

Analysis of transcriptome data and quantitative trait loci enables the identification of candidate genes responsible for fiber strength in *Gossypium barbadense*

Yajie Duan , Qin Chen , Quanjia Chen , Kai Zheng , Yongsheng Cai , Yilei Long , Jieyin Zhao , Yaping Guo , Fenglei Sun , Yanying Qu *

College of Agronomy, Xinjiang Agricultural University, Urumqi, Xinjiang 830052, China

*Corresponding author: College of Agronomy, Xinjiang Agricultural University, Urumqi, Xinjiang 830052, China. Email: xjyyq5322@126.com

Abstract

Gossypium barbadense possesses a superior fiber quality because of its fiber length and strength. An in-depth analysis of the underlying genetic mechanism could aid in filling the gap in research regarding fiber strength and could provide helpful information for *Gossypium barbadense* breeding. Three quantitative trait loci related to fiber strength were identified from a *Gossypium barbadense* recombinant inbred line (PimaS-7 × 5917) for further analysis. RNA sequencing was performed in the fiber tissues of PimaS-7 × 5917 0–35 days postanthesis. Four specific modules closely related to the secondary wall-thickening stage were obtained using the weighted gene coexpression network analysis. In total, 55 genes were identified as differentially expressed from 4 specific modules. Gene Ontology and the Kyoto Encyclopedia of Genes and Genomes were used for enrichment analysis, and *Gbar_D11G032910*, *Gbar_D08G020540*, *Gbar_D08G013370*, *Gbar_D11G033670*, and *Gbar_D11G029020* were found to regulate fiber strength by playing a role in the composition of structural constituents of cytoskeleton and microtubules during fiber development. Quantitative real-time PCR results confirmed the accuracy of the transcriptome data. This study provides a quick strategy for exploring candidate genes and provides new insights for improving fiber strength in cotton.

Keywords: *Gossypium barbadense*; fiber strength; RNA-seq; weighted gene coexpression network analysis; QTL; MPP; Multiparental Populations; Multiparent Advanced Generation Inter-Cross (MAGIC)

Introduction

Gossypium barbadense, native to South America, is one of the 4 cultivated *Gossypium* species. *G. barbadense* has the best fiber quality, and its fiber length, fineness, strength, tensile strength, and other characteristics are better than that of other cotton varieties. Due to low self-production and limited areas for the cultivation of *G. barbadense*, the main production areas in China are highly limited, and the yield accounts for only 2% of the total cotton production (Chen *et al.* 2007). Furthermore, these factors limit further study of *G. barbadense* fiber quality traits. Fiber strength (FS) is a quantitative trait controlled by multiple genes (Tu *et al.* 2007; Shi *et al.* 2015). Hence, an in-depth analysis of the genetic mechanisms affecting FS is necessary; this could aid in filling the gap in research regarding FS and could provide helpful information for improvements in *G. barbadense* breeding.

The process of cotton fiber development includes 4 stages: fiber initiation 0–3 days postanthesis (DPA), fiber elongation (3–23 DPA), secondary wall-thickening (20–40 DPA), and fiber maturity (40–50 DPA). Although there is clear demarcation between the stages, each stage has its characteristic features (Basra and Malik 1984; Haigler *et al.* 2012). Different fiber development stages are used to study different fiber quality traits. For example, fiber initiation and elongation usually affect the fiber length, while FS is more influenced by

the secondary wall-thickening stage (Lee *et al.* 2007; Chen and Burke 2015; Chen *et al.* 2019; Patel *et al.* 2020). The main components of cotton fiber are cellulose, hemicellulose, and lignin. Cellulose accumulation and structure are important factors affecting FS (Edelmann and Fry 1992; Shang *et al.* 2016). Cellulose content begins to increase 15 DPA (Potikha *et al.* 1999). During the secondary wall-thickening stage in cotton, as cellulose biosynthesis increases rapidly, complex signaling pathways related to FS, such as signal transduction (Gao *et al.* 2019), metabolic pathways (Li *et al.* 2016; Ahmed *et al.* 2020; Zhao, Dong, *et al.* 2020), and microtubule (Pydiura *et al.* 2019; Chen *et al.* 2021), cellulose, and cytoskeleton pathways are initiated (Tuttle *et al.* 2015; Zhang *et al.* 2015; Sugiyama *et al.* 2017). Therefore, analysis of molecular mechanisms underlying the development of FS is mainly focused on the secondary wall-thickening stage.

With the completion of sequencing for genomes of *Gossypium raimondii* (Wang *et al.* 2012), *Gossypium arboreum* (Li *et al.* 2014), *Gossypium hirsutum* (Hu *et al.* 2019), and *G. barbadense* (Wang *et al.* 2019), more complete reference genomes are available to study the quantitative traits of cotton. Through analysis of quantitative trait loci (QTLs), numerous QTLs and candidate genes associated with fiber quality were discovered, and the number of QTLs related to FS was found to be second only to those related to fiber length (Ali *et al.* 2018; Fan *et al.* 2018; Keerio *et al.* 2018; Tan *et al.* 2018; Zhang

Received: February 23, 2022. Accepted: June 23, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Genetics Society of America.

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

et al. 2019; Shi et al. 2020; Guo et al. 2021). Most stable QTLs related to FS were mapped using the recombinant inbred lines of *G. hirsutum* (Gu et al. 2020). *qFS-Chr.D02* was fine-mapped using a single segment introgression line (IL-D2-2 × TM-1) of *G. hirsutum* (Feng et al. 2020). Thirty-four stable QTLs associated with 5 fiber quality traits were mapped using 279 accessions of *G. barbadense*, and 6 novel QTLs were mapped using 276 accessions of *G. hirsutum* (Liu et al. 2020; Su et al. 2020). However, the number of effective molecular markers and major genes did not increase significantly. The different populations and environmental factors make it difficult to find stable QTLs. Based on the reference genome, the obtained QTL confidence interval usually covers a large physical location and includes numerous genes. Thus, it is difficult to identify genes associated with fiber quality.

With the continuous advancements in high-throughput sequencing technology, transcriptome sequencing (RNA-seq) and molecular markers developed using this technology have profoundly changed the strategies and approaches of cotton breeding (Wang, Wang, et al. 2020). Transcriptome technology effectively uses changes in gene expression to identify the genes closely related to target traits (Qin et al. 2019; Shahzad et al. 2020). Weighted gene coexpression network analysis (WGCNA), based on transcriptome data that could identify specific modules and genes related to target traits, was performed by dividing genes with similar expression patterns into the same module (Langfelder and Horvath 2008; Zou et al. 2019; Ali et al. 2020; Cheng et al. 2020). This method has been used to study high-quality fiber in upland cotton, soybean seed set and size, seed development in chickpea, and bacterial spot resistance in pepper in the past 5 years (Du et al. 2017; Garg et al. 2017; Zou et al. 2019; Feng et al. 2021; Jiang et al. 2021; Zhu, Gao, et al. 2021). The combined analysis of the transcriptome and QTLs can quickly narrow down the range of genes and accurately select genes during key developmental stages (Khattak et al. 2019; Wang, Zheng, et al. 2020; Laoue et al. 2021; Li et al. 2021). Nine genes in rapeseed were identified as candidate genes related to pod number variation (Ye et al. 2017). *KR*, *C4H*, and *FatA* were identified as candidates related to capsaicinoid biosynthesis (Park et al. 2019). *OsSAP16* was identified as a candidate related to salt tolerance (Lei et al. 2020). However, most studies focused on fiber length traits in the fiber initiation and fiber elongation stages, and thus, there is a lack of further research on FS during the secondary wall-thickening stage (Zhu, Stiller, et al. 2021).

PimaS-7 × 5917 was a recombinant inbred line population with 143 individuals constructed in 2013 to study FS traits (Fan et al. 2018). In the present study, the confidence intervals of 3 QTLs related to FS traits in PimaS-7 × 5917 population were further analyzed and their corresponding physical locations were obtained. Then, differential expression of 2 (from PimaS-7 × 5917) during whole fiber development was analyzed by transcriptome technique. A weighted gene coexpression network was constructed to screen out specific modules related to FS and combine the genes related to FS based on QTLs to screen and confirm the candidate genes. This study provides a more rapid strategy for shortening the acquisition time of target candidate genes and novel insights for studying genetic basis of cotton FS traits.

Materials and methods

Plant material

The RIL population PimaS-7 × 5917 developed using American cotton PimaS-7 and Chinese *G. barbadense* cultivar 5917 was first used to study FS and more construction processes in 2018 (Fan

et al. 2018). From 2016 to 2018, the RIL population was planted in Aksu (Xinjiang Province, 41°17'07.12 N, 80°26'50.68 E). Eight *G. barbadense* accessions (4 with high-FS and 4 with low-FS) were also planted. The experiments were conducted using a randomized group test design (double-row plots, length 5 m, width 0.8 m). All plant materials (provided by the Key Laboratory of Agricultural Biotechnology, Xinjiang Agricultural University, Urumqi, China) were considerably different in FS. The quality of fully mature fiber was determined by the Cotton Fiber Quality Supervision and Testing Center of the Chinese Ministry of Agriculture (HV11000).

Two *G. barbadense* plants from the parent varieties PimaS-7 and 5917 were used for RNA-seq. Eight *G. barbadense* accessions were planted as verification materials. The day of flowering was recorded as 0 DPA at intervals of 5 days. Cotton fiber samples (3 biological replicates) were collected at 0 DPA (ovules), 5, 10, 15, 20, 25, 30, and 35 DPA. All samples were placed on ice packs and quickly separated with sterilized tweezers and transported in liquid nitrogen. Some samples were used for RNA-seq, while others were used for quantitative real-time PCR (qRT-PCR) and stored at −80°C.

QTL region realignment

According to Fan's studies, using single nucleotide polymorphism (SNP) markers, 3 QTLs related to FS were identified from the *G. barbadense* RIL population. To further analyze the 3 QTLs, we extracted SNP marker sequences corresponding to these QTLs (Fan et al. 2018). The marker sequences were aligned to the *G. barbadense* genome (Wang et al. 2019) using BWA software (Li and Durbin 2009); the physical location of the markers was obtained, and the genes in the corresponding chromosome position were extracted for further analysis.

RNA extraction, library construction, and RNA-seq

An RNA prep pure plant kit (Tiangen, Beijing, China) was used for the total RNA extraction of all samples. RNA integrity was confirmed using 1% agarose gel electrophoresis. The concentration and the quality of RNA were determined using a Colibri Microphotometer (Titertek Berthold, Germany). A total of 48 cDNA libraries were constructed using an Illumina HiSeq2500 platform (Biomarker Technologies, Beijing, China) for the 2 *G. barbadense* parents at 8 developing times with 3 biological replicates. All RNA-seq raw data were deposited in the NCBI database with the accession number: GSE178945, at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE178945>.

Data processing and differentially expressed gene analysis

The raw data obtained by RNA-seq were provided in FASTQ format. Quality control of raw reads was performed by removing adapter sequences, low-quality reads, and reads with $N > 10\%$ (Bolger et al. 2014). The *G. barbadense* Pima 3-79 genome was designated as the reference genome, which can be found at <http://cotton.hzau.edu.cn/EN/data/download/G.barbadensegenomeHAU\v2.0.tar> (Wang et al. 2019). Reads were aligned using HISAT2 (Kim et al. 2015) and were assembled and quantified by performing a String Tie (Pertea et al. 2015). The fragments per kilobase per million (FPKM) parameter refers to the number of reads per thousand bases from the map to exon per million reads. When the gene expression satisfied $\text{FPKM} > 0.5$, the genes were identified to be expressed genes (Mortazavi et al. 2008). The DESeq2 R package (v.3.6.3) was used to identify differentially expressed genes (DEGs). When gene expression satisfied the false discovery rate (FDR) value ≤ 0.05 and $|\text{Log}_2$ -

fold change (FC) ≥ 1 , it was considered to be a DEG (Patel et al. 2020). All DEGs were annotated using Gene Ontology (GO) functional enrichment analysis. KOBAS software was used for the Kyoto Encyclopaedia of Genes and Genomes pathway enrichment analysis for all DEGs (Altschul et al. 1990; Wu et al. 2006; Xie et al. 2011).

Construction of coexpression modules using WGCNA

Coexpression modules were constructed using the WGCNA (v.1.51) package in R (v.3.6.3) (Langfelder and Horvath 2008). In the present study, for filtering out the DEGs with low dynamic variation (Lin et al. 2020), coexpression modules with a power of 15 and merCutHeight of 0.25 were identified. The membership (K_{ME}) > 0.85 was used as a standard of gene connectivity to select the DEGs related to FS. Next, the obtained DEGs that were strongly correlated in specific modules were compared with other DEGs obtained in the physical location corresponding to the marker of QTLs. GO enrichment analysis was performed on these common genes.

Confirmation of RNA-Seq results by qRT-PCR

The total RNA was extracted from all fiber samples (PimaS-7, 5917, 4 high-FS, and 4 low-FS materials at 0–35 DPA). Primer Premier v.5 was used to design related gene primers. All primers were synthesized by BGI (Beijing, China). Reverse transcription was performed using a Revert Aid First Strand cDNA Synthesis kit (Thermo Scientific, Waltham, MA, USA). qRT-PCR was performed using a Trans Start Top Green qPCR Super Mix kit (Transgen Biotech, Beijing, China). The cotton UBQ7 served as the reference gene to normalize the relative expression levels through ABI 7500 Fast Real-Time PCR System (Applied Biosystems, Waltham, MA, USA), with 3 biological repeats and 3 technical repeats. The $2^{-\Delta\Delta CT}$ method was used to analyze the relative expression of DEGs (Livak and Schmittgen 2001).

Results

Physical location of 3 QTLs related to FS

In Fan's study, a high-density genetic map was constructed using an $F_{2:6}$ RIL population (143 lines generated from PimaS-7 \times 5917) (Supplementary Tables 1–3). Based on the scaffold map of *G. barbadense* (Yuan et al. 2015), 3 QTLs (qFS -LG21-1, qFS -LG1-1, and qFS -LG4-1) related to FS were identified (Fan et al. 2018). To further discover candidate genes related to FS, the confidence intervals of the 3 QTLs were compared in the reorganized *G. barbadense* physical map in this study. qFS -LG21-1, with a limit of detection (LOD) of 3.38, explains 8.86% of the phenotypic variations, and the confidence interval located at D03:3270717–3902216 covers a physical interval of 0.6 Mb. qFS -LG1-1, with an LOD of 3.92, explains 10.42% phenotypic variations, and the confidence interval located at D08:38457462–63870127 covers a physical interval of 24.24 Mb. qFS -LG4-1, with an LOD of 4.84, explains 10.11% phenotypic variations, and the confidence interval located at D11:30147612–6797464 covers a physical interval of 36.07 Mb (Supplementary Table 4). According to the *G. barbadense* reference genome, 2,601 genes from D03, D08, and D11 were related to FS (Supplementary Table 5).

RNA-seq and quality assessment

To accurately discover genes related to FS, PimaS-7 and 5917 (Fig. 1a) were used as experimental materials for RNA-seq in 8 different fiber development times with 3 biological replications. Forty-eight cDNA libraries were constructed, and 317.57 Gb of

clean data was obtained after screening the raw data. The average clean data of each sample reached 5.78 Gb. The percentage of Q30 base was above 92.78%, and the range of GC content was between 44.17% and 46.54% (Supplementary Table 6). The mapping rate between the clean reads of samples and the designated reference genome of *G. barbadense* was 85.56–94.79%. In total, 69,338 genes were considered expressed genes with FPKM > 0.5 . Pearson's correlation coefficient was used to evaluate the correlation among the 3 biological replicates of each sample in the same fiber development times (Supplementary Fig. 1). Then, principal component analysis was used to clarify the relationship between the 2 materials in different fiber development times (Fig. 1b). In the early stages of fiber development, the 2 materials aggregate together. However, with fiber development, the degree of polymerization of the 2 gradually decreased in different fiber development times. The results indicated that the influence of different development times of fiber between the 2 materials was greater than the background difference between them.

Analysis of DEGs

In the present study, DEGs were obtained by comparing the same materials in different fiber development times and the same fiber development times of different materials. A total of 38,798 genes were selected as DEGs from all expressed genes with an FDR value ≤ 0.05 and $|\text{Log}_2(\text{FC})| \geq 1$. In PimaS-7, the DEGs between 00 and 05 DPA were highest in the early fiber development stage; 1,217 DEGs were found (6,904 upregulated, 4,101 downregulated). As fiber development progresses, the DEGs gradually decreases until between 20 and 25 DPA; then, DEGs peaked again, and 9,936 DEGs were found (5,657 upregulated, 4,279 downregulated). In 5917, the number of DEGs between 05 and 10 DPA was the highest in the early fiber development stage, and 8,659 DEGs were found (4,366 upregulated, 4,293 downregulated). After that, 7,346 DEGs were found in 25 and 30 DPA (4,067 upregulated, 3,279 downregulated). PimaS-7 had 1,217 DEGs on 00 DPA, 2,307 DEGs on 05 DPA, 1,855 DEGs on 10 DPA, 4,308 DEGs on 15 DPA, 3,147 DEGs on 20 DPA, 3,527 DEGs on 25 DPA, 3,746 DEGs on 30 DPA, and 7,407 DEGs on 35 DPA (Fig. 1c). These numbers showed that the number of DEGs changed with different fiber development times. At the beginning of the fiber elongation period until the secondary wall-thickening stage, the number of DEGs gradually increased, and it peaked at the end of the secondary wall-thickening stage. In a comparison of gene expression levels between PimaS-7 and 5917 in different fiber development times, many genes were identified as DEGs, and the total number of DEGs of PimaS-7 was similar with that of 5917. The results showed that the difference in FS was due to the gene expression levels in the 2 materials. In previous studies, according to the expression level of genes in different periods and changes in expression patterns, various genes were identified to be related to genetic diseases, yield, and resistance (Zhang et al. 2017; Cheng et al. 2020; Zhao, Bai, et al. 2020).

Construction of coexpression modules using WGCNA

To further clarify the relationship between the gene expression of the 2 materials in different fiber development times and FS, coexpression networks of DEGs were constructed using WGCNA with a power of 15 (Fig. 2a). Ten modules were identified to be related to different fiber development times; some were correlated with 1 time, while the others were correlated with more than 1 (Fig. 2b). The secondary wall-thickening stage was significantly related to the formation of FS. Then, analyzing the correlation between the

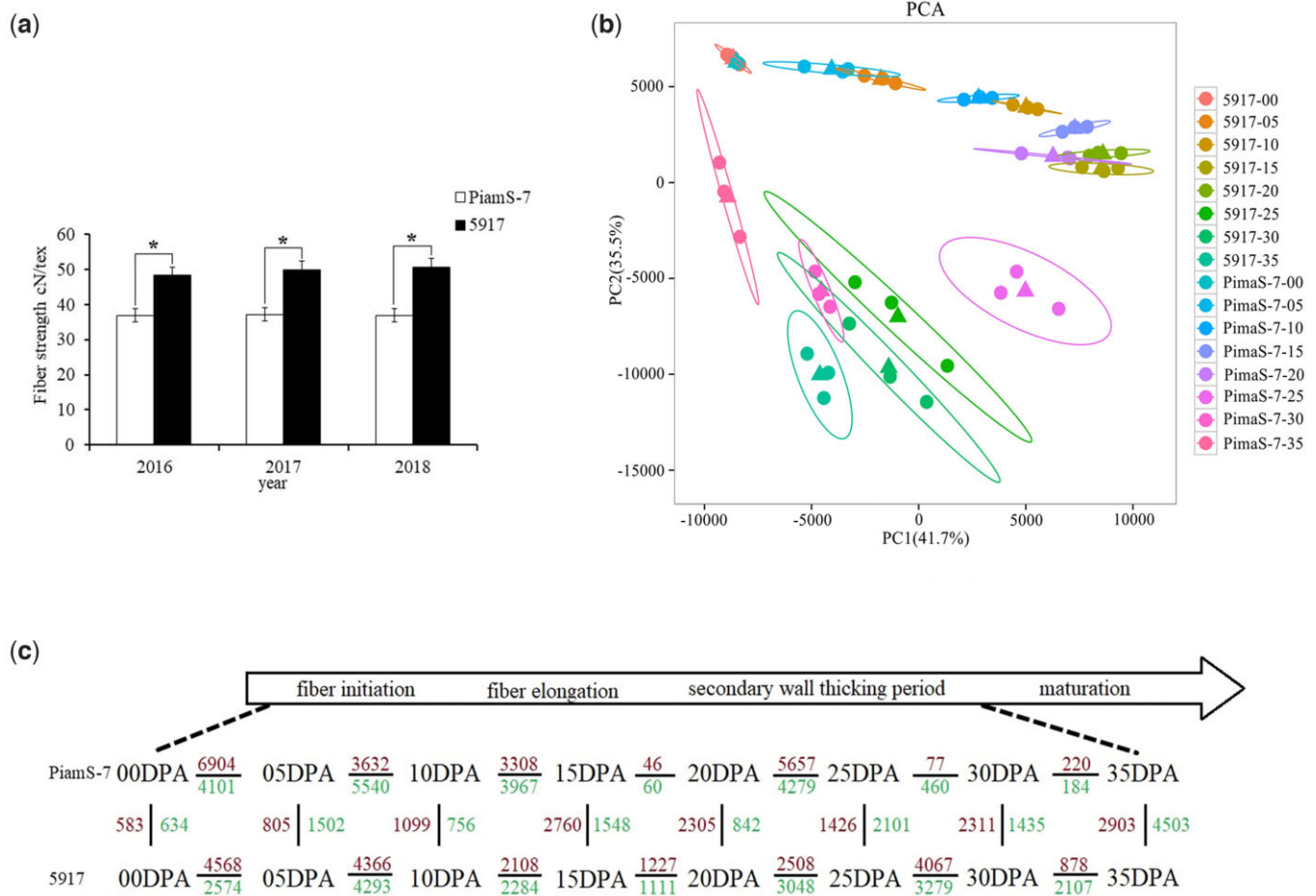


Fig. 1. Basic analysis of the PimaS-7 and 5917. a) FS data analysis between PimaS-7 and 5917. The 2 materials had significant differences in FS from 2016 to 2018 (*mean $P < 0.05$). b) PCA between PimaS-7 and 5917 in different fiber development times. c) Comparison of DEGs between PimaS-7 and 5917 in different fiber development times. Horizontal comparisons represent different developmental stages of the same gene. Vertical comparison represents the same developmental stage of different genes. Red represents upregulated genes; blue represents downregulated genes.

modules and different fiber development times, it was found that 4 specific modules were significantly related to 25, 30, and 35 DPA during the secondary wall-thickening stage. The brown module contains 1,278 genes, which were significantly correlated with 5917-35 DPA ($r^2 = 0.88$, $P = 1e-16$) and PimaS-7-30 DPA ($r^2 = 0.83$, $P = 4e-13$) (Fig. 2c). GO terms were enriched in plant-type secondary cell wall biogenesis, glucuronoxylan glucuronosyl-transferase activity, and plasma membrane (Supplementary Table 7). The green module contains 522 genes, which were highly associated with 5917-25 DPA ($r^2 = 0.88$, $P = 1e-16$), 5917-30 DPA ($r^2 = 0.88$, $P = 2e-16$) and PimaS-7-25 DPA ($r^2 = 0.85$, $P = 2e-14$). GO terms were enriched in steryl-sulfatase activity, cytoplasm and proteasome-mediated ubiquitin-dependent protein catabolic process. The MEred module contains 41 genes, which were significantly correlated with PimaS-7-35 DPA ($r^2 = 0.94$, $P = 7e-23$). GO terms were enriched in cell periphery, response to unfolded protein and unfolded protein binding. The MEblack module contains 109 genes, which were significantly related with PimaS-7-35 DPA ($r^2 = 0.83$, $P = 5e-13$). GO terms were enriched in structural constituents of ribosomes, Golgi-associated vesicle membrane and pectic galactan metabolic process. KEGG pathway analysis was used to further explore the 4 specific modules. MEbrown was enriched in amino sugar, nucleotide sugar metabolism, and phagosome (Supplementary Table 8). MEgreen was enriched in pentose phosphate pathway and carbon metabolism.

MEred was enriched in protein processing in endoplasmic reticulum, while MEblack was enriched in ribosome. This shows that DEGs from different modules may be involved in different metabolic pathways to regulate the development of FS.

Colocalization by QTLs and RNA-seq for the screening of related genes

To further narrow down the screening range of candidate genes related to FS, the expression data of 2,601 genes were extracted from RNA-seq data, of which 1,268 genes were identified as DEGs. These DEGs were used for further analysis (Supplementary Table 9). Then, $K_{ME} > 0.85$ was used as a standard of gene connectivity to select DEGs in the 4 specific modules; 1,537 DEGs were considered to participate in the regulation of FS in the 4 modules (MEbrown, MEgreen, MEred, and MEblack; Supplementary Table 10). Finally, a total of 55 DEGs were obtained by comparing them in 2 ways (Supplementary Table 11), and genes showed different expression levels in the secondary wall-thickening stage (Fig. 3a). GO enrichment analysis was performed on these DEGs (Fig. 3b); structural constituents of cytoskeletons, glutathione binding, and GTP binding were enriched in molecular function. F-actin capping protein complex, chloroplast and microtubule were enriched in cellular components. Microtubule-based process and ubiquitin-dependent protein catabolic process were enriched in biological process. Based on GO enrichment analysis, many DEGs mainly concentrated in GO term

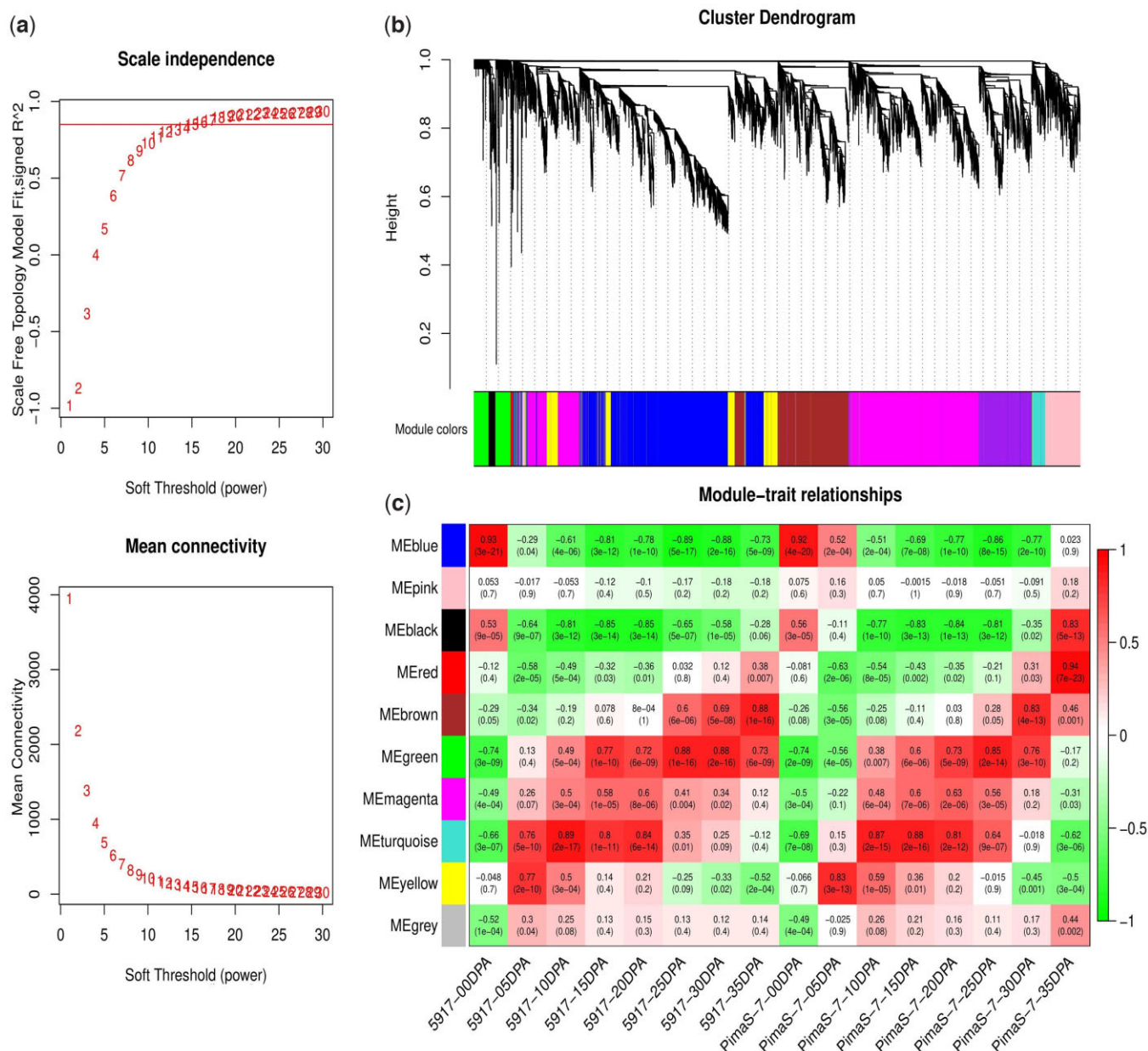


Fig. 2. Weighted coexpression network analysis of DEGs in PimaS-7 and 5917. a) Determination of soft threshold of weighted gene coexpression network. b) Hierarchical cluster diagram of coexpression modules identified by WGCNA. Each analysis represents a module, and each leaf represents a gene. c) Module related to the sample. The horizontal axis represents the different development times of the 2 materials, and the vertical axis represents the feature vectors of each module. The redder the color of the module, the higher the correlation with the corresponding fiber development stage.

related to cytoskeletal structures and microtubules. Therefore, the 5 DEGs included in the above pathway as representative genes of different terms were selected for further research. *Gbar_D08G020540* and *Gbar_D11G032910* were selected from the MEbrown module; these were significantly related to fiber development in 30 and 35 DPA. *Gbar_D08G013370* was selected from the MEred module, and it was related to fiber development in 35 DPA. *Gbar_D11G033670* and *Gbar_D11G029020* were selected from the MEgreen module, which were significantly related to fiber development in 25 and 30 DPA.

Confirmation of RNA-seq results by qRT-PCR

qRT-PCR was performed to confirm the RNA-seq data of 5 DEGs (Fig. 4a; Supplementary Table 12). The qRT-PCR results showed that

the expression trends of *Gbar_D11G032910*, *Gbar_D08G020540*, *Gbar_D11G033670*, *Gbar_D08G013370*, and *Gbar_D11G029020* in PimaS-7 and 5917 were similar with the transcriptome. The result of linear regression analysis was used to prove that the qRT-PCR results and RNA-seq were significantly correlated (Fig. 4b).

To further confirm the expression patterns of the 5 DEGs, 4 high-FS materials, and 4 low-FS materials were selected from the *G. barbadense* accessions (Table 1), and qRT-PCR was performed (Supplementary Fig. 2). During 0–15 DPA, the expression patterns of 8 materials were similar to those of the 2 parents. In 20 DPA, the upregulation of 4 gene expression levels in high-FS materials gradually became greater than those in low-FS materials. At 25–30 DPA, the expression of the 5 DEGs in all materials showed a rapid upward trend as a whole, and the rate of increase in gene

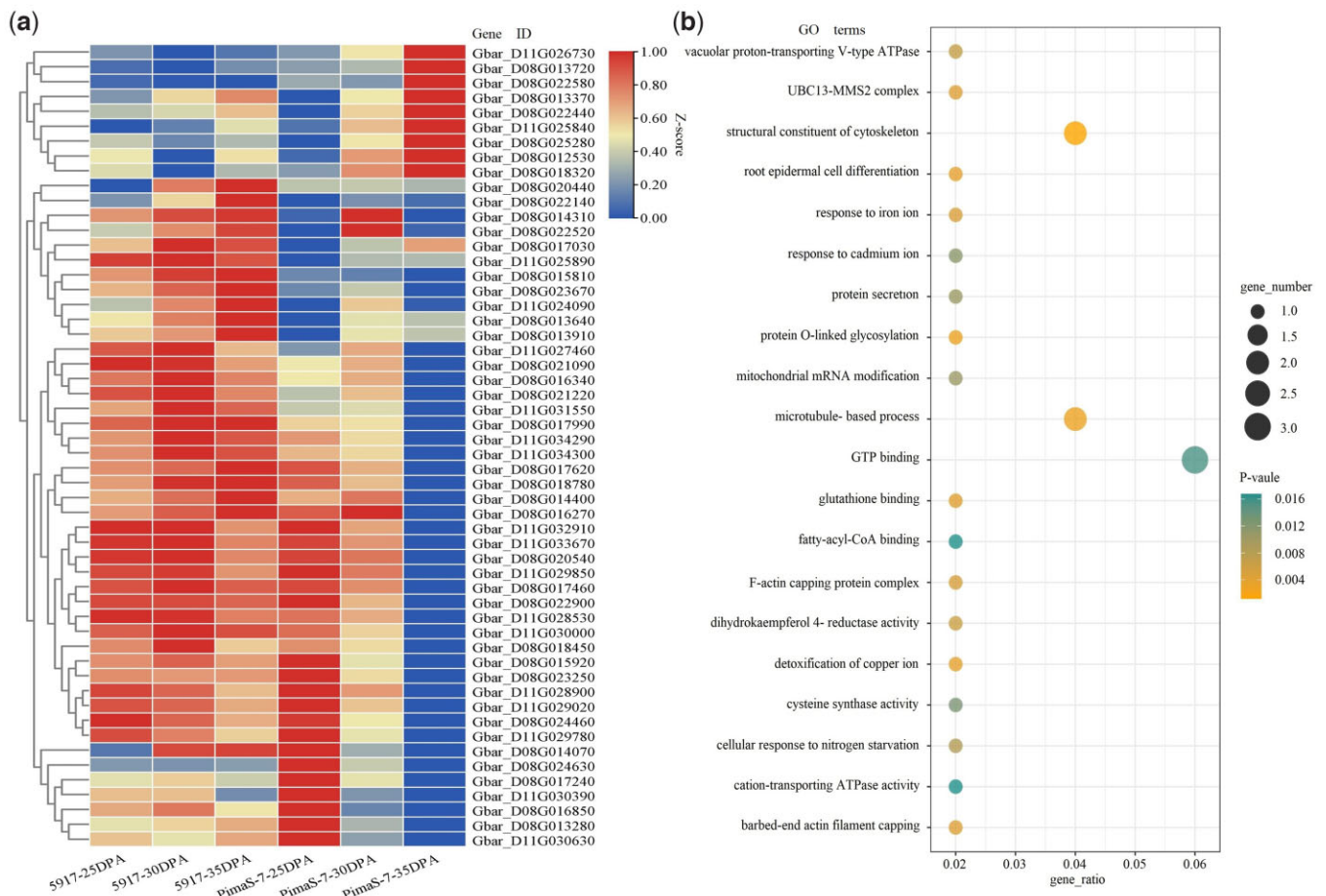


Fig. 3. Analysis of the 55 DEGs related to FS. a) Heat map of RNA-seq data of 55 DEGs in 25 DPA, 30 and 35 DPA. Red means high expression. b) GO enrichment analysis of 55 DEGs. The importance of the GO term increases as the P-value decreases, gene_number represents the number of genes enriched in each term.

expression level was increased in high-FS materials relative to that in the other group. In 35 DPA, the gene expression level of high-FS materials continued to grow, whereas that of the other group decreased. This indicates that during the secondary wall-thickening stage, the expression trends of the 5 DEGs were similar to those of the 2 parents, and the main difference lies on the difference in gene expression levels.

Correlation between the expression levels of the 5 DEGs and FS in *G. barbadense* accessions

To study the relationship between the expression levels of the 5 DEGs and FS, correlation analysis was performed on 8 materials at different fiber development times and FS (Table 2). At 5–20 DPA, there was no correlation between the expression levels of 4 DEGs (*Gbar_D11G033670*, *Gbar_D08G020540*, *Gbar_D11G029020*, and *Gbar_D08G013370*) in *G. barbadense* accessions and FS data, except for *Gbar_D11G032910*. This was consistent with the RNA-seq results, where the DEG expression level was low in the early stages of fiber development. Until the secondary wall-thickening stage, 5 DEGs were positively correlated in *G. barbadense* accessions with FS data. This result is consistent with the related gene expression levels, which showed a rapid increase at 25–30 DPA in the RNA-seq data. Therefore, the high expression of the 5 DEGs may be involved in regulating FS during the secondary wall-thickening stage.

Discussion

Combined analysis of QTL mapping and RNA-seq enables the study of FS in secondary wall-thickening stage

With continuous improvements in the sequencing of the cotton genome, many important genes related to cotton quality are being discovered. Transcriptome technology is widely used to study fiber quality traits in cotton (Lee et al. 2007; Yoo and Wendel 2014; Naoumkina et al. 2015; Li, Wang, et al. 2017) and it further combines QTL, GWAS, TWAS, and eQTL networks to provide more strategies for the improvement of fiber length and FS (Li, Cheng, et al. 2018; Li et al. 2020). Based on the RNA-seq data of the near-isogenic lines NMGA-062 and NMGA-105 combined with QTL mapping, the candidate genes were finally determined (Li, Wu, et al. 2017). Kashif Shahzad used *G. hirsutum* inbred lines and hybrids as materials; RNA-seq, QTL, and gene coexpression network analyses were performed, and *Gh_A03G1024*, *Gh_D08G1440*, and *Gh_A08G2210* were determined as candidate genes (Shahzad et al. 2020). Therefore, in the present study, the combined analysis of QTL and RNA-seq was selected to explore the genes associated with FS. In addition, numerous existing studies on fiber quality traits were mostly based on the exploration of fiber cell protrusion and fiber elongation and usually with the *G. barbadense* fragment inserted in a single background of *G. hirsutum* as the research background (Islam et al. 2016). It cannot be ignored that due to the different genetic backgrounds of the 2 genomes,

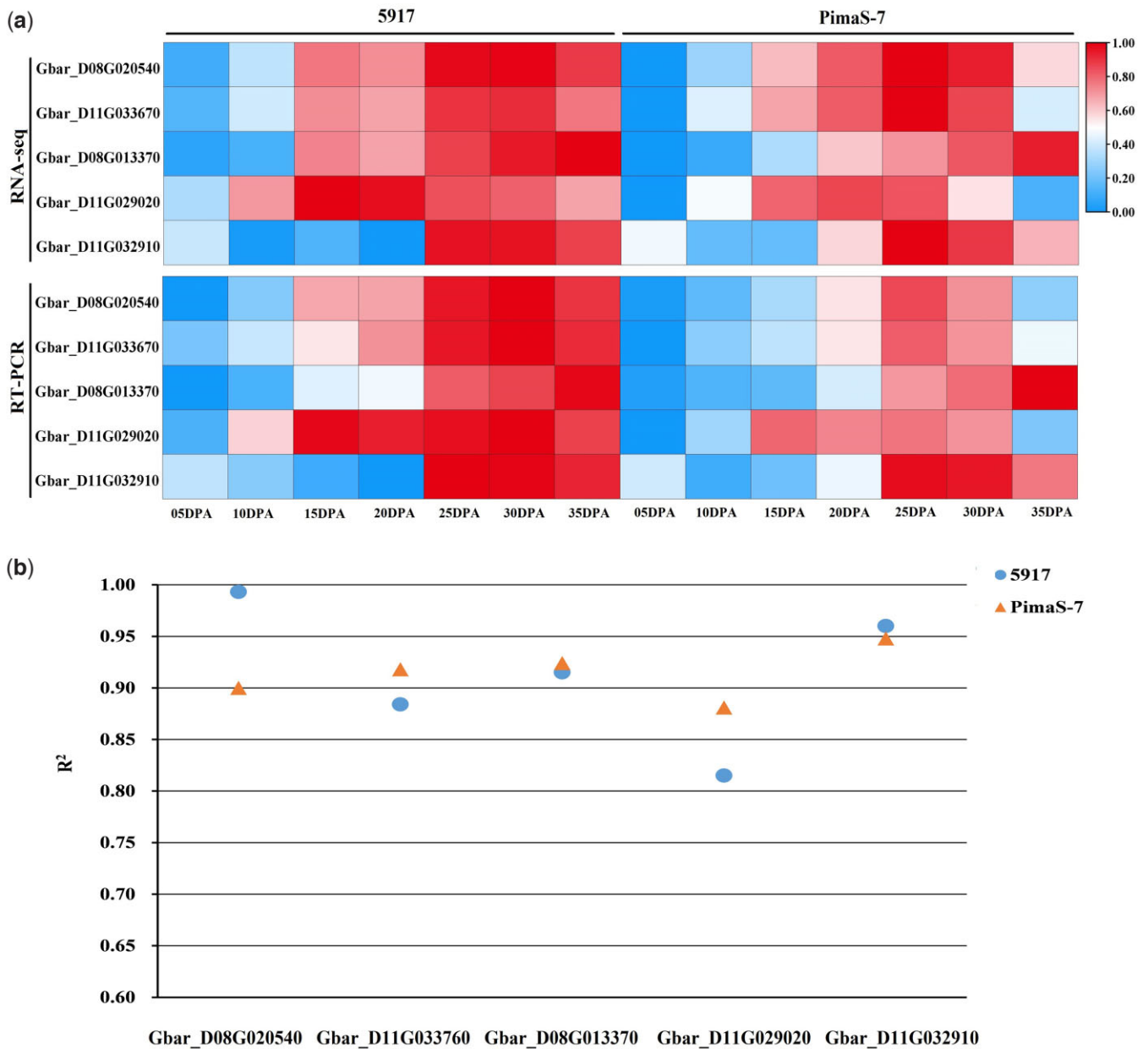


Fig. 4. Confirm of expression level of 5 DEGs by qRT-PCR. a) The heat map of qRT-PCR results and RNA-seq data of 5 DEGs in PimaS-7 and 5917. b) Linear regression analysis between qRT-PCR results and RNA-seq data of 5 DEGs. The R^2 value represents the correlation between the RNA-seq data and qRT-PCR results of the 2 materials. Each times represents the fold-change expression level compared with 00 DPA, that were converted by \log_2 standardization.

Table 1. The list of fiber quality traits of *G. barbadense* accessions.

Name	Fiber length	Fiber strength	Micronaire	Fiber elongation	Fiber uniformity	Fiber maturity	Spinning consistence index
Xinhai30	33.97 ± 0.62	52.25 ± 2.74	4.22 ± 0.26	5.50 ± 0.43	87.71 ± 0.67	0.87 ± 0.00	219.32 ± 16.76
Xinhai31	33.63 ± 0.52	52.16 ± 2.57	4.65 ± 0.04	5.49 ± 1.03	88.24 ± 0.68	0.88 ± 0.00	222.61 ± 8.49
SI	34.13 ± 1.42	52.15 ± 5.27	4.46 ± 0.21	4.90 ± 0.67	87.91 ± 1.07	0.88 ± 0.00	221.77 ± 20.00
C6015	34.82 ± 1.27	50.56 ± 3.83	4.59 ± 0.06	4.96 ± 0.86	87.55 ± 1.70	0.88 ± 00.00	213.56 ± 14.84
65-3049-6	32.06 ± 0.64	31.09 ± 5.34	4.22 ± 0.48	7.79 ± 0.78	85.54 ± 0.28	0.84 ± 0.00	154.05 ± 20.77
65-3080-2	37.93 ± 0.68	33.83 ± 3.21	3.08 ± 0.22	5.66 ± 0.73	84.65 ± 0.87	0.83 ± 0.00	178.05 ± 5.68
Ashmont	25.05 ± 1.95	29.66 ± 2.08	5.44 ± 0.58	8.25 ± 1.32	81.38 ± 0.60	0.87 ± 0.01	100.32 ± 11.13
75-86	33.98 ± 0.65	29.94 ± 1.30	4.23 ± 0.41	5.75 ± 0.95	84.64 ± 0.86	0.86 ± 0.00	146.08 ± 2.95

Table 2. Correlation analysis of gene expression level and FS of *G. barbadense* accessions with 5 DEGs.

	Gbar_D08G013370	Gbar_D11G029020	Gbar_D11G032910	Gbar_D11G033670	Gbar_D08G020540
05 DPA	0.111	0.196	0.524	0.317	0.445
10 DPA	0.240	0.622	0.297	0.090	0.648
15 DPA	0.341	0.668	0.876**	−0.240	0.135
20 DPA	0.406	0.701	0.897**	0.047	0.229
25 DPA	0.851**	0.840**	0.907**	0.366	0.748*
30 DPA	0.920**	0.938**	0.807*	0.819*	0.868**
35 DPA	0.904**	0.975**	0.912**	0.808*	0.945**

*Correlation is significant at 0.05 level (2-tailed).

**Correlation is significant at 0.01 level (2-tailed)

the growth period of *G. barbadense* was longer than that of *G. hirsutum*, and the 4 stages of fiber development were also different in *G. hirsutum* (Gilbert et al. 2014). Therefore, most genes identified as candidate genes in different research materials may be due to the differences between the species (Chen et al. 2012). The completion of gene sequencing of *G. barbadense* provides an accurate reference genome for the study of FS traits under the background of a single *G. barbadense* (Wang et al. 2019). In the present study, 2,601 genes of D03, D08, and D11 were obtained from *G. barbadense* in Fan's studies. These were believed to be related to FS. Next, transcriptome technology was used to sequence the 48 cDNA libraries with 2 types of materials from *G. barbadense*. After comparing the genes obtained by the 2 methods, a link was found between the results. These combined research methods were conducive for a more comprehensive study of the developmental mechanisms related to FS.

Analysis of candidate genes related to FS

Functional analysis of related DEGs revealed that many genes were mainly involved in terms such as the composition of structural constituents of the cytoskeleton and microtubule-based processes. Therefore, 5 DEGs were selected as representative genes in the main terms and further verified by qRT-PCR. The results of qRT-PCR were consistent with the RNA-seq data, indicating that DEGs located in the main terms have regulatory effects on FS. Previous studies showed that cellulose and pectin are important components of plant cell walls and play an important role in fiber cells (Gilbert et al. 2013; Sun et al. 2020). Cellulose synthase 6 (CesA6) mutants show a strong cellulose deficiency, and hence a short hypocotyl phenotype is observed in *Arabidopsis* (Hematy et al. 2007). Therefore, DEGs involved in the regulation of cell wall composition during the secondary wall-thickening stage were more likely to regulate FS.

The main components of cytoskeletons are microtubules and actin. *Gbar_D11G029020* encodes a capping protein with 303 amino acids, which does not depend on Ga^{2+} -bound actin. As the main component of microfilaments, actins, like tubulins, regulate the transport of intracellular substances and maintain the cytoskeleton stability (Mathur et al. 2003). In upland cotton, *GhXLIM6*, *GhWLM2*, *GhWLM5*, and *GhPLIM1* were highly expressed during the secondary cell wall-thickening stage (Zhao et al. 2010; Li et al. 2013, 2015). When the expression of *GhXLIM6* was inhibited, its association with F-actin in cotton fibers was destroyed, and the actin cytoskeletons were distorted, which seriously affected the formation of fiber length and FS (Li, Wang, et al. 2018). Other studies also proved that regulating actin can cause fiber elongation defects and affect microtubules and the direction of cellulose deposition (Li et al. 2005; Wang et al. 2010; Cao et al. 2021). Therefore, the persistent high expression of

Gbar_D11G029020 during the secondary cell wall-thickening stage may help maintain cell wall stability to enhance FS.

Gbar_D11G033670 and *Gbar_D08G020540* encode proteins with 447 and 444 amino acids, respectively, belonging to the β -tubulin group in microtubules. *Gbar_D11G032910* encodes a protein with 360 amino acids that contains a TPX2 domain belonging to microtubule-associated proteins. Microtubules are hollow tubular structures that run longitudinally along the cell wall. They play an important role in biological processes as structural components of the cytoskeleton for vesicle transport and cell wall deposition (Seagull 1998; Sedbrook and Kaloriti 2008). In *Arabidopsis*, *ATWVD2* can interact with microtubules, promote microtubule bundling, and play an important role in polar cell elongation (Perrin et al. 2007). The overexpression of *ATWDL3* and RNAi-downregulated lines showed shorter and longer hypocotyl cells than the wild-type, respectively (Liu et al. 2013). Previous research on microtubules genes in cotton mainly focused on fiber quality in the early stage of fiber development in *G. hirsutum* (Li, Wu, et al. 2017; Qin et al. 2019; Zou et al. 2019). Research on microtubules is gradually increasing. Microtubule genes are deemed to have specific expression levels in different tissues and organs (Tuttle et al. 2015), and multiple microtubule genes are highly expressed during fiber development and are regulated by hormones (Islam et al. 2016). *GhWDLA7* possibly regulates FS through its interaction with *GhTUA2* (Lei et al. 2019).

Gbar_D08G013370 encodes the protein SPIRAL1-like 2. Studies showed that the interaction of SPIRAL1-like 2 and SPIRAL1-like 1 could regulate cortical microtubules required to grow anisotropic cells (Nakajima et al. 2006). Cellulose synthase interactive protein 1 (encodes by *CSI1*) connects the cellulose synthase in vesicles to cortical microtubules (Li et al. 2012). At the same time, neatly arranged cortical microtubules under the plasma membrane guide and regulate the synthesis direction and deposition rate of cellulose fibrils on the plasma membrane (Gutierrez et al. 2009; Zhong and Ye 2015). When the genes involved in cell wall rearrangement were affected, the transcription of cellulose synthase was inhibited (Gilbert et al. 2013). From this, we can speculate that the orderly arrangement of cortical microtubules not only affects the cytoskeletal structures, but also affects the deposition of cellulose. Therefore, the continuous high expression of representative DEGs in main terms such as cytoskeletal structures and microtubules in high-FS materials provide a direction for the study of FS. Although numerous existing studies could prove that the current cellulose synthase and microtubule candidate genes were able to affect the formation of FS by regulating the pathways of fiber formation, it is necessary to further explore the specific positions of different genes in the pathways.

Conclusion

In summary, based on previous studies, 3 QTLs related to FS were relocated. The RNA-seq method was used to study the differences in gene expression at different fiber development stages between the 2 parents (PimaS-7 and 5917). Forty-eight cDNA libraries were constructed using RNA-seq, containing 2 materials and 3 biological replicates. Four specific modules closely related to the secondary wall-thickening stage were obtained using WGCNA. In genetic colocalization using QTL and RNA-seq, 55 candidate genes were identified. GO enrichment analysis showed that many DEGs mainly concentrated in GO term related to cytoskeletal structures and microtubules. The qRT-PCR results showed that the expression levels of the 5 representative genes in the main terms were significantly higher than those in low-FS materials during the secondary wall-thickening stage of fiber development. Results showed that *Gbar_D11G032910*, *Gbar_D08G020540*, *Gbar_D08G013370*, *Gbar_D11G033670*, and *Gbar_D11G029020* may regulate FS by playing a role in the composition of structural constituents of the cytoskeleton during fiber development.

Data availability

Supplementary Fig. 1 contains Pearson's correlation coefficient analysis of population parents. Supplementary Fig. 2 contains 5 DEGs were confirmed by qRT-PCR of *G. barbadense* accessions. The linkage map, RIL genotype, and RIL phenotype mean for fiber traits in Supplementary Tables 1–3. Supplementary Table 4 contains the marker intervals and the physical confidence intervals of the QTLs related FS. Supplementary Table 5 contains 2,601 genes were included in 3 QTLs physical position. Supplementary Table 6 contains summary of RNA-seq data. The RNA-seq raw data underlying this article are available in NCBI database with the accession number: GSE178945, at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE178945>. Supplementary Table 7 contains GO-term enrichment of 4 significant WGCNA modules. Supplementary Table 8 contains KEGG pathway enrichment of 4 significant WGCNA modules. Supplementary Table 9 contains 1,268 different expression genes were included in 3 QTLs physical position. Supplementary Table 10 contains the list of gene in 4 specific modules. Supplementary Table 11 contains description of 55 colocalized differentially expressed genes. Supplementary Table 12 contains primer sequences used in qRT-PCR.

Supplemental material is available at G3 online.

Acknowledgments

The authors are very grateful to Tianzhen Zhang and Jiedan Chen and his research group from Zhejiang University (Hangzhou, China).

Funding

This research was funded by Technology Major Project of Xinjiang (2020A01002-2).

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Literature cited

- Ahmed M, Iqbal A, Latif A, Din SU, Sarwar MB, Wang X, Rao AQ, Husnain T, A, Shahid, A. Overexpression of a sucrose synthase gene indirectly improves cotton fiber quality through sucrose cleavage. *Front Plant Sci.* 2020;11:476251.
- Ali I, Teng Z, Bai Y, Yang Q, Hao Y, Hou J, Jia Y, Tian L, Liu X, Tan Z, et al. A high density SLAF-SNP genetic map and QTL detection for fibre quality traits in *Gossypium hirsutum*. *BMC Genomics.* 2018; 19(1):879.
- Ali M, Cheng HL, Soomro M, Li S, Tufail BM, Nazir MF, Feng XX, Zhang YP, Zuo D, Limin L, et al. Comparative transcriptomic analysis to identify the genes related to delayed gland morphogenesis in *Gossypium bickii*. *Genes (Basel).* 2020;11(5):472.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–410.
- Basra AS, Malik CP. Development of the cotton fiber. *Int Rev Cytol.* 1984;89:65–109.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–2120.
- Cao Y, Huang H, Yu Y, Dai H, Hao H, Zhang H, Jiang Y, Ding M, Li F, Tu L, et al. A modified actin (Gly65Val Substitution) expressed in cotton disrupts polymerization of actin filaments leading to the phenotype of Ligon lintless-1 (*Li₁*) mutant. *IJMS.* 2021;22(6):3000.
- Chen B, Zhao J, Fu G, Pei X, Pan Z, Li H, Ahmed H, He S, Du X. Identification and expression analysis of tubulin gene family in upland cotton. *J Cotton Res.* 2021;4(1):20.
- Chen J, Burke JJ. Developing fiber specific promoter-reporter transgenic lines to study the effect of abiotic stresses on fiber development in cotton. *PLoS One.* 2015;10(6):e0129870.
- Chen Q, Chen QJ, Sun GQ, Zheng K, Yao ZP, Han YH, Wang LP, Duan YJ, Yu DQ, Qu YY. Genome-wide identification of cyclophilin gene family in cotton and expression analysis of the fibre development in *Gossypium barbadense*. *Int J Mol Sci.* 2019;20(2):349.
- Chen XD, Guo WZ, Liu BL, Zhang YM, Song XL, Cheng Y, Zhang LL, Zhang TZ. Molecular mechanisms of fiber differential development between *G. barbadense* and *G. hirsutum* revealed by genetical genomics. *PLoS One.* 2012;7(1):e30056.
- Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD, et al. Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol.* 2007;145(4):1303–1310.
- Cheng G, Zhang L, Wang H, Lu J, Wei H, Yu S. Transcriptomic profiling of young cotyledons response to chilling stress in two contrasting cotton (*Gossypium hirsutum* L.) genotypes at the seedling stage. *Int J Mol Sci.* 2020;21(14):5095.
- Du J, Wang S, He C, Zhou B, Ruan YL, Shou H. Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J Exp Bot.* 2017;68(8): 1955–1972.
- Edelmann HG, Fry SC. Effect of cellulose synthesis inhibition on growth and the integration of xyloglucan into pea internode cell walls. *Plant Physiol.* 1992;100(2):993–997.
- Fan L, Wang L, Wang X, Zhang H, Zhu Y, Guo J, Gao W, Geng H, Chen Q, Qu Y. A high-density genetic map of extra-long staple cotton (*Gossypium barbadense*) constructed using genotyping-by-sequencing based single nucleotide polymorphic markers and identification of fiber traits-related QTL in a recombinant inbred line population. *BMC Genomics.* 2018;19(1):489.
- Feng L, Zhou C, Su Q, Xu M, Yue H, Zhang S, Zhou B. Fine-mapping and candidate gene analysis of *qFS-Chr. D02*, a QTL for fibre strength introgressed from a semi-wild cotton into *Gossypium hirsutum*. *Plant Sci.* 2020;297:110524.

- Feng X, Liu S, Cheng H, Zuo D, Zhang Y, Wang Q, Lv L, Song G. Weighted gene co-expression network analysis reveals hub genes contributing to fuzz development in *Gossypium arboreum*. *Genes*. 2021;12(5):753.
- Gao Z, Sun W, Wang J, Zhao C, Zuo K. *GhbHLH18* negatively regulates fiber strength and length by enhancing lignin biosynthesis in cotton fibers. *Plant Sci*. 2019;286:7–16.
- Garg R, Singh VK, Rajkumar MS, Kumar V, Jain M. Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with seed development and seed size/weight determination in chickpea. *Plant J*. 2017;91(6):1088–1107.
- Gilbert MK, Kim HJ, Tang Y, Naoumkina M, Fang DD. 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*. 2014;9(4):e95554.
- Gilbert MK, Turley RB, Kim HJ, Li P, Thyssen G, Tang Y, Delhom CD, Naoumkina M, Fang DD. Transcript profiling by microarray and marker analysis of the short cotton (*Gossypium hirsutum* L.) fiber mutant *Ligon lintless-1* (*Li₁*). *BMC Genomics*. 2013;14(1):403.
- Gu Q, Ke H, Liu Z, Lv X, Sun Z, Zhang M, Chen L, Yang J, Zhang Y, Wu L, et al. A high-density genetic map and multiple environmental tests reveal novel quantitative trait loci and candidate genes for fibre quality and yield in cotton. *Theor Appl Genet*. 2020;133(12):3395–3408.
- Guo AH, Su Y, Huang Y, Wang YM, Nie HS, Zhao N, Hua JP. QTL controlling fiber quality traits under salt stress in upland cotton. *Theor Appl Genet*. 2021;134(2):661–685.
- Gutierrez R, Lindeboom JJ, Paredes AR, Emons AM, Ehrhardt DW. *Arabidopsis* cortical microtubules position cellulose synthase delivery to the plasma membrane and interact with cellulose synthase trafficking compartments. *Nat Cell Biol*. 2009;11(7):797–806.
- Haigler CH, Betancur L, Stiff MR, Tuttle JR. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Front Plant Sci*. 2012;3:104.
- Hematy K, Sado PE, Van Tuinen A, Rochange S, Desnos T, Balzergue S, Pelletier S, Renou JP, Hofte H. A receptor-like kinase mediates the response of *Arabidopsis* cells to the inhibition of cellulose synthesis. *Curr Biol*. 2007;17(11):922–931.
- Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet*. 2019;51(4):739–748.
- Islam MS, Fang DD, Thyssen GN, Delhom CD, Liu Y, Kim HJ. 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*. 2016;16:36.
- Jiang X, Fan L, Li P, Zou X, Zhang Z, Fan S, Gong J, Yuan Y, Shang H. Co-expression network and comparative transcriptome analysis for fiber initiation and elongation reveal genetic differences in two lines from upland cotton CCRI70 RIL population. *PeerJ*. 2021;9:e11812.
- Keerio A, Shen C, Nie Y, Ahmed M, Zhang X, Lin Z. QTL mapping for fiber quality and yield traits based on introgression lines derived from *Gossypium hirsutum* x *G. tomentosum*. *Int J Mol Sci*. 2018;19(1):243.
- Khattak AN, Wang T, Yu K, Yang R, Wan W, Ye B, Tian E. Exploring the basis of 2-propenyl and 3-butenyl glucosinolate synthesis by QTL mapping and RNA-sequencing in *Brassica juncea*. *PLoS One*. 2019;14(10):e0220597.
- Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–360.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559.
- Laoue J, Depardieu C, Gerardi S, Lamothe M, Bomal C, Azaiez A, Gros-Louis MC, Laroche J, Boyle B, Hammerbacher A, et al. Combining QTL mapping and transcriptomics to decipher the genetic architecture of phenolic compounds metabolism in the conifer white spruce. *Front Plant Sci*. 2021;12:675108.
- Lee JJ, Woodward AW, Chen ZJ. Gene expression changes and early events in cotton fibre development. *Ann Bot*. 2007;100(7):1391–1401.
- Lei K, Liu A, Fan S, Peng H, Zou X, Zhen Z, Huang J, Fan L, Zhang Z, Deng X, et al. Identification of *TPX2* gene family in upland cotton and its functional analysis in cotton fiber development. *Genes (Basel)*. 2019;10(7):508.
- Lei L, Zheng H, Bi Y, Yang L, Liu H, Wang J, Sun J, Zhao H, Li X, Li J, et al. Identification of a major QTL and candidate gene analysis of salt tolerance at the bud burst stage in rice (*Oryza sativa* L.) using QTL-Seq and RNA-Seq. *Rice (N Y)*. 2020;13(1):55.
- Li C, Duan Y, Miao H, Ju M, Wei L, Zhang H. Identification of candidate genes regulating the seed coat color trait in sesame (*Sesamum indicum* L.) using an integrated approach of QTL mapping and transcriptome analysis. *Front Genet*. 2021;12:700469.
- Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet*. 2014;46(6):567–572.
- Li H, Cheng X, Zhang L, Hu J, Zhang F, Chen B, Xu K, Gao G, Li H, Li L, et al. An integration of genome-wide association study and gene co-expression network analysis identifies candidate genes of stem lodging-related traits in *Brassica napus*. *Front Plant Sci*. 2018;9:796.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
- Li L, Li Y, Wang NN, Li Y, Lu R, Li XB. Cotton LIM domain-containing protein *GhPLI₁*. M1 is specifically expressed in anthers and participates in modulating F-actin. *Plant Biol*. 2015;17(2):528–534.
- Li PT, Wang M, Lu QW, Ge Q, Rashid MHO, Liu AY, Gong JW, Shang HH, Gong WK, Li JW, et al. Comparative transcriptome analysis of cotton fiber development of upland cotton (*Gossypium hirsutum*) and chromosome segment substitution lines from *G. hirsutum* x *G. barbadense*. *BMC Genomics*. 2017;18(1):705.
- Li S, Lei L, Somerville CR, Gu Y. Cellulose synthase interactive protein 1 (*CSI1*) links microtubules and cellulose synthase complexes. *Proc Natl Acad Sci U S A*. 2012;109(1):185–190.
- Li X, Wu M, Liu G, Pei W, Zhai H, Yu J, Zhang J, Yu S. Identification of candidate genes for fiber length quantitative trait loci through RNA-Seq and linkage and physical mapping in cotton. *BMC Genomics*. 2017;18(1):427.
- Li X-B, Fan X-P, Wang X-L, Cai L, Yang W-C. The cotton *ACTIN1* gene is functionally expressed in fibers and participates in fiber elongation. *Plant Cell*. 2005;17(3):859–875.
- Li Y, Jiang J, Li L, Wang XL, Wang NN, Li DD, Li XB. A cotton LIM domain-containing protein (*GhWLM5*) is involved in bundling actin filaments. *Plant Physiol Biochem*. 2013;66:34–40.
- Li Y, Tu L, Pettolino FA, Ji S, Hao J, Yuan D, Deng F, Tan J, Hu H, Wang Q, et al. *GbEXPATR*, a species-specific expansin, enhances cotton fibre elongation through cell wall restructuring. *Plant Biotechnol J*. 2016;14(3):951–963.
- Li Y, Wang NN, Wang Y, Liu D, Gao Y, Li L, Li XB. The cotton *XLIM* protein (*GhXLIM6*) is required for fiber development via maintaining dynamic F-actin cytoskeleton and modulating cellulose biosynthesis. *Plant J*. 2018;96(6):1269–1282.

- Li Z, Wang P, You C, Yu J, Zhang X, Yan F, Ye Z, Shen C, Li B, Guo K, et al. Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton. *New Phytol.* 2020;226(6):1738–1752.
- Lin JX, Yu M, Xu X, Wang YT, Xing HT, An J, Yang JP, Tang CZ, Sun D, Zhu YY. Identification of biomarkers related to CD8⁺ T cell infiltration with gene co-expression network in clear cell renal cell carcinoma. *Aging.* 2020;12(4):3694–3712.
- Liu W, Song C, Ren Z, Zhang Z, Pei X, Liu Y, He K, Zhang F, Zhao J, Zhang J, et al. Genome-wide association study reveals the genetic basis of fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *BMC Plant Biol.* 2020;20(1):395.
- Liu X, Qin T, Ma Q, Sun J, Liu Z, Yuan M, Mao T. Light-regulated hypocotyl elongation involves proteasome-dependent degradation of the microtubule regulatory protein WDL3 in *Arabidopsis*. *Plant Cell.* 2013;25(5):1740–1755.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCT} method. *Methods.* 2001;25(4):402–408.
- Mathur J, Mathur N, Kernebeck B, Hulskamp M. Mutations in actin-related proteins 2 and 3 affect cell shape development in *Arabidopsis*. *Plant Cell.* 2003;15(7):1632–1645.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5(7):621–628.
- Nakajima K, Kawamura T, Hashimoto T. Role of the SPIRAL1 gene family in anisotropic growth of *Arabidopsis thaliana*. *Plant Cell Physiol.* 2006;47(4):513–522.
- Naoumkina M, Thyssen GN, Fang DD. 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.* 2015;15:65.
- Park M, Lee JH, Han K, Jang S, Han J, Lim JH, Jung JW, Kang BC. A major QTL and candidate genes for capsaicinoid biosynthesis in the pericarp of *Capsicum chinense* revealed using QTL-seq and RNA-seq. *Theor Appl Genet.* 2019;132(2):515–529.
- Patel JD, Huang X, Lin L, Das S, Chandnani R, Khanal S, Adhikari J, Shehzad T, Guo H, Roy-Zokan EM, et al. The Ligon lintless -2 short fiber mutation is located within a terminal deletion of chromosome 18 in cotton. *Plant Physiol.* 2020;183(1):277–288.
- Perrin RM, Wang Y, Yuen CYL, Will J, Masson PH. WVD2 is a novel microtubule-associated protein in *Arabidopsis thaliana*. *Plant J.* 2007;49(6):961–971.
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290–295.
- Potikha TS, Collins CC, Johnson DI, Delmer DP, Levine A. The involvement of hydrogen peroxide in the differentiation of secondary walls in cotton fibers. *Plant Physiol.* 1999;119(3):849–858.
- Pydiura N, Pirko Y, Galinousky D, Postovoitova A, Yemets A, Kilchevsky A, Blume Y. Genome-wide identification, phylogenetic classification, and exon-intron structure characterization of the tubulin and actin genes in flax (*Linum usitatissimum*). *Cell Biol Int.* 2019;43(9):1010–1019.
- Qin Y, Sun H, Hao P, Wang H, Wang C, Ma L, Wei H, Yu S. Transcriptome analysis reveals differences in the mechanisms of fiber initiation and elongation between long- and short-fiber cotton (*Gossypium hirsutum* L.) lines. *BMC Genomics.* 2019;20(1):633.
- Seagull RW. Cytoskeletal stability affects cotton fiber initiation. *Int J Plant Sci.* 1998;159(4):590–598.
- Sedbrook JC, Kaloriti D. Microtubules, MAPs and plant directional cell expansion. *Trends Plant Sci.* 2008;13(6):303–310.
- Shahzad K, Zhang X, Guo L, Qi T, Bao L, Zhang M, Zhang B, Wang H, Tang H, Qiao X, et al. Comparative transcriptome analysis between inbred and hybrids reveals molecular insights into yield heterosis of upland cotton. *BMC Plant Biol.* 2020;20(1):239.
- Shang L, Wang Y, Wang X, Liu F, Abduweli A, Cai S, Li Y, Ma L, Wang K, Hua J. Genetic analysis and QTL detection on fiber traits using two recombinant inbred lines and their backcross populations in upland cotton. G3 (Bethesda). 2016;2717–2724. 6
- Shi Y, Li W, Li A, Ge R, Zhang B, Li J, Liu G, Li J, Liu A, Shang H, et al. Constructing a high-density linkage map for *Gossypium hirsutum* x *Gossypium barbadense* and identifying QTLs for lint percentage. *J Integr Plant Biol.* 2015;57(5):450–467.
- Shi Y, Liu A, Li J, Zhang J, Li S, Zhang J, Ma L, He R, Song W, Guo L, et al. Examining two sets of introgression lines across multiple environments reveals background-independent and stably expressed quantitative trait loci of fiber quality in cotton. *Theor Appl Genet.* 2020;133(7):2075–2093.
- Su X, Zhu G, Song X, Xu H, Li W, Ning X, Chen Q, Guo W. Genome-wide association analysis reveals loci and candidate genes involved in fiber quality traits in sea island cotton (*Gossypium barbadense*). *BMC Plant Biol.* 2020;20(1):289.
- Sugiyama Y, Wakazaki M, Toyooka K, Fukuda H, Oda Y. A novel plasma membrane-anchored protein regulates xylem cell-wall deposition through microtubule-dependent lateral inhibition of Rho GTPase domains. *Curr Biol.* 2017;27(16):2522–2528.
- Sun H, Hao P, Gu L, Cheng S, Wang H, Wu A, Ma L, Wei H, Yu S. Pectate lyase-like Gene GhPEL76 regulates organ elongation in *Arabidopsis* and fiber elongation in cotton. *Plant Sci.* 2020;293:110395.
- Tan Z, Zhang Z, Sun X, Li Q, Sun Y, Yang P, Wang W, Liu X, Chen C, Liu D, et al. Genetic map construction and fiber quality QTL mapping using the cottonSNP80K array in upland cotton. *Front Plant Sci.* 2018;9:225.
- Tu LL, Zhang XL, Liang SG, Liu DQ, Zhu LF, Zeng FC, Nie YC, Guo XP, Deng FL, Tan JF, et al. Genes expression analyses of sea-island cotton (*Gossypium barbadense* L.) during fiber development. *Plant Cell Rep.* 2007;26(8):1309–1320.
- Tuttle JR, Nah G, Duke MV, Alexander DC, Guan X, Song Q, Chen ZJ, Scheffler BE, Haigler CH. Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis, represses lignification, and prolongs elongation. *BMC Genomics.* 2015;16:477.
- Wang J, Wang HY, Zhao PM, Han LB, Jiao GL, Zheng YY, Huang SJ, Xia GX. Overexpression of a profilin (GhPFN2) promotes the progression of developmental phases in cotton fibers. *Plant Cell Physiol.* 2010;51(8):1276–1290.
- Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet.* 2012;44(10):1098–1103.
- Wang L, Wang G, Long L, Altunok S, Feng Z, Wang D, Khawar KM, Mujtaba M. Understanding the role of phytohormones in cotton fiber development through omic approaches; recent advances and future directions. *Int J Biol Macromol.* 2020;163:1301–1313.
- Wang M, Tu L, Yuan D, Zhu Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z, et al. Reference genome sequences of two cultivated allotetraploid cottons. *Nat Genet.* 2019;51(2):224–229.
- Wang X, Zheng M, Liu H, Zhang L, Chen F, Zhang W, Fan S, Peng M, Hu M, Wang H, et al. Fine-mapping and transcriptome analysis of a candidate gene controlling plant height in *Brassica napus* L. *Biotechnol Biofuels.* 2020;13:42.
- Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* 2006;34(Web Server Issue):W720–W724.

- Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li C-Y, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011; 39(Web Server Issue):W316–W22.
- Ye J, Yang Y, Chen B, Shi J, Luo M, Zhan J, Wang X, Liu G, Wang H. An integrated analysis of QTL mapping and RNA sequencing provides further insights and promising candidates for pod number variation in rapeseed (*Brassica napus* L.). *BMC Genomics.* 2017; 18(1):71.
- Yoo M-J, Wendel JF. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genet.* 2014;10(1):e1004073.
- Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, Chen L, He Y, Zhang L, Zhu L, et al. The genome sequence of sea-island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. *Sci Rep.* 2015; 5:17662.
- Zhang K, Kuruparth V, Fang H, Zhu L, Sood S, Jones DC. High-density linkage map construction and QTL analyses for fiber quality, yield and morphological traits using CottonSNP63K array in upland cotton (*Gossypium hirsutum* L.). *BMC Genomics.* 2019;20(1): 889.
- Zhang W, Zhang H, Liu K, Jian G, Qi F, Si N. Large-scale identification of *Gossypium hirsutum* genes associated with *Verticillium dahliae* by comparative transcriptomic and reverse genetics analysis. *PLoS One.* 2017;12(8):e0181609. e0181609.
- Zhang Z, Li JW, Muhammad J, Cai J, Jia F, Shi YZ, Gong JW, Shang HH, Liu AY, Chen TT, et al. 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.* 2015;10(8):e0135430.
- Zhao PM, Wang LL, Han LB, Wang J, Yao Y, Wang HY, Du XM, Luo YM, Xia GX. Proteomic identification of differentially expressed proteins in the Ligon lintless mutant of upland cotton (*Gossypium hirsutum* L.). *J Proteome Res.* 2010;2:1076–1087.
- Zhao W, Dong H, Zhou Z, Wang Y, Hu W. Potassium (K) application alleviates the negative effect of drought on cotton fiber strength by sustaining higher sucrose content and carbohydrates conversion rate. *Plant Physiol Biochem.* 2020;157:105–113.
- Zhao X, Bai S, Li L, Han X, Li J, Zhu Y, Fang Y, Zhang D, Li S. Comparative transcriptome analysis of two *Aegilops tauschii* with contrasting drought tolerance by RNA-Seq. *Int J Mol Sci.* 2020; 21(10):3595.
- Zhong R, Ye Z-H. Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant Cell Physiol.* 2015; 56(2):195–214.
- Zhu Q, Gao S, Zhang W. Identification of key transcription factors related to bacterial spot resistance in pepper through regulatory network analyses. *Genes.* 2021;12(9):1351.
- Zhu QH, Stiller W, Moncuquet P, Gordon S, Yuan Y, Barnes S, W, I. Genetic mapping and transcriptomic characterization of a new Fuzzless-tufted cottonseed mutant. *G3 (Bethesda).* 2021;11. doi: [10.1093/g3journal/jkaa042](https://doi.org/10.1093/g3journal/jkaa042).
- Zou X, Liu A, Zhang Z, Ge Q, Fan S, Gong W, Li J, Gong J, Shi Y, Tian B, et al. Co-expression network analysis and hub gene selection for high-quality fiber in upland cotton (*Gossypium hirsutum*) using RNA sequencing analysis. *Genes.* 2019;10(2):119.

Communicating editor: J. Udall