Expression profile analysis of cotton fiber secondary cell wall thickening stage

Li Liu Equal first author, 1 , **Corrinne Grover** Equal first author, 2 , **Xianhui Kong** 1 , **Josef Jareczek** 2 , **Xuwen Wang** 1 , **Aijun Si** 1 , **Juan Wang** 1 , **Yu Yu** Corresp., 1 , **Zhiwen Chen** Corresp. 3

[1] Cotton Institute, Xinjiang Academy of Agricultural and Reclamation Science, Xinjiang, China

[2] Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, USA

[3] Shanxi Datong University, Datong, China

Corresponding Authors: Yu Yu, Zhiwen Chen
Email address: xjyuyu021@sohu.com, chenzw@yazhoulab.com

To determine the genes associated with the fiber strength trait in cotton, three different cotton cultivars were selected: Sea Island cotton (Xinhai 32, with hyper-long fibers labeled as HL), and upland cotton (17-24, with long fibers labeled as L, and 62-33, with short fibers labeled as S). These cultivars were chosen to assess fiber samples with varying qualities. RNA-seq technology was used to analyze the expression profiles of cotton fibers at the secondary wall thickening stage (20, 25, and 30 DPA). The results showed that a large number of differentially expressed genes (DEGs) were obtained from the three assessed cotton cultivars at different stages of secondary wall development. For instance, at 20 DPA, Sea Island cotton (HL) had 6,215 and 5,364 DEGs compared to upland cotton 17-24 (L) and 62-33 (S), respectively. Meanwhile, there were 1,236 DEGs between two upland cotton cultivars, 17-24 (L) and 62-33 (S). GO function annotation identified 42 functions, including 20 biological processes, 11 cellular components, and 11 molecular functions. KEGG enrichment pathway analysis identified several pathways involved in secondary wall synthesis and thickening, such as glycolysis/gluconeogenesis, galactose metabolism, propanoate metabolism, biosynthesis of unsaturated fatty acids pathway, valine, leucine and isoleucine degradation, fatty acid elongation pathways, and plant hormone signal transduction. Through the identification of shared DEGs, 46 DEGs were found to exhibit considerable expressional differences at different fiber stages from the three cotton cultivars. These shared DEGs have functions including REDOX enzymes, binding proteins, hydrolases (such as GDSL thioesterase), transferases, metalloproteins (cytochromatin-like genes), kinases, carbohydrates, and transcription factors (MYB and WRKY). Therefore, RT-qPCR was performed to verify the expression levels of nine of the 46 identified DEGs, an approach which demonstrated the reliability of RNA-Seq data. Our results provided valuable molecular resources for clarifying the cell biology of secondary cell wall

biosynthesis during fiber development in cotton.

Expression profile analysis of cotton fiber secondary cell wall thickening stage

Li Liu[1#], Corrinne E. Grover[2#], Xianhui Kong[1], Josef Jareczek[2], Xuwen Wang[1], Aijun Si[1], Juan Wang[1], Yu Yu[1,*], and Zhiwen Chen[3,*]

[1]Cotton Institute, Xinjiang Academy of Agricultural and Reclamation Science/Northwest Inland Region Key Laboratory of Cotton Biology and Genetic Breeding, Xinjiang, 832000, China;

[2]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA, USA

[3]Engineering Research Center of Coal-based Ecological Carbon Sequestration Technology of the Ministry of Education, Key Laboratory of Graphene Forestry Application of National Forest and Grass Administration, Shanxi Datong University, Datong, 037009, China;

[#]These authors contributed equally to this work.

Email addresses: cottonliuli@sina.com (L. Liu), corrinne@iastate.edu (C.E. Grover), kxh920@sohu.com (X.H. Kong), jjareczek01@bellarmine.edu (J. Jareczek), wxw629@163.com (X.W. Wang), siaijun1002@163.com (A.J.Si); cottonwj@126.com (J. Wang), xjyuyu021@sohu.com (Y. Yu), chenzhiwen@sxdtdx.wecom.work (Z.W. Chen)

*Correspondence: xjyuyu021@sohu.com (Y.Y.); chenzhiwen@sxdtdx.wecom.work (Z.C.)

Running title: Cotton fiber cell expression profile

Abstract

To determine the genes associated with the fiber strength trait in cotton, three different cotton cultivars were selected: Sea Island cotton (Xinhai 32, with hyper-long fibers labeled as HL), and upland cotton (17-24, with long fibers labeled as L, and 62-33, with short fibers labeled as S). These cultivars were chosen to assess fiber samples with varying qualities. RNA-seq technology was used to analyze the expression profiles of cotton fibers at the secondary wall thickening stage (20, 25, and 30 DPA). The results showed that a large number of differentially expressed genes (DEGs) were obtained from the three assessed cotton cultivars at different stages of secondary wall development. For instance, at 20 DPA, Sea Island cotton (HL) had 6,215 and 5,364 DEGs compared to upland cotton 17-24 (L) and 62-33 (S), respectively. Meanwhile, there were 1,236 DEGs between two upland cotton cultivars, 17-24 (L) and 62-33 (S). GO function annotation identified 42 functions, including 20 biological processes, 11 cellular components, and 11 molecular functions. KEGG enrichment pathway analysis identified several pathways involved in secondary wall synthesis and thickening, such as glycolysis/gluconeogenesis, galactose metabolism, propanoate metabolism, biosynthesis of unsaturated fatty acids pathway, valine, leucine and isoleucine degradation, fatty acid elongation pathways, and plant hormone signal transduction. Through the identification of shared DEGs, 46 DEGs were found to exhibit considerable expressional differences at different fiber stages from the three cotton cultivars. These shared DEGs have functions including REDOX enzymes, binding proteins, hydrolases (such as GDSL thioesterase), transferases, metalloproteins (cytochromatin-like genes), kinases, carbohydrates, and transcription factors (MYB and WRKY). Therefore, RT-qPCR was performed to verify the expression levels of nine of the 46 identified DEGs, an approach which demonstrated the reliability of RNA-Seq data. Our results provided valuable molecular resources for clarifying the cell biology of secondary cell wall biosynthesis during fiber development in cotton.

Key words: Upland cotton, Sea Island cotton, secondary cell wall thickening, fiber quality, gene expression patterns

Introduction

Cotton is one of the seven major crops and is an essential component of the textile industry with its fiber as the primary product of cotton production (Chen et al. 2017; Huang et al. 2021a). Cotton is a model plant for studying cellulose synthesis and cell elongation (Cao et al. 2020b; Glover 2000; Guan et al. 2007). The cotton fiber is a single cell that differentiates from the epidermal cells in the outer integument of the ovule (Shan et al. 2014). Cotton fiber development comprises four stages: initiation, elongation, secondary wall thickening, and maturation (Huang et al. 2021a; Wen et al. 2023). The fiber elongation stage varies in length among cultivars and can last between 15 and 25 days post-anthesis (DPA), depending on the species and domestication status. Elongation overlaps with the transition stage (16-20 DPA) and, in domesticated cultivars, it also overlaps the beginning of the secondary wall thickening stage (20 through 40 DPA) (Mansoor & Paterson 2012). Fiber maturity occurs between 40 and 50 DPA, when the fiber cells die and the cytoplasm degrades, leaving behind hollow cells surrounded by cellulose (Jareczek et al. 2023; Kim 2018). The cross-section of the fibers shows the primary cell wall, secondary cell wall, luminal wall, and middle luminal wall. During secondary wall thickening, a helical pattern of cellulose fibers is laid down, resulting in mature fibers appearing flat and banded with a natural twist (Mansoor & Paterson 2012).

Cotton fiber properties are influenced differently during various stages of its growth. The differentiation of cotton fiber from the ovular epidermis happens during initiation (around -1 to 1 DPA) when approximately 20% to 30% of epidermal cells differentiate into fiber cells. During this stage, the fiber tips are refined, which is strongly linked with mature diameter and strength (Kelly et al. 2015). The elongation stage, on the other hand, tightly correlates with fiber length. While elongation lasts from ~3 to 20 DPA, 5 to 15 DPA comprises the most rapid elongation period, when cotton leverages fatty acids and carbohydrates to keep the primary cell wall pliable for extreme linear growth (Tian & Zhang 2021). At ~16 DPA, the fiber enters the transition stage, where the microtubules in the fiber shift to a shallow helical angle, and the fiber lays down the winding cell wall layer (Hsieh et al. 1995; Meinert & Delmer 1977). The winding cell wall layer is similar in composition to the primary cell wall, with a slight increase in cellulose content, and it is thought to impact both fiber strength and flexibility (Haigler et al. 2009; Tuttle et al. 2015; Zhang et al. 2021b). The thickening period of the secondary wall begins during the transition stage (~15DPA) when cellulose production increases substantially. The secondary wall synthesis stage starts at ~20 DPA, and it is characterized by β-1, 4-glucan chains that accumulate to facilitate cellulose accumulation and form 20-30 layers of "growing day rings". This period mainly determines the thickness and strength of the cell wall (Haigler et al. 2012). When the cell wall thickens to 3-4 microns, the cells begin to dehydrate and undergo apoptosis, and the whole fiber cells display a twisted spiral state (Hof & Saha 1997). The natural twist of cotton fiber can increase the binding force between fibers and improve the yarn strength when spinning. Mature cotton fibers contain more than 95% cellulose and less keratin, wax, inorganic matter, and other classes of protein than other plant cells (Liu 2013). Cotton fiber quality is based on several properties, such as fiber fineness, length, strength, micronaire (i.e., cell wall thickness), and yellowness (Bajwa et al. 2015; Song et al. 2021; Yang et al. 2022). Increased demand for luxury textiles has likewise increased the demand for high-quality cotton fiber; consequently, there is also increased interest in improving fiber quality in the more highly productive species/cultivars (Gao et al. 2021; Huang et al. 2021a). While breeding programs are attempting to introgress desirable fiber quality traits into these productive lines, like most crops, cotton fiber yield and quality traits are quantitatively controlled by multiple genes, limiting the success of traditional breeding techniques (Liu et al. 2023; Xu et al. 2019). Therefore, understanding the physiological and molecular basis of fiber development is paramount to improving cotton fiber quality through other techniques, such as molecular design breeding.

In recent years, the molecular mechanisms underlying cotton fiber development have been studied in depth, and a series of important advances have been made using transcriptome analysis and other high-throughput based methods (Li et al. 2022; Zang et al. 2022; Zhang et al. 2022). These studies, however, mainly focus on the initiation and elongation stages of fiber cell development (Qin et al. 2019), and the molecular mechanisms underlying fiber secondary wall synthesis and thickening (and, therefore, strength) have been rarely studied. To improve our understanding of the molecular mechanisms operating during secondary wall synthesis and their influence on fiber quality (strength), two cultivars of *Gossypium hirsutum* and one cultivar of *G. barbadense* with known differences in fiber quality were selected for comparative transcriptome analysis, and differentially expressed genes (DEGs) from the secondary wall thickening stage were comprehensively identified using RNA-seq from three timepoints. Through pairwise comparison, common enrichment pathways were identified among the DEGs in the different cultivars, and important candidate genes related to cotton fiber development were screened at different developmental stages. Our results provided a solid foundation for the analysis of the molecular mechanism of cotton fiber secondary wall development, which would be helpful for the mining and utilization of valuable gene resources.

Materials & Methods

Plant Materials

Cotton cultivars, Xinhai 32 is a high-generation inbred line of Sea Island cotton and 17-24 or 62-33 are high-generation inbred lines of Upland cotton. All of these cultivars were bred by the Cotton Institute, Xinjiang Academy of Agricultural and Reclamation Science. Xinhai 32, 17-24, and 62-33 were referred to as HL, L, and S, respectively. These three cotton cultivars were planted in the field of the Xinjiang Academy of Agricultural and Reclamation Science. Ovules (seeds) were harvested at the indicated DPA (days post-anthesis). Cotton bolls of 20 DPA, 25 DPA, and 30 DPA from the three different cotton cultivars were sampled at 10:00 AM. Five cotton bolls were selected from each plant, and the fibers were isolated from the ovules by scratching the ovule with a metal strainer in liquid nitrogen. The fiber samples were then quickly ground into powder and stored in the ultra-low temperature refrigerator at -80℃.

Fiber traits and phenotypic evaluation

Boll weight, 100 seed weight, lint index, and seed index were weighed by an analytical balance (0.0001g, BSA224S, SARTORIUS, Germany). Fiber quality traits, including the fiber length (mm), fiber uniformity ratio (%), fiber strength (cN/tex), fiber elongation, and micronaire, were measured with an HVI 900 instrument (USTER HVISPECTRUM, SPINLAB, USA) at the Cotton Fiber Quality Inspection and Test Center of Ministry of Agriculture (Anyang, China) (Shang et al. 2015). Bolls were collected for seed index analysis. For this purpose, one hundred seeds from each cultivar were randomly selected and weighed as seed index (SI, g) (Liu et al.

2023; Shang et al. 2016). To measure oil content, seeds were delinted with concentrated sulphuric acid. The oil contents of the three cotton cultivars were measured at different stages of ovular development using the Soxhlet extraction method (García-Ayuso et al. 2000). To record grain weight, one hundred cotton ovules were randomly weighed.

RNA extraction, library construction, and sequencing

Total RNA from 20 DPA, 25 DPA, and 30 DPA cotton fiber from each of the three different cotton cultivars was extracted using the RNAprep pure plant kit (TIANGEN, China) according to the manufacturer's instructions. A total of 1 μg purified mRNA was selected for cDNA library construction following a previous report (Chen et al. 2021). Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. The cDNA fragments of 240 bp length were selected, and the library fragments were purified using the AMPure XP system (Beckman Coulter, Beverly, USA). The PCR products were purified using the AMPure XP system, and the library quality was assessed on an Agilent Bioanalyzer 2100 system. After generating clusters, the libraries were sequenced using the Illumina HiSeqTM 2500 platform as 150 bp paired-end reads. Three biological replicates were performed for the nine samples.

RNA-seq reads quality control, mapping, and differentially expressed genes (DEGs) analysis

The FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was used to process the raw reads in fastq format according to a previous study (Chen et al. 2021). Clean data (clean reads) were obtained by removing reads containing adapters, reads containing poly-N sequences, and low-quality reads from the raw data. All subsequent analyses were performed based on the clean high-quality data. Among the nine fiber samples from *G. hirsutum* and *G. barbadense* (Hu et al. 2019; Wang et al. 2019), RNA-seq data were mapped to their respective reference genomes using HISAT2 software (Kim et al. 2015; Pertea et al. 2016). Reads with at most one mismatch were used to calculate the expression levels of genes. Gene expression values were calculated following the method of the previous studies (Chen et al. 2022; Chen et al. 2021). Differential expression analysis of the two groups was performed using DESeq2 and presented using fragments per kilobase of transcript per million fragments mapped (FPKM) (Love et al. 2014). The resulting *p* values were adjusted using Benjamini and Hochberg's approach for controlling the false discovery rate (FDR). Genes with an adjusted *p* value < 0.01 and twofold change (up and down) were defined as differentially expressed. TBtools (Chen et al. 2020) was used to display the gene expression patterns of the FPKM values. Clean data were available from the Genome Sequence Archive in the BIG Data Center of Sciences (https://bigd.big.ac.cn/) under accession number CRA009299. The statistical power of this experimental design, calculated in RNASeqPower is 0.84.

Gene functional annotation and enrichment analyses

The functions of differentially expressed genes were annotated using the following databases: Nr (NCBI nonredundant protein sequences, ftp://ftp.ncbi.nih.gov/blast/db/), Gene Ontology (GO) (Gene Ontology, http://www.geneontology.org/), and Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/). To analyze the enriched Gene Ontology (GO) of the DEGs, we used the GOseq R package based on Wallenius noncentral hypergeometric distribution (Young et al. 2010). KOBAS was employed to assess the statistical enrichment of the DEGs in the KEGG pathways (Mao et al. 2005).

RNA extraction, cDNA synthesis, and RT-qPCR expression analyses

These experiments were conducted according to the methods reported previously (Cao et al. 2020a; Cao et al. 2022; Cui et al. 2022; Zhang et al. 2021a). In brief, total RNAs from 20 DPA, 25 DPA, and 30 DPA fiber samples of the three different cotton cultivars were extracted using the RNAprep pure plant kit (TIANGEN, China). DNase I treatment was applied to the RNAs before synthesizing cDNA using TransScript® First-Strand cDNA Synthesis SuperMix from TransGen Biotech, China, and the resulting products were diluted fivefold before use. Specific forward and reverse gene primers (Table S1) were designed using Primer v5.0 software for real-time quantitative PCR, which was performed using SYBR-Green PCR Mastermix (TaKaRa) on a cycler (Mastercycler RealPlex; Eppendorf Ltd., China). The *G. hirsutum* and *G. barbadense* histone-3 (*GhHIS3* and *GbHIS3*) genes were used as internal references, and the relative amount of amplified product was calculated following the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen 2001).

Statistical analysis

The R package available at https://www.r-project.org/ was utilized for the analysis of variance and Student's t-test. The Shapiro-Wilk test was utilized to test for normality, confirming that the data followed a Gaussian distribution. The least significant difference (LSD) was used to test for significance at either the 1% or 5% levels. The analysis included at least three biological replicates for each sample.

Results

Physiological traits differences among the *G. barbadense* (Xinhai 32, HL) and *G. hirsutum* (17-24, L and 62-33, S) cotton cultivars

Physiological traits of the three cotton cultivars, all high-generation inbred lines developed by the Cotton Institute, Xinjiang Academy Agricultural and Reclamation Science, were characterized. Table 1 shows the main properties of these cultivars, which we have designated as S (cv 62-33), L (cv 17-24), and HL (Xinhai 32). In comparison to the upland cotton cultivars (L and S), the Sea Island cotton (HL) exhibited longer fiber, greater fiber strength, increased oil content, a higher seed index, greater fiber uniformity, and greater fiber elongation. In contrast, the upland cotton cultivars (L and S) had significantly higher boll weight, lint percentage, lint index, and micronaire than Xinhai 32 (HL). Notably, these properties, although greater in the *G. hirsutum* cultivars, are outside of the optimal range. In particular, micronaire (one of the most important measures of cotton fiber quality) was considered grade A (micronaire range: 3.7 - 4.2) in Xinhai 32, compared the B grade (micronaire range: 3.5 - 3.6 and 4.3 - 4.9) observed in the upland cotton cultivars, L and S. Between the two cultivars of upland cotton, fiber length, fiber strength, and oil content also varied, with cv 17-24 (L) exhibiting significantly higher values than cv 62-33 (S). These results indicate that major differences in agronomically important fiber properties exist between Sea Island cotton and Upland cotton, as expected, but also between the two upland cotton cultivars, with cv 17-24 (L) exhibiting better agricultural performance.

Transcriptome Data Generation of 20, 25, and 30 DPA fibers of Xinhai 32 (HL), 17-24 (L), and 62-33 (S) Cotton Cultivars

As these three cultivars showed great variance in fiber length and strength, especially fiber strength (Table 1), we evaluated gene expression at three timepoints during secondary wall synthesis (i.e., 20, 25, and 30 DPA fiber) from Xinhai 32 (HL), 17-24 (L), and 62-33 (S). RNA from at least three biological replicates was pooled at each timepoint for each accession, hereafter referred to as HL20, HL25, HL30, L20, L25, L30, S20, S25, and S30, and over 108 million high-quality reads were generated using the Illumina NovaSeq 6000 sequencing platform (Table 2). The HISAT2 software was used to align these clean reads to the reference cotton genomes (Hu et al. 2019; Wang et al. 2019) with at most one base mismatch. The ratio of mapped reads ranged from 72% in sample L30 to 77% in sample HL20 (Table 2), and the number of uniquely mapped clean reads ranged from 53% in sample L30 to 70% in sample L20. These data indicated that the RNA-seq data in this study were reliable for the subsequent analyses.

Identification of differentially expressed genes (DEGs)

Differential gene expression was surveyed for the nine cotton fiber samples. Eighteen comparisons among these nine samples were performed to capture expression differences between samples at a given DPA and within a single sample among measured timepoints (Fig. 1A). In each comparison, the number of differentially expressed genes (DEGs) varied between 761 genes in HL25-VS-L25 and 10,464 genes in HL20-VS-HL25 (Fig. 1A). It is worth noting that upregulation was more frequent than downregulation in most comparisons (Fig. 1A). Because these cultivars produce fibers with different properties, we considered the overlap in gene expression at each DPA to identify genes important for secondary cell wall biosynthesis that differed in expression among the three species/cultivars. Venn diagrams were constructed for the DEGs from each species/cultivar comparison at each surveyed stage (i.e., 20, 25, and 30

DPA stages; Fig. 1). As expected, the intraspecies cultivar comparisons (i.e., L versus S) typically exhibited fewer uniquely DEGs than the interspecies comparisons (i.e., HL versus S or L), with the exception of 25 DPA (Fig. 1). Although 25 DPA exhibited the fewest DEGs overall (Fig. 1), the intraspecies (S versus L) comparison resulted in 25 - 40% more DEGs than either interspecific comparison (Fig. 1); the fewest DEGs at this stage were between the HL and L cultivars, perhaps indicating that underlying fiber length expression differences are largely responsible for the differences in gene expression at this stage. The results showed that 423 DEGs were shared by all species/cultivar comparisons at the 20 DPA stage (Fig. 1B), possibly indicating genes that underlie the differences among cultivars. Far fewer DEGs were observed at 25 DPA (39 DEGs), although this number is commensurate with the general reduction in DEGs among cultivars at this stage (Fig. 1C). At 30 DPA, 361 DEGs were shared among the cultivar comparisons (Fig. 1D).

GO analysis for DEGs

We considered the possible biological functions of these DEGs between Sea Island and/or the Upland cottons using GO category enrichment (Fig. S1). Results from the three categories (i.e., biological process, cellular component, and molecular function), suggest enrichment of 20, 11, and 11 functional categories in the 20, 25, and 30 DPA fiber comparisons, respectively. GO terms associated with important biological processes included metabolic, cellular, developmental, and single-organism processes, biological regulation, response to stimulus, and signaling. Cellular components, such as cell, cell part, membrane, membrane part, organelle and organelle parts were enriched. Molecular function enrichment consisted of catalytic activity, transporter activity, binding, nucleic acid-binding transcription factor activity, antioxidant activity, and receptor activity.

To compare the difference between the three cultivars, the GO enrichment of functional categories in the 25 DPA fiber was analyzed (Fig. 2). The results showed that the top three GO category enrichments in biological processes among the three groups of comparisons (HL-25-vs-L-25, HL-25-vs-S-25, L-25-vs-S-25) were cellular process, metabolic process, and localization, respectively (Figs. 2A-2C). Additionally, the cellular components, cell, cell part and membrane accounted for the largest proportion in all three groups (Figs. 2A-2C). The most enriched molecular functions all consisted of catalytic activity, binding, and transporter activity between the comparisons of the three cultivars (Figs. 2A-2C). These results indicated that the GO enrichment analysis was not specific to the three cultivars.

Pathway enrichment of DEGs from the three cotton cultivars at different stages of fiber development (20 DPA, 25 DPA, and 30 DPA)

To investigate the biological functions of these DEGs during fiber development in the three cotton cultivars, we performed KEGG pathway enrichment analysis for the DEGs. At the 20 DPA fiber stage, the DEGs were assigned to 126 KEGG pathways according to the functional categorization. For the interspecific comparison HL-20-vs-L-20, the top 20 KEGG pathways of enriched DEGs were categorized into the following functional pathways (Fig. 3A). While the greatest number of DEGs mostly belonged to the plant hormone signal transduction pathways, there was also involvement of several other pathways whose RichFactor was closer to 50%, such as fatty acid elongation. Other pathways with notable enrichment are: i) carbohydrate metabolism: glycolysis/gluconeogenesis, pyruvate metabolism, galactose metabolism, fructose and mannose metabolism; ii) fatty acid metabolism: fatty acid elongation, and biosynthesis of unsaturated fatty acids; and iii) amino acid metabolism: valine, leucine and isoleucine degradation, lysine biosynthesis, and cysteine and methionine metabolism. For the other interspecific comparison, HL-20-vs-S-20, the top 20 KEGG pathways of enriched DEGs were involved in glycolysis/gluconeogenesis, galactose metabolism, fructose and mannose metabolism, fatty acid elongation, biosynthesis of unsaturated fatty acids, and leucine and isoleucine degradation (Fig. 3B); however, the greatest representation of genes was for RNA transport and the ribosome. In this comparison, the RichFactor varied more narrowly (~12%, versus 25% in HL20 versus L20). For the intraspecific comparison L-20-vs-S-20, the top 20 KEGG pathways of enriched DEGs mainly included the glycolysis/gluconeogenesis, galactose metabolism, fatty acid elongation, biosynthesis of unsaturated fatty acids, plant hormone signal transduction, and biosynthesis of secondary metabolites (Fig. 3C). The greatest numbers of genes were found in the metabolic pathways and biosynthesis of secondary metabolites. While the RichFactor for these was low, it is worth noting that the RichFactor was generally low for all pathways, possibly indicating the recruitment of few genes/pathways from these somewhat broad categories. These results suggested that the differences in the 20 DPA fiber samples between Sea Island cotton (Xinhai 32) and Upland cotton (17-24 or 62-33) were mainly caused by genes involved in the metabolic pathways of carbohydrates, fatty acids, and amino acids. However, differences in the plant hormone pathways were common to both comparisons involving L-20, and differences in secondary metabolite production were observed between the two cultivars of Upland cotton.

Inter-cultivar comparisons revealed that the 25 DPA fiber stage mainly exhibited enriched KEGG pathways for DEGs related to starch and sucrose metabolism, fatty acid elongation, and biosynthesis of secondary metabolites (Figs. 3D-3F), which could indicate that secondary metabolites, such as the biosynthesis of cellulose, was different among the three cultivars. For the fiber samples of 30 DPA, the enrichment classification of DEGs among the three cultivars did not show strong regularity, indicating the complexity of the genotypes among different cotton cultivars (Figs. 3G-3I).

Annotations of DEGs from different fiber development stages

To eliminate the effect of genotypic differences, KEGG pathway enrichment was performed on DEGs between different stages of fiber development (20 DPA, 25 DPA, and 30 DPA) within the same cotton species or cultivar (Fig. 4). The shared top 20 KEGG pathways of enriched DEGs from Xinhai 32 (HL) were mainly categorized into the following functional pathways: pyruvate metabolism, fatty acid elongation, biosynthesis of unsaturated fatty acids, valine, leucine and isoleucine degradation, as well as plant hormone signal transduction (Figs. 4A-4C). As to the upland cotton 17-24 (L) (Figs. 4D-4F) and 62-33 (S) (Figs. 4G-4I), the enriched DEGs were also mainly categorized into the fatty acid elongation, biosynthesis of unsaturated fatty acids, valine, leucine and isoleucine degradation, circadian rhythm as well as plant hormone signal transduction pathways. Consistent with the previous results, these results further supported that DEGs enriched in the pathways of carbohydrates, fatty acids, amino acids, and hormones were involved in the elongation of fiber cells or thickening of fiber secondary cell wall of different cotton species at 20-30 DPA stages.

Identification of shared DEGs from different fiber development stages in the three cotton cultivars

To further identify candidate genes contributing to the biosynthesis and thickening of cotton fiber secondary wall, the shared DEGs from different fiber development stages in the three cotton cultivars were evaluated. As can be seen from Fig. S2, there were 46 DEGs shared by the three cotton cultivars, Xinhai 32 (HL), 17-24 (L), and 62-33 (S) at different fiber development stages (20, 25, and 30DPA). These 46 recurring DEGs may be common to the fiber developmental pathway, regardless of the cultivar, perhaps suggesting they could be key genes in the regulation of cotton fiber cell elongation or secondary cell wall thickening. Their main functions include REDOX enzymes, binding proteins, hydrolases (such as GDSL thioesterase), transferases, metalloproteins (cytochromatin-like genes), kinases, carbohydrates, and transcription factors (MYB and WRKY).

Detailed analyses were conducted on these DEGs to assess their expression patterns across the nine samples (Fig. 5). In general, the DEGs can be classified into two categories (Fig. 5): low-to-high expression and high-to-low expression. Most of the DEGs (30) showed low-to-high expression, starting relatively low at 20 DPA but increasing in expression at 25 or 30 DPA in each of the three different species or cultivars (Figs. 5A, 5B). Importantly, the expression levels of these 30 DEGs in HL or L were significantly higher than that in S at 25 or 30 DPA (Figs. 5A, 5B), suggesting that these genes might play important roles in the secondary wall synthesis and thickening of fiber tissues. In contrast, nine of the remaining DEGs exhibited high-to-low expression trends, exhibiting the greatest expression at 20 DPA but reducing expression by 25 or 30 DPA (Figs. 5C, 5D). These results suggest that these genes may function in the early stages of fiber development and are downregulated as the cell commits to focused secondary wall synthesis.

Validation of candidate DEGs by RT-qPCR

We have confirmed the accuracy of these candidate gene expression profiles for nine of the differentially expressed genes (i.e., Polyphenol oxidase 9, 3-ketoacyl-CoA synthase 3, NAD(P)-binding Rossmann-fold superfamily protein, GDSL-like Lipase/Acylhydrolase superfamily protein, Caffeic acid O-methyltransferase 1, WRKY transcription factor 32, 2,4-dienoyl-CoA reductase, WRKY transcription factor 103 and Fatty acid desaturase 6) using RT-qPCR (Fig. 6). The results showed that eight candidate genes have low-to-high expression trends, presenting high levels of expression at 25 or 30 DPA compared with 20 DPA in the three different one or the other (Figs. 6A-6H). One candidate gene (3-ketoacyl-CoA synthase 3) exhibited a high-to-low expression pattern with high expression levels at 20 DPA but low expression levels at 25 or 30 DPA (Fig. 6I). Overall, these transcripts of nine genes exhibited similar expression patterns between the RT-qPCR and RNA-seq experiments (Fig. S3), and according to the results of correlation analysis (Table S2), the correlation coefficient between the RT-qPCR and RNA-seq data of nine genes ranged from 0.8562 to 0.9997. These results indicated that the RT-qPCR validation was congruent with the RNA-seq data. Both data proved that the expression profiles of 2,4-dienoyl-CoA reductase, WRKY103, and fatty acid desaturase 6 genes in HL or L were significantly higher than that in S at 25 or 30 DPA (Figs. 6F-6H), suggesting that the diverged expression patterns of these genes may be the cause of the variance in fiber strength between the three cultivars.

Discussion

Cotton fiber cell development is a complex morphogenic process regulated by the timely expression of multiple genes. Early research suggested that the timing of different fiber developmental stages is inconsistent between Sea Island and upland cotton species, specifically in the extent to which the elongation stage of fiber cell development overlaps with the secondary wall thickening stage (Zang et al. 2022; Zhang et al. 2022). In this study, we evaluated 20, 25, and 30 DPA fibers from Sea Island and upland cotton cultivars, representing developmental stages of the secondary wall biosynthesis or thickening. It is worth noting that pathways related to fatty acid metabolism were highly enriched, such as the fatty acid carbon chain extension pathway, propionic acid metabolic pathway, and unsaturated fatty acid biosynthesis pathway. Their differential expression during the synthesis process of the fiber secondary wall suggested that fatty acid metabolism was closely related to fiber quality.

Further, a total of 46 genes were screened as candidates that were commonly differentially expressed during the different development stages from the three cotton cultivars, including cytochrome P450 enzyme gene, glycohydrolase and glycosyltransferase, binding protein gene, WRKY transcription factor (Fig. 5). Plant cytochrome P450s are involved in the biosynthetic pathways of fatty acid hydroxylation, epoxidation and cleavage of hydrogen peroxide functional groups of unsaturated fatty acids (Davidson et al. 2006). Previous reports have reported that ethylene biosynthesis, cytoskeleton, signaling pathway, fatty acid biosynthesis and fatty acid carbon chain extension pathway (Shi et al. 2006) were significantly up-regulated during fiber development (Gou et al. 2007; Ruan et al. 2004). These results indicated that lipid metabolism was significantly correlated with fiber development.

It is well-known that the MYB transcription factor was confirmed to promote the secondary cell-wall biosynthesis (Xiao et al. 2021). GhMYBL1, an R2R3-MYB transcription factor, was specifically expressed at the stage of secondary wall deposition in cotton fibers and participated in modulating the process of secondary wall biosynthesis (Sun et al. 2015). Additionally, GhMYB7 has been shown to regulate the biosynthesis of secondary cell walls both in *Arabidopsis thaliana* and upland cotton (Huang et al. 2021b; Huang et al. 2016). In our study, a MYB transcription factor was identified in the highly expressed genes of the three cultivars at 25 and 30 DPA fiber, and may be involved in the secondary cell wall (SCW) cellulose biosynthesis. WRKY members have also been reported to be involved in fiber development, such as GhWRKY16, and GhWRKY53 (Wang et al. 2021; Yang et al. 2021). Two WRKY genes

(WRKY32 and 103) were also detected in our data, and their potential biological regulations in SCW biosynthesis of fiber cells were still not clear, which is worth further research to explore their functions. Thus, we speculated that one factor that cause the difference in fiber strength of the three cultivars might be the distinct expression patterns of genes related to the SCW biosynthesis.

Many other important genes were also identified to be specifically differentially expressed in cotton fibers, such as the cotton sucrose synthetase gene (Zhang et al. 2017), transcription factor GhMYB2 (Wang et al. 2004), cytoskeletal proteins GhTUB1 and GhACT1, which have been shown to participate in the elongation process of fiber cells (Li et al. 2005). GhACT1 and GhTUB1, which encode actin and tubulin, also affect cytoskeleton assembly and fiber elongation (Li et al. 2002; Li et al. 2005). LIM domain protein GhWLIM1a can promote secondary wall synthesis by binding to tubulin (Han et al. 2013). Transcription factors were also involved in regulating the secondary wall synthesis of cotton fiber cells. Overexpression of a cotton NAC transcription factor (GhFSN1) resulted in thicker fiber secondary walls but shorter fibers (Zhang et al. 2018). Subsequently, a primary GhTCP4 transcription factor was found to play an important role in balancing cotton fiber cell elongation and secondary wall thickening (Cao et al. 2020b). Transcriptomic and promoter activity analysis showed that GhTCP4 activated GhFSN1 transcription factor and cellulose synthase genes responsible for secondary wall synthesis. The transcriptional activity of GhCESA8 (GhCESA8) accelerated the biosynthetic pathway of the secondary walls of fiber cells, resulting in shorter fibers and thicker cell walls (Cao et al. 2020b). The time-course analysis of fiber samples in *G. hirsutum* and *G. barbadense* cultivars revealed that the glycosyltransferase was involved in the synthesis of glucuronoxylan hemicellulose and cell wall morphogenesis during SCW formation (Zhang et al. 2022). A recent study on the genetic regulation of fiber development in *G. hirsutum* based on 2,215 time-series transcriptomes also revealed that a NAD(P)-linked oxidoreductase protein, was positively correlated with fiber strength development (You et al. 2023). In our study, the NAC52 transcription factor, and NAD(P)-binding Rossmann-fold superfamily protein, have been identified to be involved in the regulation of cotton fiber secondary wall development. These results indicated that multiple pathway-related genes play roles in the biosynthesis of fiber secondary cell walls.

Supplementary Materials: The following supporting information can be downloaded at:

Author contributions

LL, YY, and ZWC conceived and designed this experiment. LL, XHK, XWW, AJS and JW collected samples and performed the study. LL, CEG, JJ, and ZWC participated in the acquisition and analysis of the data. LL and ZWC wrote the manuscript. LL and YY participated in the discussion draft of the manuscript. ZWC revised the final manuscript. All authors read and approved the final manuscript.

Acknowledgments: We are deeply indebted to Professor Lida Zhang for helpful suggestions and comments on bioinformatic analyses, and we thank Professor Yi Huang for valuable comments on previous versions of the manuscript. We are also grateful to two anonymous reviewers for their helpful suggestions and comments.

Abbreviations

DEGs, differentially expressed genes; FPKM, fragments per kilobase of transcript per million mapped fragments; *G. hirsutum*, *Gossypium hirsutum*; *G. barbadense*, *Gossypium barbadense*; *A. thaliana*, *Arabidopsis thaliana*; RT-qPCR, real-time quantitative polymerase chain reaction; SCW, secondary cell wall.

References

Bajwa KS, Shahid AA, Rao AQ, Bashir A, Aftab A, and Husnain T. 2015. Stable transformation and expression of GhEXPA8 fiber expansin gene to improve fiber length and micronaire value in cotton. *Front Plant Sci* 6:838. 10.3389/fpls.2015.00838

Cao J-F, Huang J-Q, Liu X, Huang C-C, Zheng Z-S, Zhang X-F, Shangguan X-X, Wang L-J, Zhang Y-G, Wendel JF, Grover CE, and Chen Z-W. 2020a. Genome-wide characterization of the GRF family and their roles in response to salt stress in Gossypium. *BMC Genomics* 21:575. 10.1186/s12864-020-06986-0

Cao J, Huang C, Liu J, Li C, Liu X, Zheng Z, Hou L, Huang J, Wang L, Zhang Y, Shangguan X, and Chen Z. 2022. Comparative genomics and functional studies of putative m(6)A methyltransferase (METTL) genes in cotton. *Int J Mol Sci* 23. 10.3390/ijms232214111

Cao JF, Zhao B, Huang CC, Chen ZW, Zhao T, Liu HR, Hu GJ, Shangguan XX, Shan CM, Wang LJ, Zhang TZ, Wendel JF, Guan XY, and Chen XY. 2020b. The miR319-targeted GhTCP4 promotes the transition from cell elongation to wall thickening in cotton fiber. *Mol Plant* 13:1063-1077. 10.1016/j.molp.2020.05.006

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, and Xia R. 2020. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant* 13:1194-1202. 10.1016/j.molp.2020.06.009
Chen Z, Zhao J, Qiao J, Li W, Guan Z, Liu Z, Bai X, Xing B, Zhang J, Li J, Yin W, and Zhu H. 2022. Graphene-mediated antioxidant enzyme activity and respiration in plant roots. *ACS Agricultural Science & Technology* 2:646-660. 10.1021/acsagscitech.2c00074
Chen Z, Zhao J, Song J, Han S, Du Y, Qiao Y, Liu Z, Qiao J, Li W, Li J, Wang H, Xing B, and Pan Q. 2021. Influence of graphene on the multiple metabolic pathways of Zea mays roots based on transcriptome analysis. *PLoS One* 16:e0244856. 10.1371/journal.pone.0244856
Chen ZW, Cao JF, Zhang XF, Shangguan XX, Mao YB, Wang LJ, and Chen XY. 2017. Cotton genome: challenge into the polyploidy. *Sci Bull (Beijing)* 62:1622-1623. 10.1016/j.scib.2017.11.022
Cui W, Chen Z, Shangguan X, Li T, Wang L, Xue X, and Cao J. 2022. TRY intron2 determined its expression in inflorescence activated by SPL9 and MADS-box genes in Arabidopsis. *Plant Science* 321:111311. https://doi.org/10.1016/j.plantsci.2022.111311
Davidson SE, Reid JB, and Helliwell CA. 2006. Cytochromes P450 in gibberellin biosynthesis. *Phytochemistry Reviews* 5:405-419. 10.1007/s11101-006-9005-5
Gao Y, Chen Y, Song Z, Zhang J, Lv W, Zhao H, Huo X, Zheng L, Wang F, Zhang J, and Zhang T. 2021. Comparative dynamic transcriptome reveals the delayed secondary-cell-wall thickening results in altered lint percentage and fiber elongation in a chromosomal segment substitution line of cotton (*Gossypium hirsutum* L.). *Front Plant Sci* 12:756434. 10.3389/fpls.2021.756434
García-Ayuso LE, Velasco J, Dobarganes MC, and Luque de Castro MD. 2000. Determination of the oil content of seeds by focused microwave-assisted soxhlet extraction. *Chromatographia* 52:103-108. 10.1007/BF02490801
Glover BJ. 2000. Differentiation in plant epidermal cells. *J Exp Bot* 51:497-505. 10.1093/jexbot/51.344.497
Gou JY, Wang LJ, Chen SP, Hu WL, and Chen XY. 2007. Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Res* 17:422-434. 10.1038/sj.cr.7310150
Guan X, Yu N, Shangguan X, Wang S, Lu S, Wang L-J, and Chen X-Y. 2007. Arabidopsis trichome research sheds light on cotton fiber development mechanisms. *Chinese Science Bulletin* 52:1734-1741.
Haigler CH, Betancur L, Stiff MR, and Tuttle JR. 2012. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Front Plant Sci* 3:104. 10.3389/fpls.2012.00104
Haigler CH, Singh B, Wang G, and Zhang D. 2009. Genomics of cotton fiber secondary wall deposition and cellulose biogenesis. In: Paterson AH, ed. *Genetics and Genomics of Cotton*. New York, NY: Springer US, 385-417.
Han LB, Li YB, Wang HY, Wu XM, Li CL, Luo M, Wu SJ, Kong ZS, Pei Y, Jiao GL, and Xia GX. 2013. The dual functions of WLIM1a in cell elongation and secondary wall formation in developing cotton fibers. *Plant Cell* 25:4421-4438. 10.1105/tpc.113.116970
Hof J, and Saha S. 1997. Cotton fibers can undergo cell division. *Am J Bot* 84:1231.
Hsieh Y-L, Honik E, and Hartzell MM. 1995. A developmental study of single fiber strength: greenhouse grown SJ-2 Acala cotton. *Textile Research Journal* 65:101-112. 10.1177/004051759506500206
Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, Baruch K, Fang D, Liu X, Ruan YL, Rahman MU, Han J, Wang K, Wang Q, Wu H, Mei G, Zang Y, Han Z, Xu C, Shen W, Yang D, Si Z, Dai F, Zou L, Huang F, Bai Y, Zhang Y, Brodt A, Ben-Hamo H, Zhu X, Zhou B, Guan X, Zhu S, Chen X, and Zhang T. 2019. Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of

allotetraploid cotton. *Nat Genet* 51:739-748. 10.1038/s41588-019-0371-5
Huang G, Huang JQ, Chen XY, and Zhu YX. 2021a. Recent advances and future perspectives in cotton research. *Annu Rev Plant Biol* 72:437-462. 10.1146/annurev-arplant-080720-113241
Huang J, Chen F, Guo Y, Gan X, Yang M, Zeng W, Persson S, Li J, and Xu W. 2021b. GhMYB7 promotes secondary wall cellulose deposition in cotton fibres by regulating GhCesA gene expression through three distinct cis-elements. *New Phytol* 232:1718-1737. 10.1111/nph.17612
Huang J, Chen F, Wu S, Li J, and Xu W. 2016. Cotton GhMYB7 is predominantly expressed in developing fibers and regulates secondary cell wall biosynthesis in transgenic Arabidopsis. *Sci China Life Sci* 59:194-205. 10.1007/s11427-015-4991-4
Jareczek JJ, Grover CE, and Wendel JF. 2023. Cotton fiber as a model for understanding shifts in cell development under domestication. *Front Plant Sci* 14:1146802. 10.3389/fpls.2023.1146802
Kelly BR, Abidi N, Ethridge D, and Hequet EF. 2015. Fiber to Fabric.
Kim D, Langmead B, and Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12:357-360. 10.1038/nmeth.3317
Kim HJ. 2018. Cotton fiber biosynthesis. In: Fang DD, ed. *Cotton Fiber: Physics, Chemistry and Biology*. Cham: Springer International Publishing, 133-150.
Li S, Geng S, Pang B, Zhao J, Huang Y, Rui C, Cui J, Jiao Y, Zhang R, and Gao W. 2022. Revealing genetic differences in fiber elongation between the offspring of Sea Island cotton and Upland cotton backcross populations based on transcriptome and weighted gene coexpression networks. *Genes (Basel)* 13. 10.3390/genes13060954
Li X-B, Cai L, Cheng N-H, and Liu J-W. 2002. Molecular Characterization of the Cotton GhTUB1 Gene That Is Preferentially Expressed in Fiber. *Plant Physiology* 130:666-674. 10.1104/pp.005538
Li X-B, Fan X-P, Wang X-L, Cai L, and Yang W-C. 2005. The cotton ACTIN1 gene is functionally expressed in fibers and participates in fiber elongation. *The Plant Cell* 17:859-875. 10.1105/tpc.104.029629
Liu L, Wang D, Hua J, Kong X, Wang X, Wang J, Si A, Zhao F, Liu W, Yu Y, and Chen Z. 2023. Genetic and morpho-physiological differences among transgenic and no-transgenic cotton cultivars. *Plants* 12:3437.
Liu Y. 2013. Recent progress in fourier transform infrared (FTIR) spectroscopy study of compositional, structural and physical attributes of developmental cotton fibers. *Materials (Basel)* 6:299-313. 10.3390/ma6010299
Livak KJ, and Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* 25:402-408. 10.1006/meth.2001.1262
Love MI, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. 10.1186/s13059-014-0550-8
Mansoor S, and Paterson AH. 2012. Genomes for jeans: cotton genomics for engineering superior fiber. *Trends Biotechnol* 30:521-527. S0167-7799(12)00094-7 [pii]
10.1016/j.tibtech.2012.06.003
Mao X, Cai T, Olyarchuk JG, and Wei L. 2005. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21:3787-3793. 10.1093/bioinformatics/bti430
Meinert MC, and Delmer DP. 1977. Changes in biochemical composition of the cell wall of the cotton fiber during development. *Plant Physiol* 59:1088-1097. 10.1104/pp.59.6.1088
Pertea M, Kim D, Pertea GM, Leek JT, and Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq

experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11:1650-1667. 10.1038/nprot.2016.095

Qin Y, Sun H, Hao P, Wang H, Wang C, Ma L, Wei H, and Yu S. 2019. Transcriptome analysis reveals differences in the mechanisms of fiber initiation and elongation between long- and short-fiber cotton (Gossypium hirsutum L.) lines. *BMC Genomics* 20:633. 10.1186/s12864-019-5986-5

Ruan YL, Xu SM, White R, and Furbank RT. 2004. Genotypic and developmental evidence for the role of plasmodesmatal regulation in cotton fiber elongation mediated by callose turnover. *Plant Physiol* 136:4104-4113. 10.1104/pp.104.051540

Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ, Zhou BL, Hu GJ, Guan XY, Chen ZJ, Wendel JF, Zhang TZ, and Chen XY. 2014. Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. *Nat Commun* 5:5519. ncomms6519 [pii] 10.1038/ncomms6519

Shang L, Abduweli A, Wang Y, and Hua J. 2016. Genetic analysis and QTL mapping of oil content and seed index using two recombinant inbred lines and two backcross populations in Upland cotton. *Plant Breeding* 135:224-231. https://doi.org/10.1111/pbr.12352

Shang L, Liang Q, Wang Y, Wang X, Wang K, Abduweli A, Ma L, Cai S, and Hua J. 2015. Identification of stable QTLs controlling fiber traits properties in multi-environment using recombinant inbred lines in Upland cotton (*Gossypium hirsutum* L.). *Euphytica* 205:877-888. 10.1007/s10681-015-1434-z

Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, Cheng J, Wei LP, Wang ZY, and Zhu YX. 2006. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18:651-664. 10.1105/tpc.105.040303

Song J, Pei W, Ma J, Yang S, Jia B, Bian Y, Xin Y, Wu L, Zang X, Qu Y, Zhang J, Wu M, and Yu J. 2021. Genome-wide association study of micronaire using a natural population of representative upland cotton (Gossypium hirsutum L.). *Journal of Cotton Research* 4:14. 10.1186/s42397-021-00089-1

Sun X, Gong S-Y, Nie X-Y, Li Y, Li W, Huang G-Q, and Li X-B. 2015. A R2R3-MYB transcription factor that is specifically expressed in cotton (Gossypium hirsutum) fibers affects secondary cell wall biosynthesis and deposition in transgenic Arabidopsis. *Physiologia Plantarum* 154:420-432. https://doi.org/10.1111/ppl.12317

Tian Y, and Zhang T. 2021. MIXTAs and phytohormones orchestrate cotton fiber development. *Curr Opin Plant Biol* 59:101975. 10.1016/j.pbi.2020.10.007

Tuttle JR, Nah G, Duke MV, Alexander DC, Guan X, Song Q, Chen ZJ, Scheffler BE, and Haigler CH. 2015. Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis, represses lignification, and prolongs elongation. *BMC Genomics* 16:477. 10.1186/s12864-015-1708-9

Wang M, Tu L, Yuan D, Zhu D, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z, Huang H, Yan F, Ma Y, Zhang L, Liu M, You J, Yang Y, Liu Z, Huang F, Li B, Qiu P, Zhang Q, Zhu L, Jin S, Yang X, Min L, Li G, Chen LL, Zheng H, Lindsey K, Lin Z, Udall JA, and Zhang X. 2019. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. *Nat Genet* 51:224-229. 10.1038/s41588-018-0282-x

Wang NN, Li Y, Chen YH, Lu R, Zhou L, Wang Y, Zheng Y, and Li XB. 2021. Phosphorylation of WRKY16 by MPK3-1 is essential for its transcriptional activity during fiber initiation and elongation in cotton (*Gossypium hirsutum*). *Plant Cell* 33:2736-2752. 10.1093/plcell/koab153

Wang S, Wang JW, Yu N, Li CH, Luo B, Gou JY, Wang LJ, and Chen XY. 2004. Control of plant trichome development

by a cotton fiber MYB gene. *Plant Cell* 16:2323-2334. 10.1105/tpc.104.024844
Wen X, Chen Z, Yang Z, Wang M, Jin S, Wang G, Zhang L, Wang L, Li J, Saeed S, He S, Wang Z, Wang K, Kong Z, Li F, Zhang X, Chen X, and Zhu Y. 2023. A comprehensive overview of cotton genomics, biotechnology and molecular biological studies. *Sci China Life Sci* 66:2214-2256. 10.1007/s11427-022-2278-0
Xiao R, Zhang C, Guo X, Li H, and Lu H. 2021. MYB transcription factors and its regulation in secondary cell wall formation and lignin biosynthesis during xylem development. *Int J Mol Sci* 22. 10.3390/ijms22073560
Xu Y, Magwanga RO, Cai X, Zhou Z, Wang X, Wang Y, Zhang Z, Jin D, Guo X, Wei Y, Li Z, Wang K, and Liu F. 2019. Deep transcriptome analysis reveals reactive oxygen species (ROS) network evolution, response to abiotic stress, and regulation of fiber development in cotton. *Int J Mol Sci* 20. 10.3390/ijms20081863
Yang D, Liu Y, Cheng H, Wang Q, Lv L, Zhang Y, Song G, and Zuo D. 2021. Identification of the Group III WRKY subfamily and the functional analysis of GhWRKY53 in *Gossypium hirsutum* L. Plants.
Yang P, Sun X, Liu X, Wang W, Hao Y, Chen L, Liu J, He H, Zhang T, Bao W, Tang Y, He X, Ji M, Guo K, Liu D, Teng Z, Liu D, Zhang J, and Zhang Z. 2022. Identification of candidate genes for lint percentage and fiber quality through QTL mapping and transcriptome analysis in an allotetraploid interspecific cotton CSSLs population. *Front Plant Sci* 13:882051. 10.3389/fpls.2022.882051
You J, Liu Z, Qi Z, Ma Y, Sun M, Su L, Niu H, Peng Y, Luo X, Zhu M, Huang Y, Chang X, Hu X, Zhang Y, Pi R, Liu Y, Meng Q, Li J, Zhang Q, Zhu L, Lin Z, Min L, Yuan D, Grover CE, Fang DD, Lindsey K, Wendel JF, Tu L, Zhang X, and Wang M. 2023. Regulatory controls of duplicated gene expression during fiber development in allotetraploid cotton. *Nature Genetics* 55:1987-1997. 10.1038/s41588-023-01530-8
Young MD, Wakefield MJ, Smyth GK, and Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* 11:R14. 10.1186/gb-2010-11-2-r14
Zang Y, Hu Y, Dai F, and Zhang T. 2022. Comparative transcriptome analysis reveals the regulation network for fiber strength in cotton. *Biotechnol Lett* 44:547-560. 10.1007/s10529-022-03236-z
Zhang J, Huang GQ, Zou D, Yan JQ, Li Y, Hu S, and Li XB. 2018. The cotton (*Gossypium hirsutum*) NAC transcription factor (FSN1) as a positive regulator participates in controlling secondary cell wall biosynthesis and modification of fibers. *New Phytol* 217:625-640. 10.1111/nph.14864
Zhang J, Mei H, Lu H, Chen R, Hu Y, and Zhang T. 2022. Transcriptome time-course analysis in the whole period of cotton fiber development. *Front Plant Sci* 13:864529. 10.3389/fpls.2022.864529
Zhang X, Cao J, Huang C, Zheng Z, Liu X, Shangguan X, Wang L, Zhang Y, and Chen Z. 2021a. Characterization of cotton ARF factors and the role of GhARF2b in fiber development. *BMC Genomics* 22:202. 10.1186/s12864-021-07504-6
Zhang X, Xue Y, Guan Z, Zhou C, Nie Y, Men S, Wang Q, Shen C, Zhang D, Jin S, Tu L, Yin P, and Zhang X. 2021b. Structural insights into homotrimeric assembly of cellulose synthase CesA7 from Gossypium hirsutum. *Plant Biotechnol J* 19:1579-1587. 10.1111/pbi.13571
Zhang Z, Ruan YL, Zhou N, Wang F, Guan X, Fang L, Shang X, Guo W, Zhu S, and Zhang T. 2017. Suppressing a putative sterol carrier gene reduces plasmodesmal permeability and activates sucrose transporter genes during cotton fiber elongation. *Plant Cell* 29:2027-2046. 10.1105/tpc.17.00358

Legends:

Figure 1 Differentially expressed genes (DEGs) identified among three different cotton cultivars at 20, 25, and 30 DPA fiber samples. (A) DEGs identified among 18 paired comparisons, (B) Venn diagram comparisons of DEGs among cultivars at 20 DPA, (C) Venn diagram comparisons of DEGs among cultivars at 25 DPA, (D) Venn diagram comparisons of DEGs among cultivars at 30 DPA. Note: Comparison of sample A with B was designated as A-VS-B, in which A was the control and B was the treatment. HL, L, and S represent accessions and 20, 25, and 30 represent the days post-anthesis (DPA).

Figure 2 Gene ontology (GO) enrichment analysis of DEGs in 25 DPA fiber samples between Sea Island or Upland cottons. (A) GO enrichment analysis between HL-25-vs-L-25, (B) GO enrichment analysis between HL-25-vs-S-25, (C) GO enrichment analysis between L-25-vs-S-25. The X-axis represents the biological functions (molecular function, biological process, and cellular component) of these DEGs. The Y-axis represents the percentage or number of genes categorized into different functional pathways.

Figure 3 KEGG pathway analysis of enriched differentially expressed genes. The "RichFactor" (x-axis) represents the ratio of differentially expression genes versus all genes in that pathway. Circle sizes correspond to gene number, and the q-value is given for each analysis. (A) Top 20 pathways of significantly enriched DEGs from HL-20 vs L-20, (B) Top 20 pathways of significantly enriched DEGs from HL-20 vs S-20, (C) Top 20 pathways of significantly enriched DEGs from L-20 vs S-20, (D) Top 20 pathways of significantly enriched DEGs from HL-25 vs L-25, (E) Top 20 pathways of significantly enriched DEGs from HL-25 vs S-25, (F) Top 20 pathways of significantly enriched DEGs from L-25 vs S-25, (G) Top 20 pathways of significantly enriched DEGs from HL-30 vs L-30, (H) Top 20 pathways of significantly enriched DEGs from HL-30 vs S-30, (I) Top 20 pathways of significantly enriched DEGs from L-30 vs S-30.

Figure 4 KEGG pathway analysis of enriched differentially expressed genes. (A) Top 20 pathways of significantly enriched DEGs from HL-20 vs HL-25, (B) Top 20 pathways of significantly enriched DEGs from HL-20 vs HL-30, (C) Top 20 pathways of significantly enriched DEGs from HL-25 vs HL-30, (D) Top 20 pathways of significantly enriched DEGs from L-20 vs L-25, (E) Top 20 pathways of significantly enriched DEGs from L-20 vs L-30, (F) Top 20 pathways of significantly enriched DEGs from L-25 vs L-30, (G) Top 20 pathways of significantly enriched DEGs from S-20 vs S-25, (H) Top 20 pathways of significantly enriched DEGs from S-20 vs S-30, (I) Top 20 pathways of significantly enriched DEGs from S-25 vs S-30.

Figure 5 Expression patterns of DEGs across the nine fiber samples. (A) 23 DEGs with low expression at 20 DPA but high expression at 25 or 30 DPA (i.e., low-to-high expression), (B) 7 additional DEGs with low-to-high expression, (C) 5 DEGs with high expression at 20 DPA but low expression at 25 or 30 DPA (i.e., high-to-low expression), (D) 4 additional DEGs showing high-to-low expression. Note: HL represents Xinhai 32 (Sea Island cotton), L represents 17-24 (Upland cotton), and S represents 62-33 (Upland cotton). DPA: days post anthesis.

Figure 6 Real-time quantitative PCR validation of DEGs from RNA-seq data. Relative expression of (A) *Polyphenol oxidase-9*, (B) *WRKY32*, (C) *GDSL-like lipase*, (D) *Caffeic acid* O-*methyltransferase 1*, (E) *NAD(P)-binding Rossmann-fold superfamily protein*, (F) *2,4-dienoyl-CoA reductase*, (G) *WRKY103*, (H) *Fatty acid desaturase 6*, (I) *3-ketoacyl-CoA synthase 3* genes at 20, 25, or 30 DPA fiber cells of the three cotton cultivars, and the expression level in the HL-20 sample was set to 1 (means of triplicates ± SD). Note: HL represents Xinhai 32 (Sea Island cotton), L represents 17-24 (Upland cotton), and S represents 62-33 (Upland cotton). Relative gene expression levels are normalized to histone-3 gene values. Error bars indicate SD (n = 3). Statistically significant differences ("a" is different from "b", "c", "d", "e" "f", "g", "h", or "i", α = 0.05 level) of expression values are indicated with different letters with analysis of variance in R (https://www.r-project.org/).

Table 1. The main properties of different cotton cultivars.

Table 2. Mapping results of RNA-seq clean reads from nine fiber samples.

Supporting information

TABLE S1. List of forward and reverse primers used for this study.

TABLE S2. The correlation coefficient between the RT-qPCR and RNA-seq data of nine genes.

Fig. S1 Gene ontology (GO) enrichment analysis of DEGs in 20-30 DPA fiber samples between Sea Island or Upland cottons. The X-axis represents the biological functions (molecular function, biological process, and cellular component) of these DEGs. The Y-axis represents the percentage or number of genes categorized into different functional pathways.

Fig. S2 Venn diagram comparison of differentially expressed genes (DEGs) from different fiber development stages in the three cotton cultivars.

Fig. S3 The expression patterns of nine candidate genes from RNA-seq data. Note: HL represents Xinhai 32 (Sea Island cotton), L represents 17-24 (Upland cotton), and S represents 62-33 (Upland cotton).

Figure 1

Differentially expressed genes (DEGs) identified among three different cotton cultivars at 20, 25, and 30 DPA fiber samples.

(A) DEGs identified among 18 paired comparisons, (B) Venn diagram comparisons of DEGs among cultivars at 20 DPA, (C) Venn diagram comparisons of DEGs among cultivars at 25 DPA, (D) Venn diagram comparisons of DEGs among cultivars at 30 DPA. Note: Comparison of sample A with B was designated as A-VS-B, in which A was the control and B was the treatment. HL, L, and S represent accessions and 20, 25, and 30 represent the days post anthesis (DPA).

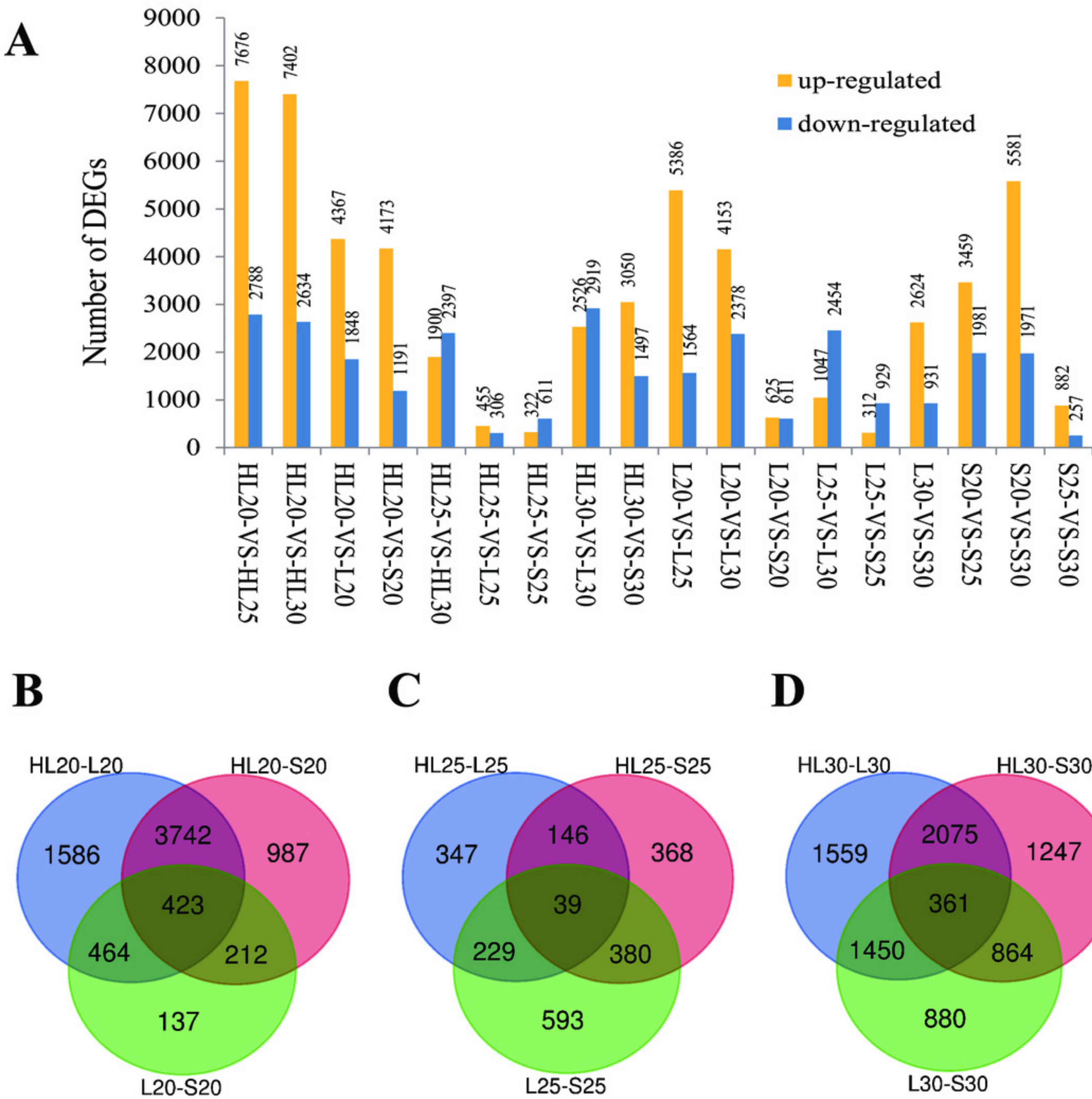
A
Number of DEGs
up-regulated
down-regulated
HL20-VS-HL25
HL20-VS-HL30
HL20-VS-L20
HL20-VS-S20
HL25-VS-HL30
HL25-VS-L25
HL25-VS-S25
HL30-VS-L30
HL30-VS-S30
L20-VS-L25
L20-VS-L30
L20-VS-S20
L25-VS-L30
L25-VS-S25
L30-VS-S30
S20-VS-S25
S20-VS-S30
S25-VS-S30
B
HL20-L20
HL20-S20
L20-S20
C
HL25-L25
HL25-S25
L25-S25
D
HL30-L30
HL30-S30
L30-S30

Figure 2

Gene ontology (GO) enrichment analysis of DEGs in 25 DPA fiber samples between Sea Island or Upland cottons.

(A) GO enrichment analysis between HL-25-vs-L-25, (B) GO enrichment analysis between HL-25-vs-S-25, (C) GO enrichment analysis between L-25-vs-S-25. The X-axis represents the biological functions (molecular function, biological process, and cellular component) of these DEGs. The Y-axis represents the percentage or number of genes categorized into different functional pathways.

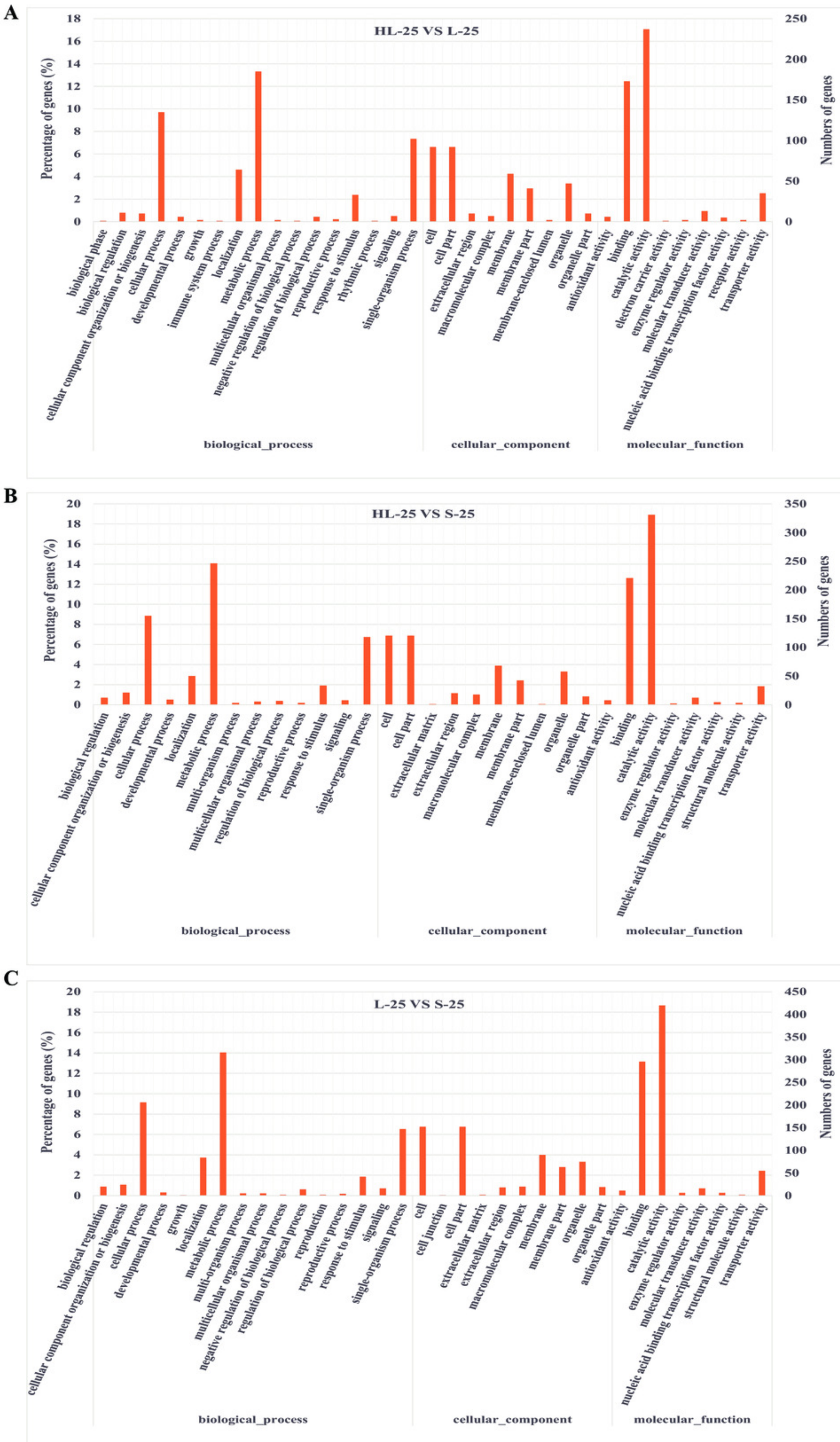
A
HL-25 VS L-25
Percentage of genes (%)
Numbers of genes
0
2
4
6
8
10
12
14
16
18
0
50
100
150
200
250
biological phase
biological regulation
cellular component organization or biogenesis
cellular process
developmental process
growth
immune system process
localization
metabolic process
multicellular organismal process
negative regulation of biological process
regulation of biological process
reproductive process
response to stimulus
rhythmic process
signaling
single-organism process
cell
cell part
extracellular region
macromolecular complex
membrane
membrane part
membrane-enclosed lumen
organelle
organelle part
antioxidant activity
binding
catalytic activity
electron carrier activity
enzyme regulator activity
molecular transducer activity
nucleic acid binding transcription factor activity
receptor activity
transporter activity
biological_process
cellular_component
molecular_function
B
HL-25 VS S-25
Percentage of genes (%)
Numbers of genes
0
2
4
6
8
10
12
14
16
18
20
0
50
100
150
200
250
300
350
biological regulation
cellular component organization or biogenesis
cellular process
developmental process
localization
metabolic process
multi-organism process
multicellular organismal process
regulation of biological process
reproductive process
response to stimulus
signaling
single-organism process
cell
cell part
extracellular matrix
extracellular region
macromolecular complex
membrane
membrane part
membrane-enclosed lumen
organelle
organelle part
antioxidant activity
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
nucleic acid binding transcription factor activity
structural molecule activity
transporter activity
biological_process
cellular_component
molecular_function
C
L-25 VS S-25
Percentage of genes (%)
Numbers of genes
0
2
4
6
8
10
12
14
16
18
20
0
50
100
150
200
250
300
350
400
450
biological regulation
cellular component organization or biogenesis
cellular process
developmental process
growth
localization
metabolic process
multi-organism process
multicellular organismal process
negative regulation of biological process
regulation of biological process
reproduction
reproductive process
response to stimulus
signaling
single-organism process
cell
cell junction
cell part
extracellular matrix
extracellular region
macromolecular complex
membrane
membrane part
organelle
organelle part
antioxidant activity
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
nucleic acid binding transcription factor activity
structural molecule activity
transporter activity
biological_process
cellular_component
molecular_function

Figure 3

KEGG pathway analysis of enriched differentially expressed genes.

The “RichFactor” (x-axis) represents the ratio of differentially expression genes versus all genes in that pathway. Circle sizes correspond to gene number, and the q-value is given for each analysis. (A) Top 20 pathways of significantly enriched DEGs from HL-20 vs L-20, (B) Top 20 pathways of significantly enriched DEGs from HL-20 vs S-20, (C) Top 20 pathways of significantly enriched DEGs from L-20 vs S-20, (D) Top 20 pathways of significantly enriched DEGs from HL-25 vs L-25, (E) Top 20 pathways of significantly enriched DEGs from HL-25 vs S-25, (F) Top 20 pathways of significantly enriched DEGs from L-25 vs S-25, (G) Top 20 pathways of significantly enriched DEGs from HL-30 vs L-30, (H) Top 20 pathways of significantly enriched DEGs from HL-30 vs S-30, (I) Top 20 pathways of significantly enriched DEGs from L-30 vs S-30.

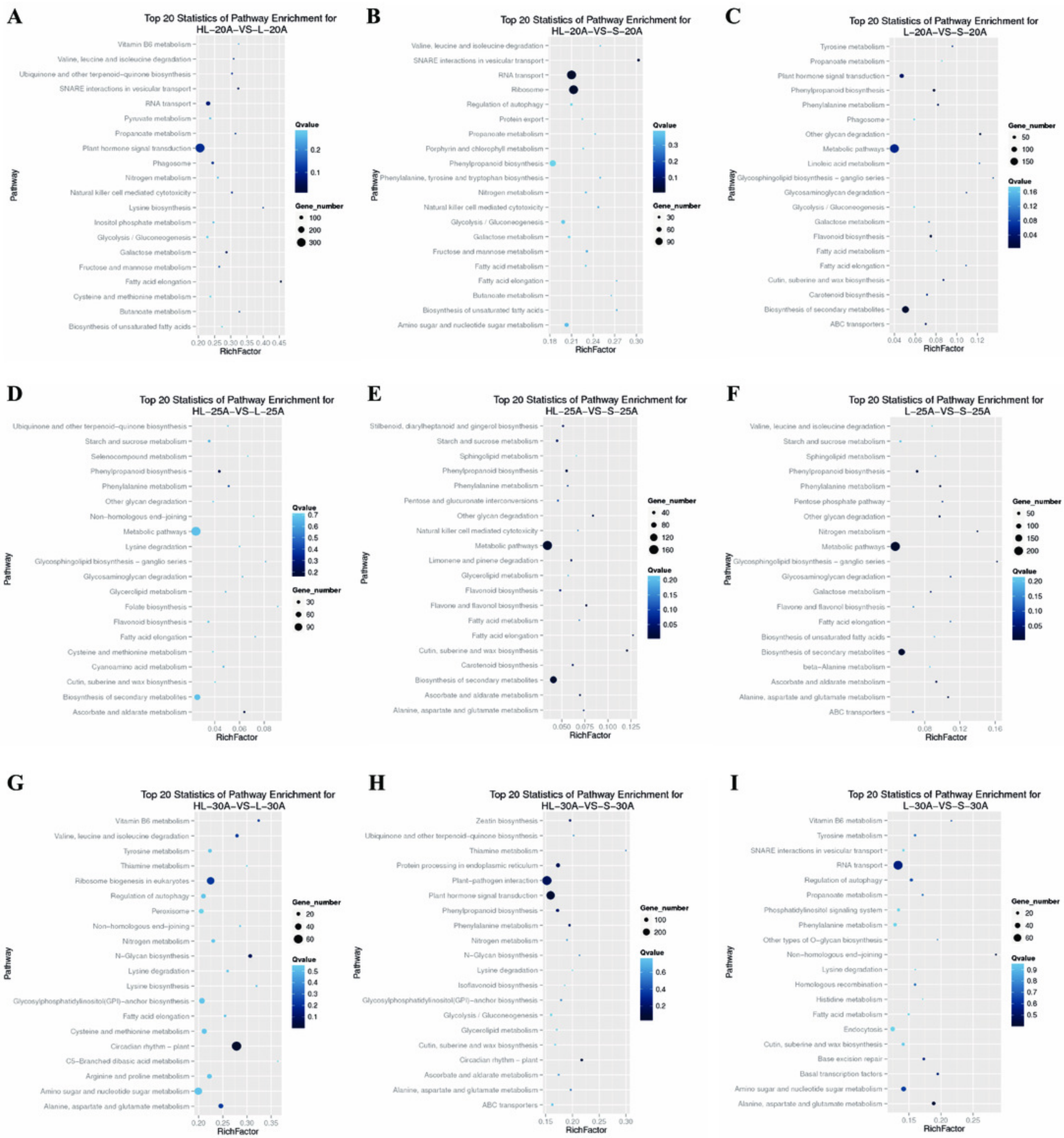
A
Top 20 Statistics of Pathway Enrichment for HL-20A-VS-L-20A
B
Top 20 Statistics of Pathway Enrichment for HL-20A-VS-S-20A
C
Top 20 Statistics of Pathway Enrichment for L-20A-VS-S-20A
D
Top 20 Statistics of Pathway Enrichment for HL-25A-VS-L-25A
E
Top 20 Statistics of Pathway Enrichment for HL-25A-VS-S-25A
F
Top 20 Statistics of Pathway Enrichment for L-25A-VS-S-25A
G
Top 20 Statistics of Pathway Enrichment for HL-30A-VS-L-30A
H
Top 20 Statistics of Pathway Enrichment for HL-30A-VS-S-30A
I
Top 20 Statistics of Pathway Enrichment for L-30A-VS-S-30A
Pathway
RichFactor
Qvalue
Gene_number

Figure 4

KEGG pathway analysis of enriched differentially expressed genes.

(A) Top 20 pathways of significantly enriched DEGs from HL-20 vs HL-25, (B) Top 20 pathways of significantly enriched DEGs from HL-20 vs HL-30, (C) Top 20 pathways of significantly enriched DEGs from HL-25 vs HL-30, (D) Top 20 pathways of significantly enriched DEGs from L-20 vs L-25, (E) Top 20 pathways of significantly enriched DEGs from L-20 vs L-30, (F) Top 20 pathways of significantly enriched DEGs from L-25 vs L-30, (G) Top 20 pathways of significantly enriched DEGs from S-20 vs S-25, (H) Top 20 pathways of significantly enriched DEGs from S-20 vs S-30, (I) Top 20 pathways of significantly enriched DEGs from S-25 vs S-30.

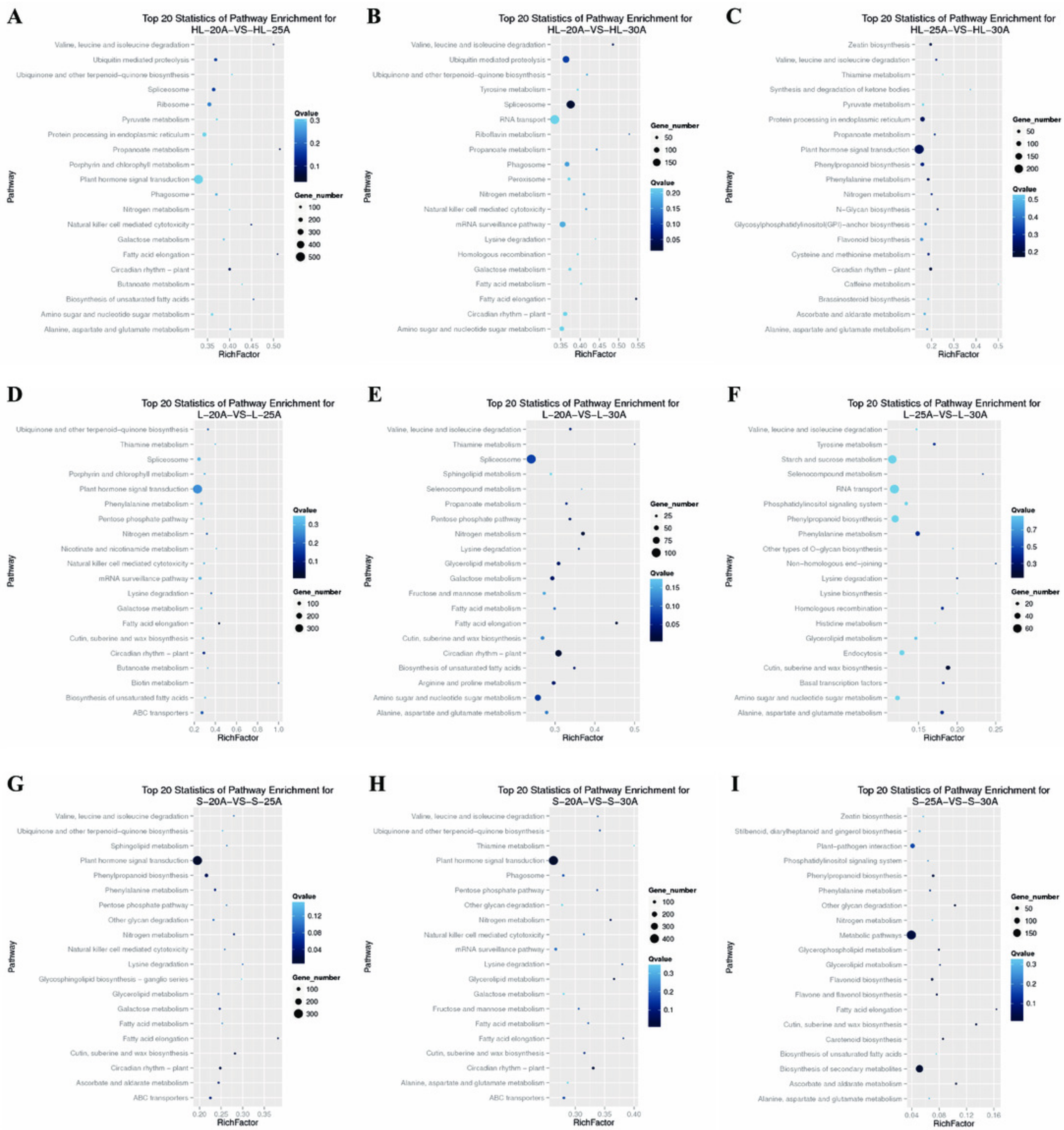
A
Top 20 Statistics of Pathway Enrichment for HL−20A−VS−HL−25A
Valine, leucine and isoleucine degradation
Ubiquitin mediated proteolysis
Ubiquinone and other terpenoid−quinone biosynthesis
Spliceosome
Ribosome
Pyruvate metabolism
Protein processing in endoplasmic reticulum
Propanoate metabolism
Porphyrin and chlorophyll metabolism
Plant hormone signal transduction
Phagosome
Nitrogen metabolism
Natural killer cell mediated cytotoxicity
Galactose metabolism
Fatty acid elongation
Circadian rhythm − plant
Butanoate metabolism
Biosynthesis of unsaturated fatty acids
Amino sugar and nucleotide sugar metabolism
Alanine, aspartate and glutamate metabolism
Pathway
RichFactor
Qvalue
Gene_number
B
Top 20 Statistics of Pathway Enrichment for HL−20A−VS−HL−30A
Valine, leucine and isoleucine degradation
Ubiquitin mediated proteolysis
Ubiquinone and other terpenoid−quinone biosynthesis
Tyrosine metabolism
Spliceosome
RNA transport
Riboflavin metabolism
Propanoate metabolism
Phagosome
Peroxisome
Nitrogen metabolism
Natural killer cell mediated cytotoxicity
mRNA surveillance pathway
Lysine degradation
Homologous recombination
Galactose metabolism
Fatty acid metabolism
Fatty acid elongation
Circadian rhythm − plant
Amino sugar and nucleotide sugar metabolism
Pathway
RichFactor
Gene_number
Qvalue
C
Top 20 Statistics of Pathway Enrichment for HL−25A−VS−HL−30A
Zeatin biosynthesis
Valine, leucine and isoleucine degradation
Thiamine metabolism
Synthesis and degradation of ketone bodies
Pyruvate metabolism
Protein processing in endoplasmic reticulum
Propanoate metabolism
Plant hormone signal transduction
Phenylpropanoid biosynthesis
Phenylalanine metabolism
Nitrogen metabolism
N−Glycan biosynthesis
Glycosylphosphatidylinositol(GPI)−anchor biosynthesis
Flavonoid biosynthesis
Cysteine and methionine metabolism
Circadian rhythm − plant
Caffeine metabolism
Brassinosteroid biosynthesis
Ascorbate and aldarate metabolism
Alanine, aspartate and glutamate metabolism
Pathway
RichFactor
Gene_number
Qvalue
D
Top 20 Statistics of Pathway Enrichment for L−20A−VS−L−25A
Ubiquinone and other terpenoid−quinone biosynthesis
Thiamine metabolism
Spliceosome
Porphyrin and chlorophyll metabolism
Plant hormone signal transduction
Phenylalanine metabolism
Pentose phosphate pathway
Nitrogen metabolism
Nicotinate and nicotinamide metabolism
Natural killer cell mediated cytotoxicity
mRNA surveillance pathway
Lysine degradation
Galactose metabolism
Fatty acid elongation
Cutin, suberine and wax biosynthesis
Circadian rhythm − plant
Butanoate metabolism
Biotin metabolism
Biosynthesis of unsaturated fatty acids
ABC transporters
Pathway
RichFactor
Qvalue
Gene_number
E
Top 20 Statistics of Pathway Enrichment for L−20A−VS−L−30A
Valine, leucine and isoleucine degradation
Thiamine metabolism
Spliceosome
Sphingolipid metabolism
Selenocompound metabolism
Propanoate metabolism
Pentose phosphate pathway
Nitrogen metabolism
Lysine degradation
Glycerolipid metabolism
Galactose metabolism
Fructose and mannose metabolism
Fatty acid metabolism
Fatty acid elongation
Cutin, suberine and wax biosynthesis
Circadian rhythm − plant
Biosynthesis of unsaturated fatty acids
Arginine and proline metabolism
Amino sugar and nucleotide sugar metabolism
Alanine, aspartate and glutamate metabolism
RichFactor
Gene_number
Qvalue
F
Top 20 Statistics of Pathway Enrichment for L−25A−VS−L−30A
Valine, leucine and isoleucine degradation
Tyrosine metabolism
Starch and sucrose metabolism
Selenocompound metabolism
RNA transport
Phosphatidylinositol signaling system
Phenylpropanoid biosynthesis
Phenylalanine metabolism
Other types of O−glycan biosynthesis
Non−homologous end−joining
Lysine degradation
Lysine biosynthesis
Homologous recombination
Histidine metabolism
Glycerolipid metabolism
Endocytosis
Cutin, suberine and wax biosynthesis
Basal transcription factors
Amino sugar and nucleotide sugar metabolism
Alanine, aspartate and glutamate metabolism
Pathway
RichFactor
Qvalue
Gene_number
G
Top 20 Statistics of Pathway Enrichment for S−20A−VS−S−25A
Valine, leucine and isoleucine degradation
Ubiquinone and other terpenoid−quinone biosynthesis
Sphingolipid metabolism
Plant hormone signal transduction
Phenylpropanoid biosynthesis
Phenylalanine metabolism
Pentose phosphate pathway
Other glycan degradation
Nitrogen metabolism
Natural killer cell mediated cytotoxicity
Lysine degradation
Glycosphingolipid biosynthesis − ganglio series
Glycerolipid metabolism
Galactose metabolism
Fatty acid metabolism
Fatty acid elongation
Cutin, suberine and wax biosynthesis
Circadian rhythm − plant
Ascorbate and aldarate metabolism
ABC transporters
Pathway
RichFactor
Qvalue
Gene_number
H
Top 20 Statistics of Pathway Enrichment for S−20A−VS−S−30A
Valine, leucine and isoleucine degradation
Ubiquinone and other terpenoid−quinone biosynthesis
Thiamine metabolism
Plant hormone signal transduction
Phagosome
Pentose phosphate pathway
Other glycan degradation
Nitrogen metabolism
Natural killer cell mediated cytotoxicity
mRNA surveillance pathway
Lysine degradation
Glycerolipid metabolism
Galactose metabolism
Fructose and mannose metabolism
Fatty acid metabolism
Fatty acid elongation
Cutin, suberine and wax biosynthesis
Circadian rhythm − plant
Alanine, aspartate and glutamate metabolism
ABC transporters
Pathway
RichFactor
Gene_number
Qvalue
I
Top 20 Statistics of Pathway Enrichment for S−25A−VS−S−30A
Zeatin biosynthesis
Stilbenoid, diarylheptanoid and gingerol biosynthesis
Plant−pathogen interaction
Phosphatidylinositol signaling system
Phenylpropanoid biosynthesis
Phenylalanine metabolism
Other glycan degradation
Nitrogen metabolism
Metabolic pathways
Glycerophospholipid metabolism
Glycerolipid metabolism
Flavonoid biosynthesis
Flavone and flavonol biosynthesis
Fatty acid elongation
Cutin, suberine and wax biosynthesis
Carotenoid biosynthesis
Biosynthesis of unsaturated fatty acids
Biosynthesis of secondary metabolites
Ascorbate and aldarate metabolism
Alanine, aspartate and glutamate metabolism
Pathway
RichFactor
Gene_number
Qvalue

Figure 5

Expression patterns of DEGs across the nine fiber samples

(A) 23 DEGs with low expression at 20 DPA but high expression at 25 or 30 DPA (i.e., low-to-high expression), (B) 7 additional DEGs with low-to-high expression, (C) 5 DEGs with high expression at 20 DPA but low expression at 25 or 30 DPA (i.e., high-to-low expression), (D) 4 additional DEGs showing high-to-low expression. Note: HL represents Xinhai 32 (Sea Island cotton), L represents 17-24 (Upland cotton), and S represents 62-33 (Upland cotton). DPA: days post anthesis.

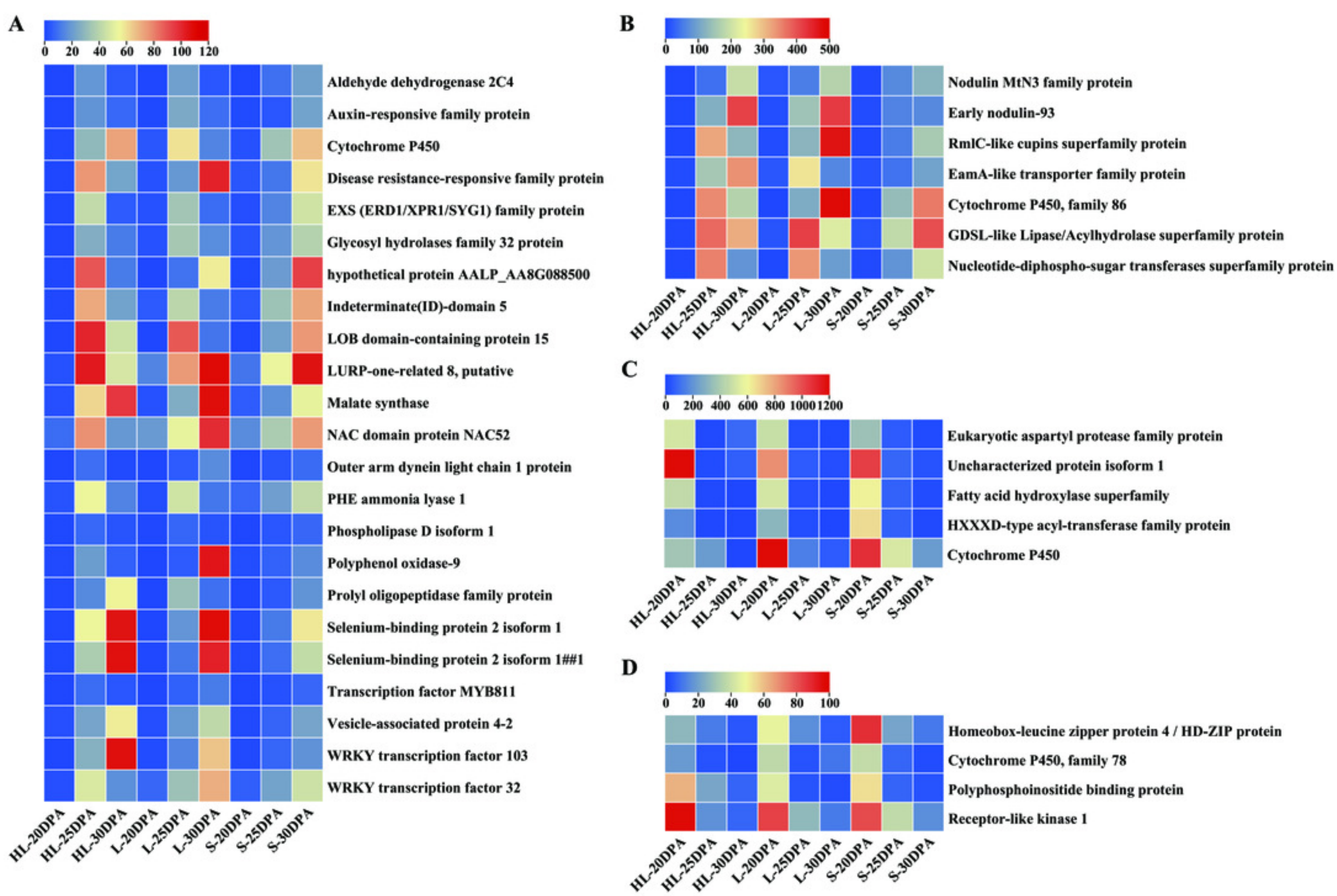

Figure 6

Real-time quantitative PCR validation of DEGs from RNA-seq data

Relative expression of (A) *Polyphenol oxidase-9*, (B) *WRKY32*, (C) *GDSL-like lipase*, (D) *Caffeic acid* O-*methyltransferase 1*, (E) *NAD(P)-binding Rossmann-fold superfamily protein*, (F) *2,4-dienoyl-CoA reductase*, (G) *WRKY103*, (H) *Fatty acid desaturase 6*, (I) *3-ketoacyl-CoA synthase 3* genes at 20, 25, or 30 DPA fiber cells of the three cotton cultivars, and the expression level in the HL-20 sample was set to 1 (means of triplicates ± SD). Note: HL represents Xinhai 32 (Sea Island cotton), L represents 17-24 (Upland cotton), and S represents 62-33 (Upland cotton). Relative gene expression levels are normalized to histone-3 gene values. Error bars indicate SD (n = 3). Statistically significant differences (“a” is different from “b”, “c”, “d”, “e” “f”, “g” , “h”, or “i”, α = 0.05 level) of expression values are indicated with different letters with analysis of variance in R (https://www.r-project.org/).

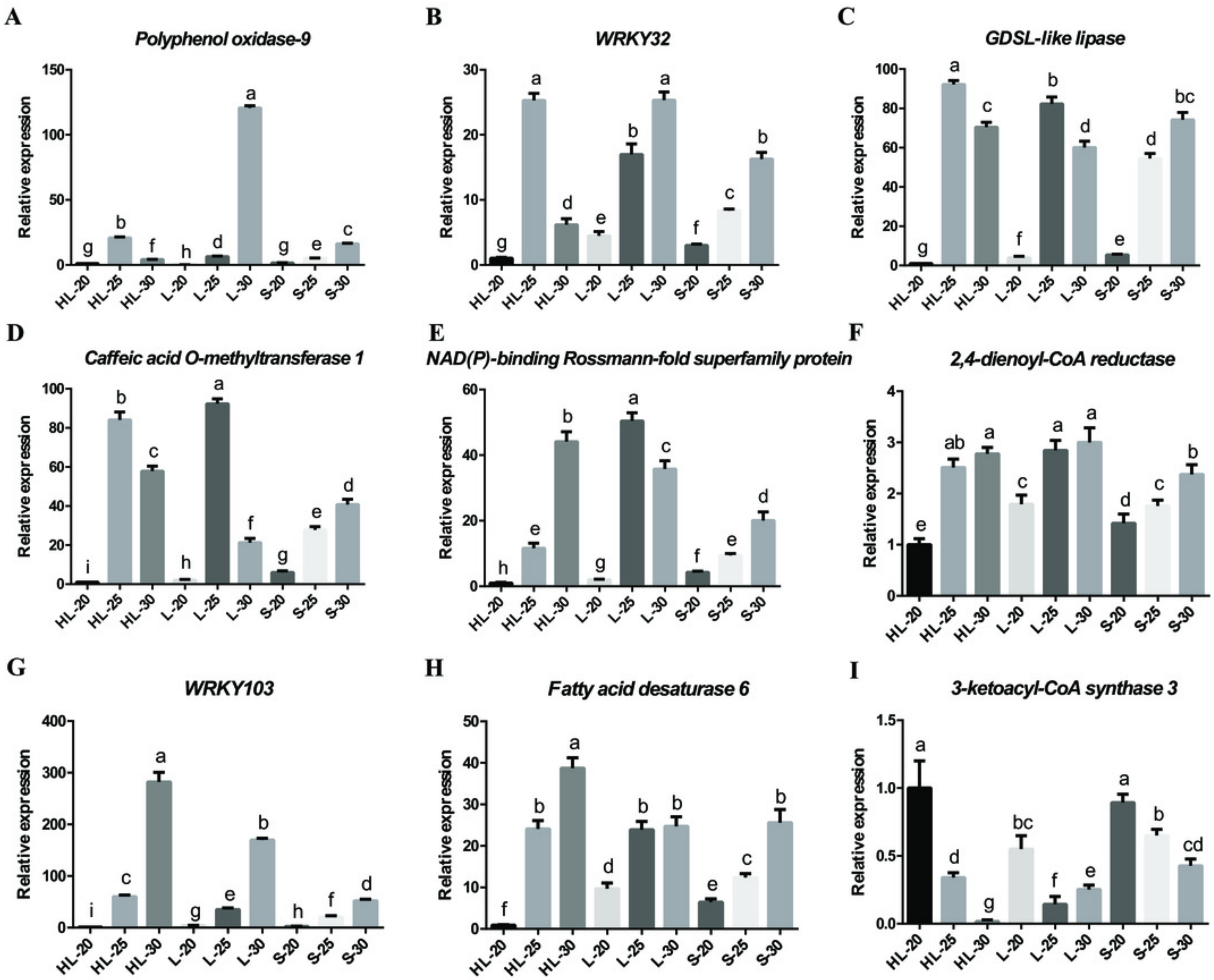
A
Polyphenol oxidase-9
B
WRKY32
C
GDSL-like lipase
D
Caffeic acid O-methyltransferase 1
E
NAD(P)-binding Rossmann-fold superfamily protein
F
2,4-dienoyl-CoA reductase
G
WRKY103
H
Fatty acid desaturase 6
I
3-ketoacyl-CoA synthase 3
Relative expression
HL-20
HL-25
HL-30
L-20
L-25
L-30
S-20
S-25
S-30

Table 1 (on next page)

The main properties of different cotton cultivars.

The main properties of different cotton cultivars.

Table 1. The main properties of different cotton cultivars.

Traits	**62-33（S）**	**17-24（L）**	**Xinhai 32（HL）**
Boll weight (g)	5.67±0.07^{a}	5.81±0.07^{a}	2.69±0.32^{b}
Lint percentage (%)	39.48±0.36^{a}	39.50±0.21^{a}	30.44±0.57^{b}
100 seed weight (g)	18.50±0.28^{a}	19.42±0.31^{a}	19.46±0.28^{a}
Lint index (g)	7.43±0.21^{a}	7.50±0.26^{a}	6.10±0.30^{a}
Seed index (g)	11.10±0.36^{a}	11.87±0.06^{a}	13.33±0.31^{b}
Fiber length (mm)	**28.91±0.15^{a}**	**33.45±0.44^{b}**	**39.84±0.22^{c}**
Fiber uniformity ratio (%)	87.47±0.56^{a}	87.63±0.15^{a}	90.15±0.25^{b}
Micronaire (MIC)	4.61±0.13^{a}	4.50±0.14^{a}	3.83±0.12^{b}
Fiber strength (cN/tex)	**31.57±1.90^{a}**	**37.73±0.90^{b}**	**59.93±0.90^{c}**
Fiber elongation (%)	7.07±0.06^{a}	7.03±0.06^{a}	7.72±0.12^{b}
Oil content (%)	28.43±0.22^{a}	30.39±0.18^{b}	35.35±0.13^{c}

Note: Error bars indicate SD (n = 3). Statistically significant differences ("a" is different from "b" or "c", α = 0.05 level) of values are indicated with different letters with analysis of variance in R (https://www.r-project.org/). Micronaire value is a comprehensive index reflecting the fineness and maturity of cotton fiber. Micronaire is divided into three levels: A, B and C, with B being the standard level. A grade values range from 3.7 to 4.2 with the best quality; Grade B values range from 3.5 to 3.6 and 4.3 to 4.9. Grade C is below to 3.4 or above to 5.0, showing the worst quality.

Table 2(on next page)

Mapping results of RNA-seq clean reads from nine fiber samples.

Mapping results of RNA-seq clean reads from nine fiber samples.

Table 2. Mapping results of RNA-seq clean reads from nine fiber samples.

Samples	Map to Genome	20DPA number	percentage	25DPA number	percentage	30DPA number	percentage
HL	Total Reads	12.1 M	100%	12.3 M	100.%	11.6 M	100%
	Total Base Pairs	590.9 Mbp	100%	600.8 Mbp	100 %	569.0 Mbp	100%
	Total Mapped Reads	9.3 M	77%	9.3 M	76 %	8.7 M	75 %
	perfect match	5.2 M	43%	5.1 M	41.86%	4.8 M	40.99%
	<=2bp mismatch	4.1 M	34%	4.2 M	34 %	4.0 M	34 %
	unique match	8.1 M	67%	8.4 M	68 %	7.7 M	65.96%
	multi-position match	1.2 M	10%	0.97 M	7.91%	1.1 M	9 %
	Total Unmapped Reads	2.8 M	23%	2.9 M	23.82%	2.9 M	24.82%
L	Total Reads	12.1 M	100 %	11.7 M	100 %	11.9 M	100 %
	Total Base Pairs	593.0 Mbp	100 %	573.5 Mbp	100 %	583.3 Mbp	100 %
	Total Mapped Reads	9.2 M	76%	8.9 M	76%	8.5 M	72%
	perfect match	5.0 M	41 %	4.9 M	42%	4.9 M	41 %
	<=2bp mismatch	4.1 M	34 %	4.0 M	34 %	3.6 M	30 %
	unique match	8.5 M	70%	8.1 M	69 %	6.3 M	53%
	multi-position match	0.68 M	6%	0.79 M	7%	2.2 M	19%
	Total Unmapped Reads	2.9 M	24 %	2.8 M	24 %	3.4 M	28 %
S	Total Reads	12.6 M	100 %	11.7 M	100 %	12.2 M	100 %
	Total Base Pairs	616.5 Mbp	100 %	573.5 Mbp	100 %	598.9 Mbp	100 %
	Total Mapped Reads	9.4 M	75%	8.9 M	76%	9.1 M	75%
	perfect match	5.2 M	41 %	4.9 M	42%	5.0 M	41%
	<=2bp mismatch	4.2 M	34%	4.0 M	34 %	4.2 M	34 %
	unique match	8.7 M	69 %	8.2 M	70%	8.3 M	68 %
	multi-position match	0.69 M	5 %	0.72 M	6 %	0.80 M	7%
	Total Unmapped Reads	3.2 M	25 %	2.8 M	24 %	3.1 M	25 %