI7561), but not in the low quality CSSL (MBI7285). Many of these genes were associated with various metabolic pathways that affect fiber strength. Upregulated DEGs were associated with polysaccharide metabolic regulation, single-organism localization, cell wall organization, and biogenesis, while the downregulated DEGs were associated with microtubule regulation, the cellular response to stress, and the cell cycle. Further analyses indicated that three genes, XLOC_036333 [mannosyl-oligosaccharide- α -mannosidase (*MNS1*)], XLOC_029945 (*FLA8*), and XLOC_075372 (*snakin-1*), were potentially important for the regulation of cotton fiber strength. Our results suggest that these genes may be good candidates for future investigation of the molecular mechanisms of fiber strength formation and for the improvement of cotton fiber quality through molecular breeding.

KEYWORDS

cotton
fiber strength
transcriptome
DEG
secondary cell
wall synthesis
GenPred
Shared Data
Resources
Genomic
Selection

Cotton is one of the largest commercial fiber crops worldwide, and represents the most prevalent source of renewable natural fiber for the global textile industry (Wendel and Cronn 2003). The simultaneous improvement of fiber quality and yield is currently the primary goal of cotton breeding programs (Constable *et al.* 2015). Upland cotton (*Gossypium hirsutum*) produces high yields of relatively low-quality fibers, whereas Sea Island cotton (*G. barbadense*) produces relatively low yields of high-quality fibers (Shi *et al.* 2015). Upland cotton plants are usually crossed with Sea Island cotton plants to generate introgression lines that produce improved yields of longer, stronger, finer fibers (Fang *et al.* 2014b; Gilbert *et al.* 2014).

Cotton fibers develop from single epidermal cells of the developing ovule (Gao *et al.* 2012); these epidermal cells serve as excellent models for biological studies of plant cell elongation, as well as of the mechanisms of biosynthesis of the cell wall and cellulose (Liu *et al.* 2015). From initiation [\sim 3 d postanthesis (DPA)] until 23 DPA, the fiber cells in *G. hirsutum* rapidly elongate, with a peak elongation rate of >2 mm/d (Fang *et al.* 2014b). Overlapping with this elongation stage, from 15 to 40 DPA, fiber cells deposit the secondary wall; maturation occurs between 40 and 50 DPA (Kim and Triplett 2001).

Fiber strength is a key indicator of cotton fiber quality (Shang *et al.* 2016). Cellulose is the primary component of cotton fibers ($>90\%$),

■ **Table 1 Polymorphism and fiber strength of the three CSSLs and the recurrent parent CCRI45**

Cotton	Number of Differential Makers	Proportion of Polymorphic Markers ^a	Upper Half Mean Length (mm)	Mironaire	Fiber Strength (cN/tex)
MBI7561	27	5.13%	32.10 ± 1.65 ^b	5.10 ± 0.45	34.30 ± 1.89 ^b
MBI7747	11	2.10%	33.80 ± 1.64 ^b	4.30 ± 0.41	36.50 ± 1.83 ^b
MBI7285	39	7.41%	27.10 ± 1.34	4.40 ± 0.45	25.90 ± 1.76
CCRI45	0	0	28.13 ± 0.71	5.03 ± 0.49	27.97 ± 0.93

^a526 SSR markers used for genetic testing.

^bCSSLs significantly different from CCRI45 ($P < 0.01$).

providing not only fiber strength, but also mechanical support and protection for the plant body (Edelmann and Fry 1992). The biosynthesis of cellulose increases over 100-fold during the deposition of the secondary cell wall (Zhong and Ye 2015). Rapid and dynamic changes in gene expression during this stage indicate that an amalgamation of complex processes may contribute to cotton fiber development, including transcriptional regulation, signal transduction, and metabolic pathway induction (Pan *et al.* 2007; Kim and Triplett 2001; Tuttle *et al.* 2015).

Several of the multiple genes that are involved in the deposition of the secondary cell wall and resultant fiber strength have been identified (Z. Zhang *et al.* 2015). For example, in *G. hirsutum*, the over-expression of *GbEXPA2* produces fibers with slightly thicker walls and higher crystalline cellulose content, while the overexpression of *GbEXPATR* produces longer, finer, stronger fibers with significantly thinner cell walls (Y. Li *et al.* 2016).

In addition to secondary cell wall biosynthesis, hormone responses might also influence fiber strength by improving fiber-to-fiber interactions (Islam *et al.* 2016a): receptor-like kinase signaling pathways might regulate cotton-fiber cell-wall assembly and strength by coordinating cell elongation and secondary cell wall biosynthesis (Islam *et al.* 2016a). However, the molecular mechanisms that regulate the strength and formation of the secondary cell walls of individual cotton fibers remain unclear. A better understanding of these mechanisms would facilitate the development of cotton cultivars with higher quality fibers.

The genetic and molecular basis of the traits of high quality fibers can be effectively investigated with genome-wide transcriptome profiles of relevant genes and pathways (Tuttle *et al.* 2015; Fang *et al.* 2014a,b); transcriptome analyses have provided insight into the dynamic regulation of fiber development (Tu *et al.* 2007; Islam *et al.* 2016b; Ma *et al.* 2016; Hu *et al.* 2013; Liang *et al.* 2015). Indeed, recent comparative transcriptomes analyses have suggested several genes as candidates for molecular breeding, genetic manipulation, and the regulation of cell-wall assembly and strength (Islam *et al.* 2016b; Wu *et al.* 2016). However, few studies have focused on introgression lines carrying *G. barbadense* fragments in a *G. hirsutum* background (Fang *et al.* 2014a,b).

To investigate the molecular mechanisms underlying cotton fiber strength regulation and the production of superior quality fibers, with a particular focus on introgression lines, we first aimed to create a population of CSSLs. We then aimed to identify CSSLs with higher and lower fiber strength than the recurrent parent, and sequence their transcriptomes. Finally, we aimed to compare, in developing fibers, the genes expressed differently in the higher fiber strength CSSLs as compared to the lower fiber strength CSSLs. We hypothesized that genes associated with different metabolic pathways would affect fiber strength and, in different introgression lines, gene expression in the same pathway would change over time.

MATERIALS AND METHODS

Genotyping and fiber strength testing

At the Institute of Cotton Research of the Chinese Academy of Agricultural Sciences (CRICAAS; Anyang, Henan province, China), we derived a population of 332 CSSLs by backcrossing the donor parent Hai1 with the recurrent parent CCRI45 for four generations, followed by three consecutive generations of selfing. Hai1 is a conventional cultivar of *G. barbadense* with a dominant glandless gene, high resistance to *Verticillium* wilt, and excellent fiber quality (Shi *et al.* 2015). CCRI45 is a widely grown upland cotton cultivar with high yield developed by CRICAAS (Shi *et al.* 2015). The 332 CSSLs and the two parent cultivars were genotyped. We then tested fiber strength across three environments: Anyang, China (2010, 2011), and Aksu, Xinjiang Uyghur Autonomous Region (2011; Supplemental Material, Figure S1; (Shi *et al.* 2015; Ma 2013; B. Li *et al.* 2016). Based on these results, we selected two CSSLs with significantly higher fiber strength than the recurrent parent CCRI45, and one with lower fiber strength. Using the previously constructed genetic linkage map (Shi *et al.* 2015), we selected 526 pairs of simple sequence repeats (SSRs) at intervals of 10–20 cM to screen for introgressed *G. barbadense* chromosomal segments and polymorphisms between CCRI45 and each of the three selected CSSLs.

Planting and harvesting

The selected CSSLs as well as the recurrent parent CCRI45 were planted at the Experimental Farm at CRICAAS in 2014. We tagged all flowers on the day of anthesis, and harvested 3–10 bolls from the tagged flowers by 10:00 AM at 15, 20, 25, and 28 DPA. Harvested bolls were immediately immersed in ice. To minimize experimental error, all samples from the same developmental stage were tagged and collected on the same day. Fiber samples from the ovules of 15–28 developing bolls were dissected, frozen into liquid nitrogen, and stored at -70° . We sampled each developmental stage of each genotype three times.

RNA isolation and calibration

We isolated total RNA from each fiber sample replicate heavier than 1 g using a modified CTAB extraction protocol (Jiang and Zhang 2003): a total of 45 fiber samples. After RNA extraction, RNA degradation and

Copyright © 2017 Lu *et al.*

doi: <https://doi.org/10.1534/g3.117.300108>

Manuscript received August 2, 2017; accepted for publication August 24, 2017; published Early Online September 5, 2017.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material is available online at www.g3journal.org/lookup/suppl/doi:10.1534/g3.117.300108/-/DC1.

¹These authors contributed equally to this work.

²Corresponding authors: State Key Laboratory of Cotton Biology, Key Laboratory of Biological and Genetic Breeding of Cotton, The Ministry of Agriculture, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan, China. E-mail: yylicri@126.com; and College of Agriculture, Shanxi Agricultural University, Taiyuan, Shanxi, China. E-mail: huangjl@sxau.edu.cn

contamination were monitored on 1% agarose gels. We measured RNA purity with a NanoPhotometer spectrophotometer (IMPLEN, CA); RNA concentration with a Qubit RNA Assay Kit (Life Technologies, CA) using a Qubit 2.0 Fluorometer (Life Technologies); and RNA integrity with a RNA Nano 6000 Assay Kit (Agilent Technologies, CA) using a Bioanalyzer 2100 (Agilent Technologies). Equal aliquots of the RNA isolated from each of the three replicates were mixed to yield one total RNA sample from each time point.

Transcriptome sequencing

To systematically analyze transcriptome profiles relating to cotton fiber development, we first constructed sequence libraries with a NEBNext Ultra RNA Library Prep Kit for Illumina (NEB) following the manufacturer's instructions; index codes were added to attribute sequences to each sample. Briefly, 150 ng of each total RNA sample was sheared, converted to cDNA, and ligated to sequencing adaptors following the manufacturer's protocol. The library fragments were purified using AMPure XP (Beckman Coulter, Beverly). We preferentially selected 150–200 bp fragments. We constructed a total of 15 cDNA libraries from the total RNA samples: four collected at 15 DPA, four collected at 20 DPA, four collected at 25 DPA, and three collected at 28 DPA. Transcriptomes were sequenced on an Illumina HiSeq2000 V4 by Berry Genomics Co., Ltd. (Beijing, China), and 125 bp paired-end reads were generated.

Data processing and statistical evaluation

Sequences were filtered and cleaned with the NGS QC Toolkit (<http://www.nipgr.res.in/ngsqualitytoolkit.html>; (Patel and Jain 2012)). Cleaned sequences were mapped to the reference genome of *G. hirsutum* (T. Zhang *et al.* 2015) with TopHat v2 (<http://ccb.jhu.edu/software/tophat/index.shtml>; (Kim *et al.* 2013)), allowing single mismatches but discarding multiple mismatches. Gene expression levels are subject to transcript abundance, but this can be controlled by the calculation of reads mapped to reference genes or exons (Anders and Huber 2010). Thus, expression levels are expressed as reads per kilobase per million mapped reads (RPKM).

Cluster and functional analyses

We calculated the Pearson correlation coefficient (PCC) for all genes. We used the “Self-Organizing Tree Algorithm” (SOTA; MeV 4.9.0; (Wang *et al.* 1998)) to cluster the log₂-transformed RPKM values of all genes at 15, 20, 25, and 28 DPA. We clustered CSSL genes in CSSLs at 15, 20, 25, and 28 DPA with Cluster3.0 (<http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>; (de Hoon *et al.* 2004)). Analysis of gene ontology (GO) was performed using BLAST2GO (<http://www.blast2go.com/b2gohome>; (Conesa *et al.* 2005)) with *E*-value set to 1E-5 and max_target_seqs set to 20. We used Fisher's exact test to test for statistically significant differences, and the Benjamini and Hochberg false discovery rate (FDR) procedure to adjust for the effects of multiple tests (Benjamini *et al.* 2001). GO terms were considered significant when FDR < 0.01.

Selection of DEGs

To identify the DEGs in the secondary cell wall formation stage of fiber development, gene expression levels of all three CSSLs were compared with those of the recurrent parent CCRI45 at 15, 20, 25, and 28 DPA. DEGs were identified using the generalized fold change algorithm (GFOLD, <http://www.tongji.edu.cn/~zhanglab/GFOLD/index.html>; Feng *et al.* 2012). Genes with a GFOLD value ≥ 0.5 were considered significantly upregulated DEGs, and genes with GFOLD value ≤ -0.5

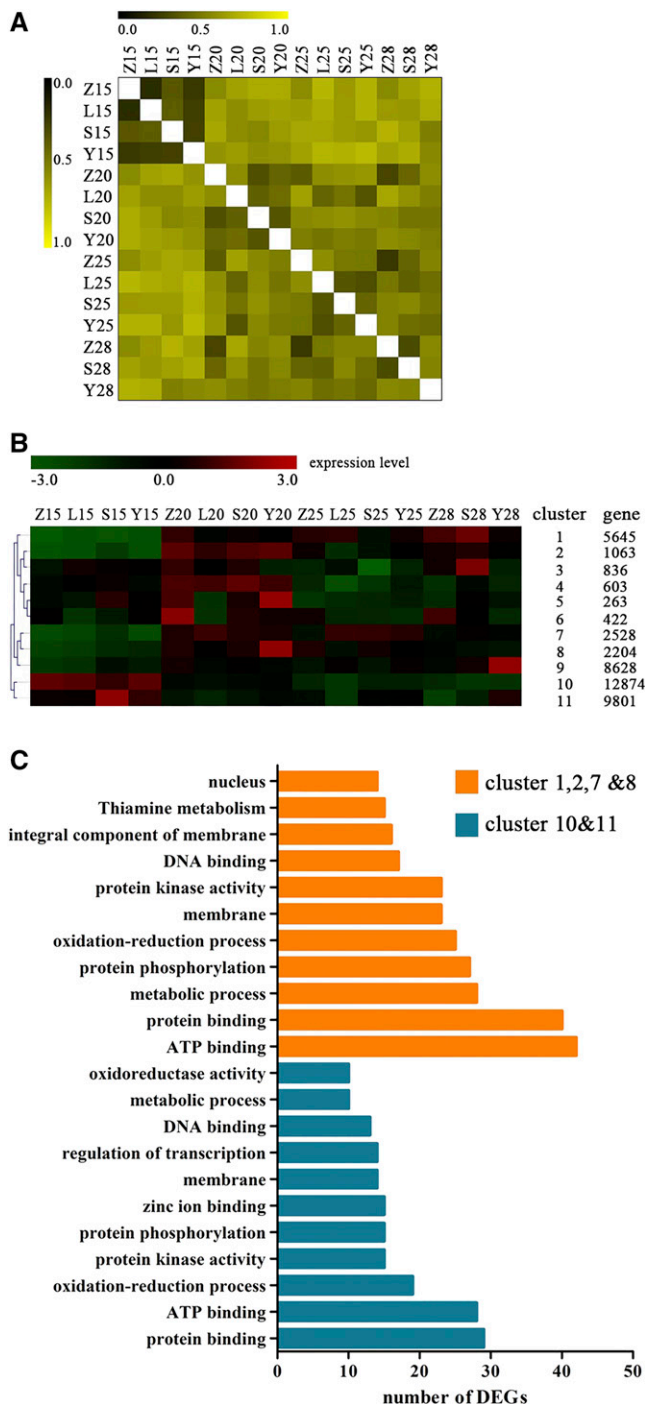


Figure 1 Statistical analysis of transcript profiling data. Z, S, L, and Y indicate CCRI45, MBI7561, MBI7747, and MBI7285, respectively. 15, 15 DPA; 20, 20 DPA; 25, 25 DPA; and 28, 28 DPA. (A) Pearson correlation coefficients of all samples. (B) Self-Organizing Tree Algorithm clustering of all genes across all samples. (C) Functional annotations of genes in more highly expressed at 15 DPA than later (orange bars) and those more highly expressed at 28 DPA than earlier (blue bars). DEG, differentially-expressed gene; DPA, days postanthesis.

were considered significantly downregulated DEGs. DEGs found in both high quality lines (MBI7747 and MBI7561) but not in the low-quality line (MBI7285) were defined as common DEGs. We considered DEGs where $|Gfold\ value| \geq 1$ “significant DEGs.” To better

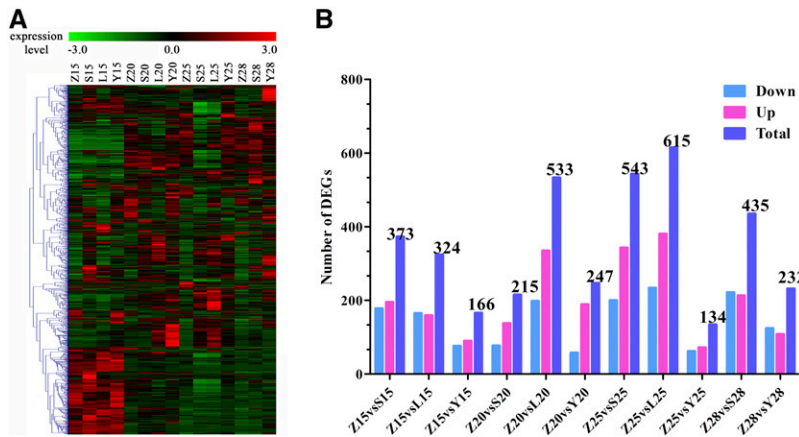


Figure 2 Overview of DEGs. Z, S, L, and Y indicate CCRI45, MBI7561, MBI7747, and MBI7285, respectively. 15, 15 DPA; 20, 20 DPA; 25, 25 DPA; and 28, 28 DPA. (A) Heatmap of all common DEGs at 15, 20, 25, and 28 DPA. (B) Expression pattern of common DEGs over the four time points as compared to the recurrent parent CCRI45. DEG, differentially-expressed gene; DPA, days postanthesis.

understand the different gene expression patterns in the three introgression lines as compared with the recurrent parent CCRI45, we identified the significant DEGs ($|\text{Gfold value}| \geq 1$) found in each CSSL from the 2200 common DEGs, and compared the number of significant DEGs found in each CSSL to the number found in the recurrent parent CCRI45. We next focused on the functions of the genes differently expressed by MBI17561, clustering the DEGs following the procedure outlined above. In addition, we performed a Venn analysis with the SAB lab web tool (<http://sablab.net/venn.php>; Yang *et al.* 2010).

DEGs associated with fiber strength

Phytohormones play an important regulatory role in various plant growth and developmental processes, factors directly related to fiber strength (Sun *et al.* 2015). Transcription factors also regulate fiber strength (Salih *et al.* 2016). We identified common DEGs associated with these processes (*e.g.*, hormone signal transduction and biosynthesis of auxin, brassinosteroid, and ethylene) and integrated the expression profiles. We then organized these DEGs by biological process. Next, to study changes in the biological processes identified in MBI7561 but absent in CCRI45, we identified DEGs involved in the same biological processes across at least two time points. We used BLASTP to search for proteins encoded by these genes.

Colocalization of common DEGs

Based on the genomic position of the introgressed *G. barbadense* chromosome segments identified above, we identified the common DEGs in those introgressed regions from the transcriptome data using the gene name. We annotated the colocalized common DEGs with GO as detailed above.

Quantitative real-time PCR (qRT-PCR)

To test the reliability of transcriptome data and to validate the gene expression data, we performed qRT-PCR analyses on three DEGs that were not only expressed in MBI7561 across all four time points, but were also differentially expressed between at least two of those time points. In addition, we used seven genes selected randomly. We performed qRT-PCR on all DEGs with a 7500 Fast Real-Time PCR System (Applied Biosystems, San Francisco, CA). We designed gene-specific primers with Primer Premier 5.0 (Premier BioSoft International, Palo Alto, CA) and checked their specificity with BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>; Altschul *et al.* 1990). Our novel primers (Table S1) were synthesized by Sangon Biotech Co., Ltd. (Shanghai, China). All DEGs

were analyzed three times. We used the histone3 gene of *G. hirsutum* (forward primer: 5'-GGTGGTGTGAAGAAGCCTCAT-3'; reverse primer: 5'-AATTTACGAACAAGCCTCTGGAA-3') as the internal reference. Relative gene expression levels were calculated as $2^{-\Delta\text{CT}}$.

Data availability

The data sets supporting the results of this article are included within the article and its Supplemental Materials. Figure S1, development of the CSSLs. Figure S2, graphical representation of CSSL genotypes. Figure S3, GO of the common DEGs anchored to quantitative trait loci. Table S1, the specific primer sequences. Table S2, annotation of genes differentially expressed. Table S3, the common DEGs located in common introgressed *G. barbadense* chromosome segments.

RESULTS

CSSLs genotypes and fiber strength

The mean fiber strengths of two of the selected CSSLs, MBI7561 (34.30 cN/tex) and MBI7747 (36.50 cN/tex), were significantly higher than that of the recurrent parent CCRI45 (27.97 cN/tex), while the mean strength of the third selected CSSL, MBI7285 (25.90 cN/tex), was lower (Table 1). Our CSSLs contained some chromosomal segments of *G. barbadense* in a *G. hirsutum* background. The introgressed *G. barbadense* chromosomal segments were different in each of the three CSSLs. Among the 526 SSRs screened, we found 27 polymorphisms in MBI7561, 11 in MBI7747, and 39 in MBI7285 as compared to CCRI45 (Figure S2 and Table 1): <7.41% of the total chromosome.

Functional clusters

All the raw data we sequenced was deposited in the NCBI short read archives (SRA; accession number SRP084203). We identified 44,687 genes across all cDNA libraries. The PCC was low for CCRI45 at all DPAs tested, and was also low for all three CSSLs at 15 DPA. However, higher PCCs were observed at 20 DPA, 25 DPA, and 28 DPA (Figure 1A). The 44,678 identified genes were grouped into 11 clusters based on expression patterns (Figure 1B). Genes in clusters 10 and 11 were more highly expressed at 15 DPA than at later stages, while genes in clusters 1, 2, 7, and 8 had the opposite pattern (Figure 1B). GO analysis suggested that genes in all six of these clusters shared some similar functions, such as the regulation of protein and ATP binding, but differed substantially as well: genes in clusters 10 and 11 were integral components of membrane and thiamine metabolisms, while genes in clusters

■ **Table 2 All significant DEGs in diverse comparisons**

Comparison	Down	Up	Total
CCRI45-15DPA vs. MBI7561-15DPA	178	195	373
CCRI45-15DPA vs. MBI7747-15DPA	165	159	324
CCRI45-15DPA vs. MBI7285-15DPA	76	90	166
CCRI45-20DPA vs. MBI7561-20DPA	77	138	215
CCRI45-20DPA vs. MBI7747-20DPA	198	335	533
CCRI45-20DPA vs. MBI7285-20DPA	58	189	247
CCRI45-25DPA vs. MBI7561-25DPA	200	343	543
CCRI45-25DPA vs. MBI7747-25DPA	234	381	615
CCRI45-25DPA vs. MBI7285-25DPA	62	72	134
CCRI45-28DPA vs. MBI7561-28DPA	222	213	435
CCRI45-28DPA vs. MBI7285-28DPA	124	108	232

DPA, days postanthesis.

1, 2, 7, and 8 were mainly related to the regulation of transcription (e.g., zinc ion binding; Figure 1C).

Analysis of common DEGs

In total, we obtained 2200 common DEGs. Most of these were expressed at low levels (Figure 2A). More significant DEGs were identified in MBI7747 at 15, 20, and 25 DPA than in CCRI45 (Table 2). When compared to CCRI45 at 25 DPA, 615 significant DEGs were identified in MBI7747 and 543 were identified in MBI7561, many more than were identified in MBI7285 (Figure 2B). The common DEGs that were significantly upregulated (GFOLD > 0.5) were associated with polysaccharide metabolic processes, regulation of hormone levels, and regulation of transcription (Figure 3A). The common DEGs that were significantly downregulated (GFOLD < 0.5) were associated with microtubule-based processes, cellular response to stress, and the cell cycle (Figure 3B).

Cluster analysis of significant DEGs in high fiber strength CSSLs

Our Venn analysis indicated 23 common DEGs to be significantly up- or downregulated in MBI7561 but not CCRI45 at 15, 20, 25, and 28 DPA (Figure 4A). These same 23 common DEGs were also significantly up- or downregulated in MBI7747 but not CCRI45 at 15, 20, and 25 DPA

(28 DPA not available; Figure 4B). Over all four time points, we found 960 genes that were significantly up- or downregulated in MBI7561 as compared to CCRI45. These 960 significant DEGs were sorted into 11 clusters by expression pattern (Figure 4C). Genes in clusters 5 and 6 (regulation of protein binding and xylan biosynthesis) were highly expressed at 15 DPA; genes in cluster 8 (regulation of integral components of membranes and oxidation–reduction processes) were highly expressed at 20 DPA; genes in clusters 10 and 11 (regulation of oxidation–reduction processes and protein binding) were highly expressed at 25 DPA; and genes in cluster 9 (regulation of metal ion binding and oxidation–reduction processes) were highly expressed at 28 DPA (Figure 4, D and E).

Common DEGs related to fiber strength

We found 31 common DEGs associated with phytohormones: 18 in hormone signal transduction and 13 in the biosynthesis of auxin, brassinosteroid, and ethylene. In addition, 75 common DEGs involved in cell wall and fatty acid biosynthesis were observed. Finally, 59 DEGs were associated with carbohydrate metabolism. Our results indicated that most of the genes associated with these biological processes were minimally expressed (Figure 5A). After sorting by biological process, it was clear that most DEGs regulated the cell wall, carbohydrate metabolism, and transcription factor activation (Figure 5B).

Colocalization of common DEGs

We colocalized 16 common DEGs in two *G. barbadense* chromosome segments, present in both MBI7561 and MBI7747. Among these 16 DEGs, 10 were downregulated and 6 were upregulated; four were associated with cellular processes, five were associated with single-organism processes, four were associated with catalytic activity, and four were associated with the membrane and the membrane-enclosed lumen (Figure S3).

Validation of DEGs with qRT-PCR

We selected three DEGs expressed in MBI7561 but absent in MBI7285 at all four time points, with differing expression for at least two time points: XLOC_036333, XLOC_029945, and XLOC_075372 (Figure 6). XLOC_036333 [encodes *MNS1*] was upregulated at 20, 25, and 28 DPA, while XLOC_029945 (encodes fasciclin-like arabinogalactan

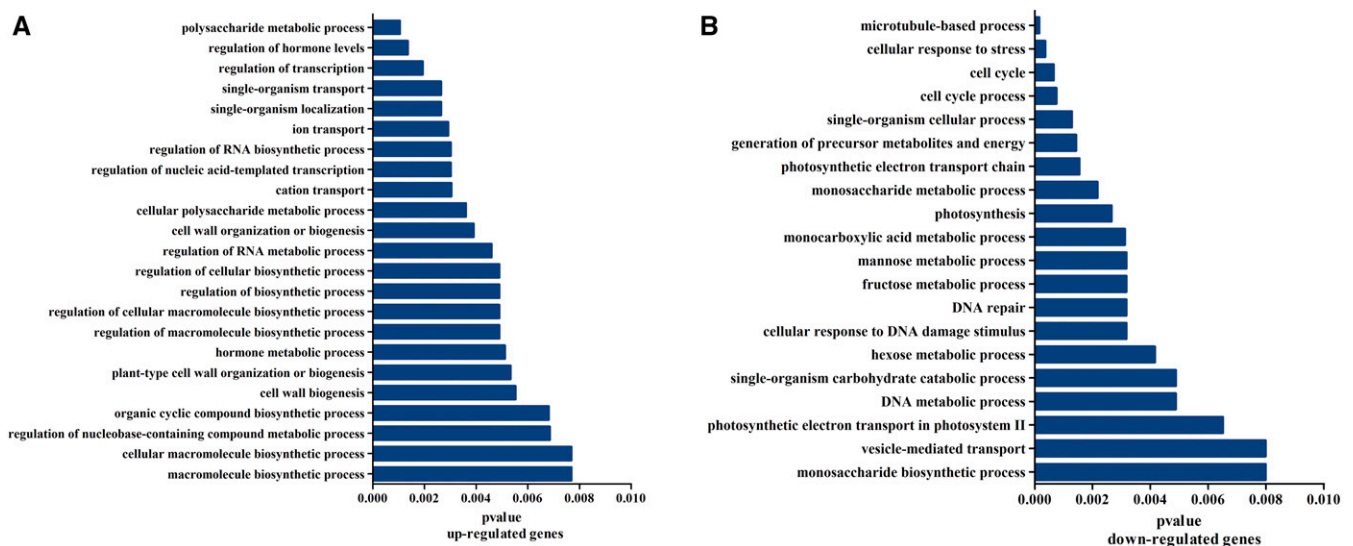


Figure 3 Functional associations of common differentially-expressed genes (DEGs). (A) Biological processes associated with upregulated common DEGs. (B) Biological processes associated with downregulated common DEGs.

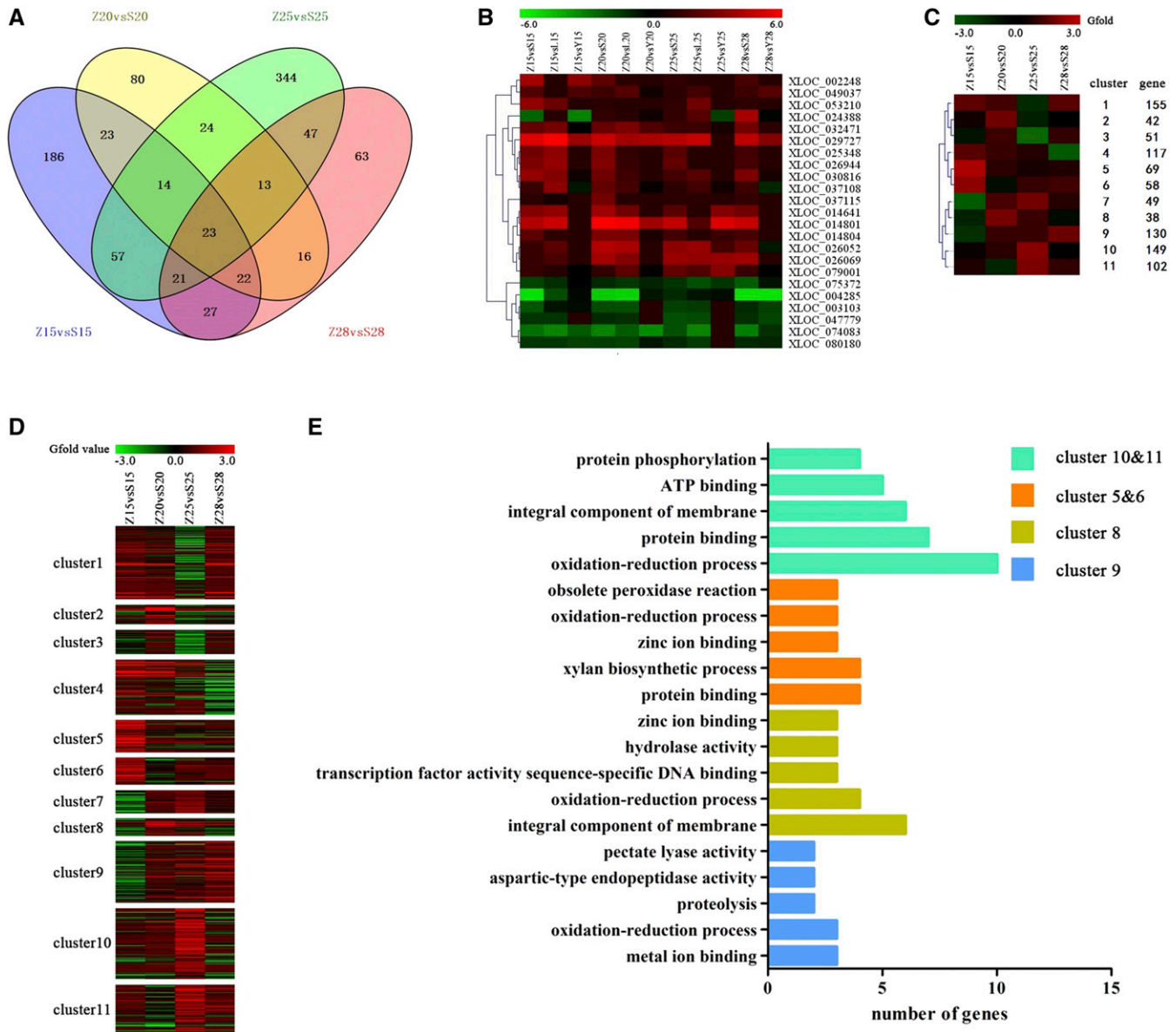


Figure 4 Cluster analysis of the common DEGs. Z, S, L, and Y indicate CCRI45, MBI7561, MBI7747, and MBI7285, respectively. 15, 15 DPA; 20, 20 DPA; 25, 25 DPA; and 28, 28 DPA. (A) Venn diagram of significant DEGs found in MBI7561 and CCRI45 across various time points. (B) Heatmap of common DEGs. (C) Self-Organizing Tree Algorithm clustering of all significant DEGs between CCRI45 and MBI7561 at all four time points. (D) Heatmap analysis of the expression of all significant DEGs between CCRI45 and MBI7561 at all four time points. (E) Functional annotations of significant DEGs: green bars indicate genes highly expressed at 15 DPA; orange bars indicate genes highly expressed at 20 DPA; gold bars indicate genes highly expressed at 25 DPA; and blue bars indicate genes highly expressed at 28 DPA. ATP, adenosine triphosphate; DEG, differentially-expressed gene; DPA, days postanthesis.

protein 8, FLA8) was upregulated only at 25 DPA. XLOC_075372 (encodes *snakin-1*) was downregulated at all time points.

The DEG expression patterns recovered by our qRT-PCR analysis of the three DEGs identified above (XLOC_036333, XLOC_029945, and XLOC_075372) and the seven other randomly selected genes were consistent with the results of transcriptome data (Figure 7).

DISCUSSION

Cotton fiber strength, primarily determined during the secondary cell wall formation stage, is a critical component of fiber quality (Lee *et al.* 2007). Genetic screening has identified a number of genes involved in

fiber development (Z. Zhang *et al.* 2015), but the mechanisms by which these genes affect fiber strength and development remain poorly understood. Recently, several large-scale genomic studies have identified genes that are associated with cotton fiber development and involved in related signaling pathways (Yang *et al.* 1982; Liang *et al.* 2015; Naoumkina *et al.* 2015; Fang *et al.* 2014a,b). Transcriptome analyses suggested that gene expression patterns and functional distributions vary significantly across different stages of fiber development due to the complexities of the mechanisms involved and the regulatory components in the process (Fang *et al.* 2014a,b; Islam *et al.* 2016b; Tuttle *et al.* 2015).

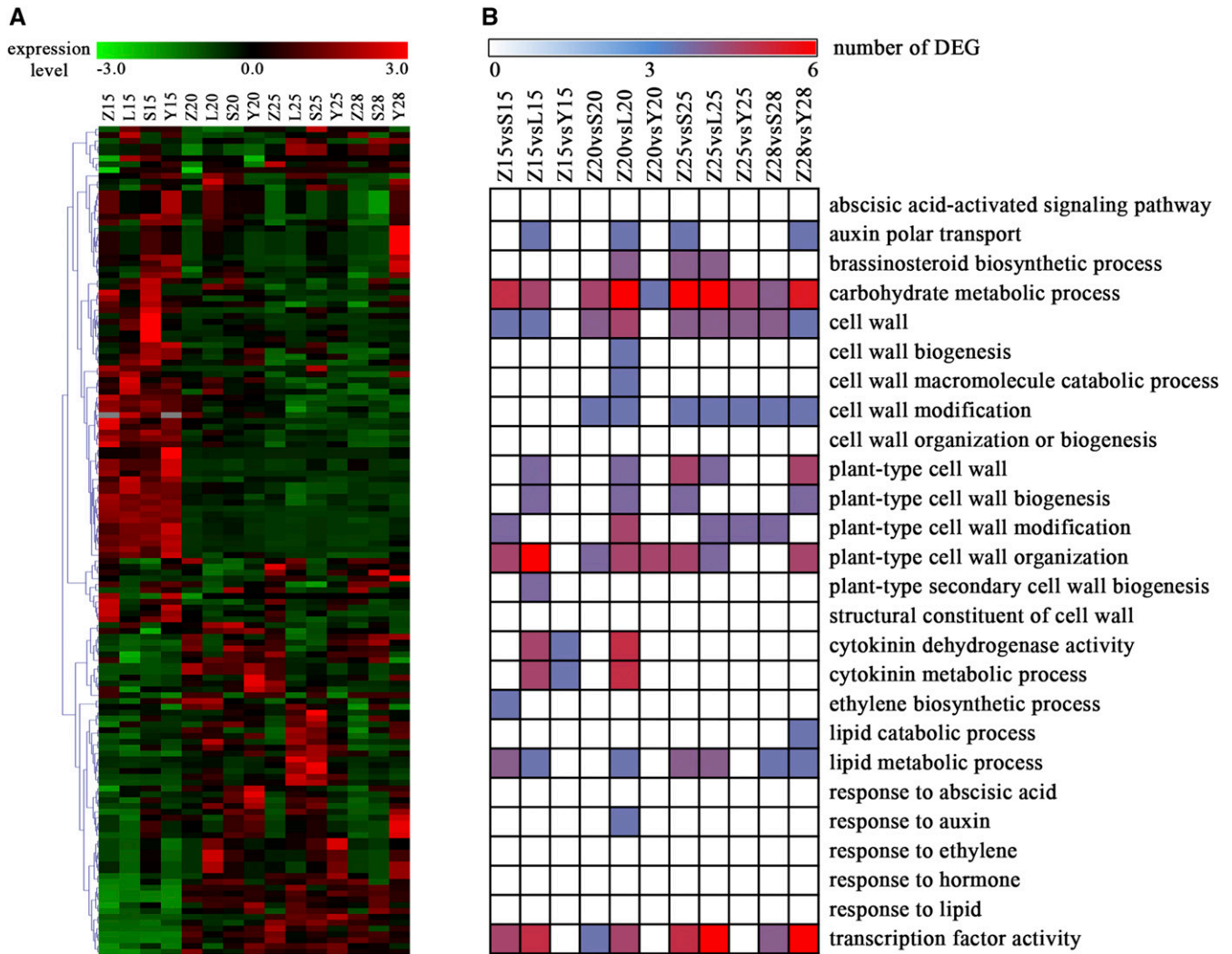


Figure 5 Common DEGs related to fiber strength. Z, S, L, and Y indicate CCRI45, MBI7561, MBI7747, and MBI7285, respectively. 15, 15 DPA; 20, 20 DPA; 25, 25 DPA; and 28, 28 DPA. (A) Heatmap of common DEGs related to fiber strength. (B) Numbers of common DEGs involved in processes related to fiber strength. DEG, differentially-expressed gene; DPA, days postanthesis.

We found 2200 common DEGs in the two high fiber strength CSSLs, as compared to the recurrent parent CCRI45 (Figure 2B), indicating differences in gene expression patterns between the CSSLs and the recurrent parent. Additionally, these common DEGs clustered into functional groups that were modulated differently over time (Figure 1A), suggesting that gene expression profiles in the same pathway, as well as the genes involved in that pathway, may be altered differently at various stages of fiber development. We identified several functional candidate genes involved in secondary cell wall formation, mainly related to biological processes such as polysaccharide metabolism, regulation of hormone levels, regulation of transcription, and cell wall biogenesis (Figure 1C). Thus, consistent with our hypothesis, our results indicated that genes associated with different metabolic pathways affect fiber strength.

Most of the 2200 common DEGs identified in this study are involved in cell wall biogenesis. We propose the DEGs that we identified as being implicated in secondary wall synthesis and metabolism as candidates for the further development of cotton cultivars with superior fiber quality. However, genes involved in the regulation

of polysaccharide metabolic processes were also upregulated in our high fiber strength CSSLs. It has been shown that polysaccharide components such as pectin also play an important role in secondary cell wall synthesis (Gou *et al.* 2007), and that an increase in non-cellulosic polysaccharides may trigger a delay in the transition to secondary wall synthesis, resulting in improved fiber quality (Y. Li *et al.* 2016).

We also recovered several common DEGs involved in hormone signal transduction and the biosynthesis of auxin, brassinosteroid, and ethylene. Previous studies have suggested that ethylene might promote cell elongation by increasing the expression of sucrose synthase, tubulin, and expansin genes (Shi *et al.* 2006); activation of ethylene biosynthesis by saturated very-long-chain fatty acids significantly promoted fiber cell elongation in *G. hirsutum* (Qin *et al.* 2007).

Our colocalization indicated that 16 of the common DEGs were located segments known to harbor quantitative traits for fiber strength (Ma 2013), further suggesting a relationship between these common DEGs and fiber strength.

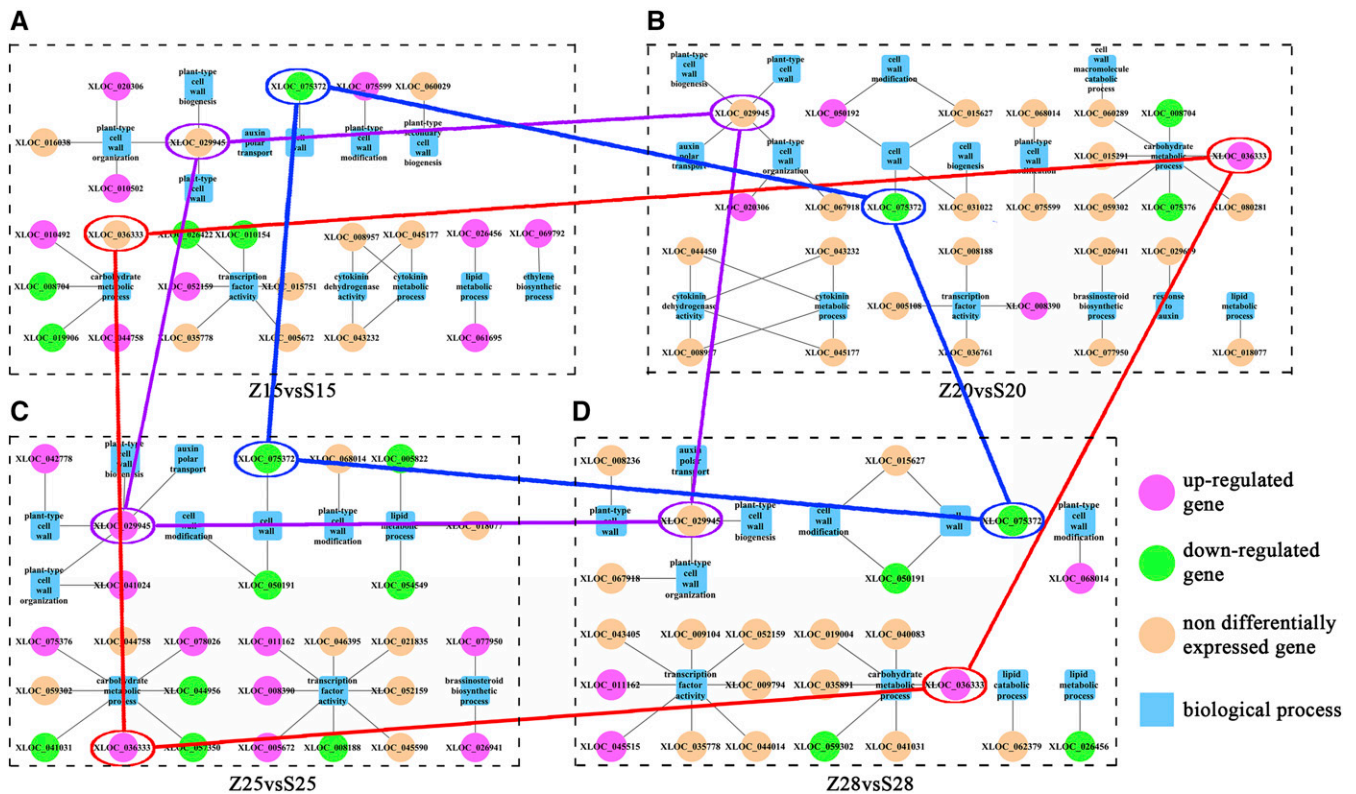


Figure 6 Biological functions associated with selected DEGs in MBI7561. Z, S, L, and Y indicate CCRI45, MBI7561, MBI7747, and MBI7285, respectively. 15, 15 DPA; 20, 20 DPA; 25, 25 DPA; and 28, 28 DPA. (A) Biological processes of genes at 15 DPA. (B) Biological processes of genes at 20 DPA. (C) Biological processes of genes at 25 DPA. (D) Biological processes of genes at 28 DPA. DEG, differentially-expressed gene; DPA, days postanthesis.

Based on these analyses, we selected three genes putatively associated with fiber strength, expressed in MBI7561 at all four time points. *FLA8* (XLOC_029945) was upregulated at 25 DPA in the high fiber strength CSSLs, suggesting that *FLA8* may have a tissue-specific expression preference for the secondary cell wall developmental stage, and may contribute to fiber strength by affecting cellulose synthesis and the orientation of microfibril deposition. *FLA8* is involved in moderating and organizing the cell wall, anchoring components of the plasma membrane, regulation of cell size, and auxin polar transport (Lee *et al.* 2005). *FLA8* has been previously described as a cell wall-associated protein regulating plant growth and development, including cell wall pectin plasticity, cell expansion, and cell wall architecture (Lee *et al.* 2005); *FLA8* might play a critical role in the regulation of fiber strength (Liu *et al.* 2013). Genes encoding *FLA8* were differentially expressed across various plant tissues, in developing fibers, and between *G. hirsutum* and *G. barbadense* (Pan *et al.* 2007; Liu *et al.* 2008, 2013).

The gene *snakin-1* (XLOC_075372), which was downregulated at all time points, regulates the extracellular region, cell wall, and defense response (Nahirnak *et al.* 2012). The silencing of *snakin-1* has been reported to affect cell division, primary leaf metabolism, and cell wall composition in potato plants (Nahirnak *et al.* 2012). Therefore, the downregulation of *snakin-1* we observed indicates that it may play an important role in fiber strength by affecting the composition and structure of the cell wall.

The gene *MNS1* (XLOC_036333) was upregulated at 20 and 28 DPA. This gene is likely to be related to fiber strength regulation

as it encodes an enzyme belonging to the glycosyl hydrolase family, involved in the synthesis of glycoproteins (Zhou *et al.* 2012). It also regulates mannosyl-oligosaccharide 1,2- α -mannosidase activity, membrane and calcium ion binding, and carbohydrate metabolism (Hosomi *et al.* 2010). Therefore, these three genes might regulate important biological processes involved in cell wall biogenesis and might mediate secondary cell wall deposition, contributing to the development of cotton fiber strength. As such, these three genes may be particularly good candidates for future investigation of the molecular mechanisms of fiber strength formation, and for the improvement of cotton fiber quality through molecular breeding.

ACKNOWLEDGMENTS

This work was funded by the Natural Science Foundation of China (3110118) and the National High Technology Research and Development Program of China (2012AA101108). The funders had no role in study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

Author contributions: Q.L., X.X., P.L., W.S., S.L., Z.Z., M.H., and R.P. performed the experiments for library construction and sequencing. Y.S., J.G., W.G., A.L., H.S., J.L., and Q.G. collected and analyzed the data from the field. Y.Y. and J.H. designed the experiments and provided the materials. All authors have read, edited, and approved the current version of the manuscript. The authors declare that they have no competing interests.

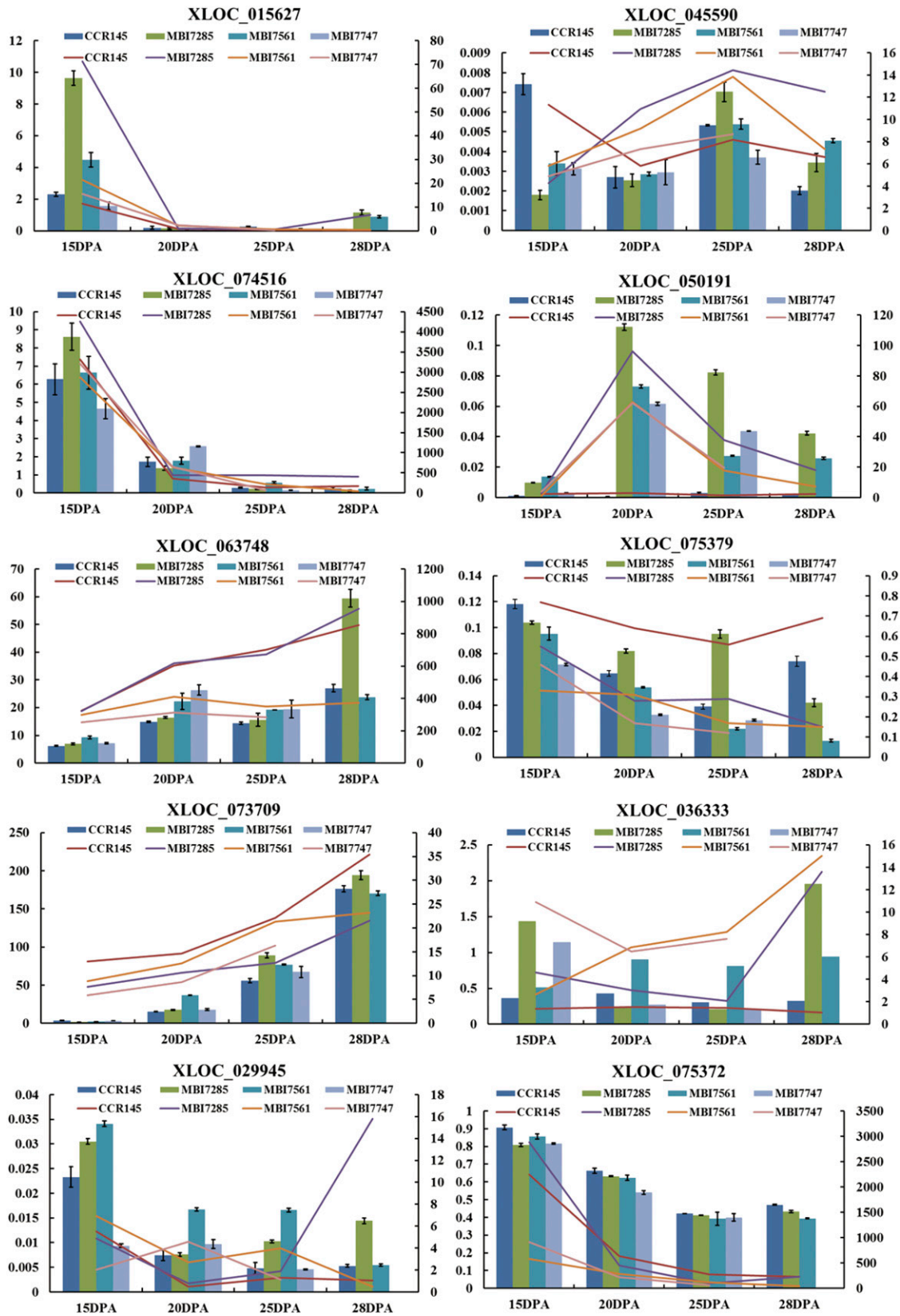


Figure 7 Quantitative real-time polymerase chain reaction validation of 10 DEGs. DPA, days post-anthesis.

LITERATURE CITED

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.
- Anders, S., and W. Huber, 2010 Differential expression analysis for sequence count data. *Genome Biol.* 11: 1–12.
- Benjamini, Y., D. Drai, G. Elmer, N. Kafkafi, and I. Golani, 2001 Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* 125: 279–284.
- Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon *et al.*, 2005 Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
- Constable, G., D. Llewellyn, S. A. Walford, and J. D. Clement, 2015 Cotton breeding for fiber quality improvement, pp. 191–232 in *Industrial Crops: Breeding for BioEnergy and Bioproducts*, edited by Cruz, V. M. V., and D. A. Dierig. Springer, New York.
- de Hoon, M. J., S. Imoto, J. Nolan, and S. Miyano, 2004 Open source clustering software. *Bioinformatics* 20: 1453–1454.
- Edelmann, H. G., and S. C. Fry, 1992 Effect of cellulose synthesis inhibition on growth and the integration of xyloglucan into pea internode cell walls. *Plant Physiol.* 100: 993–997.
- Fang, L., R. Tian, J. Chen, S. Wang, X. Li *et al.*, 2014a Transcriptomic analysis of fiber strength in upland cotton chromosome introgression lines carrying different *Gossypium barbadense* chromosomal segments. *PLoS One* 9: e94642.
- Fang, L., R. Tian, X. Li, J. Chen, S. Wang *et al.*, 2014b Cotton fiber elongation network revealed by expression profiling of longer fiber lines introgressed with different *Gossypium barbadense* chromosome segments. *BMC Genomics* 15: 1–15.
- Feng, J., C. A. Meyer, Q. Wang, J. S. Liu, X. Shirley Liu *et al.*, 2012 GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* 28: 2782–2788.
- Gao, W., S. Saha, D. P. Ma, Y. Guo, J. N. Jenkins *et al.*, 2012 A cotton-fiber-associated cyclin-dependent kinase gene: characterization and chromosomal location. *Int. J. Plant Genomics* 2012: 613812.
- Gilbert, M. K., H. J. Kim, Y. Tang, M. Naoumkina, and D. D. Fang, 2014 Comparative transcriptome analysis of short fiber mutants Ligon-lintless 1 and 2 reveals common mechanisms pertinent to fiber elongation in cotton (*Gossypium hirsutum* L.). *PLoS One* 9: e95554.
- 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.
- Hosomi, A., K. Tanabe, H. Hirayama, I. Kim, H. Rao *et al.*, 2010 Identification of an Htm1 (EDEM)-dependent, Mns1-independent endoplasmic reticulum-associated degradation (ERAD) pathway in *Saccharomyces cerevisiae*: application of a novel assay for glycoprotein ERAD. *J. Biol. Chem.* 285: 24324–24334.
- Hu, G., J. Koh, M. J. Yoo, K. Grupp, S. Chen *et al.*, 2013 Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. *New Phytol.* 200: 570–582.
- Islam, M. S., D. D. Fang, G. N. Thyssen, C. D. Delhom, Y. Liu *et al.*, 2016a Comparative fiber property and transcriptome analyses reveal key genes potentially related to high fiber strength in cotton (*Gossypium hirsutum* L.) line MD52ne. *BMC Plant Biol.* 16: 1–19.
- Islam, M. S., L. Zeng, G. N. Thyssen, C. D. Delhom, H. J. Kim *et al.*, 2016b Mapping by sequencing in cotton (*Gossypium hirsutum*) line MD52ne identified candidate genes for fiber strength and its related quality attributes. *Theor. Appl. Genet.* 129: 1071–1086.
- Jiang, J. X., and T. Z. Zhang, 2003 Extraction of total RNA in cotton tissues with CTAB-acidic phenolic method. *Mianhua Xuebao* 15: 166–167.
- Kim, D., G. Perte, C. Trapnell, H. Pimentel, R. Kelley *et al.*, 2013 TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14: R36.
- Kim, H. J., and B. A. Triplett, 2001 Cotton fiber growth in planta and in vitro. models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* 127: 1361–1366.
- Lee, J. J., A. W. Woodward, and Z. J. Chen, 2007 Gene expression changes and early events in cotton fibre development. *Ann. Bot.* 100: 1391.
- Lee, K. J., Y. Sakata, S. L. Mau, F. Pettolino, A. Bacic *et al.*, 2005 Arabinogalactan proteins are required for apical cell extension in the moss *Physcomitrella patens*. *Plant Cell* 17: 3051–3065.
- Li, B., Y. Shi, J. Gong, J. Li, A. Liu *et al.*, 2016 Genetic effects and heterosis of yield and yield component traits based on *Gossypium barbadense* chromosome segment substitution lines in two *Gossypium hirsutum* backgrounds. *PLoS One* 11: e0157978.
- Li, Y., L. Tu, F. A. Pettolino, S. Ji, J. Hao *et al.*, 2016 GbEXPATR, a species-specific expansin, enhances cotton fibre elongation through cell wall restructuring. *Plant Biotechnol. J.* 14: 951–963.
- Liang, W., L. Fang, D. Xiang, Y. Hu, H. Feng *et al.*, 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.
- Liu, B., Y. Zhu, and T. Zhang, 2015 The R3-MYB gene GhCPC negatively regulates cotton fiber elongation. *PLoS One* 10: e0116272.
- Liu, D., L. Tu, Y. Li, L. Wang, L. Zhu *et al.*, 2008 Genes encoding fasciclin-like arabinogalactan proteins are specifically expressed during cotton fiber development. *Plant Mol. Biol. Rep.* 26: 98–113.
- Liu, H., R. Shi, X. Wang, Y. Pan, Z. Li *et al.*, 2013 Characterization and expression analysis of a fiber differentially expressed Fasciclin-like arabinogalactan protein gene in Sea Island cotton fibers. *PLoS One* 8: e70185.
- Ma, L., Y. Wang, G. Yan, S. Wei, D. Zhou *et al.*, 2016 Global analysis of the developmental dynamics of *Gossypium hirsutum* based on strand-specific transcriptome. *Physiol. Plant.* 158: 106–121.
- Ma, L. J., Y. Shi, M. Lan, Z. Yang, J. Zhang *et al.*, 2013 Evaluation of chromosome segment substitution lines related to fiber yield and quality traits from *Gossypium hirsutum* × *Gossypium barbadense*. *Cotton Science.* 25: 486–495.
- Nahirnak, V., N. I. Almasia, P. V. Fernandez, H. E. Hopp, J. M. Estevez *et al.*, 2012 Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. *Plant Physiol.* 158: 252–263.
- Naoumkina, M., G. N. Thyssen, and D. D. Fang, 2015 RNA-seq analysis of short fiber mutants Ligon-lintless-1 (*Li*₁) and -2 (*Li*₂) revealed important role of aquaporins in cotton (*Gossypium hirsutum* L.) fiber elongation. *BMC Plant Biol.* 15: 65.
- Pan, Y., J. Ma, G. Zhang, G. Han, X. Wang *et al.*, 2007 cDNA-AFLP profiling for the fiber development stage of secondary cell wall synthesis and transcriptome mapping in cotton. *Chin. Sci. Bull.* 52: 2358–2364.
- Patel, R. K., and M. Jain, 2012 NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7: e30619.
- Qin, Y. M., C. Y. Hu, Y. Pang, A. J. Kastaniotis, J. K. Hiltunen *et al.*, 2007 Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis. *Plant Cell* 19: 3692–3704.
- Salih, H., W. Gong, S. He, G. Sun, J. Sun *et al.*, 2016 Genome-wide characterization and expression analysis of MYB transcription factors in *Gossypium hirsutum*. *BMC Genet.* 17: 129.
- Shang, L., Y. Wang, X. Wang, F. Liu, A. Abdugheni *et al.*, 2016 Genetic analysis and QTL detection on fiber traits using two recombinant inbred lines and their backcross populations in upland cotton. *G3 (Bethesda)* 6: 2717–2724.
- Shi, Y., S. Zhu, X. Mao, J. Feng, Y. Qin *et al.*, 2006 Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18: 651–664.
- Shi, Y., W. Li, A. Li, R. Ge, B. Zhang *et al.*, 2015 Constructing a high-density linkage map for *Gossypium hirsutum* × *Gossypium barbadense* and identifying QTLs for lint percentage. *J. Integr. Plant Biol.* 57: 450–467.
- Sun, R., K. Wang, T. Guo, D. C. Jones, J. Cobb *et al.*, 2015 Genome-wide identification of auxin response factor (ARF) genes and its tissue-specific prominent expression in *Gossypium raimondii*. *Funct. Integr. Genomics* 15: 481–493.
- Tu, L. L., X. L. Zhang, S. G. Liang, D. Q. Liu, L. F. Zhu *et al.*, 2007 Genes expression analyses of sea-island cotton (*Gossypium barbadense* L.) during fiber development. *Plant Cell Rep.* 26: 1309–1320.

- Tuttle, J. R., G. Nah, M. V. Duke, D. C. Alexander, X. Guan *et al.*, 2015 Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis, represses lignification, and prolongs elongation. *BMC Genomics* 16: 477.
- Wang, H. C., J. Dopazo, D. L. F. Lg, Y. P. Zhu, and J. M. Carazo, 1998 Self-organizing tree-growing network for the classification of protein sequences. *Protein Sci.* 7: 2613–2622.
- Wendel, J. F., and R. C. Cronn, 2003 Polyploidy and the evolutionary history of cotton, pp. 139–186 in *Advances in Agronomy*. Academic Press, San Diego, CA.
- Wu, M., L. Zhang, X. Li, X. Xie, W. Pei *et al.*, 2016 A comparative transcriptome analysis of two sets of backcross inbred lines differing in lint yield derived from a *Gossypium hirsutum* × *Gossypium barbadense* population. *Mol. Genet. Genomics* 291: 1–19.
- Yang, J., W. Hu, W. Zhao, B. Chen, Y. Wang *et al.*, 2016 Fruiting Branch K (+) Level Affects Cotton Fiber Elongation Through Osmoregulation. Yang J, Wei H, Zhao W, *et al.* Fruiting Branch K+ Level Affects Cotton Fiber Elongation Through Osmoregulation. *Front Plant Sci.* 7(e56315):13.
- Yang, K., G. Zhang, J. Mei, D. Chen, and M. Wu, 2010 Screening and analysis of pathogenic genes during DMBA-induced buccal mucosa carcinogenesis in golden hamsters. *Oncol. Rep.* 23: 1619–1624.
- Zhang, T., Y. Hu, W. Jiang, L. Fang, X. Guan *et al.*, 2015 Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33: 531–537.
- Zhang, Z., J. Li, J. Muhammad, J. Cai, F. Jia *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.
- Zhong, R., and Z. H. Ye, 2015 Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant Cell Physiol.* 56: 195.
- Zhou, J., F. Yang, N. A. Leu, and P. J. Wang, 2012 MNS1 is essential for spermiogenesis and motile ciliary functions in mice. *PLoS Genet.* 8: e1002516.

Communicating editor: J. Ma