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 identify the candidate genes related to the fiber strength trait, three cotton cultivars: Sea Island cotton (Xinhai 32, hyper-long fiber defined as HL) and upland cotton (17-24, long fiber defined as L and 62-33, short fiber defined as S) fiber samples with different qualities were selected and the expression profiles of cotton fiber at secondary wall thickening stage (20, 25 and 30 DPA) were analyzed by RNA-seq technology. The results showed that a large number of differentially expressed genes (DEGs) were obtained from three cotton cultivars at different stages of secondary wall development. For instance, at 20 DPA, 6,215 and 5,364 DEGs were obtained from Sea Island cotton (HL) compared with upland cotton 17-24 (L) or 62-33 (S), respectively, meanwhile, there were 1,236 DEGs between two upland cotton cultivars 17-24 (L) and 62-33 (S). Through GO function annotation, 42 functions are mainly annotated, among which 20 were biological processes, 11 were cellular components and 11 were molecular functions. KEGG enrichment pathway analysis identified 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 were involved in secondary wall synthesis and thickening. Identification of shared DEGs showed that there were 46 DEGs existed huge expression divergence at different fiber stages from three cotton cultivars, with functions including REDOX enzymes, binding proteins, hydrolases (such as GDSL thioesterase), transferases, metalloproteins (cytochromatin-like genes), kinases, carbohydrates, and transcription factors (MYB and WRKY). Finally, RT-qPCR was performed to verify the expression levels of 11 genes, which proved the reliability of RNA-Seq data. Our results provided valuable gene 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 identify the candidate genes related to the fiber strength trait, three cotton cultivars: Sea Island cotton (Xinhai

25 32, hyper-long fiber defined as HL) and upland cotton (17-24, long fiber defined as L and 62-33, short fiber

26 defined as S) fiber samples with different qualities were selected and the expression profiles of cotton fiber at

27 secondary wall thickening stage (20, 25 and 30 DPA) were analyzed by RNA-seq technology. The results

28 showed that a large number of differentially expressed genes (DEGs) were obtained from three cotton cultivars

29 at different stages of secondary wall development. For instance, at 20 DPA, 6,215 and 5,364 DEGs were obtained

30 from Sea Island cotton (HL) compared with upland cotton 17-24 (L) or 62-33 (S), respectively, meanwhile, there

31 were 1,236 DEGs between two upland cotton cultivars 17-24 (L) and 62-33 (S). Through GO function

32 annotation, 42 functions are mainly annotated, among which 20 were biological processes, 11 were cellular

33 components and 11 were molecular functions. KEGG enrichment pathway analysis identified

34 glycolysis/gluconeogenesis, galactose metabolism, propanoate metabolism, biosynthesis of unsaturated fatty

35 acids pathway, valine, leucine and isoleucine degradation, fatty acid elongation pathways, and plant hormone

36 signal transduction were involved in secondary wall synthesis and thickening. Identification of shared DEGs

37 showed that there were 46 DEGs existed huge expression divergence at different fiber stages from three cotton

38 cultivars, with functions including REDOX enzymes, binding proteins, hydrolases (such as GDSL thioesterase),

39 transferases, metalloproteins (cytochromatin-like genes), kinases, carbohydrates, and transcription factors (MYB

40 and WRKY). Finally, RT-qPCR was performed to verify the expression levels of 11 genes, which proved the

41 reliability of RNA-Seq data. Our results provided valuable gene resources for clarifying the cell biology of

42 secondary cell wall biosynthesis during fiber development in cotton.

43 **Key words:** Upland cotton, Sea Island cotton, secondary cell wall thickening, fiber quality, gene

44 expression patterns

45

46

47 # Introduction

48 As one of the seven major crops, cotton plays an important role in the textile industry with its fiber

49 as the main product of cotton production (Chen et al. 2017; Huang et al. 2021a). Cotton is also a

50 model plant for cellulose synthesis and cell elongation studies (Cao et al. 2020b; Glover 2000;

51 Guan et al. 2007). The cotton fiber is a single cell, which is differentiated from epidermal cells in

52   the outer integument of the ovule (Shan et al. 2014). Cotton fiber development can be divided into

53   four stages: initiation, elongation, secondary wall thickening, and maturation (Huang et al. 2021a;

54   Wen et al. 2023). The fiber elongation stage can vary in length among accessions, lasting until 15-

55   25 days post anthesis (DPA) depending on species and domestication status. Elongation typically

56   overlaps the transition stage (16-20 DPA) and, in the domesticated accessions, it also overlaps the

57   beginning of the secondary wall thickening stage (20 through 40 DPA (Mansoor & Paterson 2012).

58   Fiber maturity occurs between 40 to 50 DPA, when the fiber cells die and the cytoplasm is

59   degraded, leaving behind the hollow cells surrounded by cellulose (Jareczek et al. 2023; Kim

60   2018). The primary cell wall, secondary cell wall, luminal wall and middle luminal wall can be

61   observed in the cross section of the fibers. Because secondary wall thickening occurs by laying

62   down a helical pattern of cellulose fibers, mature fibers appear flat and banded, and have a natural

63   twist (Mansoor & Paterson 2012).

64   Different fiber properties are influenced most strongly in the different morphogenic stages.

65   Differentiation of cotton fiber from the ovular epidermis occurs during initiation (around -1 to 1

66   DPA), when approximately 20% to 30% of epidermal cells differentiate into fiber cells. During

67   this stage, fiber tips are refined, which strongly correlates with both mature diameter and strength

68   (Kelly et al. 2015). The elongation stage, on the other hand, is strongly correlated with fiber length.

69   While elongation lasts from ~3 to 20 DPA, 5 to 15 DPA comprises the most rapid elongation

70   period, when cotton leverages fatty acids and carbohydrates to keep the primary cell wall pliable

71   for extreme linear growth (Tian & Zhang 2021). At ~16 DPA, the fiber enters the transition stage,

72   where the microtubules in the fiber shift to a shallow helical angle and the fiber lays down the

73   winding cell wall layer (Hsieh et al. 1995; Meinert & Delmer 1977). The composition of the

74   winding cell wall layer is similar to that of the primary cell wall, with a slight increase in cellulose

75   content. The deposition of the winding cell wall layer is thought to impact both fiber strength and

76   flexibility (Haigler et al. 2009; Tuttle et al. 2015; Zhang et al. 2021b). The thickening period of

77   the secondary wall begins during the transition stage (~15DPA), when cellulose production

78   increases substantially. The secondary wall synthesis stage begins at ~20 DPA, and is

79   characterized by β-1, 4-glucan chains that accumulate to facilitate cellulose accumulation and form

80   20-30 layers of "growing day rings". This period mainly determines the thickness and strength of

81   the cell wall (Haigler et al. 2012). When the cell wall thickens to 3-4 microns, the cells began to

82   dehydrate and undergo apoptosis, and the whole fiber cells showed a twisted spiral state (Hof &

83   Saha 1997). The natural twist of cotton fiber can increase the binding force between fibers and

84   improve the yarn strength when spinning. Mature cotton fibers contain more than 95% cellulose

85   and less keratin, wax, inorganic matter, and protein than other plant cells (Liu 2013). From an

86   agronomic standpoint, cotton fiber quality is based on several properties, such as fiber fineness,

87   length, strength, micronaire (i.e., cell wall thickness), and yellowness (Bajwa et al. 2015; Song et

88   al. 2021; Yang et al. 2022). Increased demand for luxury textiles has likewise increased the demand

89   for high-quality cotton fiber; consequently, there is also increased interest in improving fiber

90   quality in the more highly productive species/cultivars (Gao et al. 2021; Huang et al. 2021a). While

91   breeding programs are actively attempting to introgress desirable fiber quality traits into these

92   productive lines, like most crops, cotton fiber yield and quality traits are quantitatively controlled

93   by multiple genes, limiting the success of traditional breeding techniques (Liu et al. 2023; Xu et

94   al. 2019). Therefore, understanding the physiological and molecular basis of fiber development is

95   paramount to improving cotton fiber quality through other techniques, such as molecular design

96   breeding.

97   In recent years, the molecular mechanisms underlying cotton fiber development have been

98   studied in depth, and a series of important advances have been made using transcriptome analysis

99   and other high-throughput based methods (Li et al. 2022; Zang et al. 2022; Zhang et al. 2022).

100   These studies, however, mainly focus on the initiation and elongation stages of fiber cell

101   development (Qin et al. 2019), and the molecular mechanisms underlying fiber secondary wall

102   synthesis and thickening (and, therefore, strength) have been rarely studied. In order to improve

103   our understanding of the molecular mechanisms operating during secondary wall synthesis and

104   their influence on fiber quality (strength), two cultivars of *Gossypium hirsutum* and one cultivar

105   of *G. barbadense* with known differences in fiber quality were selected for comparative

106 transcriptome analysis, and differentially expressed genes (DEGs) from the secondary wall

107 thickening stage were comprehensively identified using RNA-seq from three timepoints. Through

108 pairwise comparison, common enrichment pathways were identified using the DEGs among

109 cultivars, and important candidate genes related to cotton fiber development were screened at

110 different developmental stages. Our results provided a good basis for the analysis of the molecular

111 mechanism of cotton fiber secondary wall development, which would be helpful for the mining

112 and utilization of the valuable gene resources.

Materials & Methods

Plant Materials

115 Cotton cultivars, Xinhai 32 is a high-generation inbred line of Sea Island cotton and 17-24 or 62-

116 33 are high-generation inbred lines of Upland cotton. All cultivars were bred by Cotton Institute,

117 Xinjiang Academy of Agricultural and Reclamation Science. Xinhai 32, 17-24, and 62-33 were

118 defined as HL, L, and S, respectively. These three cotton cultivars were planted in the field of

119 Xinjiang Academy of Agricultural and Reclamation Science. Ovules (seeds) were harvested at the

120 indicated DPA (days post anthesis). Cotton bolls of 20 DPA, 25 DPA, and 30 DPA from three

121 different cotton cultivars were sampled at 10:00 AM. Five cotton bolls from each plant were

122 selected, and the fibers were isolated from the ovules by scratching the ovule with a metal strainer

123 in liquid nitrogen. Then, these fiber samples were quickly ground into powder, and stored in the

124 ultra-low temperature refrigerator at -80℃.

Fiber traits and phenotypic evaluation

126 Boll weight, 100 seed weight, lint index, and seed index were weighed by an analytical balance

127 (0.0001g, BSA224S, SARTORIUS, Germany). Fiber quality traits, including the fiber length

128 (mm), fiber uniformity ratio (%), fiber strength (cN/tex), fiber elongation and micronaire, were

129 measured with an HVI 900 instrument (USTER HVISPECTRUM, SPINLAB, USA) at the Cotton

130 Fiber Quality Inspection and Test Center of Ministry of Agriculture (Anyang, China) (Shang et al.

131 2015). Bolls were collected for seed index analysis. For this purpose, one hundred seeds from each

132 cultivar were randomly selected and weighed as seed index (SI, g) (Liu et al. 2023; Shang et al.

133 2016). To measure oil content, seeds were delinted with concentrated sulphuric acid. The oil

134 contents of three cotton cultivars were measured at different stages of ovular development using

135 the Soxhlet extraction method (García-Ayuso et al. 2000). To record grain weight, one hundred

136 cotton ovules were weighed randomly.

RNA extraction, library construction, and sequencing

138 Total RNA from 20 DPA, 25 DPA, and 30 DPA cotton fiber from each of the three different cotton

139 cultivars was extracted using the RNAprep pure plant kit (TIANGEN, China) according to the

140 manufacturer's instructions. A total of 1 μg purified mRNA was selected for cDNA library

141 construction following a previous report (Chen et al. 2021). Briefly, mRNA was purified from

142 total RNA using poly-T oligo-attached magnetic beads. cDNA fragments 240 bp in length were

143 preferentially selected, and the library fragments were purified with an AMPure XP system

144 (Beckman Coulter, Beverly, USA). Finally, the PCR products were purified (AMPure XP system),

145 and library quality was assessed on an Agilent Bioanalyzer 2100 system. After cluster generation,

146 the libraries were sequenced on an Illumina HiSeqTM 2500 platform as 150 bp paired-end reads.

147 Three biological replicates were performed for the nine samples.

RNA-seq reads quality control, mapping and differentially expressed gene (DEG) analysis

150 FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was used to process the raw reads in

151 fastq format according to a previous study (Chen et al. 2021). Clean data (clean reads) were

152 obtained by removing reads containing adapters, reads containing poly-N sequences, and low-

153 quality reads from the raw data. All the subsequent analyses were performed based on the clean

154 high-quality data. Among the nine fiber samples from *G. hirsutum* and *G. barbadense* (Hu et al.

155 2019; Wang et al. 2019), RNA-seq data were mapped to their reference genomes using HISAT2

156 software, respectively (Kim et al. 2015; Pertea et al. 2016). Reads with at most one mismatch were

157 used to calculate the expression levels of genes. Gene expression values were calculated following

158   the method of the previous study (Chen et al. 2021). Differential expression analysis of the two

159   groups was performed using DESeq2 and presented using fragments per kilobase of transcript per

160   million fragments mapped (FPKM) (Love et al. 2014). The resulting *p* values were adjusted using

161   Benjamini and Hochberg's approach for controlling the false discovery rate (FDR). Genes with an

162   adjusted *p* value < 0.01 and twofold change (up and down) were defined as differentially

163   expressed. TBtools (Chen et al. 2020) was used to display the gene expression patterns of the

164   FPKM values. Clean data were available from the Genome Sequence Archive in the BIG Data

165   Center of Sciences (https://bigd.big.ac.cn/) under accession number CRA009299. The statistical

166   power of this experimental design, calculated in RNASeqPower is 0.84.

Gene functional annotation and enrichment analyses

168   Differentially expressed gene functions were annotated based on the following databases: Nr

169   (NCBI nonredundant protein sequences, ftp://ftp.ncbi.nih.gov/blast/db/), Gene Ontology (GO)

170   (Gene Ontology, http://www.geneontology.org/), and Kyoto Encyclopedia of Genes and Genomes

171   (KEGG) (http://www.genome.jp/kegg/). Gene Ontology (GO) enrichment analysis of the DEGs

172   was implemented with the GOseq R package based on Wallenius noncentral hypergeometric

173   distribution (Young et al. 2010). KOBAS was used to test the statistical enrichment of the DEGs

174   in the KEGG pathways (Mao et al. 2005).

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

176   These experiments were conducted according to the methods reported previously (Cao et al.

177   2020a; Cao et al. 2022; Cui et al. 2022; Zhang et al. 2021a). In brief, total RNAs from 20 DPA,

178   25 DPA, and 30 DPA fiber samples of three different cotton cultivars were extracted using a

179   RNAprep pure plant kit (TIANGEN, China). The resulting RNAs were treated with DNase I prior

180   to synthesizing cDNA with TransScript® First-Strand cDNA Synthesis SuperMix (TransGen

181   Biotech, China); these products were diluted fivefold before using. For real-time quantitative PCR,

182   Primer v5.0 software was used to design specific forward and reverse gene primers (Table S1).

183   Analyses were performed with SYBR-Green PCR Mastermix (TaKaRa) on a cycler (Mastercycler

184   RealPlex; Eppendorf Ltd., China). The *G. hirsutum* and *G. barbadense* histone-3 (*GhHIS3* and

185 *GbHIS3*) genes were used as internal references, and the relative amount of amplified product was

186 calculated following the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen 2001).

Statistical analysis

188 The R package available at https://www.r-project.org/ was utilized for the analysis of variance and

189 Student's t test. The normality test was conducted using the Shapiro-Wilk test to confirm that the

190 data followed a Gaussian distribution. The significance was tested using the least significant

191 difference (LSD) at the 1% or 5% levels. Each sample included in the analysis was based on at

192 least three biological replicates.

Results

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

196 Physiological traits of the three cotton cultivars, all high-generation inbred lines developed by the

197 Cotton Institute, Xinjiang Academy Agricultural and Reclamation Science, were characterized.

198 Table 1 shows the main properties of these three cotton cultivars, which we have designated as S

199 (cv 62-33), L (cv 17-24), and HL (Xinhai 32). Compared with the upland cotton cultivars (i.e., L

200 and S), the Sea Island cotton (HL) exhibited longer fiber, greater fiber strength, increased oil

201 content, a higher seed index, greater fiber uniformity, and greater fiber elongation. In contrast, the

202 boll weight, lint percentage, lint index, and micronaire were significantly higher in the upland

203 cottons (L and S) than in Xinhai 32 (HL). Notably, these properties, although greater in the *G.*

204 *hirsutum* cultivars, are outside of the optimal range. In particular, micronaire (one of the most

205 important measures of cotton fiber quality) was considered grade A (micronaire range: 3.7 - 4.2)

206 in Xinhai 32, compared the B grade (micronaire range: 3.5 - 3.6 and 4.3 - 4.9) observed in the

207 upland cotton cultivars, L and S. Between the two cultivars of upland cotton, fiber length, fiber

208 strength, and oil content also varied, with cv 17-24 (L) exhibiting significantly higher values than

209 cv 62-33 (S). These results indicate that major differences in agronomically important fiber

210 properties exist between Sea Island cotton and Upland cotton, as expected, but also between the

211 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

214 As these three cultivars showed great variance in fiber length and strength, especially fiber strength

215 (Table 1), we evaluated gene expression at three timepoints during secondary wall synthesis (i.e.,

216 20, 25, and 30 DPA fiber) from Xinhai 32 (HL), 17-24 (L), and 62-33 (S). RNA from at least three

217 biological replicates was pooled at each timepoint for each accession, hereafter referred to as

218 HL20, HL25, HL30, L20, L25, L30, S20, S25, and S30, and over 108 million high-quality reads

219 were generated using the Illumina NovaSeq 6000 sequencing platform (Table 2). The HISAT2

220 software was used to align these clean reads to the reference cotton genomes (Hu et al. 2019; Wang

221 et al. 2019) with at most one base mismatch. The ratio of mapped reads ranged from 72% in sample

222 L30 to 77% in sample HL20 (Table 2), and the number of uniquely mapped clean reads ranged

223 from 53% in sample L30 to 70% in sample L20. These data indicated that the RNA-seq data in

224 this study were reliable for the subsequent analyses.

Identification of differentially expressed genes (DEGs)

226 Differential gene expression was surveyed for the 9 cotton fiber samples. Eighteen comparisons

227 among these 9 samples were performed to capture expression differences between samples at a

228 given DPA and within a single sample among measured timepoints (Fig. 1A). These comparisons

229 identified 85,207 DEGs in total, among which 55,940 were up-regulated and 29,267 were down-

230 regulated (Fig. 1A). In each comparison, the number of differentially expressed genes (DEGs)

231 varied between 761 genes in HL25-VS-L25 and 10,464 genes in HL20-VS-HL25 (Fig. 1A). It is

232 worth noting that upregulation was more frequent than downregulation in most comparisons,

233 except for the comparison between 25 versus 30 DPA in the two longer and stronger fiber

234 accessions, Xinhai 32 (HL) and 17-24 (L) (Fig. 1A).

235 　　Because these cultivars produce fibers with different properties, we considered the overlap in

236 gene expression at each DPA to identify genes important for secondary cell wall biosynthesis that

237 differed in expression among the three species/cultivars. Venn diagrams were constructed for the

238 DEGs from each species/cultivar comparison at each surveyed stage (i.e., 20, 25 and 30 DPA

239 stages; Fig. 1). As expected, the intraspecies cultivar comparisons (i.e., L versus S) typically

240 exhibited fewer uniquely DEG than the interspecies comparisons (i.e., HL versus S or L), with the

241 exception of 25 DPA (Fig. 1). Although 25 DPA exhibited the fewest DEGs overall (Fig. 1), the

242 intraspecies (S versus L) comparison resulted in 25 - 40% more DEGs than either interspecific

243 comparison (Fig. 1); the fewest DEGs at this stage were between the HL and L cultivars, perhaps

244 indicating that underlying fiber length expression differences are largely responsible for the

245 differences in gene expression at this stage. The results showed that 423 DEGs were shared by all

246 species/cultivar comparisons at the 20 DPA stage (Fig. 1B), possibly indicating genes that underlie

247 the differences among cultivars. Far fewer DEGs were observed at 25 DPA (39 DEGs), although

248 this is number is commensurate with the general reduction in DEGs among cultivars at this stage

249 (Fig. 1C). At 30 DPA, 361 DEGs were shared among the cultivar comparisons (Fig. 1D).

GO analysis for DEGs

251 We considered the possible biological functions of these DEGs between Sea Island and/or the

252 Upland cottons using GO category enrichment (Fig. S1). Results from the three categories (i.e.,

253 biological process, cellular component, and molecular function), suggest enrichment of 20, 11, and

254 11 functional categories in the 20, 25, and 30 DPA fiber comparisons, respectively. GO terms

255 associated with important biological processes included metabolic, cellular, developmental, and

256 single-organism processes, biological regulation, response to stimulus, and signaling. Cellular

257 components, such as cell, cell part, membrane, membrane part, organelle and organelle parts were

258 enriched. Molecular function enrichment consisted of catalytic activity, transporter activity,

259 binding, nucleic acid-binding transcription factor activity, antioxidant activity, and receptor

260 activity.

261 To compare the difference between the three cultivars, the GO enrichment of functional

262 categories in the 25 DPA fiber was analyzed (Fig. 2). The results showed that the top three GO

263 category enrichment in biological processes among three groups of comparisons (HL-25-vs-L-25,

264   HL-25-vs-S-25, L-25-vs-S-25) were all cellular process, metabolic process, and localization,

265   respectively (Figs. 2A-2C). As for the cellular components, cell, cell part and membrane accounted

266   for the largest proportion in all three groups (Figs. 2A-2C). Finally, the most enriched molecular

267   function all consisted of catalytic activity, binding, and transporter activity between the

268   comparisons of three cultivars (Figs. 2A-2C). These results indicated that GO enrichment analysis

269   was not specific in the three cultivars.

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

272   To investigate the biological functions of these DEGs from three cotton cultivars at different

273   stages of fiber (20 DPA, 25 DPA, and 30 DPA), we performed KEGG pathway enrichment

274   analysis for the DEGs. At 20 DPA fiber stage, the DEGs were assigned to 126 KEGG pathways

275   according to the functional categorization. For the interspecific comparison HL-20-vs-L-20, the

276   top 20 KEGG pathways of enriched DEGs were categorized into the following functional

277   pathways (Fig. 3A). While greatest number of DEGs mostly belonged to the plant hormone signal

278   transduction pathways, there was also involvement of several other pathways whose RichFactor

279   (i.e., the proportion of genes in that pathway that are DE) was closer to 50%, such as fatty acid

280   elongation. Other pathways with notable enrichment are: i) carbohydrate metabolism:

281   glycolysis/gluconeogenesis, pyruvate metabolism, galactose metabolism, fructose and mannose

282   metabolism; ii) fatty acid metabolism: fatty acid elongation, and biosynthesis of unsaturated fatty

283   acids; and iii) amino acid metabolism: valine, leucine and isoleucine degradation, lysine

284   biosynthesis, and cysteine and methionine metabolism. For the other interspecific comparison,

285   HL-20-vs-S-20, the top 20 KEGG pathways of enriched DEGs were also involved in

286   glycolysis/gluconeogenesis, galactose metabolism, fructose and mannose metabolism, fatty acid

287   elongation, biosynthesis of unsaturated fatty acids, and leucine and isoleucine degradation (Fig.

288   3B); however, the greatest representation of genes was for RNA transport and the ribosome. In

289   this comparison, the RichFactor varied more narrowly (~12%, versus 25% in HL20 versus L20).

290   For the intraspecific comparison L-20-vs-S-20, the top 20 KEGG pathways of enriched DEGs

291 mainly included the glycolysis/gluconeogenesis, galactose metabolism, fatty acid elongation,

292 biosynthesis of unsaturated fatty acids, plant hormone signal transduction, and biosynthesis of

293 secondary metabolites (Fig. 3C). The greatest numbers of genes were found in the metabolic

294 pathways and biosynthesis of secondary metabolites. While the RichFactor for these was low, it is

295 worth noting that the RichFactor was generally low for all pathways in, possibly indicating

296 recruitment of few genes/pathways from these somewhat broad categories. These results suggested

297 that the differences in 20 DPA fiber samples between Sea Island cotton (Xinhai 32) and Upland

298 cotton (17-24 or 62-33) were mainly caused by genes involved in the metabolic pathways of

299 carbohydrates, fatty acids, and amino acids; however, differences in the plant hormone pathways

300 were common to both comparisons involving L-20, and differences in secondary metabolite

301 production were seen between the two cultivars of Upland cotton.

302   Inter-cultivar comparisons revealed that the 25 DPA fiber stage mainly exhibited enriched

303 KEGG pathways for DEGs in pathways relating to starch and sucrose metabolism, fatty acid

304 elongation, or biosynthesis of secondary metabolites (Figs. 3D-3F), which could indicate that

305 secondary metabolites, such as biosynthesis of cellulose, was different among the three cultivars.

306 For the fiber samples of 30 DPA, the enrichment classification of DEGs among the three cultivars

307 did not show strong regularity, indicating the complexity of the genotypes among different cotton

308 cultivars (Figs. 3G-3I).

Annotations of DEGs from different fiber development stages

310   To eliminate the effect of genotypic differences, KEGG pathway enrichment was performed

311 on DEGs between different stages of fiber (20 DPA, 25 DPA and 30 DPA) within the same cotton

312 species or cultivar (Fig. 4). The shared top 20 KEGG pathways of enriched DEGs from Xinhai 32

313 (HL) were mainly categorized into the following functional pathways: pyruvate metabolism, fatty

314 acid elongation, biosynthesis of unsaturated fatty acids, valine, leucine and isoleucine degradation,

315 as well as plant hormone signal transduction (Figs. 4A-4C). As to the upland cotton 17-24 (L)

316 (Figs. 4D-4F) and 62-33 (S) (Figs. 4G-4I), the enriched DEGs were also mainly categorized into

317 the fatty acid elongation, biosynthesis of unsaturated fatty acids, valine, leucine and isoleucine

318 degradation, circadian rhythm as well as plant hormone signal transduction pathways. Consistent

319 with the previous results, these results further supported that DEGs enriched in the pathways of

320 carbohydrate, fatty acid and amino acid and hormone were involved in the elongation of fiber cell

321 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 three cotton cultivars

324   To further identify candidate genes contributing to the biosynthesis and thickening of cotton

325 fiber secondary wall, the shared DEGs from different fiber development stages in three cotton

326 cultivars were evaluated. As can be seen from Fig. S2, there were 46 DEGs shared by the three

327 cotton cultivars, Xinhai 32 (HL), 17-24 (L), and 62-33 (S) at different fiber development stages

328 (20, 25, and 30DPA). These 46 recurring DEGs may common to the fiber developmental pathway,

329 regardless of cultivar, perhaps suggesting they could be key genes in the regulation of cotton fiber

330 cell elongation or secondary cell wall thickening. Their main functions include REDOX enzymes,

331 binding proteins, hydrolases (such as GDSL thioesterase), transferases, metalloproteins

332 (cytochromatin-like genes), kinases, carbohydrates, and transcription factors (MYB and WRKY).

333   Detailed analyses were performed on these 46 DEGs according to their expression patterns

334 among the nine samples (Fig. 5). In general, these 46 DEGs can be divided into two categories

335 (Fig. 5): low-to-high expression and high-to-low expression. Most of the DEGs (30) showed low-

336 to-high expression, starting relatively low at 20 DPA but increasing in expression at 25 or 30 DPA

337 in each of the three different species or cultivars (Figs. 5A, 5B). Importantly, the expression levels

338 of these 30 DEGs in HL or L were significantly higher than that in S at 25 or 30 DPA (Figs. 5A,

339 5B), suggesting that these genes might play important roles in the secondary wall synthesis and

340 thickening of fiber tissues. In contrast, 9 of the remaining DEGs exhibited high-to-low expression

341 trends, exhibiting greatest expression at 20 DPA but reducing expression by 25 or 30 DPA (Figs.

342 5C, 5D). These results suggest that these genes may function earlier in fiber development and are

343 downregulated as the cell commits to focused secondary wall synthesis.

Validation of candidate DEGs by RT-qPCR

We confirmed the accuracy of these candidate gene profiles for 11 of the differentially expressed genes (i.e., Polyphenol oxidase 9, 3-ketoacyl-CoA synthase 3, 3-ketoacyl-CoA synthase 4, 3-hydroxyacyl-CoA dehydratase, 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 also showed 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 high-to-low expression pattern with high expression levels at 20 DPA but low expression levels at 25 or 30 DPA (Fig. 6I). The expression patterns of two genes showed no obvious trends (Figs. 6J, 6K). Overall, these transcripts of 11 genes exhibited similar expression patterns between the RT-qPCR and RNA-seq experiments (Fig. 6L), and according to the results of correlation analysis (Table S2), the correlation coefficient between the RT-qPCR and RNA-seq data of 11 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 the Sea Island and upland cotton species, specifically in the extent to which the elongation stage of fiber cell development overlaps with secondary wall thickening stage (Zang et al. 2022; Zhang et al. 2022). The time-course analysis of fiber samples

371 in *G. hirsutum* and *G. barbadense* cultivars also revealed that the glycosyltransferase was involved

372 in the synthesis of glucuronoxylan hemicellulose and cell wall morphogenesis during SCW (Zhang

373 et al. 2022). In this study, we evaluated 20, 25, and 30 DPA fibers from Sea Island and upland

374 cotton cultivars, representing developmental stages of the secondary wall biosynthesis or

375 thickening. We identified DEGs among different cultivars with different fiber properties for these

376 stages of secondary wall synthesis, finding that the differences in gene expression among these

377 cultivars were derived from multiple pathways such as metabolic process, carbohydrate

378 metabolism, fatty acid metabolism and transport. It is worth noting that pathways related to fatty

379 acid metabolism were highly enriched, such as fatty acid carbon chain extension pathway,

380 propionic acid metabolic pathway and unsaturated fatty acid biosynthesis pathway. Their

381 differential expression during the synthesis process of fiber secondary wall suggested that fatty

382 acid metabolism was closely related to fiber quality.

383 Further, a total of 46 genes were screened as candidates that were commonly differentially

384 expressed during the different development stages from three cotton cultivars, including

385 cytochrome P450 enzyme gene, glycohydrolase and glycosyltransferase, binding protein gene,

386 WRKY transcription factor (Fig. 5). Studies have shown that plant cytochrome P450s are involved

387 in the biosynthetic pathways of fatty acid hydroxylation and epoxidation and cleavage of hydrogen

388 peroxide functional groups of unsaturated fatty acids (Davidson et al. 2006). Many previous

389 reports have also reported that ethylene biosynthesis, cytoskeleton, signaling pathway, fatty acid

390 biosynthesis and fatty acid carbon chain extension pathway (Shi et al. 2006) were significantly up-

391 regulated during fiber development (Gou et al. 2007; Ruan et al. 2004). These results further

392 indicated that lipid metabolism was significantly correlated with fiber development.

393 It is well-known that the MYB transcription factor was confirmed to promote the secondary

394 cell-wall biosynthesis (Xiao et al. 2021). A R2R3-MYB transcription factor, GhMYBL1 was

395 specifically expressed in cotton fibers at the stage of secondary wall deposition and participated in

396 modulating the process of secondary wall biosynthesis (Sun et al. 2015). GhMYB7 has been shown

397 to regulate biosynthesis of secondary cell walls of *Arabidopsis thaliana* and upland cotton (Huang

398 et al. 2021b; Huang et al. 2016). In our study, a MYB transcription factor was identified in the

399 highly expressed genes of three cultivars at 25 and 30 DPA fiber, and may be involved in the SCW

400 cellulose biosynthesis. In addition, members of WRKY have been reported to be involved in the

401 fiber development, such as GhWRKY16, and GhWRKY53 (Wang et al. 2021; Yang et al. 2021).

402 Two WRKY genes (WRKY32 and 103) were also detected in our data, and their potential

403 biological regulations in SCW biosynthesis of fiber cells were still not clear, which is worth further

404 research to explore their functions. Thus, we speculated that one factor that cause the difference in

405 fiber strength of the three cultivars might be the distinct expression patterns of genes related to the

406 SCW biosynthesis.

407 Many other important genes were also identified to be specifically differentially expressed in

408 cotton fibers, such as cotton sucrose synthetase gene (Zhang et al. 2017), transcription factor

409 GhMYB2 (Wang et al. 2004), cytoskeletal proteins GhTUB1 and GhACT1, which have been

410 shown to participate in the elongation process of fiber cells (Li et al. 2005). GhACT1 and

411 GhTUB1, which encode actin and tubulin, also affect cytoskeleton assembly and affect fiber

412 elongation (Li et al. 2002; Li et al. 2005). LIM domain protein GhWLIM1a can promote secondary

413 wall synthesis by binding to tubulin (Han et al. 2013). Transcription factors were also involved in

414 regulating secondary wall synthesis of cotton fiber cells. In addition, overexpression of a cotton

415 NAC transcription factor (GhFSN1) resulted in an increase in the thickness of the fiber secondary

416 wall but a decrease in the fiber length (Zhang et al. 2018). Subsequently, a primary GhTCP4

417 transcription factor was found to play an important role in balancing cotton fiber cell elongation

418 and secondary wall thickening (Cao et al. 2020b). Transcriptomic and promoter activity analysis

419 showed that GhTCP4 activated GhFSN1 transcription factor and cellulose synthase genes

420 responsible for secondary wall synthesis. The transcriptional activity of GhCESA8 (GhCESA8)

421 accelerated the biosynthetic pathway of the secondary walls of fiber cells, resulting in shorter fibers

422 and thicker cell walls (Cao et al. 2020b). Recent study about the genetic regulation of fiber

423 development in *G. hirsutum* based on 2,215 time-series transcriptomes also revealed that a

424 NAD(P)-linked oxidoreductase protein, was positively correlated with fiber strength development

425 (You et al. 2023). In our study, NAC52 transcription factor, and NAD(P)-binding Rossmann-fold

426 superfamily protein, have also been identified to might be involved in the regulation of cotton fiber

427 secondary wall development. These results indicated that multiple pathway related genes play roles

428 in the biosynthesis of fiber secondary walls.

429

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

Author contributions

432 LL, YY, and ZWC conceived and designed this experiment. LL, XHK, XWW, AJS and JW collected samples

433 and performed the study. LL, CEG, JJ, and ZWC participated in the acquisition and analysis of the data. LL and

434 ZWC wrote the manuscript. LL and YY participated in the discussion draft of the manuscript. ZWC revised the

435 final manuscript. All authors read and approved the final manuscript.

439 **Abbreviations**

440 DEGs, differentially expressed genes; FPKM, fragments per kilobase of transcript per million mapped fragments; *G. hirsutum*,

441 *Gossypium hirsutum*; *G. barbadense*, *Gossypium barbadense*; *A. thaliana*, *Arabidopsis thaliana*; RT-qPCR, real-time quantitative

442 polymerase chain reaction.

443

References

445 Bajwa KS, Shahid AA, Rao AQ, Bashir A, Aftab A, and Husnain T. 2015. Stable transformation and expression of

446 GhEXPA8 fiber expansin gene to improve fiber length and micronaire value in cotton. *Front Plant Sci*

447 6:838. 10.3389/fpls.2015.00838

448 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

449 CE, and Chen Z-W. 2020a. Genome-wide characterization of the GRF family and their roles in response to

450 salt stress in Gossypium. *BMC Genomics* 21:575. 10.1186/s12864-020-06986-0

451 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.

452 Comparative genomics and functional studies of putative m(6)A methyltransferase (METTL) genes in

453 cotton. *Int J Mol Sci* 23. 10.3390/ijms232214111

454   Cao JF, Zhao B, Huang CC, Chen ZW, Zhao T, Liu HR, Hu GJ, Shangguan XX, Shan CM, Wang LJ, Zhang TZ, Wendel JF,
455   Guan XY, and Chen XY. 2020b. The miR319-targeted GhTCP4 promotes the transition from cell elongation
456   to wall thickening in cotton fiber. *Mol Plant* 13:1063-1077. 10.1016/j.molp.2020.05.006
457   Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, and Xia R. 2020. TBtools: An Integrative Toolkit Developed
458   for Interactive Analyses of Big Biological Data. *Mol Plant* 13:1194-1202. 10.1016/j.molp.2020.06.009
459   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
460   graphene on the multiple metabolic pathways of Zea mays roots based on transcriptome analysis. *PLoS*
461   *One* 16:e0244856. 10.1371/journal.pone.0244856
462   Chen ZW, Cao JF, Zhang XF, Shangguan XX, Mao YB, Wang LJ, and Chen XY. 2017. Cotton genome: challenge into
463   the polyploidy. *Sci Bull (Beijing)* 62:1622-1623. 10.1016/j.scib.2017.11.022
464   Cui W, Chen Z, Shangguan X, Li T, Wang L, Xue X, and Cao J. 2022. TRY intron2 determined its expression in
465   inflorescence activated by SPL9 and MADS-box genes in Arabidopsis. *Plant Science* 321:111311.
466   https://doi.org/10.1016/j.plantsci.2022.111311
467   Davidson SE, Reid JB, and Helliwell CA. 2006. Cytochromes P450 in gibberellin biosynthesis. *Phytochemistry*
468   *Reviews* 5:405-419. 10.1007/s11101-006-9005-5
469   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
470   dynamic transcriptome reveals the delayed secondary-cell-wall thickening results in altered lint
471   percentage and fiber elongation in a chromosomal segment substitution line of cotton (*Gossypium*
472   *hirsutum* L.). *Front Plant Sci* 12:756434. 10.3389/fpls.2021.756434
473   García-Ayuso LE, Velasco J, Dobarganes MC, and Luque de Castro MD. 2000. Determination of the oil content of
474   seeds by focused microwave-assisted soxhlet extraction. *Chromatographia* 52:103-108.
475   10.1007/BF02490801
476   Glover BJ. 2000. Differentiation in plant epidermal cells. *J Exp Bot* 51:497-505. 10.1093/jexbot/51.344.497
477   Gou JY, Wang LJ, Chen SP, Hu WL, and Chen XY. 2007. Gene expression and metabolite profiles of cotton fiber
478   during cell elongation and secondary cell wall synthesis. *Cell Res* 17:422-434. 10.1038/sj.cr.7310150
479   Guan X, Yu N, Shangguan X, Wang S, Lu S, Wang L-J, and Chen X-Y. 2007. Arabidopsis trichome research sheds light
480   on cotton fiber development mechanisms. *Chinese Science Bulletin* 52:1734-1741.
481   Haigler CH, Betancur L, Stiff MR, and Tuttle JR. 2012. Cotton fiber: a powerful single-cell model for cell wall and
482   cellulose research. *Front Plant Sci* 3:104. 10.3389/fpls.2012.00104
483   Haigler CH, Singh B, Wang G, and Zhang D. 2009. Genomics of cotton fiber secondary wall deposition and cellulose
484   biogenesis. In: Paterson AH, ed. *Genetics and Genomics of Cotton*. New York, NY: Springer US, 385-417.
485   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
486   of WLIM1a in cell elongation and secondary wall formation in developing cotton fibers. *Plant Cell* 25:4421-
487   4438. 10.1105/tpc.113.116970
488   Hof J, and Saha S. 1997. Cotton fibers can undergo cell division. *Am J Bot* 84:1231.
489   Hsieh Y-L, Honik E, and Hartzell MM. 1995. A developmental study of single fiber strength: greenhouse grown SJ-2
490   Acala cotton. *Textile Research Journal* 65:101-112. 10.1177/004051759506500206
491   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
492   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
493   F, Bai Y, Zhang Y, Brodt A, Ben-Hamo H, Zhu X, Zhou B, Guan X, Zhu S, Chen X, and Zhang T. 2019.
494   *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of

495   allotetraploid cotton. *Nat Genet* 51:739-748. 10.1038/s41588-019-0371-5

496   Huang G, Huang JQ, Chen XY, and Zhu YX. 2021a. Recent advances and future perspectives in cotton research.

497   *Annu Rev Plant Biol* 72:437-462. 10.1146/annurev-arplant-080720-113241

498   Huang J, Chen F, Guo Y, Gan X, Yang M, Zeng W, Persson S, Li J, and Xu W. 2021b. GhMYB7 promotes secondary

499   wall cellulose deposition in cotton fibres by regulating GhCesA gene expression through three distinct cis-

500   elements. *New Phytol* 232:1718-1737. 10.1111/nph.17612

501   Huang J, Chen F, Wu S, Li J, and Xu W. 2016. Cotton GhMYB7 is predominantly expressed in developing fibers and

502   regulates secondary cell wall biosynthesis in transgenic Arabidopsis. *Sci China Life Sci* 59:194-205.

503   10.1007/s11427-015-4991-4

504   Jareczek JJ, Grover CE, and Wendel JF. 2023. Cotton fiber as a model for understanding shifts in cell development

505   under domestication. *Front Plant Sci* 14:1146802. 10.3389/fpls.2023.1146802

506   Kelly BR, Abidi N, Ethridge D, and Hequet EF. 2015. Fiber to Fabric.

507   Kim D, Langmead B, and Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature*

508   *Methods* 12:357-360. 10.1038/nmeth.3317

509   Kim HJ. 2018. Cotton fiber biosynthesis. In: Fang DD, ed. *Cotton Fiber: Physics, Chemistry and Biology*. Cham:

510   Springer International Publishing, 133-150.

511   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

512   in fiber elongation between the offspring of Sea Island cotton and Upland cotton backcross populations

513   based on transcriptome and weighted gene coexpression networks. *Genes (Basel)* 13.

514   10.3390/genes13060954

515   Li X-B, Cai L, Cheng N-H, and Liu J-W. 2002. Molecular Characterization of the Cotton GhTUB1 Gene That Is

516   Preferentially Expressed in Fiber. *Plant Physiology* 130:666-674. 10.1104/pp.005538

517   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

518   and participates in fiber elongation. *The Plant Cell* 17:859-875. 10.1105/tpc.104.029629

519   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-

520   physiological differences among transgenic and no-transgenic cotton cultivars. *Plants* 12:3437.

521   10.3390/plants12193437

522   Liu Y. 2013. Recent progress in fourier transform infrared (FTIR) spectroscopy study of compositional, structural

523   and physical attributes of developmental cotton fibers. *Materials (Basel)* 6:299-313. 10.3390/ma6010299

524   Livak KJ, and Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and

525   the 2(T)(-Delta Delta C) method. *Methods* 25:402-408. 10.1006/meth.2001.1262

526   Love MI, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with

527   DESeq2. *Genome Biol* 15:550. 10.1186/s13059-014-0550-8

528   Mansoor S, and Paterson AH. 2012. Genomes for jeans: cotton genomics for engineering superior fiber. *Trends*

529   *Biotechnol* 30:521-527. S0167-7799(12)00094-7 [pii]

530   10.1016/j.tibtech.2012.06.003

531   Mao X, Cai T, Olyarchuk JG, and Wei L. 2005. Automated genome annotation and pathway identification using the

532   KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21:3787-3793.

533   10.1093/bioinformatics/bti430

534   Meinert MC, and Delmer DP. 1977. Changes in biochemical composition of the cell wall of the cotton fiber during

535   development. *Plant Physiol* 59:1088-1097. 10.1104/pp.59.6.1088

536 Pertea M, Kim D, Pertea GM, Leek JT, and Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq
537   experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11:1650-1667. 10.1038/nprot.2016.095
538 Qin Y, Sun H, Hao P, Wang H, Wang C, Ma L, Wei H, and Yu S. 2019. Transcriptome analysis reveals differences in
539   the mechanisms of fiber initiation and elongation between long- and short-fiber cotton (Gossypium
540   hirsutum L.) lines. *BMC Genomics* 20:633. 10.1186/s12864-019-5986-5
541 Ruan YL, Xu SM, White R, and Furbank RT. 2004. Genotypic and developmental evidence for the role of
542   plasmodesmatal regulation in cotton fiber elongation mediated by callose turnover. *Plant Physiol*
543   136:4104-4113. 10.1104/pp.104.051540
544 Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ, Zhou BL, Hu GJ,
545   Guan XY, Chen ZJ, Wendel JF, Zhang TZ, and Chen XY. 2014. Control of cotton fibre elongation by a
546   homeodomain transcription factor GhHOX3. *Nat Commun* 5:5519. ncomms6519 [pii]
547 10.1038/ncomms6519
548 Shang L, Abduweli A, Wang Y, and Hua J. 2016. Genetic analysis and QTL mapping of oil content and seed index
549   using two recombinant inbred lines and two backcross populations in Upland cotton. *Plant Breeding*
550   135:224-231. https://doi.org/10.1111/pbr.12352
551 Shang L, Liang Q, Wang Y, Wang X, Wang K, Abduweli A, Ma L, Cai S, and Hua J. 2015. Identification of stable QTLs
552   controlling fiber traits properties in multi-environment using recombinant inbred lines in Upland cotton
553   (*Gossypium hirsutum* L.). *Euphytica* 205:877-888. 10.1007/s10681-015-1434-z
554 Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, Cheng J, Wei LP, Wang ZY, and Zhu YX. 2006. Transcriptome
555   profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber
556   cell elongation. *Plant Cell* 18:651-664. 10.1105/tpc.105.040303
557 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
558   association study of micronaire using a natural population of representative upland cotton (Gossypium
559   hirsutum L.). *Journal of Cotton Research* 4:14. 10.1186/s42397-021-00089-1
560 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
561   specifically expressed in cotton (Gossypium hirsutum) fibers affects secondary cell wall biosynthesis and
562   deposition in transgenic Arabidopsis. *Physiologia Plantarum* 154:420-432.
563   https://doi.org/10.1111/ppl.12317
564 Tian Y, and Zhang T. 2021. MIXTAs and phytohormones orchestrate cotton fiber development. *Curr Opin Plant Biol*
565   59:101975. 10.1016/j.pbi.2020.10.007
566 Tuttle JR, Nah G, Duke MV, Alexander DC, Guan X, Song Q, Chen ZJ, Scheffler BE, and Haigler CH. 2015.
567   Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis,
568   represses lignification, and prolongs elongation. *BMC Genomics* 16:477. 10.1186/s12864-015-1708-9
569 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,
570   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,
571   Lindsey K, Lin Z, Udall JA, and Zhang X. 2019. Reference genome sequences of two cultivated
572   allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. *Nat Genet* 51:224-229.
573   10.1038/s41588-018-0282-x
574 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
575   essential for its transcriptional activity during fiber initiation and elongation in cotton (*Gossypium*
576   *hirsutum*). *Plant Cell* 33:2736-2752. 10.1093/plcell/koab153

577   Wang S, Wang JW, Yu N, Li CH, Luo B, Gou JY, Wang LJ, and Chen XY. 2004. Control of plant trichome development
578   by a cotton fiber MYB gene. *Plant Cell* 16:2323-2334. 10.1105/tpc.104.024844
579   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,
580   Zhang X, Chen X, and Zhu Y. 2023. A comprehensive overview of cotton genomics, biotechnology and
581   molecular biological studies. *Sci China Life Sci* 66:2214-2256. 10.1007/s11427-022-2278-0
582   Xiao R, Zhang C, Guo X, Li H, and Lu H. 2021. MYB transcription factors and its regulation in secondary cell wall
583   formation and lignin biosynthesis during xylem development. *Int J Mol Sci* 22. 10.3390/ijms22073560
584   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.
585   Deep transcriptome analysis reveals reactive oxygen species (ROS) network evolution, response to abiotic
586   stress, and regulation of fiber development in cotton. *Int J Mol Sci* 20. 10.3390/ijms20081863
587   Yang D, Liu Y, Cheng H, Wang Q, Lv L, Zhang Y, Song G, and Zuo D. 2021. Identification of the Group III WRKY
588   subfamily and the functional analysis of GhWRKY53 in *Gossypium hirsutum* L. Plants.
589   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,
590   Liu D, Zhang J, and Zhang Z. 2022. Identification of candidate genes for lint percentage and fiber quality
591   through QTL mapping and transcriptome analysis in an allotetraploid interspecific cotton CSSLs
592   population. *Front Plant Sci* 13:882051. 10.3389/fpls.2022.882051
593   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
594   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
595   Wang M. 2023. Regulatory controls of duplicated gene expression during fiber development in
596   allotetraploid cotton. *Nature Genetics* 55:1987-1997. 10.1038/s41588-023-01530-8
597   Young MD, Wakefield MJ, Smyth GK, and Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for
598   selection bias. *Genome Biology* 11:R14. 10.1186/gb-2010-11-2-r14
599   Zang Y, Hu Y, Dai F, and Zhang T. 2022. Comparative transcriptome analysis reveals the regulation network for fiber
600   strength in cotton. *Biotechnol Lett* 44:547-560. 10.1007/s10529-022-03236-z
601   Zhang J, Huang GQ, Zou D, Yan JQ, Li Y, Hu S, and Li XB. 2018. The cotton (*Gossypium hirsutum*) NAC transcription
602   factor (FSN1) as a positive regulator participates in controlling secondary cell wall biosynthesis and
603   modification of fibers. *New Phytol* 217:625-640. 10.1111/nph.14864
604   Zhang J, Mei H, Lu H, Chen R, Hu Y, and Zhang T. 2022. Transcriptome time-course analysis in the whole period of
605   cotton fiber development. *Front Plant Sci* 13:864529. 10.3389/fpls.2022.864529
606   Zhang X, Cao J, Huang C, Zheng Z, Liu X, Shangguan X, Wang L, Zhang Y, and Chen Z. 2021a. Characterization of
607   cotton ARF factors and the role of GhARF2b in fiber development. *BMC Genomics* 22:202.
608   10.1186/s12864-021-07504-6
609   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.
610   Structural insights into homotrimeric assembly of cellulose synthase CesA7 from Gossypium hirsutum.
611   *Plant Biotechnol J* 19:1579-1587. 10.1111/pbi.13571
612   Zhang Z, Ruan YL, Zhou N, Wang F, Guan X, Fang L, Shang X, Guo W, Zhu S, and Zhang T. 2017. Suppressing a
613   putative sterol carrier gene reduces plasmodesmal permeability and activates sucrose transporter genes
614   during cotton fiber elongation. *Plant Cell* 29:2027-2046. 10.1105/tpc.17.00358
615
616

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

644   DEGs from S-20 vs S-30, (I) Top 20 pathways of significantly enriched DEGs from S-25 vs S-30.

645   **Figure 5 Expression patterns of DEGs among nine fiber samples.** (A) 23 DEGs with low expression at 20

646   DPA but high expression at 25 or 30 DPA (i.e., low-to-high expression), (B) 7 additional DEGs with low-to-

647   high expression, (C) 5 DEGs with high expression at 20 DPA but low expression at 25 or 30 DPA (i.e., high-to-

648   low expression), (D) 4 additional DEGs showing high-to-low expression. Note: HL represents Xinhai 32 (Sea

649   Island cotton), L represents 17-24 (Upland cotton), and S represents 62-33 (Upland cotton). DPA: days post

650   anthesis.

651   **Figure 6 Real-time quantitative PCR validation of DEGs from RNA-seq data.** Relative expression of (A)

652   *Polyphenol oxidase-9*, (B) *WRKY32*, (C) *GDSL-like lipase*, (D) *Caffeic acid* O-*methyltransferase 1*, (E) *NAD(P)-*

653   *binding Rossmann-fold superfamily protein*, (F) *2,4-dienoyl-CoA reductase*, (G) *WRKY103*, (H) *Fatty acid*

654   *desaturase 6*, (I) *3-ketoacyl-CoA synthase 3*, (J) *3-ketoacyl-CoA synthase 4*, (K) *3-hydroxyacyl-CoA dehydratase*

655   genes at 20, 25 or 30 DPA fiber cells of three cotton cultivars, and the expression level in the HL-20 sample was

656   set to 1 (means of triplicates ± SD), (L) The expression patterns of 11 candidate genes from RNA-seq data.

657   Note: HL represents Xinhai 32 (Sea Island cotton), L represents 17-24 (Upland cotton), and S represents 62-33

658   (Upland cotton). Relative gene expression levels are normalized to histone-3 gene values. Error bars indicate SD

659   (n = 3). Statistically significant differences ("a" is different from "b", "c", "d", "e" "f", "g", "h", or "i", α = 0.05

660   level) of expression values are indicated with different letters with analysis of variance in R (https://www.r-

661   project.org/).

662   **Table 1.** The main properties of different cotton cultivars.

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

Supporting information

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

666   **TABLE S2**. The correlation coefficient between the RT-qPCR and RNA-seq data of 11 genes.

667   Fig. S1 Gene ontology (GO) enrichment analysis of DEGs in 20-30 DPA fiber samples between Sea Island or

668   Upland cottons. The X-axis represents the biological functions (molecular function, biological process, and

669   cellular component) of these DEGs. The Y-axis represents the percentage or number of genes categorized into
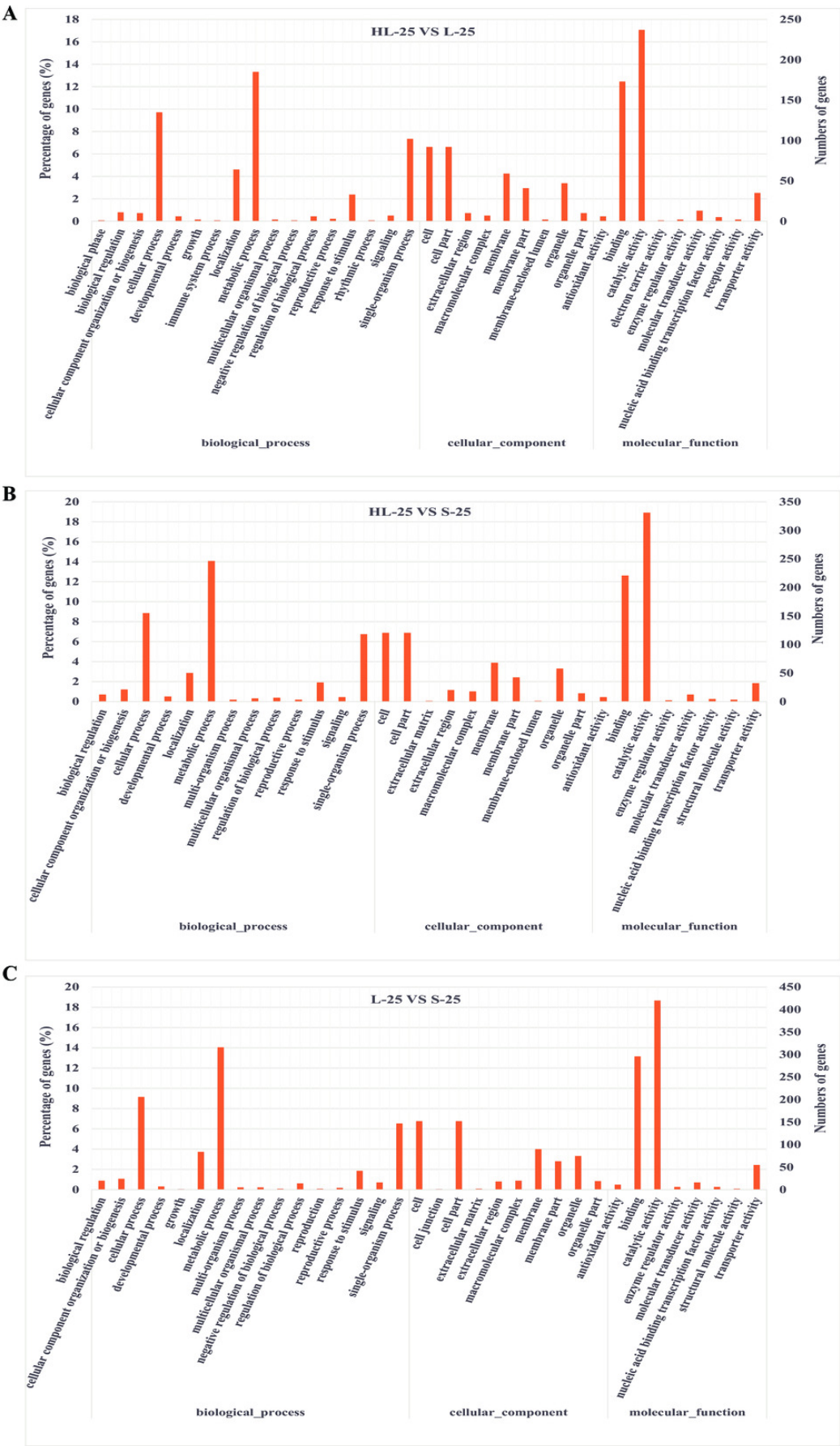
670   different functional pathways.

671   Fig. S2 Venn diagram comparison of differentially expressed genes (DEGs) from different fiber development

672   stages in three cotton cultivars.

673

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.
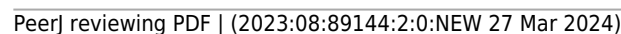
Figure 4

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.
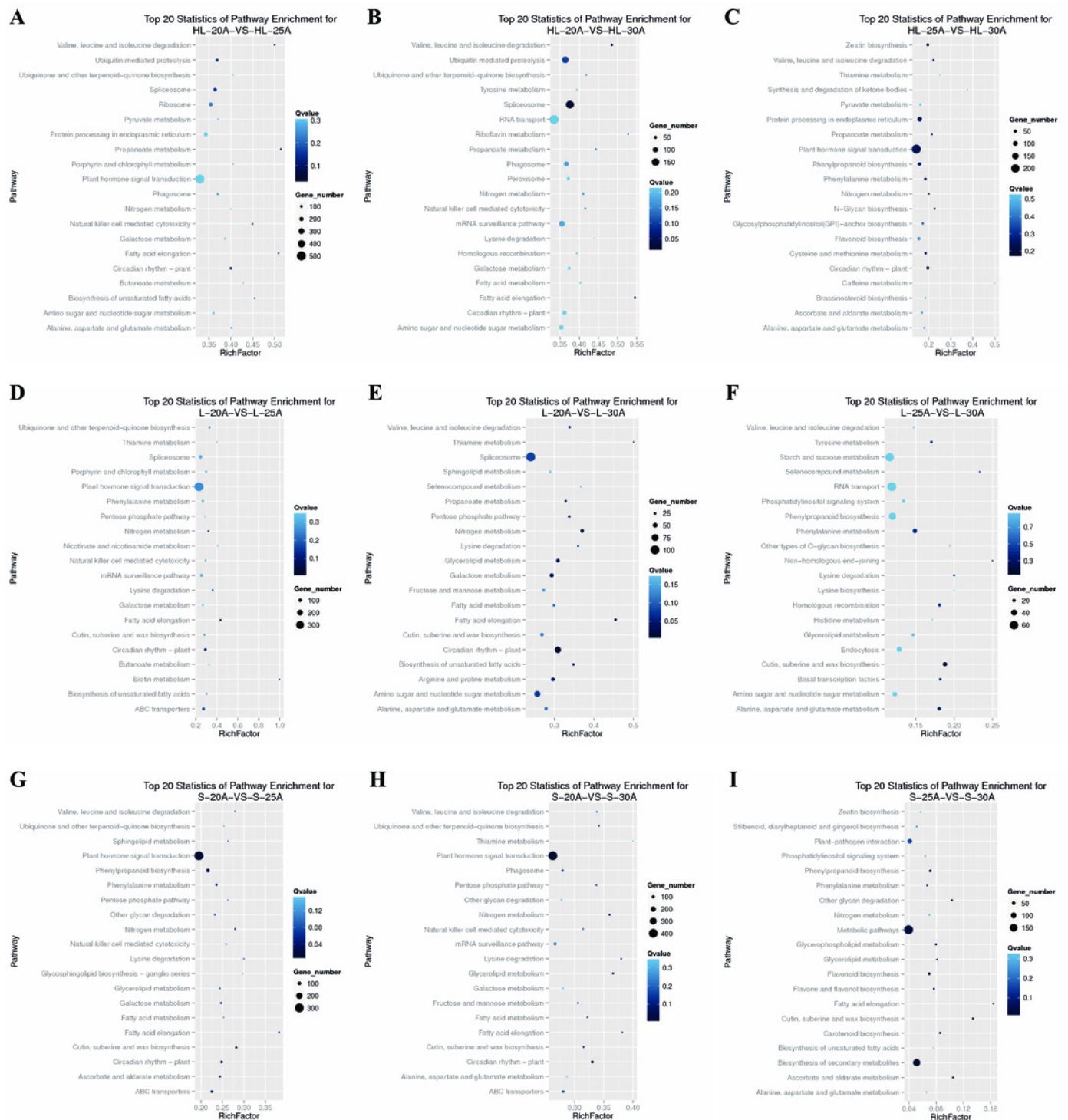
Figure 5

Expression patterns of DEGs among 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.
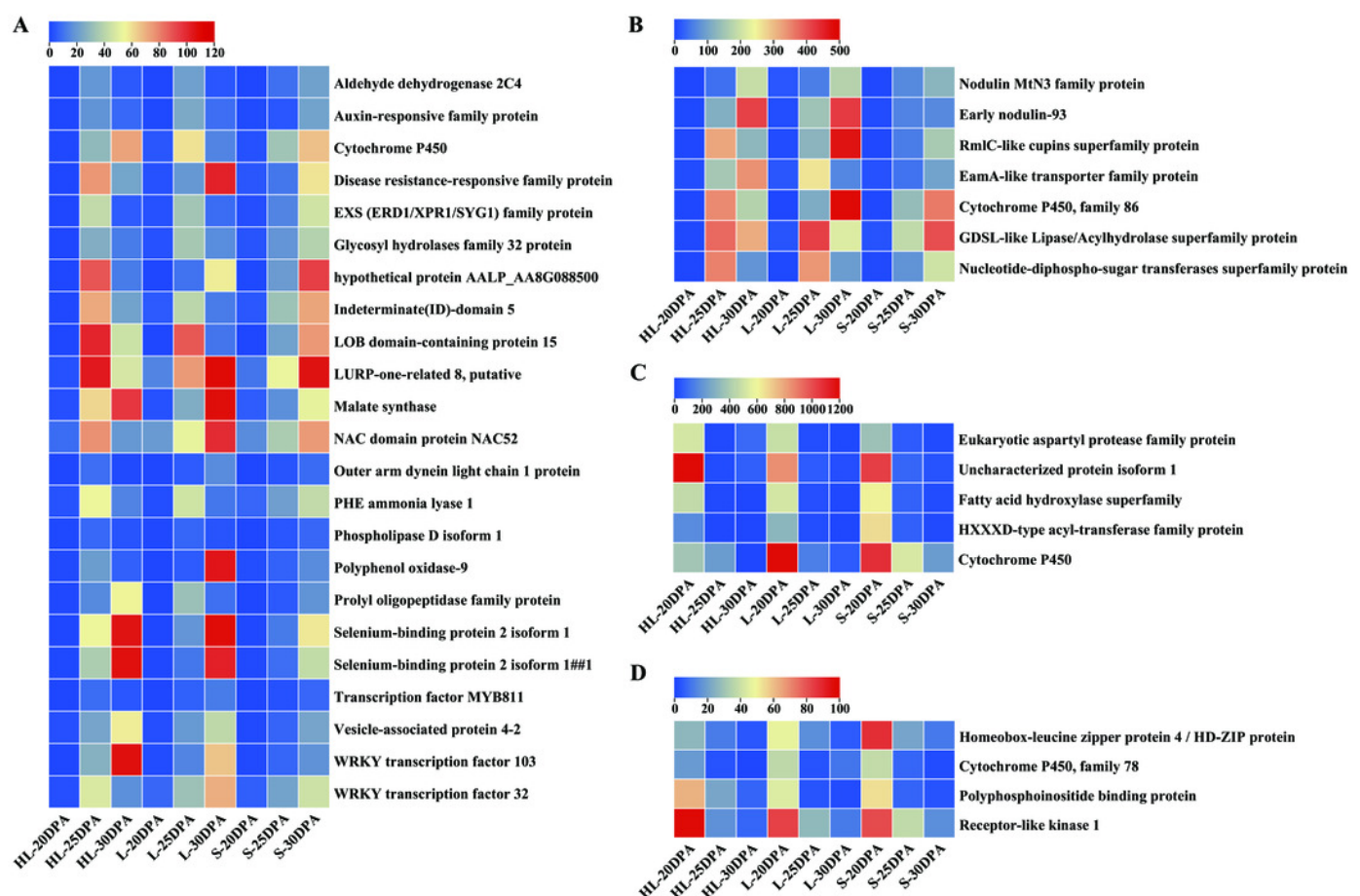
Figure 6

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*, (J) *3-ketoacyl-CoA synthase 4*, (K) *3-hydroxyacyl-CoA dehydratase* genes at 20, 25 or 30 DPA fiber cells of three cotton cultivars, and the expression level in the HL-20 sample was set to 1 (means of triplicates ± SD), (L) The expression patterns of 11 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). 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/).
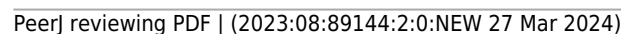
Table 1(on next page)

The main properties of different cotton cultivars.

The main properties of different cotton cultivars.

1 **Table 1.** The main properties of different cotton cultivars.

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

2 Note: Error bars indicate SD ($n = 3$). Statistically significant differences ("a" is different from "b" or "c", $\alpha = 0.05$ level) of values

3 are indicated with different letters with analysis of variance in R (https://www.r-project.org/). Micronaire value is a comprehensive

4 index reflecting the fineness and maturity of cotton fiber. Micronaire is divided into three levels: A, B and C, with B being the

5 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.

6 Grade C is below to 3.4 or above to 5.0, showing the worst quality.

7

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.

1   **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 %

2
3