A global survey of the gene network and key genes for oil accumulation in cultivated tetraploid cottons

De Zhu, Yu Le, Ruiting Zhang, Xiaojing Li and Zhongxu Lin*

National Key Laboratory of Crop Genetic Improvement, College of Plant Sciences & Technology, Huazhong Agricultural University, Wuhan, Hubei, China

Received 14 October 2020; revised 3 December 2020; accepted 20 December 2020.

*Correspondence (Tel: +86-02787280196; fax: +86-02787280196; email: linzhongxu@mail.hzau.edu.cn)

Summary
To enrich our knowledge about gene network of fatty acid biosynthesis in cottonseed, we conducted comparative transcriptome to reveal the differences in gene expression between Gossypium hirsutum and Gossypium barbadense during cottonseed development. The prolonged expression period and increased expression abundance of oil-related genes are the main reasons for producing high seed oil content (SOC) in G. barbadense, which manifested as the bias of homeologous gene expression in Dt-subgenome after 25 day postanthesis (DPA). The dynamic expression profile showed that SAD6 and FATA are more important for oil biosynthesis in G. barbadense than that in G. hirsutum. Three key transcription factors, WR11, NF-YB6 and DDBF2, showed their elite roles in regulating seed oil in cotton. We observed that sequence variations in the promoter region of BCCP2 genes might contribute to its divergence in expression level between the two species. Based on the quantitative trait loci (QTL) information of the seed oil content and utilizing additional G. barbadense introgression lines (ILs), we propose 21 candidate genes on the basis of their differential expression level, of which the GbsSWEET and the GbACBP6 showed the potential functional to improve the oil content. Taken together, studying the different expression of oil-related genes and their genetic regulation mechanisms between G. hirsutum and G. barbadense provide new insights to understanding the mechanism of fatty acid biosynthesis network and fatty acid genetic improving breeding in cotton.

Introduction
Worldwide increasing demand for vegetable oil, especially those with high nutritional profile, requires continuous breeding efforts to broaden the sources and increase the yield of oil crop. Cotton (Gossypium), as an important cash crop in the world, provides not only the main natural fibre for the textile industry, but also the 2nd major source of vegetable oil (Sharif et al., 2019). Cottonseed oil contains a large amount of unsaturated fatty acids, of which more than 50% belongs to linoleic acid (C18:2), and also rich in palmitic acid (C16:0) and oleic acid (C18:1) (Konuskan et al., 2017). The higher proportion of polyunsaturated fatty acids in cottonseed oil is beneficial to human health. In addition, the suitable length of carbon chain in cottonseed oil is considered to be an ideal raw material of biofuel (Yesilyurt and Aydin, 2020). Modern upland cotton (G. hirsutum) cultivars are the result of long-term selection and domestication (Grover et al., 2015; Senchina et al., 2003), which contain great variability of cottonseed oil content (SOC) ranging from 12.60% to 39.33% (Hinze et al., 2015; Zhao et al., 2019). However, low genetic diversity limits the potential for genetic breeding within species to improve SOC (Kaur et al., 2017; Tyagi et al., 2014). The G. barbadense has often been used as an important germplasm donor to improve yield (Li et al., 2016), fibre quality (Jenkins et al., 2018) and disease resistance in G. hirsutum (Zhao et al., 2018a).

In cotton, the interspecific introgression line (IL) population of G. hirsutum × G. barbadense has frequently been used to identify genomic regions associated with interested quantitative traits (Zhang et al., 2014). However, few studies to date have been aimed at using the G. barbadense to improve the SOC in G. hirsutum background. Studying the correlations between cottonseed traits, Wu et al. (2009) and Yu et al. (2012) showed the potential value of G. barbadense in improving SOC in G. hirsutum background. In addition, the survey of seed traits in cotton germplasm resources performed by Hinze et al. (2015) and Shockey et al. (2016) also confirmed that the G. barbadense accessions have the higher SOC characteristic. Hence, studying the biosynthesis in G. barbadense and revealing its genetic regulation mechanism represent an interesting goal in cotton oil breeding.

Lipids and fatty acids (FAs) are a large family in plant, while most edible vegetable oil consists of a few common FAs, including saturated, monounsaturated, and polyunsaturated fatty acids, and stored in seeds in form of triacylglycerol (TAG) on usual (Troncoso et al., 2011). The TAG biosynthesis pathway has become one of the hallmarks of lipid biochemistry, which controlled by a series of functional genes (Bates et al., 2013). As next-generation high-throughput sequencing has become routine, several studies in oil crop shed new light on understanding the TAG biochemical pathways involved in oilseeds at genomic and transcriptome levels (Gusev et al., 2016). De novo transcriptome sequencing analysis in Arabidopsis (Belmonte et al.,...
rapeseed (Deng et al., 2015; Wan et al., 2017), soybean (Jones and Vodkin, 2013), camelina (Abdullah et al., 2016), peanut (Liu et al., 2019) and sesame (Wang et al., 2014) not only depicts the expression patterns of lipid genes during the development of oilseeds, but reveals the molecular mechanisms of TAG metabolism and highlights the key enzymes for lipid storage with different oil components (Baud and Lepiniec, 2010). Genetic engineering of those rate-limiting enzymes involved in TAG has been successful in regulating the levels and composition of FAs in upland cotton. RNA interference (RNAi) of D12 fatty acid desaturase (FAD2), D9 stearoyl-ACP desaturase (SAD1) and β-ketoacyl-acyl carrier protein synthase (KASII) in cotton dramatically increased the accumulation of oleic acids, stearic acids and palmitic acid in cotton seed, respectively (Liu et al., 2002, 2017b). In addition, RNAi down-regulation of phosphoenolpyruvate carboxyl (PEPC2) involved in tricarboxylic acid (TCA) cycle in cottonseeds up-regulated most lipid synthesis-related genes and increased 7.3% of oil content in seed kernel (Zhao et al., 2018b). The global transcriptome profile of development seeds further enriched our understanding of the important role of these functional genes in regulating lipids biosynthesis in cotton (Hovav et al., 2015; Zhao et al., 2018c). In forward genetic analysis, a large number of oil-related genes also have been identified by genome-wide association study, such as fatty acyl-ACP thioesterase B (FATB), acyl carrier protein S (ACPS) (Yuan et al., 2018) and KAS III (Du et al., 2018). Additionally, the calcium-dependent lipid-binding (CaLB) family protein (Zhao et al., 2019) and the peroxidase gene (PRXR1) (Ma et al., 2019) were also considered to be involved in improving SOC in cotton, which implied that there may be more complex of the genetic regulation of SOC in cotton.

However, previous studies mainly focused on the upland cotton, the biosynthesis of lipids and their regulatory mechanisms associated with high SOC formation in G. barbadense remain unclear. Elucidation of high-quality cotton genome (G. hirsutum and G. barbadense) (Wang et al., 2019), and the availability of IL population derived from interspecific crossing allowed the unprecedented opportunity to study the lipids biosynthesis to understand its genetic mechanism of higher SOC in G. barbadense. In our previous research (Zhu et al., 2020), via analysing the genetic effects of interspecific IL population, we found the amazing performance of SOC and identified a lot of related QTL. Therefore, the aim of present study was to explore the mechanism of higher SOC in G. barbadense by analysing the differentially expressed genes during seed development and to propose associated candidate genes using interspecific IL population.

Results
Phenotype variation and transcriptome assembly
Based on the genotype of IL population identified in our previous research (Zhu et al., 2020), we selected the high SOC lines with inconsistent introgression fragments. The phenotype of SOC in these lines showed significantly differences from their recipient parent G. hirsutum cv. ‘Emian22’ (P < 0.001), but similar to the donor parent G. barbadense acc. 3-79 (3-79) (Figure S1), which indicated that there may be multiple genetic loci controlling high SOC in G. barbadense.

In order to understand the underlying molecular differences of SOC formation between G. hirsutum and G. barbadense, we selected three stages of whole seeds for transcriptome sequencing to obtain a global view of gene expression during seed development. About 9 Gb transcriptome data were generated from each library, and at least 53 million paired-end reads were obtained (Table S1). Clean reads from Emian22 and 3-79 were mapped to the G. hirsutum and G. barbadense reference genome, respectively. Nine libraries of gene expression data from different stages were clustered together concordantly in the principal component analysis (PCA), which showed that these data were appropriated for comparative analysis in developing seed (Figure S2a).

A total of 26,528 and 29,306 genes were expressed in G. hirsutum and G. barbadense during seed development, respectively, accounting for 37.79% and 39.70 of total genes in each genome (Figure S2b). Notably, the number of expressed genes decreased during seed development in both species, but the time-specifically expressed genes in G. hirsutum was significantly higher than that in G. barbadense at 30 DPA (1361 vs 789) (Figure S2c).

Estimation of differentially expressed genes (DEGs)
The time-series differential expression analysis helps us to understand how the different gene expression levels affect embryo development. During seed development, a total of 25,074 and 22,390 genes were differentially expressed in G. hirsutum and G. barbadense, respectively. There was no significant difference in the up- or down-regulated genes between the two species in the comparison of ‘20 DPA vs 10 DPA’. Notably, the huge difference was observed in the ‘30 DPA vs 20 DPA’ comparison, only 673 DEGs were up-regulated in G. barbadense comparing 3082 DEGs in G. hirsutum, and 1,860 DEGs were down-regulated in G. barbadense comparing 4885 DEGs in G. hirsutum (Figure 1a). This indicated that a large number of genes may still had higher expression levels from 20 DPA to 30 DPA in G. barbadense compared with G. hirsutum, which is consistent with the result that the number of up-regulated genes in G. barbadense is more than that in G. hirsutum in the comparison of ‘30 DPA vs 10 DPA’. A previous study reported that the oil content in cottonseed rapidly increased from 20 DPA to 30 DPA (Hovav et al., 2015), so we analysed the specific DEGs in the comparison ‘30 DPA vs 20 DPA’. We found that only 352 DEGs and 766 DEGs were specifically up- and down-regulated in the G. barbadense from 20 DPA to 30 DPA. However, a total of 3775 DEGS with 1662 DEGs up-regulated and 2,113 DEGs down-regulated were specifically detected in G. hirsutum from 20 DPA to 30 DPA (Figure 1b). Down-regulated expression of a large number of genes during oil accumulation may result in the low SOC in the G. hirsutum.

To comprehend the seed development processes, Gene Ontology (GO) categories were performed to classify the functional interpretation and gene product attributes of these DEGs. Under the biological processes, most of the DEGs were concentrated on the similar potential functions between species during seed development. A total of 1104 and 1118 genes were annotated to the lipid metabolic process (GO: 0006629) in the G. hirsutum and G. barbadense reference genome, respectively. However, no significantly enrichment was observed that the DEGs were involved in lipid-related biological process, suggesting that the high oil content in cottonseed was regulated by a few key genes (Tables S2 and S3). Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was used to categorize function annotations for all DEGs and summarized in Table S4 and Table S5. During the time phase from 10 DPA to 20 DPA, 91 DEGs classified in the lipid metabolism were up-regulated in...
G. barbadense, and only 61 up-regulated DEGs were detected in G. hirsutum (Figure S3). Notably, the types of lipid metabolism pathway are not different with the number of expressed genes increased (Table S5). Interestingly, in the group of down-regulated DEGs in the ‘30 DPA vs 20 DPA’, more lipid-related DEGs down-regulated in G. hirsutum than that in G. barbadense (79 vs 16) (Figure S4). Among those, we found four significant enrichment pathways involved in lipid metabolism in G. hirsutum, including fatty acid biosynthesis, fatty acid metabolism, glycerolipid metabolism and biosynthesis of unsaturated fatty acids, while no relevant pathway was observed in G. barbadense (Figure S5). The prolonged expression period of lipid-related genes may played critical roles in determining the high SOC in G. barbadense.

To further categorize the DEGs of these oil-related genes, a hierarchical clustering analysis was performed on the expression value fragments per kilobase of exon per million mapped reads (FPKM) and visualized in a heatmap. Indeed, functional enrichment revealed that most of genes enriched with fatty acid elongation pathway and fatty acid biosynthesis processes continued to be highly expression from 20 DPA to 30 DPA in G. barbadense (Figure S6). In the meantime, the alpha-linolenic acid metabolism- and linolenic acid metabolism-related genes were highly activated at late stage in both species, which emphasized the prevalence of the high unsaturated fatty acid synthesis content in cottonseed.

Dynamic changes of expression profiling during oil accumulation

To investigate the potential mechanism underlying the variation in SOC, we compared the oil-related genes expression during seed development between the two species, with more attention being paid to the TAG biosynthesis. According to the previous report (You et al., 2016), a total of 374 genes involved in fatty acid metabolism were identified in both species (Tables S6 and S7). Yet, only 29.68% and 28.17% of them were highly expressed (FPKM ≥ 10 in at least one stage) during seed development in G. hirsutum and G. barbadense, respectively. And among these genes, most of highly expressed genes are on orthologous positions in the genome between the two species.

Based on the reference oil biosynthesis model of upland cotton (Hovav et al., 2015; Zhao et al., 2018c), the dynamic expression profile was modified and is shown in Figure 2. Overall, the results indicated higher regulated levels of oil-related genes involved in the entire FA biosynthesis in G. barbadense compared with G. hirsutum. In the process of de novo formation acyl chains in plastids, the activity of the subunit (α-PHD) of the pyruvate dehydrogenase complex (PDHC) and 2-oxoacid dehydrogenase acyltransferase (DHLAT) at 30 DPA may provide more acetyl-CoA precursor for continuous FA synthesis. Intriguingly, as a key regulatory step of the formation of malonyl-CoA from acetyl-CoA, the biotin carboxyl carrier protein 2 (BCCP2) displayed a
more important role in *G. barbadense*, while the *BCCP1* showed a slightly higher expression level in *G. hirsutum*, like to what has been reported in the upland cotton with high oil content (Zhao et al., 2018c). The three types of ketoacyl-AcP synthase (*KASI*, *KASII*, and *KASIII*) were poorly expressed in cottonseed except *KASI*, which was abundant at 10 DPA and 20 DPA. Both enoyl-ACP reductase (*ENR*) and polyketide synthase (*ER*) showed highly expressed in *G. barbadense* than that in *G. hirsutum* in the cycling of fatty acid synthase (FAS). As a crucial desaturation enzyme for C18:1 synthesis, the SAD showed a constantly high expression level in developing cottonseed in both species. Surprisingly, the SAD6 may played a decisive role in determining high oil content in *G. barbadense* with an increased in expression after 20 DPA.

The significant differential expression pattern of the TAG biosynthesis was observed in the endoplasmic reticulum (ER), with most enzymes showing low expression levels in *G. barbadense* in comparison with *G. hirsutum*. For example, four enzymes involving in the assembly of TAG from glycerol-3-phosphate (G3P) and acyl-CoAs in *G. hirsutum*, including glycerol-3-phosphate acyltransferase (GPAT), lysophosphatidic acid acyltransferase (LPAA), diacylglycerol acyltransferase (DGAT) and phospholipid diacylglycerol acyltransferase (PDAT), displayed a slightly higher expression level during seed development, excepting PDAT, which was barely expressed in *G. barbadense*. The decisive enzyme of formation linoleic acid from oleic acid, FAD2 showed extensive expression abundance at 20 PDA and 30 DPA, with higher in *G. barbadense*. Finally, a series of enzymes related to TAG storage showed high activity at 20 DPA and 30 DPA in both species, which mirrored oil accumulation with temporal in the seed development.

A series of the transcription factors involved in regulation of seed development and maturation have been reported (Santos-mendoza et al., 2008). A total of 61 and 59 genes belonging to

![Gene expression network for fatty acid biosynthesis in development cottonseed.](image_url)
14 transcription factors (TFs) with their expression pattern were picked out from *G. hirsutum* and *G. barbadense*, respectively (Table S8). The time-series expression levels of 13 expressed TFs were higher in *G. barbadense* than in *G. hirsutum*, except ABI3 and DOF2.1 (Figure 3a). Surprisingly, three TFs showed high and exclusive activity in development seed, especially at later stage in *G. barbadense*, including WRI1, NF-YB6 and DPBF2. WRI1, an AP2-type transcription factor, which was considered to be the most important regulator in seed storage metabolism, could regulate many enzymes in the FA metabolism (Maeo et al., 2009). Gene expression bias is prevalent in the polyploidy genome, and previous study has shown that about 20% of genes showed subgenome-specific expression in upland cotton (Hovav et al., 2015). The expression pattern observed for WRI1 exhibited homeologous expression bias between the two species, although most of them in genome were barely expressed in general. The expression levels of two WRI1 genes located on chromosomes A10 and D10 in *G. barbadense* rapidly increased after 10 DPA; however, increased expression in *G. hirsutum* was only observed for Dt allele and decreased rapidly after 20 DPA (Figure 3b). In addition, the similar expression trend was detected for another two TFs (NF-YB6 and DPBF2). The prolonged expression period and asymmetric homologous expression of these TFs may contributed to regulate the continuous activity of enzymes involving in FA biosynthesis to improve the oil content in *G. barbadense*.

**Genetic variation of oil-related homologous TF genes**

In this study, significant differential expression levels and prolonged expression time of the oil-related homologous genes between *G. hirsutum* and *G. barbadense* were observed, which may be the direct cause of the difference SOC between the two species. In general, driving the expression of a gene is considered to be a function of its promoter, especially the core promoter region closely associated with the transcription start site (TSS) (usually less than 1 kb), which is enriched in cis-elements/motifs for the binding of regulatory TFs (Ye et al., 2018). Therefore, comparing the variation of the core promoter region of homologous genes can quickly reveal the mechanism of differences in their expression levels. In tetraploid cotton, there are two (Gbar_A05G038350 and Gbar_D04G004710) and three (Gbar_A05G037550, Gbar_A09G001130 and Gbar_D04G004660) BCCP2 genes in *G. hirsutum* and *G. barbadense* genome, respectively. Unlike all three BCCP2 genes expressed in high SOC of *G. barbadense*, only one on A05 expressed with a declining trend during seed development in *G. hirsutum* (Figure 4a). Via alignment of the relevant promoter sequences of the orthologous BCCP2 genes, a 12 bp deletion located ~38-bp upstream of the TSS was found in the putative promoter in the corresponding regions of Gbar_A05G037550 (Figure 4b). Interestingly, two predicted TF-motif were located in this missing sequence, including AP2-motif (TAGAT) and NF-YB-motif (ATTGA), which were closely related to the TFs that regulate the expressions of fatty acid genes shown above. Another BCCP2 gene located on D04 was barely expressed in *G. hirsutum*, while its orthologous gene in *G. barbadense* had a higher expression level. Moreover, we found that the basis for the differential expression might result from three adjacent deletions in the putative promoter region of the 3-79 allele of this gene (Figure 4c). Deletion of the motif in the promoter may cause its declining ability to be regulated by TFs and can explain the altered expression of BCCP2 genes between two species (Figure 4d). The high expression of BCCP2 genes provided abundant precursor for FA biosynthesis, which improved the SOC in *G. barbadense*. However, no significant structure variation was discovered in the core promoter region of the highly expressed oil-related genes (Table S9), such as FAD and SAD gene. These results imply that there should be

---

**Figure 3** Gene expression trend of oil-related TFs in cottonseed development. (a) Heatmap showing the expression of oil-related TFs during seed development (10 DPA, 20 DPA and 30 DPA); (b) homolog expression bias analysis of three key TFs in developing seed.
more complicated genetic mechanism of regulating SOC in cotton.

Identification of candidate genes for SOC

In previous research, we have identified fifteen QTL for SOC in the G. hirsutum cv. Emian22 × G. barbadense acc. 3-79 interspecific IL population, which provided a good opportunity to deeply analyse the genetic mechanism of SOC. Conducting gene expression profile in embryo allowed us to identify tissue-specific genes and to reveal different genetic regulation of lipid metabolism between species. To identify causal genes within the QTL intervals, we examined genes located in the confidence intervals by referring to the TM-1 genome annotation. A total of 1573 genes were identified in these QTL intervals, ranging from 12 genes in q-SOCD01-2 to 491 genes in q-SOCD01-1. We then compared the expression pattern of these genes and exacted those that were differentially expressed between the two parental lines. Among these genes, eleven candidate genes associated with 8 QTLs can be proposed on the basis of their gene annotation. However, only four of these candidate genes were associated with the introgression chromosome interval carried by the high SOC ILs. To avoid errors caused by QTL detection, we searched all the genes located in the introgression intervals of high SOC ILs, based on sequence homology, as well as differential expression and potentially lipid metabolism-related. Finally, another ten genes were picked out (Table 1). Most of the candidate genes showed significant differential expression levels during the synthesis and accumulation of cottonseed oil between G. hirsutum and G. barbadense, especially in the late stage of cottonseed oil accumulation (25 DPA to 35 DPA) (Figure 5). Among these candidate genes, eleven genes were reported to be involved in FA metabolism or central metabolism. On chromosome A01, for instance, the candidate gene pyruvate dehydrogenase kinase (PDK) in the q-SOCA01-1 is only carried by IL N159, which has been described in the previous research (Zhu et al., 2020). Another gene, Ghir_A01G002250, encoding a member of the sweet sucrose efflux transporter family protein, has been identified as key player in intercellular transport of sugars (Eom et al., 2015). The significantly higher expression level of this gene may increased the efficiency of sugar transporting to promote seed oil synthesis in G. barbadense during the 20 DPA to 35 DPA (Figure 5a). On the same chromosome, a gene was identified in the q-SOCA01-3 with significant differential expression from 25 DPA to 35 DPA between Emian22 and 3-79 (Figure 5b). This gene encodes a phosphoglycerate kinase (PGK), which is a highly conserved reversible enzyme that participates in both glycolysis and photosynthesis (Li et al., 2019). The production of dihydroxyacetone-phosphate (DHAP) from the glycolysis pathway is the intermediate carbon donor for FA synthesis, and previous research has confirmed that high activities of PGK can increase the FA content in sunflower seeds (Troncoso-Ponce et al., 2010). In additional, the candidate gene, Ghir_A05G025600, classified in a GDSL-motif acyltransferase, was identified in the region of q-SOCA05. The expression pattern of this gene showed a high degree of coincidence with the period of seed oil accumulation in G. hirsutum (Figure 5c), while several researches reported that overexpression of the GDSL-type esterase could reduce the total seed fatty acid content and change the seed fatty acid composition in Arabidopsis (Chen et al., 2012; Huang et al., 2015). Higher expression levels of this

![Figure 4](https://example.com/figure4.png)

**Figure 4** Compression of BCP2 homolog alleles. (a) Expression levels of these two homolog genes in developing seed; (b) the promoter of the Emian22 allele of Ghir_A05G038350 features a 12-bp insertion compared with the 3-79 allele Gbar_A05G037550; (c) the promoter of the Emian22 allele of Ghir_D04G004710 features three insertion compared with the 3-79 allele Gbar_D04G004660; (d) expression levels of these two homolog genes in developing seed.
The gene, Ghir D03G009400, shared in both ILs R104 and R106, encodes a cytochrome p450 monooxygenase (CYP78A9) and was predicted to involve in the seed development. Although overexpression of this gene induced large and seedless fruits in G. barbadense (Ito and Meyerowitz, 2000), the analysis of genes coexpressed with CYP78A9 showed that most of them were involving in FA metabolism or FA metabolism process (Sotelo-Silveira et al., 2013). Moreover, examination of the promoter region of this allele gene between parental lines revealed a 21 bp deletion from the −187 bp of the TAG in the Emian22 (Figure 6a), which may explained the differential expression (Figure 5h). Moreover, we also identified three other genes whose expression patterns were consistent with the accumulation of seed oil (Figure 5i, j and k), such as Ghir A05G025470, Ghir A12G012590 and Ghir D06G002090. No evidence has been reported that these genes are directly involved in the oil biosynthesis pathway, but indispensable genes for seed development. (Fiume et al., 2016; Mentzen et al., 2008). Furthermore, genes involved in the redox reactions thought to be related FA metabolism have also been identified in this study. In q-SOCA01-4, the candidate gene Gbar A01G009640, as a gene insertion event in the homology region of G. barbadense, coding a peroxisomal catalase 2, may improve the SOC by regulating the metabolism of H2O2 in cottonseed, based on the report that the formation of high oleic was closely related to the H2O2 content in cottonseed (Liu et al., 2019). However, significantly different expression level was detected at the maturation stage of the cottonseed (30 DPA to 35 DPA) (Figure 5i), which indicated that the biological regulation of cottonseed oil may be more complicated.
No annotated biosynthetic gene likely involved in the production of the FA synthesis was found in the q-SOCA07. Incidentally, an altered expression gene, Ghir_A07G019980, was found in this region (Figure 5m). This gene was described as a small ubiquitin-like modifier (SUMO) polypeptide, involved in the post-translational modification process of protein, which can regulate embryo development (Liu et al., 2017a). Previous studies reported that the retrograde signalling of the candidate gene GhGUN1 (Ghir_A12G013320) may be involved in regulating acetyl-CoA carboxylase 2 in Arabidopsis (Wang et al., 2018), which is essential for the first step of fatty acid synthesis. The significant difference in the expression pattern of this gene between G. hirsutum and G. barbadense indicated that it may play an important role in the formation of the higher SOC in the G. barbadense (Figure 5n). Four candidate genes with different expression were selected from the introgression region of IL line N45 (Figure 5o, p, q and r). One significant frameshift mutation occurs in the coding sequence between Ghir_D07G012450 and its allele gene in G. barbadense (Gbar_D07G012560) (Figure 6b). This allele gene encodes a short-chain dehydrogenase/reductase in G. barbadense, which plays a role in detoxifying reactive carbonyls in plants (Yamauchi et al., 2011). On the contrary, there is no clear function reported of the homolog allele gene in upland cotton. Moving on to the next gene, Ghir_D07G013130, this gene showed extensive expression abundance, especially after 20 DPA (Figure 5r). Therefore, it is speculated to play dual roles as functional inhibitor for lipid-transfer and seed storage.

Potential functional roles of candidate genes in fatty acid synthesis

A total of 17 candidate genes were selected to construct the overexpression vector pYES2 in the INVScL1 strain of Saccharomyces cerevisiae. Only two genes (GbSWEET and GbACBP6)
have the potential to regulate the FA biosynthesis in the yeast (Figure 6c). Overexpression of the \textit{GbSWEET} in yeast increased about 17.54%, 22.17% and 40.07% of the OC in the three transgenic strains compared with that in the negative control (CK), which indicated that higher expression level of the \textit{GbSWEET} in \textit{G. barbadense} played a crucial role in the enhancing the SOC. Meanwhile, the oil content in another transgenic yeast strains carried \textit{GbACBP6} also showed significantly higher than the empty-vector.

**Discussion**

Cotton, beyond its leading nature fibre, seed oil is also a valuable source for food and biodiesel. Although an increasing number of works have reported the biochemical and biological functions of oil-related genes in upland cotton, less attention was paid to identify corresponding genes in sea-island cotton. Comparative transcriptome analysis of high- and low-SOC cottonseeds provides an effective approach for studying gene differential expression pattern and dissecting candidate genes in oil synthesis. Here, based on deep RNA sequencing, a high-resolution gene expression network in developing cottonseed was constructed to analyse the genetic regulation of lipid metabolism in \textit{G. barbadense}. Different from previous study (Zhao et al., 2018c), the present study found that the number of expressed genes decreased during seed development. Compared with the DEGs in \textit{G. hirsutum}, significantly reduced number of DEGs appeared in the comparison of ‘30 DPA vs 20 DPA’ in \textit{G. barbadense} with higher SOC, which meant that most of genes still maintain similar expression pattern in 30 DPA as in 20 DPA. Then, GO and KEGG analysis were performed to identify the functional interpretation and pathway enrichment of those DEGs. The pathway enrichment analysis also confirmed that more up-regulated and less down-regulated genes in lipid metabolism pathway were detected during the critical period of oil accumulation in \textit{G. barbadense} than that in \textit{G. hirsutum} (Tables S4 and S5). Furthermore, several genes related unsaturated fatty acid synthesis showed significantly down-regulated in \textit{G. hirsutum}, such as \textit{SAD6} and \textit{FATA}. In short, the high expression level maintained by a large number of genes at 30 DPA is the direct reason for the higher SOC in \textit{G. barbadense} compared with \textit{G. hirsutum}.

During the \textit{de novo} biosynthesis of FA, the difference was observed in the expressed gene type between the two species. For example, \(\alpha\)-CT showed high expression levels in \textit{G. barbadense}, while \(\beta\)-CT is more active in \textit{G. hirsutum}. Similar result was observed for the two types of \textit{BCCP} genes. DNA sequencing of the orthologous promoter regions revealed deletions in the \textit{GbBCCP2}, moving the regulation binding motif away from the TSS. The \textit{WRI1} as an AP2 type TF has been shown to positively regulate the expression of the \textit{BCCP2} genes (Maeo et al., 2009). As a carrier of FAs, upland cotton relied on more \textit{FATB} at the later period. Different from the previous studies in upland cotton, the \textit{LACS} and \textit{PDCT} seem no value for the TAG formation in sea-island cotton, because of its barely expressed level. Via overexpression of \textit{LACS} and \textit{PDCT} in sea-island cotton may be a good idea to improve the SOC in \textit{G. barbadense} cultivars in the future.

**Figure 6** Candidate genes in this study. (a) The promoter variation of candidate gene \textit{Ghir}_D03G009400; (b) structure variation in the coding sequence of the candidate gene \textit{Ghir}_D07G012450; (c) function analysis of \textit{GbSWEET} and \textit{GbACBP6} in the fatty acid synthesis in yeast.
The LPAAT2 gene was regarded as a rate-limiting enzyme in the Kennedy pathway in higher plants, and previous study showed that this gene highly expressed in upland cotton (Wang et al., 2017b). On the contrary, decreasing expression trend was observed for this gene in G. barbadense. The different expression pattern of genes related to FA synthesis implies that different properties or mechanisms of oil biosynthesis exist in the two distantly related cotton species.

TFs have been proved to play an important role in controlling SOC in this study. Most of the oil-related TFs showed abundant expression level from 20 DPA to 30 DPA in G. barbadense and thus may be the cause of significantly higher expression of FA synthesis genes than G. hirsutum. However, no apparently visible sequence mutation was detected in the core promoter region of these TFs. It is possible either that we cannot detect mutations that exit in these genes or that an unknown regulatory mechanism exists. Therefore, a higher quality reference genome or cloning of gene is needed for future analysis. An example, a reported gene, Gbar_A05G017010, encoding a peroxidase superfamily protein (PRX), was found in the introgression region of IL line N45. Previous study has confirmed that overexpression of this gene could significant improve the oil content in cottonseed (Ma et al., 2019), while no gene allele was annotated in the homology sequence in the G. hirsutum. Interestingly, this allele can be detected in the recently updated G. hirsutum reference genome data (Huang et al., 2020).

Since upland cotton and sea-island cotton are allotetraploid species, homolog expression bias of genes is a phenomenon worthy of attention. Via comparing the oil-related gene expression pattern between published data and this study, we found that slightly higher level of expression bias towards the D-genome and those genes high expressed in At-subgenome were more likely to be regulated (Table S8). We highlighted three TFs with expression patterns significantly related to the oil synthesis in this study, which homolog alleles in At-subgenome showed lower expression levels in G. hirsutum. This result is similar to previous report for cotton fibre evolution, which showed that domestication has dramatically changed the transcriptome during fibre development (Yoo and Wandel, 2014) and more Dt-bias domestication selection (Wang et al., 2017a).

To investigate the potential genetic basis of SOC variability between G. hirsutum and G. barbadense, we evaluated the genome variation and gene expression patterns in the candidate intervals of previously identified SOC-related QTL and the introgression chromosome segments carried by the ILs with high SOC. Genes involved in lipid metabolism, glucose metabolism, redox metabolism and embryo development were identified as candidate genes related to the seed oil accumulation. Upon further analysis by quantitative real-time PCR between G. hirsutum and G. barbadense, most of the candidate genes were highly consistent with the oil accumulation in cotton. In this study, relying on the fatty acid synthesis system, we identified two genes (SWEET and ACPBP6) may influence oil accumulation in cottonseed. These results suggested that the method of selecting the candidate genes used in this study was credible. Correspondingly, in many cases, regarding differentially expressed genes as candidate genes can also be supported by fine-mapping (Che et al., 2015). Nevertheless, future studies are required for validation of candidate genes identified in this study. The use of omics, especially metabolomics combined with IL resources, to analyse the genetic regulation of lipids is also considered to be a feasible solution (Fernandez-Moreno et al., 2017; Garbowicz et al., 2018).

**Conclusion**

The current study provides a first comprehensive view of the genetic regulation of FA metabolism in G. barbadense. The interspecific IL population provided an excellent resource to compare analysis of oil-related genes differential expression and confirmed in this study. Furthermore, some candidate genes have been reported in association with oil accumulation. This work further enriches our understanding of the gene expression network of FA biosynthesis in cotton and guides us future genetic breeding of cotton oil.

**Experimental procedures**

**Plant materials**

The interspecific IL population was derived from the cross between G. hirsutum cv. ‘Emian22’ (low SOC) and G. barbadense acc. 3-79 (high SOC). The SOC data set presented is based on field-grown over three years in five environments, which has been described in our previous study (Zhu et al., 2020). The genome sequencing data and genotypic data of the ILs also have been described in that study (Zhu et al., 2020). Of this IL population, 8 with high SOC (N45, N58, N70, N156, N159, R104, R106 and R120) and their parents were grown at an experimental field in Huazhong Agricultural University, Wuhan, China (30.4°N,114.2°E), under normal farming conditions. The SOC traits were phenotyped by measuring the total oil percentage using nuclear magnetic resonance (NMR) analyser following the protocol of Zhao et al. (2019).

**RNA-sequencing dada analysis and oil accumulation gene network construction**

The embryos were collected at four developmental stages (10, 20, 30 and 40 day postanthesis (DPA)), representing different stages of seed filling and maturation. The samples were quickly frozen in liquid nitrogen for RNA isolation, and three biological replicates were taken from each time point. High-quality RNA extraction was performed using the RNAprep Pure Plant Kit (Tiangen, Beijing, China). Total RNA at three stages (10, 20 and 30 DPA) for parents was sequenced with the Illumina HiSeq 2000 system (paired-end 150 bp). The clean RNA-seq reads of Emian22 and 3-79 were mapped to the G. hirsutum and G. barbadense reference genome using HISAT 2.0 software (http://ccb.jhu.edu/software.shtml), respectively. Fragments per kilobase of exon per million mapped reads (FPKM) value were calculated by StringTie for gene expression levels (Pertea et al., 2016), and only criterion of FPKM ≥ 1 in each library was considered an expressed gene. The differentially expressed genes (DEGs) were identified using an R package of DESeq (Anders and Huber, 2010). The genes were defined as differentially expressed following these rules: (1) significance P-value ≤ 0.05 and false discovery rate (FDR) threshold ≤ 0.05; (2) log2(FoldChange) ≥ 2. Gene Ontology (GO) function annotation and Kyoto Encyclopaedia of Genes and Genomes (KEGG) annotation were carried out using the online data of the new reference genome project (http://www.cottongen.org), which was updated by our laboratory in previously.

Genes involved in FAs biosynthesis were detected and referenced to Arabidopsis acyl-lipid metabolism (http://aralip.plantb
The gene network of lipid biosynthesis was constructed and modified by referencing previous reported in upland cotton (Hoav et al., 2015; Zhao et al., 2018c). The expression abundance of genes is counted by their putative functions, independent from their relative expression levels.

Candidate gene selection

The candidate genes for SOC were selected from two parts. In the gene network of lipid biosynthesis, the differential expression gene has any function mutation between G. hirsutum and G. barbadense was considered as a candidate gene, including gene doubling and deletion, structural variation in the promoter region and SNP differences in the coding sequences producing non-conserved protein changes. In the chromosome regions carried by SOC-related QTLs and ILs with high SOC, candidate genes selected on the basis of: (i) the corresponding gene annotation being related to the lipids biosynthetic process; (ii) the differential gene expression level between G. hirsutum and G. barbadense being consistent with seed developing stages; (iii) gene function mutation requirements being consistent with above. We named candidate genes in the interest interval that satisfied two or three of as described.

Quantitative PCR

Total RNA was isolated using the RNeasy Prep Pure Plant Kit (Tiangen, Beijing, China) from cottonseeds at 10, 15, 20, 25, 30, 35 and 40 DPA. High-quality RNA was reverse-transcribed using the SuperScript III reverse transcriptase (Invitrogen, Cat. No. 18080-093, USA). The ABI Prism 7500 system was used to perform qRT-PCR, with Ghuubo7 (DQ116441) as the internal control. The real-time PCR primers are listed in Table S10.

Expression of the candidate genes in yeast fatty acid system

The open reading frame (ORF) of these candidate genes was PCR-amplified from a G. barbadense acc. 3-79 ovules cDNA library and inserted into pYeS2 (Invitrogen) vector to construct overexpression system in the INVSc1 strain of Saccharomyces cerevisiae (Weidi Biotechnology Co., Shanghai, China). The primers used for vector construction are listed in Table S10. The empty-vector pYeS2 was used as the negative control (CK). The transgenic yeast and CK was induced by the method according to the Ma et al. (2019). The oil content of the transgenic yeast and CK was measured by a tissue triglyceride assay kit (Applygen Technologies Inc., Beijing, China).

Acknowledgements

We thank Dr. Jiwen Yu from Cotton Institute of the Chinese Academy of Agricultural Sciences, Anyang, Henan, China, and Dr. ShaoWu Xue from Huazhong Agricultural University, Wuhan, China, for kindly sharing the yeast vector. This work was financially supported by the Genetically Modified Organisms Breeding Major Project of China (No. 2016ZX08009001).

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Authors’ contributions

Zhongxu Lin conceived and designed the project. Xiaojing Li performed the field experiments and investigated phenotypic data. De Zhu, Yu Le and Ruiting Zhang conducted the experiments. De Zhu analysed the data and wrote the manuscript draft. Zhongxu Lin revised the manuscript. All authors discussed the results and approved the final manuscript.

Consent for publication

Not applicable.

Data availability statement

The clean raw sequencing data in this manuscript have been deposited in NCBI Sequence Read Archive under accession number PRJNA667195.

References

Abdulrah, H.M., Akbari, P., Paulose, B., Schnell, D., Qi, W., Park, Y., Pareek, A. et al. (2016) Transcriptome profiling of Camellia sativa to identify genes involved in triacylglycerol biosynthesis and accumulation in the developing seeds. Biotechnol. Biofuels, 9, 136.

Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. Genome Biol. 11, 1–12.

Bates, P.D., Stymne, S. and Ohlrogge, J. (2013) Biochemical pathways in seed oil synthesis. Curr. Opin. Plant Biol. 16, 358–364.

Baud, S. and Lepinec, L. (2010) Physiological and developmental regulation of seed oil production. Prog. Lipid Res. 49, 235–249.

Belmonte, M.F., Kirkbride, R.C., Stone, S.L., Pelletier, J.M., Bui, A.Q., Yeung, E.C., Hashimoto, M. et al. (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. Proc. Natl. Acad. Sci. USA, 110, 20122061.

Che, R., Tong, H., Shi, B., Liu, Y., Fang, S., Liu, D., Xiao, Y. et al. (2015) Control of grain size and rice yield by GL2-mediated brassinosteroid responses. Nat. Plants, 2, 2055–0278.

Chen, M., Du, X., Zhu, Y., Wang, Z., Hua, S., Li, Z., Guo, W. et al. (2012) Seed fatty acid reducer acts downstream of gibberellic signalling pathway to lower seed fatty acid storage in Arabidopsis. Plant Cell Environ. 35, 2155–2169.

Deng, W., Yan, F., Zhang, X., Tang, Y. and Yuan, Y. (2015) Transcriptional profiling of canola developing embryo and identification of the important roles of BnDof5.6 in embryo development and fatty acids synthesis. Plant Cell Physiol. 56, 1624–1640.

Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., Li, N. et al. (2018) Resequencing of 243 diploid cotton accessions based on an updated a genome identifies the genetic basis of key agronomic traits. Nat. Genet. 50, 796–802.

Egli, B., Kolling, K., Kohler, C., Zeeman, S.C. and Streb, S. (2010) Loss of cytosolic phosphoglucomutase compromises genophyte development in Arabidopsis. Plant Physiol. 154, 1659–1671.

Eom, J.S., Chen, L.Q., Sosso, D., Julius, B.T., Lin, I.W., Qu, X.Q., Braun, D.M. et al. (2015) SWEETs, transporters for intracellular and intercellular sugar translocation. Curr. Opin. Plant Biol. 25, 53–62.

Fernandez-Moreno, J.P., Levy-Samoha, D., Monforte, A.J., Orzaez, D., Aharoni, A. and Granell, A. (2017) Uncovering tomato quantitative trait loci and candidate genes for fruit cuticular lipid composition using the Solanum pennelli introgression line population. J. Exp. Bot. 68, 2703–2716.

Fume, E., Guyon, V., Remoue, C., Magnani, E., Miquel, M., Grain, D. and Lepinec, L. (2016) TWST1, a novel small protein, regulates various aspects of seed and plant development. Plant Physiol. 172, 1732–1745.

Garbowicz, K., Liu, Z., Alseekh, S., Tieman, D., Taylor, M., Kuhlitskaya, A., Ofner, I. et al. (2018) Quantitative trait loci analysis identifies a prominent
gene involved in the production of fatty acid-derived flavor volatiles in tomato. Mol. Plant. 11, 1147–1165.

Grover, C.E., Gallagher, J.P., Jareczek, J.J., Page, J.T., Udall, J.A., Gore, M.A. et al. (2015) Re-evaluating the phylogeny of allopolyploid Gossypium L. Mol. Phylogenet. Evol. 92, 45–52.

Gusev, A., Ko, A., Shi, H., Bhatia, G., Wang, W., Pennix, B.W.J.H., Jansen, R. et al. (2016) Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. 48, 245–252.

Hinze, L.L., Horn, P.I., Kothari, N., Dever, J.K., Freilichowski, J., Chapman, K.D. and Percy, R.G. (2015) Nondestructive measurements of cottonseed nutritional trait diversity in the U.S. national cotton Germplasm collection. Crop Sci. 55, 770–782.

Hovat, R., Faienborn-Doron, A., Kadmon, N., Hu, G., Zhang, X., Gallagher, J.P. and Wendel, J.F. (2015) A transcriptome profile for developing seed of polyploid cotton. Plant Genome. 8, 19403372.

Huang, L.M., Lai, C.P., Chen, L.F.O., Chan, M. and Shaw, J. (2015) Arabidopsis SFAAR is a novel GDSL-type esterase involved in fatty acid degradation and glucose tolerance. Bot Stud. 56, 33.

Huang, G., Wu, Z., Percy, R.G., Bai, M., Li, Y., Freilichowski, I.E., Hu, J. et al. (2020) Genome sequence of Gossypium herbaceum and genome updates of Gossypium arboreum and Gossypium hirsutum provide insights into cotton A-genome evolution. Nat. Genet. 52, 516–524.

Ito, T. and Meyerowitz, E.M. (2009) Overexpression of a gene encoding a cytochrome F450, CYP78A9, induces large and seedless fruit in Arabidopsis. Plant Cell. 12, 1541–1550.

Jenkins, J.N., McCarty, J.C., Deng, D., Geng, L., Hayes, R.W., Jones, D.C. and Mamwooda, R. (2018) Introgression of Gossypium barbadense L. into upland cotton germplasm RMBU-451. Euphytica. 214, 0014–2336.

Jones, S. and Vodkin, L.O. (2013) Using RNA-seq to profile soybean seed development from fertilization to maturity. Plant Cell. 25, 611–622.

Kakandes, D.B., Yilmaztekin, M., Mert, M. and Gencer, O. (2017) Physico-chemical characteristics of fatty acid compositions of cottonseed oils. Tarım Bilim. Derg. 23, 253–259.

Li, R., Qiu, Z., Wang, X., Gong, P., Xu, Q., Yu, Q. and Guan, Y. (2019) Pooled CRISPR/Cas9 reveals redundant roles of plastidial phospholipase kinas in carbon fixation and metabolism. Plant J. 98, 1078–1089.

Li, B., Shi, Y., Gong, J., Li, J., Liu, A., Shang, H., Gong, W. et al. (2016) Genetic effects and heterosis of yield and yield component traits based on Gossypium barbadense chromosome segment substitution lines in two Gossypium hirsutum backgrounds. PLoS One. 11, e0157978.

Liu, H., Gu, J., Lu, Q., Li, H., Hong, Y., Chen, X., Ren, L. et al. (2019) Transcriptomic analysis reveals the high-oleic acid feedback regulating the homologous gene expression of stearyl-ACP desaturase 2 (SAD2) in peanuts. Int. J. Mol. Sci. 20, 3091.

Liu, L., Jiang, Y., Zhang, X., Wang, X., Wang, Y., Han, Y., Coupland, G. et al. (2017a) Two SUMO proteases SUMO PROTEASE RELATED TO FERTILITY1 and 2 are required for fertility in Arabidopsis. Plant Physiol. 175, 1703–1719.

Liu, Q., Singh, S.P. and Green, A.G. (2002) High-steaic and high-oleic cottonseed oils produced by hairpin RNA-mediated gene silencing. Plant Physiol. 129, 1732–1743.

Liu, Q., Wu, M., Zhang, B., Shrestha, P., Petrie, J., Green, A.G. and Singh, S.P. (2017b) Genetic enhancement of palmitic acid accumulation in cotton seed oil through RNAi down-regulation of ghKAS2 encoding beta-ketoacyl-ACP synthase II (KASII). Plant Biotechnol J. 15, 132–143.

Ma, J., Liu, J., Pei, W., Ma, Q., Wang, N., Zhang, X., Cui, Y. et al. (2019) Genome-wide association study of the oil content in upland cotton (Gossypium hirsutum L.) and identification of GHRKPR, a candidate gene for a stable QTL-QC-D1S1. Plant Sci. 286, 89–97.

Mao, K., Tokuda, T., Ayame, A., Mitsu, N., Kawai, T., Tsukagoshi, H., Ishiguro, S. et al. (2009) An AP2-type transcription factor, WRNKLE1, of Arabidopsis thaliana binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. Plant J. 60, 476–487.

Mentzen, W.J., Peng, J., Ransoms, N., Nikolau, B.J. and Wuurtele, E.S. (2008) Articulation of three core metabolic processes in Arabidopsis: fatty acid biosynthesis, leucine catabolism and starch metabolism. BMC Plant Biol. 8, 76.
compression-ignition engine fueled with cottonseed oil biodiesel/diethyl ether/diesel fuel blends. Energy Convers. Manage. 205, 22.

Yoo, M. and Wendel, J.F. (2014) Comparative evolutionary and developmental dynamics of the cotton (Gossypium hirsutum) fiber transcriptome. PloS Genet. 10, 1–19.

You, Q., Xu, W., Zhang, K., Zhang, L., Yi, X., Yao, D., Wang, C. et al. (2016) ccNET: Database of co-expression networks with functional modules for diploid and polyploid Gossypium. Nucleic Acids Res. 45, 5625–5626.

Yu, J., Yu, S., Fan, S., Song, M., Zhai, H., Li, X. and Zhang, J. (2012) Mapping quantitative trait locus for cottonseed oil composition in cotton. J. Plant Physiol. 169, 101–112.

Zhu, D., Li, X., Wang, Z., You, C., Nie, X., Sun, J., Zhang, X. et al. (2020) Genetic dissection of an allotetraploid interspecific CSSLs guides interspecific genetics and breeding in cotton. BMC Genom. 21, 1471–2164.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 SOC of ILs and their parents in 2019. (One ANOVA analysis for two lines and Dunnett’s multiple comparison for multiple lines. *** indicated P < 0.001).

Figure S2 Gene expression cluster and summary of expressed genes at different stages. (a) PCA was performed for all library in each species; (b) summary of expressed gene number in each stage; (c) Venn diagram implies that genes are expressed in specific stages.

Figure S3 KEGG enrichment analysis of up-regulated DEGs in ‘20 DPA vs 10 DPA’ comparison.

Figure S4 KEGG enrichment analysis of down-regulated DEGs in ‘30 DPA vs 20 DPA’ comparison.

Figure S5 KEGG enrichment analysis of down-regulated DEGs in the comparison of ‘30 DPA vs 20 DPA’. The red frames mark the pathways involved in lipid metabolism.

Figure S6 Hierarchical clustering of oil-related DEGs. Enriched gene ontology categories were shown by each cluster (FDR < 0.1). Value of the colour key refers to the log2 of gene expression (fragments per kilobase of exon per million mapped reads, FPKM).

Table S1 Summary of RNA_seq data quality in different libraries.

Table S2 GO enrichment analysis of DEGs in cottonseed developmental stages of G. hirsutum.

Table S3 GO enrichment analysis of DEGs in cottonseed developmental stages of G. barbadense.

Table S4 KEGG pathways enrichment analysis of different categories of DEGs in G. hirsutum.

Table S5 KEGG pathways enrichment analysis of different categories of DEGs in G. barbadense.

Table S6 Expression analysis of genes involved in fatty acid metabolism in developing cottonseed in G. hirsutum.

Table S7 Expression analysis of genes involved in fatty acid metabolism in developing cottonseed in G. barbadense.

Table S8 Oil-related TFs in cottonseed development between G. hirsutum and G. barbadense.

Table S9 Summary of possible genomic variations of the significant differentially expressed oil-related genes.

Table S10 Primers used in the research.