Identification and analysis of oil candidate genes revealed the molecular basis of oil accumulation in *Gossypium hirsutum* L. for cottonseed biofuels development

zhibin zhang
Chinese Academy of Agricultural Sciences Cotton Research Institute
Juwu Gong
Chinese Academy of Agricultural Sciences Cotton Research Institute
Zhen Zhang
Chinese Academy of Agricultural Sciences Cotton Research Institute
Wankui Gong
Chinese Academy of Agricultural Sciences Cotton Research Institute
Junwen Li
Chinese Academy of Agricultural Sciences Cotton Research Institute
Yuzhen Shi
Chinese Academy of Agricultural Sciences Cotton Research Institute
Aiying Liu
Chinese Academy of Agricultural Sciences Cotton Research Institute
Qun Ge
Chinese Academy of Agricultural Sciences Cotton Research Institute
Jingtao Pan
Chinese Academy of Agricultural Sciences Cotton Research Institute
Senmiao Fan
Chinese Academy of Agricultural Sciences Cotton Research Institute
Xiaoying Deng
Chinese Academy of Agricultural Sciences Cotton Research Institute
Shaoqi Li
Chinese Academy of Agricultural Sciences Cotton Research Institute
Quanjia Chen
Xinjiang Agricultural University
Youlu Yuan
Chinese Academy of Agricultural Sciences Cotton Research Institute
Haihong Shang (shanghaihong@caas.cn)
Chinese Academy of Agricultural Sciences Cotton Research Institute

Research

Keywords: Gossypium hirsutum L., Seed oil content, QTL mapping, microRNA, Transcription factor, Regulatory network, Biodiesel

Posted Date: June 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-34332/v1

Abstract

Background

Cottonseed oil is increasingly becoming a promising target for biodiesel production with its high content of unsaturated fatty acids as greenhouse gases emission and global warming become more severe. However, the molecular regulatory basis of cottonseed oil accumulation is still unclear so far, and it is necessary to identify vital genes and regulators involved in oil accumulation during developing cottonseed ovules.

Results

In this study, a recombinant inbred line (RIL) population, developed from a cross between upland cotton cultivars 0-153 and sGK9708, was used to detect quantitative trait loci (QTLs) associated with cottonseed oil content and further identify candidate factors regulating cottonseed lipid synthesis. A total of 39 QTLs located on eighteen different chromosomes were identified across eight different environments. Of these, five QTLs were stable in at least three environments. By integrating candidate gene approach and physical mapping data, we preliminary obtained 43 candidate genes potentially involved in carbon metabolism, fatty acid (FA) synthesis and transcription, and triacylglycerol (TAG) synthesis around the stable QTLs. KEGG pathway enrichment analysis and local BLAST with *Arabidopsis* oil-related genes, as well as transcriptome analysis further showed that 19 candidate genes, expressed during developing cottonseed ovules, could influence cottonseed oil accumulation. Transcription factors (TFs) and microRNAs (miRNAs) regulatory network analyses suggested six genes, two core miRNAs (ghr-miR2949b and ghr-miR2949c), and one TFs *GhHSL1* were considered to be closely associated with cottonseed lipid content.

Conclusions

The study provides an unprecedented level of insight from QTL mapping and regulatory network analysis to reveal the oil accumulation mechanism in developing cottonseed ovules through the construction of a detailed oil accumulation model. Moreover, the present study of cottonseed oil also lays a foundation for further oil production improvement in oil crops, and contributes to renewable energy production for solving the energy shortage and stabilizing greenhouse gases at a certain extent.

Background

The *Gossypium* genus, as one of the most important oilseed crops worldwide, not only provides numbers of natural fiber for global textile industries, but also is a vital source of edible oil and biodiesel for its high content of unsaturated fatty acids [1, 2]. Cottonseed, the second major product from the cotton plant, generally has an oil content ranging from 16–25% [1] accounting for 10–15% of the total value of cotton crops. According to the United States National Cottonseed Products Association (http://www.cottonseed.com/), cotton provides the sixth largest source of vegetable oil in the world, as well as renewable raw materials for various industrial products such as biofuels and lubricants oils [3–5]. Moreover, cottonseed oil has become increasingly acceptable in a wide range of processed foods due to its significant cost advantage and flavor stability compared to canola or olive oil [6]. The history of cottonseed oil-related researches could be traced back to ancient times, and the modern cottonseed crushing and refinery technology have obtained a rapid development with the development of epoch [7]. Given the importance of cottonseed as a good source of edible oil and alternative energy, the relative dearth of cottonseed oil research is extremely surprising, especially when compared with the extensive research devoted to cotton fibers.

Up to now many QTLs have been detected to be associated with cottonseed oil content by linkage mapping. For example, Shang *et al.* [8] detected 40 QTLs associated with cottonseed oil using two backcross inbred lines (BILs) and two recombinant inbred lines (RILs) populations; Yu *et al.* [9] obtained 17 cottonseed oil QTLs using the BIL population of 146 lines. Similarly, Alfred *et al.* [10] found 4 QTLs using an immortalized F_2 population (IF_2); Liu *et al.* [11] acquired 15 QTLs using an $F_{2:7}$ cotton RIL population. To overcome the low genotypic variation and limited recombination events in specific genetic populations, genome-wide association studies (GWAS) were further adopted. Such as Du *et al.* [12], Yuan *et al.* [13], and Badigannavar *et al.* [14] detected 16, 28 and 6 significant SNPs related to cottonseed oil, respectively. Moreover, plant seed oil accumulation, as a complex biological process, includes at least conversion of sucrose to pyruvate, *de novo* FA synthesis in plastids, endoplasmic TAG synthesis, and oil-body assembly [15, 16]. In the past several decades, some genes and TFs in oil crops also have been verified to be associated with lipid synthesis based on the *Arabidopsis thaliana* oil-related genes through multi-omics analyses. For instance, phosphoenolpyruvate carboxylase (*PEPC*) in upland cotton for the formation of pyruvate [17]; acetyl-CoA carboxylase (*ACCase*) in rapeseed [18], cotton [19], and potato [20] participate in *de novo* FA synthesis; fatty acylthioesterase B (*FATB*) in soybean [21] and cotton [22], and β-ketoacyl-ACP synthase I (*KASI*) and long chain acyl-CoA synthetase 9 (*LACS9*) in soybean [23], and *KASII* [24], Stearyl-ACPdesaturase (*SAD*) [25], and *GhWRI1* [26] in cotton are involved in FA synthesis; glycerol-3-phosphate dehydrogenase (*GPDH*) in rapeseed [27], glycerol-3-phosphate acyltransferase (*GPAT*) [28] and *GhDof1* [29] in cotton, 2-lysophosphatidic acid acyltransferase (*LPAAT*) in cotton [30], and *OLEs* in soybean [31] are reported to be related to TAG synthesis. Recently, Liu *et al.* [32] identified *GmPDAT*, *GmAGT*, *GmACP4*, *GmZF351* and *GmPgs1* five soybean seed oil related genes using an innovative three-dimension network construction approach, as well as genome-wide association studies and multi-omics analyses. However, few gene regulatory network studies about cottonseed oil synthetic have been reported, and the genetic basis under cottonseed lipid storage is far from being understood.

The present work developed a 196 lines intra-specific RILs population from the cross between 0-153 and sGK9708. The cottonseed oil content trait was evaluated across eight environments, and a whole-genome-based high-density genetic linkage map constructed previously by our lab was used for QTL mapping. A total 39 QTLs for cottonseed oil content trait were identified in these environments, of which 5 stable QTLs were identified in more than three environments. There were 832 genes in the confidence intervals of all the stable QTLs, of which 19 genes expressed during developing cottonseed ovules were ultimately considered as cottonseed oil-related potential candidate genes. Moreover, several important miRNAs and TFs participated in cottonseed lipid synthesis were also obtained through gene regulation network analysis. To clarify the molecular regulatory mechanisms of cottonseed oil accumulation, the carbon metabolism, along with FA and TAG biosynthesis pathways in developing cottonseed ovules were constructed to form a lipid accumulation model that includes detailed information on where and how important genes are expressed and affected seed oil accumulation. The results will provide a useful resource for cottonseed oil research and lay a foundation to convert cottonseed oil into biofuels in the future.

Results

Phenotype analysis of cottonseed oil content traits

The descriptive statistics of phenotypic data for cottonseed oil content from the parents and the RIL populations are summarized in Table 1. There is a significant difference between the average cottonseed oil content of parents 0-153 and sGK9708 ($P < 0.01$), which ranged from 28.09% to 31.64% (mean ± SD, 30.57% ± 1.27), and 31.67% to 35.74% (mean ± SD, 32.86% ± 1.71), respectively. While the average cottonseed oil content of the RIL population increased from 25.09% to 28.49%. Through comparison of the mean values of cottonseed oil content, we observed a relatively great variation between the parents and the RIL population, and the mean values of oil content for RIL population existed transgressive segregations lower than the low parent. Variability in the RIL population according to the standard

deviation values was slight. The phenotypic data of seed oil content trait in the RIL population showed continuous distribution, and the absolute values of skewness were all less than one, which demonstrates that the segregations of cottonseed oil content trait fit a normal distribution. Additionally, the ANOVA indicated that the genotype, environment, repeat, and genotype × environment / environment × repeat interaction had significant effects on seed oil content ($P < 0.01$), and the estimated h^2 was 91.61% (Additional file 1: Table S1). These results suggested that the seed oil content is highly inherited in upland cotton, although environmental factor should be considered in oil accumulation.

Mapping QTLs for cottonseed oil content traits across multiple environments

Detection of main- and small-effect quantitative trait loci (QTLs) for cottonseed oil content Based on the consensus genetic map of *G. hirsutum* L. in our previous studies (8295 markers spanning a total distance of 5197.17 cM) and the phenotype data of cottonseed oil content across eight environments, a total of 39 QTLs located on eighteen different chromosomes were identified in the upland cotton 0-153 × sGK9708 RIL population by CIM method in WinQTLCart2.5 and GCIM method in QTL.gCIMapping. The LOD values of these QTLs ranged from 3.02 to 5.96, and the individual phenotypic variance (r^2) ranged from 4.43 to 14.15 (%) (Additional file 2: Table S2). Of these, five QTLs were stably detected in at least three environments (Table 2). And these stable QTLs were mainly distributed on upland cotton chromosomes At11, At12 and Dt4, Dt9, and Dt11. Intriguingly, we found that *qOC-chr17-3* is stable in four environments, while other four stable QTLs are in three environments. Moreover, the additive effects of QTLs *qOC-chr12-1* and *qOC-chr22-1* detected by GCIM method were relatively smaller than that of other QTLs identified by CIM method.

Detection of QTL-by-environment interactions for cottonseed oil content The above datasets in QTL mapping were also used to detect QTN-by-environment interactions (QEI) effects using mix model-based composite interval mapping (MCIM) module in QTLNetwork 2.1 software. Among the stable QTLs, a total of 3 QTLs showed significant Q × E interactions ($P < 0.05$), of which *qOC-chr11-1* at 15ale (QEI = -0.002, $P < 0.05$), 14ay (QEI = -0.026, $P < 0.05$), and 15shz (QEI = -0.197, $P < 0.05$), *qOC-chr12-1* at 14ale (QEI = 0.0346, $P < 0.05$), 14kel (QEI = -0.0361, $P < 0.05$), and 15kel (QEI = 0.044, $P < 0.05$), and *qOC-chr17-3* at 15kel (QEI = -0.080, $P < 0.05$), 14shz (QEI = 0.074, $P < 0.05$), 15ay (QEI = -0.119, $P < 0.05$) and 14ale (QEI = -0.092, $P < 0.05$). *qOC-chr12-1* and *qOC-chr17-3* had antagonistic pleiotropic effects in different environments (Additional file 2: Table S2). Moreover, the locus *qOC-chr17-3* was found to be significantly associated with cottonseed oil content in four different environments ($P < 0.05$).

Identification of candidate genes related to cottonseed lipid synthesis

To further mine candidate genes associated with cottonseed oil accumulation, we adopted an integrated method that combines the orthologs sequence alignment with the *Arabidopsis* oil-related genes and KEGG pathway enrichment analysis for those candidate genes within stable QTL regions. A total of 832 genes were acquired within the confidence interval of these stable QTLs (Additional file 3: Table S3). A KEGG pathway enrichment analysis was subsequently carried out to obtain the top 20 KEGG pathways of those genes (Fig. 1a). Of these, six enriched pathways with 43 candidate genes were found to be associated with oil synthesis. These enriched pathways included “biosynthesis of amino acids”, “valine, leucine and isoleucine biosynthesis”, “glycine, serine and threonine metabolism”, “biosynthesis of unsaturated fatty acids”, “fatty acid metabolism”, and “fatty acid biosynthesis” (Fig. 1a; Additional file 4: Table S4). The network relationships analysis of enriched pathway terms completed with metascape (http://metascape.org/gp/index.html) also showed that above-mentioned six pathways have interconnectedness, where terms with a similarity > 0.3 (Fig. 1b).

To date more than 700 *Arabidopsis* acyl-lipid metabolism genes have been detected, of which 135 are directly involved in the processes of *de novo* FA synthesis, TAG synthesis and lipid droplet formation. We further compared the 43 potential oil candidate genes of cottonseed with the 135 *Arabidopsis* oil-related genes using BLAST tool to validate

these common significant potential candidate genes associated with oil content in upland cotton from a different perspective. The results showed that there are 24 potential candidate genes that are involved in cottonseed oil accumulation based on *Arabidopsis* orthologous oil-related genes identified previously (Additional file 5: Table S5).

Dynamic expression patterns of candidate genes in developing cottonseed ovules

To investigate the dynamic expression patterns of 24 candidate lipid genes in developing cottonseed ovules, we analyzed the genes expression level at ten different ovules developmental stages (−3, −1, 0, 1, 3, 5, 10, 20, 30 and 35 DPA) of TM-1 using transcriptome sequencing data (Additional file 6: Table S6; Fig. 2). Among them, 5 genes were not expressed during cottonseed ovules developmental stages. The gene expression levels heatmap of other 19 potential candidate genes showed that they were clustered into three branches based on gene expression profiles after *Z*-score standard normalization (Fig. 2). It was noted that the gene expression levels in branches with red dot were relatively lower during early stage (−3 DPA to 1 DPA), while higher at middle stage (3 DPA to 10 DPA). Surprisingly, most genes showed a sharp decline in expression at mature stage (30DPA to 35 DPA) (Fig. 2).

Transcription factors and microRNAs regulatory network construction of oil candidate genes

Determining gene partners by gene network is an essential step toward understanding protein-coding genes function and identifying relevant biological pathways. To clarify the regulatory networks of potential candidate genes in upland cotton, regulatory associations between microRNAs (miRNAs) and their target genes, transcription factors (TFs) and their host genes, as well as TFs and miRNAs were macroscopically investigated. We directly acquired TFs and miRNAs of *G. hirsutum* L. from PlantTFDB v5.0 (http://plantregmap.cbi.pku.edu.cn/) and miRBase v22.0 (http://www.mirbase.org/), respectively. Together with the results of candidate oil genes mentioned above, total four cotton TFs are obtained, namely *GhLEC1* (*Gh_A11G1560*), *GhHSL1* (*Gh_A12G0518* and *Gh_D04G0979*), *GhPKL* (*Gh_D04G1092*), and *GhABI4* (*Gh_D04G1143*) (Additional file 7: Table S7). In addition to cleaving mRNA, plant miRNA reportedly inhibits the translation of target genes. The miRNAs target genes among 19 cotton lipid candidate genes were predicted using psRNATarget, a plant small RNA target analysis server (http://plantgrn.noble.org/psRNATarget/) (Additional file 8: Table S8). A regulatory network of miRNAs and its target genes that defined the regulation in the complex process of lipid synthesis was constructed based on those results, which showed the interactions between seventeen miRNAs and seven target genes, which including one TFs (*GhHSL1*, *Gh_A12G0518*) (Fig. 3, Additional file 8: Table S8). There are 6 hub nodes with connectivity no less than 3 in the regulatory network, of which 2 miRNAs (ghr-miR2949b and ghr-miR2949c) and 4 candidate target genes (*GhACP2*, *Gh_D11G2438*; *GhLPEAT2*, *Gh_D04G1019*; *GhCICDH*, *Gh_A11G1562*; and *GhHSL1*, *Gh_A12G0518*) (Fig. 3). Of those, *GhHSL1* was targeted by nine microRNAs with translation or cleavage mRNA. As a putative transcription factor, it is also involved in protein phosphorylation and ultimately having high expression level during early-middle stages of developing cotton ovules. *Gh_D11G2438* encodes *ACP2*, a member of acyl carrier protein family, which is involved in fatty acid synthetic process. Gene *Gh_D04G1019* encodes acyl-CoA:lysophosphatidylethanolamine acyltransferase 2 (*LPEAT2*), a member of the *LPLAT* family implicated in very long-chain fatty acid metabolic process. Gene *Gh_A11G1562* encodes cytosolic NADP+-dependent isocitrate dehydrogenase (*CICDH*), which involves in carbon metabolism. The genes *GhACP2*, *GhLPEAT2* and *GhHSL1* were directly target of ghr-miR2949b and ghr-miR2949c. Meanwhile, *GhHSL1* was targeted by additional seven miRNAs at the same time (Fig. 3). The expression level of *GhHSL1* was high (~ 6.99) during the early-middle stages in developing cottonseed ovules, while during the late stage come down (~ 1.69) (Fig. 2). The results indicated that transcription factor *GhHSL1* is specific to lipid synthesis in developing cottonseed ovules.

Discussion

Cottonseed oil has a great potential as an alternative commercial source of biofuels production [33, 34] and edible oil [35] rich in polyunsaturated fatty acids (PUFAs). The investigation of molecular regulatory mechanism of fatty acid synthesis in developing cotton ovules is pivotal for cottonseed biodiesel development. Although some studies have been carried out to identify QTLs/genes for cottonseed oil synthesis so far, the molecular genetic mechanisms are still unclear. In this study, we used an integrated approach, including QTL mapping and gene identification, RNA-seq, and regulatory network, to identify and analyze candidate genes involved in cottonseed oil accumulation.

With the development of SNP arrays, sequencing and genotyping technologies, an increasing number of QTL mapping about cottonseed oil studies recently have been carried out [36–40]. This study developed an intra-specific RIL population in upland cotton for QTL mapping, which consisting of 196 lines with the parents 0-153 and sGK9708. The consensus genetic map was constructed by Zhang *et al.* [41] in previous study, which covered the whole genome of upland cotton with a high saturation and was a valuable tool for QTL mapping across the whole genome. 39 QTLs for cottonseed oil content were identified in the study. As the quantitative trait influenced by environment, some QTLs identified in multiple specific environments or generations (≥ 3) were named as stable QTLs [42]. Based on this definition, this study therefore identified five stable QTLs. To further determine whether the stable QTLs obtained in our study were new or had been identified previously, we compared our results with those QTLs from cotton QTL database based on their physical confidence intervals and previous QTL mapping reports. It was found that stable QTL *qOC-chr24-2* shared the overlapping confidence intervals with QTLs identified in previous GWAS studies [8, 43], while other four were newly identified which could provide more information about the mechanism of cottonseed oil accumulation and accelerate biofuels development of cottonseed oil. Stable QTL *qOC-chr17-3* was also identified to be connected with seed oil content in other upland cotton RIL population (Zhong 70) which was unpublished nowadays. Moreover, phenotypic plasticity refers to the ability of a single genotype to exhibit variable phenotypes in different environments. When phenotypic plasticity differs among genotypes, it can be classified as a GEI. A mixed model methodology with terms for QEI can be used to reveal the genetic basis of complex traits showing GEI. In this study, 3 stable QTLs with cottonseed oil traits were identified with significant QEI (Additional file 2: Table S2), which demonstrated that average temperature and the coordinate of latitude and longitude could influence cottonseed oil content across years and locations.

To further explore the mechanism of cottonseed oil accumulation from the standpoint of lipid candidate genes, total 832 genes located in stable QTL (*qOC-chr11-1*, *qOC-chr12-1*, *qOC-chr17-3*, *qOC-chr22-1*, and *qOC-chr24-2*) were extracted based on the physical location. However, the analysis and interpretation of these candidate genes is often a major challenge for many researchers which requires an impractically large amount of manual literature searching to interpret. A standard approach to addressing this problem is pathway enrichment analysis, which summarizes the large gene list as a smaller list of more easily interpretable pathways [44]. In this study, we obtained the top 20 KEGG enrichment pathway of 832 genes using pathway enrichment analysis database KOBAS (http://kobas.cbi.pku.edu.cn/index.php) (Fig. 1a). Acyl-lipid synthesis, as a complex biological process, includes at least the conversion of sucrose to pyruvate, plastid *de novo* FA synthesis, endoplasmic TAG synthesis, and oil-body assembly [45]. Moreover, an integrated omics analysis of rapeseed and soybean previously showed that some genes involved in photosynthesis and plant hormone signal transduction were related to lipid metabolism [46], as well as substrate competition between seed oil and protein synthesis exists in oil crops [47]. Therefore, the six pathways, “biosynthesis of amino acids”, “valine, leucine and isoleucine biosynthesis”, “glycine, serine and threonine metabolism”, “biosynthesis of unsaturated fatty acids”, “fatty acid metabolism”, and “fatty acid biosynthesis”, were ultimately selected for subsequent identification of candidate oil genes based on the results of KEGG enrichment analysis and network analyses of enriched terms (Fig. 1a,b). Among them, 43 genes were found and further analyzed by BLAST with *Arabidopsis* oil-related genes. As a result, 24 *Arabidopsis*-ortholog oil genes in upland cotton were obtained, five of which were not expressed during developing cotton ovules (Additional file 5: Table S5). Integration of KEGG, BLAST and

RNA-seq analysis results, 19 credible candidate genes related to cottonseed oil accumulation were ultimately obtained (Additional file 5: Table S5). Of these, transcription factor *HSL1* were found to regulate a series of reactions in oil synthesis in previous researches for oilseed crops [29, 48]. This indicated that *GhHSL1* may play a significant role for oil synthesis in developing cotton ovules.

The expression profiles of these 19 candidate genes were analyzed to investigate the cottonseed oil accumulation mechanism in developing cotton ovules. Ma *et al.* [39] reported that cottonseed oil accumulates rapidly from approximately 20 DPA in cotton ovules. Therefore, we focused on genes with significantly higher expression levels at the early-middle stages (-3–20 DPA) of ovule development than those in the late stages (30–35 DPA). In the end, total 19 candidate genes expressed in developing cotton ovules were considered to be related to cottonseed oil accumulation (Additional file 6: Table S6). The heatmap clusters of gene expression levels showed that branches with red hot in Fig. 2 basically matched the condition of cottonseed oil rapid accumulation nearly 20 DPA.

Gene regulatory network analysis is a powerful approach to identify the key regulators underlying important biological processes. In this study, we identified one TFs regulated by nine miRNAs, and three core genes (*GhCICDH*, *GhACP2* and *GhLPEAT2*) that all have sustained up-regulation expression in early-middle stage of developing cottonseed ovules (Figs. 2,3,4; Additional file 6: Table S6), and two hub miRNAs (ghr-miR2949b and ghr-miR2949c) that both have regulated three genes in the regulatory network (Figs. 3). Previous researches have showed that these miRNAs appeared to play a critical role in regulating cottonseed oil accumulation by regulating the accumulation of auxin [49]. Moreover, it was formerly noted that ghr-miR2949b was predicted to act as a regulator of sugar transporter protein (STP) and the down-regulation of ghr-miR2949b could increase STP expression levels [50]. Meanwhile, there is a lack of soluble sugar in the plastids, and most of soluble sugar is mainly transported from cytoplasm via STPs [51]. Additionally, *Arabidopsis* transcription factors *AtHSL1* could repress the sugar-inducible expression of seed maturation program in seedlings and play an essential role in regulating the transition from seed maturation to seedling growth, which contributes to the seed oil accumulation [52]. Moreover, *HSL1*-dependent histone methylation plays critical roles in regulation of seed dormancy during seed germination and early seedling growth, which will affect the oil content accumulated in seed [53]. Similarly, *AtLPEAT2* [54], *AtLACS2* [55] and *ChACP4* [56] were found to be associated with PI (34:3), linolenic acid (P = 2.63e-07), and pyruvate (LOD = 14.68) in oil synthesis process, respectively. Ma *et al.* [39] also found that the cotton genes *GhHSL1*, *GhACP2* and *GhLPEAT2* were differentially expressed between the upland cotton (*G. hirsutum* L.) high oil content genotype (3008) and low oil content genotype (3012) in 10, 20 and 30 DPA ovules. These results indicated that *GhHSL1*, *GhACP2* and *GhLPEAT2* may play vital roles in cottonseed oil accumulation. Of course, further experiments are needed for the other lipid candidate genes to confirm that they can indeed affect the oil content of cotton seeds.

A detailed cottonseed oil accumulation model was constructed based on our results of the entire study (Fig. 4). It starts with the transportation of sucrose, followed by glycolysis and fatty acid synthesis in the cytosol, and results in the formation of TAG in the endoplasmic reticulum. Pyruvate and acetyl-CoA are the substrate for FA synthesis. The putative lipid candidate genes, participating in the pathway of pyruvate and acetyl-CoA formation, showed similar expression patterns during developing cottonseed ovules (Fig. 2), which indicated that these genes may play an essential role in cottonseed oil accumulation. Additionally, it is noteworthy that these genes (such as *GhCICDH*, *GhACP2*, *GhFAD2*, *GhACBP*, *GhLPEAT2*, *GhLACS4* and *GhPAH1*) were up-regulated expressed at early-middle development stages of cotton ovules (Fig. 4) and primarily involved in carbohydrate metabolism and lipid synthesis (Additional file 5: Table S5), indicated that they play a vital role in carbon source supply, FA synthesis, and TAG synthesis. As acyl-CoA-binding protein (*ACBP*) plays an important housekeeping role in lipid metabolism by maintaining the intracellular acyl-CoA pool, and involved in lipid synthesis and transport, gene expression, and membrane biogenesis [57]. Genes *ACBP4*, *ACBP6*, *FAD2* and *PAH1* were also found to have catalyzed substrate Gly-3-P

to generate TAG. Furthermore, the deletion of 3-phosphate glycerol dehydrogenase isomer (*GPDH*) in *Yarrowia lipolytica* can redirect carbon flux to lipid synthesis and significantly increase lipid content [58]. *LACS4* was found to be involved in the synthesis of very long-chain fatty acid [55]. This investigation will help to understand the molecular regulatory basis of cottonseed oil accumulation, and provide a foundation to increase cottonseed oil content for biofuels development.

Conclusion

In the present study, a total of five stable QTLs related to cottonseed oil content were identified, of which four QTLs were identified for the first time and one QTL was verified in previous studies. Upon further analysis of these genes around the five stable QTLs by means of KEGG enrichment, sequence alignment and TFs/microRNAs regulatory network, we obtained 19 oil candidate genes in cottonseed (including four TFs) and two core microRNAs related to cottonseed oil accumulation during the developing ovules. Of these, genes *GhACP2*, *GhLPEAT2* and transcription factor *GhHSL1*, along with ghr-miR2949, as the hub nodes of gene regulation network, may be key candidate factors on increasing seed oil content in *G. hirsutum* L. These results provide a fundamental resource for cottonseed genetic research and breeding, and encourage the prospects of the cottonseed oil used for biofuels in the future.

Methods

Plant materials and seed oil trait

A RIL population (n = 196) developed between two upland cotton parental lines, 0-153 and sGK9708, in our lab was used for QTL mapping in this study. Briefly, the cross was made in 2001, and F_1 plants were self-pollinated in Anyang, Henan Province in 2002 (02AY), and 250 F_2 seeds harvested were grown in 10 rows (each 8 m long and 0.8 m apart in 03AY), then 196 $F_{2:3}$ families were randomly selected to be planted in single-row plots, with the parents and F_1 plants in two-row plots (each 8 m long and 0.8 m apart in 04AY). Finally, a segregation population consisting of 196 $F_{6:8}$ RILs were developed in 2007 [42]. In year 2014 and 2015, the RIL population and its parents were planted in 4 various ecological locations (Anyang, Alaer, Shihezi, and Kuerle) with each row 5 m long and 0.8 m apart, and 20 plants in each row. Field management was performed according to the local farming practices.

In 2014 and 2015 years, the mature open boll samples of RIL population and its parents were harvested by hand and ginned by machines based on per plant in the plot, the seed oil contents were subsequently investigated. In this study, we randomly chose 100 healthy seeds (with hull) of each sample in Anyang, Alaer, Shihezi and Kuerle respectively and shelled seeds by hand to measure the oil content with 3 repeats per genotype (2014 and 2015). The Niumag Imaging and Analyzing System (NMI20-Analyst, Niumag Electric Corporation, Shanghai, China) was used to detect the total oil content of mature cottonseeds.

Data statistical analysis

Student's t test was used to detect the statistical significance of the phenotypes. A variance (ANOVA) analysis of the oil content for the 196 RIL population across eight environments were performed using SPSS. The software QTL IciMapping was used to estimate the combined broad-sense heritability (h^2) of oil content among different environments [59].

QTL mapping and QTL-by-environment interaction

The consensus high-density genetic map used in this study was constructed with three types of markers (8295 markers, 5197.17 cM) by Zhang *et al.* [41]. We could determine the physical positions of linkage groups using the map, which could supply powerful information for subsequently selecting candidate genes. In this study, composite interval mapping (CIM) method in WinQTLCart2.5 [60], and genome-wide composite interval mapping (GCIM) method in QTL.gCIMapping [61-63] were applied to detect main- and small-effect QTLs, and estimate the effects which were included as co-factors in the stepwise selection model used for background marker selection in the CIM. QTLs that are consistently identified in at least three environments are considered stable [42]. Positive additive effects indicated that 0-153 alleles increased the phenotypic trait values, and negative scores indicated that 0-153 alleles decreased the values, while the sGK9708 alleles had the opposite effects. QTL were named as follows: (q + trait abbreviation) + chromosome/linkage groups + QTL number. QTL for the same trait across different genetic backgrounds were considered common QTL when their confidence intervals overlapped. The variance ratio, additive (a) effects of each particular QTL interpretation, and QTL-by-environment interactions were obtained with the MCIM method in QTLNetwork-2.1 [64]. To produce an accurate LOD profile and adjust the background mark effect, a 10 cM window size, a 0.5 cM walking speed and the critical LOD score of 3 for significant QTL were used to declare the linkage [65,66]. Additionally, a critical *P*-value of 0.05 was used to avoid over-fitting of the model [67].

Identification and annotation of genes around QTL regions

To identify all genes in the confidence intervals of QTLs regions for cottonseed oil content, we used Cotton Functional Genomics Database (CottonFGD, https://cottonfgd.org/) [68] to search for *G. hirsutum* L. genes based on genomic coordinate or genetic marker. We annotated all the genes around QTL regions through Gene Ontology (GO) annotation and Kyoto Encyclopedia of Gene and Genome (KEGG) pathway enrichment analysis with agriGO 2.0 (http://bioinfo.cau.edu.cn/agriGO/) [69] and KOBAS 3.0 (http://kobas.cbi.pku.edu.cn/index.php) [70], respectively. The *P*-values for each KEGG biological process was calculated by Fisher's exact test. To control the false discovery rate (FDR $\leq$ 0.05), the Benjamin-Hochberg method was used to conduct multiple testing correction. In addition, the small term cut-off value was set at 5.

Construction and visualization of lipid-related genes regulatory network

TFs and miRNAs of *G. hirsutum* L. were directly acquired from databases PlantTFDB v5.0 (http://plantregmap.cbi.pku.edu.cn/) [71] and miRBase v22.0 (http://www.mirbase.org/) [72], respectively. The target genes of miRNAs were predicted via psRNATarget (http://plantgrn.noble.org/psRNATarget) [73] with default parameters. Graphical visualization of the TFs and miRNAs regulation network was performed using Cytoscape 3.6.1 (http://www.cytoscape.org/) [74].

Abbreviations

ACP: acyl carrier protein; ANOVA: analysis of variance; cM: centimorgan; DGAT: diacylglycerol acyltransferase; ER: endoplasmic reticulum; FAD: fatty acyl desaturase; LOD: logarithm of odds; PAP: phosphatidate phosphatase; QTL: quantitative trait locus / loci; SNP: single nucleotide polymorphism; FA: fatty acid; TAG: triacylglycerol; TF: Transcription factors; miRNA: microRNAs; FATB: fatty acylthioesterase B; LPAAT; 2-lysophosphatidic acid acyltransferase; OC: oil content; Gly-3-P: glycerate 3-phosphate; PI: phosphatidylinositol; CICDH: NADP+-dependent isocitrate dehydrogenase regulatory subunit; KASII: ketoacyl-ACP synthase II; ENR: enoyl-ACP reductase; PAH: phosphatidate phosphatase; ACBP: acyl CoA binding protein; FAD2: oleate desaturase 2; LPEAT: 1-acylglycerol-3-phosphoethanolamine acyltransferase; DAG-CPT: diacylglycerol choline phosphotransferase; CCT1: choline-phosphate cytidylyltransferase 1; LACS: long-chain acyl-CoA synthetase; LEC: leafy cotyledon 1; HSL1: high-level expression of sugar-inducible gene-like 1; PKL: chromatin remodeling factor of the CHD3 group; ABI4: abscisic acid-insensitive

Declarations

Acknowledgements

Not applicable.

Funding

This study was supported by the National Natural Science Foundation of China (31621005), the National Agricultural Science and Technology Innovation project for CAAS (CAAS-ASTIP-2016-ICR), and the Central Level of the Scientific Research Institutes for Basic R & D Special Fund Business (1610162019010101).

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that there are no competing interests.

Authors' contributions

ZZ, SH and YY conceived the study; ZZ contributed to the data processing, analysis and wrote the manuscript; GJ collected and analyzed the phenotype data; ZZ, GW, LJ and SY managed the RIL population; LA, GQ and PJ collected the phenotype data in Anyang and Shihezi; FS and DX managed and collected the phenotype data in Kuerle and Alaer; LS and CQ helped with manuscript reviewing; SH and YY provided the resources, designed the experiment and revised the manuscript. All authors read and approved the final manuscript.

References

1. Lukonge E, Labuschagne MT, Hugo A. The evaluation of oil and fatty acid composition in seed of cotton accessions from various countries. J Sci Food Agric. 2007;87:340–7.
2. Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, Chen L, He Y, Zhang L, Zhu L, Li Y, Liang Q, Lin Z, Yang X, Liu N, Jin S, Lei Y, Ding Y, Li G, Ruan X, Ruan Y, Zhang X. The genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the allopolyploidization and development of superior spinnable fibres. Sci Rep. 2015;5:17662.
3. Yan X, Wang B, Zhang J. Liquefaction of cotton seed in sub-critical water/ethanol with modified medical stone for bio-oil. Bioresour Technol. 2015;197:120–7.
4. Jiao X, Zhao X, Zhou XR, Green AG, Fan Y, Wang L, Singh SP, Liu Q. Comparative transcriptomic analysis of developing cotton cotyledons and embryo axis. PLoS One. 2013;8(8):e71756.

5. Sinha D, Murugavelh S. Biodiesel production from waste cotton seed oil using low cost catalyst: engine performance and emission characteristics. Perspect Sci. 2016;8:237–40.
6. Ma M, Ren Y, Xie W, Zhou D, Tang S, Kuang M, Wang Y, Du SK. Physicochemical and functional properties of protein isolate obtained from cottonseed meal. Food Chem. 2018;240:856–62.
7. Desrochers P, Szurmak J. Insights from the history of the American cottonseed industry. Sustainability. 2017;9:579.
8. Shang L, Abduweli A, Wang Y, Hua JP. Genetic analysis and QTL mapping of oil content and seed index using two recombinant inbred lines and two backcross populations in Upland cotton. Plant Breed. 2016;135:224–31.
9. Yu J, Yu S, Fan S, Song M, Zhai H, Li X, Zhang J. Mapping quantitative trait loci for cottonseed oil, protein and gossypol content in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. Euphytica. 2012;187:191–201.
10. Alfred Q, Hai YL, Hai MX, Jin LR, Guo WJ, Zhu SJ, Shi CH. Mapping of quantitative trait loci for oil content in cottonseed kernel. J Genet. 2012;91:289–95.
11. Liu D, Liu F, Shan X, Zhang J, Tang S, Fang X, Liu X, Wang W, Tan Z, Teng Z, Zhang Z, Liu D. Construction of a high-density genetic map and lint percentage and cottonseed nutrient trait QTL identification in upland cotton (*Gossypium hirsutum* L.). Mol Genet Genomics. 2015;290:1683–700.
12. Du X, Liu S, Sun J, Zhang G, Jia Y, Pan Z, Xiang H, He S, Xia Q, Xiao S, Shi W, Quan Z, Liu J, Ma J, Pang B, Wang L, Sun G, Gong W, Jenkins JN, Lou X, Zhu J, Xu H. Dissection of complicate genetic architecture and breeding perspective of cottonseed traits by genome-wide association study. BMC Genom. 2018;19(1):451.
13. Yuan Y, Wang X, Wang L, Xing H, Wang Q, Saeed M, Tao J, Feng W, Zhang G, Song XL, Sun XZ. Genome-wide association study identifies candidate genes related to seed oil composition and protein content in *Gossypium hirsutum* L. Front Plant Sci. 2018;9:1359.
14. Badigannavar A, Myers GO. Genetic diversity, population structure and marker trait associations for seed quality traits in cotton (*Gossypium hirsutum*). J Genet. 2015;94(1):87–94.
15. Hovav R, Faigenboim-Dorona A, Kadmona N, Hub G, Zhang X, Gallagherb JP, Wendelb JF. A Transcriptome profile for developing seed of polyploid cotton. Plant Genome. 2015;8.
16. Li X, Teitgen AM, Shirani A, Ling J, Busta L, Cahoon RE, Zhang W, Li Z, Chapman KD, Berman D, Zhang C, Minto RE, Cahoon EB. Discontinuous fatty acid elongation yields hydroxylated seed oil with improved function. Nat Plants. 2018;4(9):711–20.
17. Zhao Y, Huang Y, Wang Y, Cui Y, Liu Z, Hua J. RNA interference of GhPEPC2 enhanced seed oil accumulation and salt tolerance in Upland cotton. Plant Sci. 2018;271:52–61.
18. Roesler K, Shintani D, Savage L, Boddupalli S, Ohlrogge J. Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiol. 1997;113(1):75–81.
19. Cui Y, Liu Z, Zhao Y, Wang Y, Huang Y, Li L, Wu H, Xu S, Hua JP. Overexpression of heteromeric *GhACCase* subunits enhanced oil accumulation in Upland cotton. Plant Mol Bio Rep. 2017;35:287–97.
20. Klaus D, Ohlrogge JB, Neuhaus HE, Dörmann P. Increased fatty acid production in potato by engineering of acetyl-CoA carboxylase. Planta. 2004;219(3):389–96.
21. Murad AM, Vianna GR, Machado AM, da Cunha NB, Coelho CM, Lacerda VA, Coelho MC, Rech EL. Mass spectrometry characterisation of fatty acids from metabolically engineered soybean seeds. Anal Bioanal Chem. 2014;406:2873–83.
22. Liu F, Zhao YP, Zhu HG, Zhu QH, Sun J. Simultaneous silencing of *GhFAD2-1* and *GhFATB* enhances the quality of cottonseed oil with high oleic acid. J Plant Physiol. 2017;215:132–9.

23. Zhou L, Luo L, Zuo JF, Yang L, Zhang L, Guang X, Niu Y, Jian J, Geng QC, Liang L, Song Q, Dunwell JM, Wu Z, Wen J, Liu YQ, Zhang YM. Identification and validation of candidate genes associated with domesticated and improved traits in soybean. Plant Genome. 2016; 9(2).
24. Liu Q, Wu M, Zhang B, Shrestha P, Petrie J, Green AG, Singh SP. Genetic enhancement of palmitic acid accumulation in cotton seed oil through RNAi down-regulation of ghKAS2 encoding β-ketoacyl-ACP synthase II (*KASII*). Plant Biotechnol J. 2017;15(1):132–43.
25. Shang X, Cheng C, Ding J, Guo W. Identification of candidate genes from the *SAD* gene family in cotton for determination of cottonseed oil composition. Mol Genet Genomics. 2017;292(1):173–86.
26. Zang X, Pei W, Wu M, Geng Y, Wang N, Liu G, Ma J, Li D, Cui Y, Li X, Zhang J, Yu J. Genome-scale analysis of the WRI-Like family in Gossypium and functional characterization of *GhWRI1a* controlling triacylglycerol content. Front Plant Sci. 2018;9:1516.
27. Vigeolas H, Waldeck P, Zank T, Geigenberger P. Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. Plant Biotechnol J. 2007;5(3):431–41.
28. Cui Y, Ma J, Liu G, Wang N, Pei W, Wu M, Li X, Zhang J, Yu J. Genome-wide identification, sequence variation, and expression of the Glycerol-3-Phosphate Acyltransferase (*GPAT*) gene family in *Gossypium*. Front Genet. 2019;10:116.
29. Su Y, Liang W, Liu Z, Wang Y, Zhao Y, Ijaz B, Hua J. Overexpression of *GhDof1* improved salt and cold tolerance and seed oil content in *Gossypium hirsutum*. J Plant Physiol. 2017;218:222–34.
30. Wang N, Ma J, Pei W, Wu M, Li H, Li X, Yu S, Zhang J, Yu J. A genome-wide analysis of the lysophosphatidate acyltransferase (*LPAAT*) gene family in cotton: organization, expression, sequence variation, and association with seed oil content and fiber quality. BMC Genom. 2017;18(1):218.
31. Zhang L, Wang SB, Li QG, Song J, Hao YQ, Zhou L, Zheng HQ, Dunwell JM, Zhang YM. An integrated bioinformatics analysis reveals divergent evolutionary pattern of oil biosynthesis in high- and low-oil plants. PLoS One. 2016;11(5):e0154882.
32. Liu JY, Li P, Zhang YW, Zuo JF, Li G, Han X, Dunwell JM, Zhang YM. Three-dimension genetic networks among seed oil-related traits, metabolites and genes reveal the genetic foundations of oil synthesis in soybean. Plant J. 2020. doi:10.1111/tpj.14788.
33. Panagiotopoulos IA, Pasias S, Bakker RR, de Vrije T, Papayannakos N, Claassen PA, Koukios EG. Biodiesel and biohydrogen production from cotton-seed cake in a biorefinery concept. Bioresour Technol. 2013;136:78–86.
34. Costa GB, Fernandes DD, Almeida VE, Araújo TS, Melo JP, Diniz PH, Véras G. Digital image-based classification of biodiesel. Talanta. 2015;139:50–5.
35. Yang A, Qi M, Wang X, Wang S, Sun L, Qi D, Zhu L, Duan Y, Gao X, Ali Rajput S, Zhang N. Refined cottonseed oil as a replacement for soybean oil in broiler diet. Food Sci Nutr. 2019;7(3):1027–34.
36. Ma Z, He S, Wang X, Sun J, Zhang Y, Zhang G, Wu L, Li Z, Liu Z, Sun G, Yan Y, Jia Y, Yang J, Pan Z, Gu Q, Li X, Sun Z, Dai P, Liu Z, Gong W, Wu J, Wang M, Liu H, Feng K, Ke H, Wang J, Lan H, Wang G, Peng J, Wang N, Wang L, Pang B, Peng Z, Li R, Tian S, Du X. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat Genet. 2018;50(6):803–13.
37. Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z, Shen C, Li J, Zhang L, Zhou X, Nie X, Li Z, Guo K, Ma Y, Huang C, Jin S, Zhu L, Yang X, Min L, Yuan D, Zhang Q, Lindsey K, Zhang X. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. Nat Genet. 2017;49(4):579–87.
38. Fang L, Wang Q, Hu Y, Jia Y, Chen J, Liu B, Zhang Z, Guan X, Chen S, Zhou B, Mei G, Sun J, Pan Z, He S, Xiao S, Shi W, Gong W, Liu J, Ma J, Cai C, Zhu X, Guo W, Du X, Zhang T. Genomic analyses in cotton identify signatures of

selection and loci associated with fiber quality and yield traits. Nat Genet. 2017;49(7):1089–98.

39. Ma J, Liu J, Pei W, Ma Q, Wang N, Zhang X, Cui Y, Li D, Liu G, Wu M, Zang X, Song J, Zhang J, Yu S, Yu J. Genome-wide association study of the oil content in upland cotton (*Gossypium hirsutum* L.) and identification of *GhPRXR1*, a candidate gene for a stable QTL qOC-Dt5-1. Plant Sci. 2019;286:89–97.
40. Wang W, Sun Y, Yang P, Cai X, Yang L, Ma J, Ou Y, Liu T, Ali I, Liu D, Zhang J, Teng Z, Guo K, Liu D, Liu F, Zhang Z. A high density SLAF-seq SNP genetic map and QTL for seed size, oil and protein content in upland cotton. BMC Genom. 2019;20(1):599.
41. Zhang Z, Li J, Jamshed M, Shi Y, Liu A, Gong J, Wang S, Zhang J, Sun F, Jia F, Ge Q, Fan L, Zhang Z, Pan J, Fan S, Wang Y, Lu Q, Liu R, Deng X, Zou X, Jiang X, Liu P, Li P, Iqbal MS, Zhang C, Zou J, Chen H, Tian Q, Jia X, Wang B, Ai N, Feng G, Wang Y, Hong M, Li S, Lian W, Wu B, Hua J, Zhang C, Huang J, Xu A, Shang H, Gong W, Yuan Y. Genome-wide quantitative trait loci reveal the genetic basis of cotton fibre quality and yield-related traits in a *Gossypium hirsutum* recombinant inbred line population. Plant Biotechnol J. 2020;18(1):239–53.
42. Sun FD, Zhang JH, Wang SF, Gong WK, Shi YZ, Liu AY, Li JW, Gong JW, Shang HH, Yuan YL. QTL mapping for fiber quality traits across multiple generations and environments in upland cotton. Mol Breeding. 2012;30:569.
43. Said JI, Knapka JA, Song M, Zhang J. Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between *Gossypium hirsutum* and *G. hirsutum* × *G. barbadense* populations. Mol Genet Genomics. 2015;290(4):1615–25.
44. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, Merico D, Bader GD. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019;14(2):482–517.
45. McGlew K, Shaw V, Zhang M, Kim RJ, Yang W, Shorrosh B, Suh MC, Ohlrogge J. An annotated database of Arabidopsis mutants of acyl lipid metabolism. Plant Cell Rep. 2015;34(4):519–32.
46. Zhang Z, Dunwell JM, Zhang YM. An integrated omics analysis reveals molecular mechanisms that are associated with differences in seed oil content between *Glycine max* and *Brassica napus*. BMC Plant Biol. 2018;18(1):328.
47. Voelker T, Kinney AJ. Variations in the biosynthesis of seed-storage lipids. Annu Rev Plant Physiol Plant Mol Biol. 2001;52:335–61.
48. Xia B, Cai GH, Yang H, Wang SP, Mitchell GA, Wu JW. Adipose tissue deficiency of hormone-sensitive lipase causes fatty liver in mice. PLoS Genet. 2017;13(12):e1007110.
49. Zhang H, Wan Q, Ye W, Lv Y, Wu H, Zhang T. Genome-wide analysis of small RNA and novel microRNA discovery during fiber and seed initial development in *Gossypium hirsutum*. L PLoS One. 2013;8(7):e69743.
50. Zhang B, Zhang X, Liu G, Guo L, Qi T, Zhang M, Li X, Wang H, Tang H, Qiao X, Pei W, Shahzad K, Xing C, Zhang J, Wu J. A combined small RNA and transcriptome sequencing analysis reveal regulatory roles of miRNAs during anther development of Upland cotton carrying cytoplasmic male sterile *Gossypium harknessii* (D2) cytoplasm. BMC Plant Biol. 2018;18(1):242.
51. Wang S, Wang C, Zhang XX, Chen X, Liu JJ, Jia XF, Jia SQ. Transcriptome de novo assembly and analysis of differentially expressed genes related to cytoplasmic male sterility in cabbage. Plant Physiol Biochem. 2016;105:224–32.
52. Tsukagoshi H, Morikami A, Nakamura K. Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in *Arabidopsis* seedlings. Proc Natl Acad Sci USA. 2007;104(7):2543–7.
53. Chen N, Wang H, Abdelmageed H, Veerappan V, Tadege M, Allen RD. HSI2/VAL1 and HSL1/VAL2 function redundantly to repress DOG1 expression in *Arabidopsis* seeds and seedlings. New Phytol. 2020. doi:10.1111/nph.16559.

54. Jasieniecka-Gazarkiewicz K, Lager I, Carlsson AS, Gutbrod K, Peisker H, Dörmann P, Stymne S, Banaś A. Acyl-CoA:Lysophosphatidylethanolamine acyltransferase activity regulates growth of *Arabidopsis*. Plant Physiol. 2017;174(2):986–98.
55. Katavic V, Shi L, Yu Y, Zhao L, Haughn GW, Kunst L. Investigation of the contribution of oil biosynthetic enzymes to seed oil content in *brassica napus* and *Arabidopsis thaliana*. Can J Plant Sci. 2014;94:1109–12.
56. Feng Y, Zhang Y, Wang Y, Liu J, Liu Y, Cao X, Xue S. Tuning of acyl-ACP thioesterase activity directed for tailored fatty acid synthesis. Appl Microbiol Biotechnol. 2018;102:3173–82.
57. Yurchenko OP, Weselake RJ. Involvement of low molecular mass soluble acyl-CoA-binding protein in seed oil biosynthesis. N Biotechnol. 2011;28(2):97–109.
58. Beopoulos A, Mrozova Z, Thevenieau F, Le Dall M-T, Hapala I, Papanikolaou S, et al. Control of lipid accumulation in the Yeast *Yarrowia lipolytica*. Appl Environ Microbiol. 2008;74:7779.
59. Meng L, Li H, Zhang L, Wang J. QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J. 2015;3:269–83.
60. Wang S, J.Basten C, Zeng ZB. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. 2012; https://brcwebportal.cos.ncsu.edu/qtlcart/index.php.
61. Wang SB, Wen YJ, Ren WL, Ni YL, Zhang J, Feng JY, Zhang YM. Mapping small-effect and linked quantitative trait loci for complex traits in backcross or DH populations via a multi-locus GWAS methodology. Sci Rep. 2016;6:29951.
62. Wen YJ, Zhang YW, Zhang J, Feng JY, Dunwell JM, Zhang YM. An efficient multi-locus mixed model framework for the detection of small and linked QTLs in F_2. Brief Bioinform. 2019;20(5):1913–24.
63. Zhang YW, Wen YJ, Dunwell JM, Zhang YM. QTL.gCIMapping.GUI v2.0: An R software for detecting small-effect and linked QTLs for quantitative traits in bi-parental segregation populations. Comput Struct Biotechnol J. 2019;18:59–65.
64. Yang J, Hu C, Hu H, Yu R, Xia Z, Ye X, Zhu Y. QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. Bioinformatics. 2008;24:721–3.
65. Hwang S, Ray JD, Cregan PB, King CA, Davies MK, Purcell LC. Genetics and mapping of quantitative traits for nodule number, weight, and size in soybean (*Glycine max* L. [Merr.]). Euphytica. 2014;195:419–34.
66. Du Q, Gong C, Wang Q, Zhou D, Yang H, Pan W, Li B, Zhang D. Genetic architecture of growth traits in Populus revealed by integrated quantitative trait locus (QTL) analysis and association studies. New Phytol. 2016;209:1067–82.
67. Williams B, Williams R, Kantartzi S, Meksem K, Barakat A, Lightfoot DA, Kassem A. Genetic analysis of root and shoot traits in the soybean [*Glycine max* (L.) Merr.] Essex' by Forrest' recombinant inbred line population. J Plant Genome Sci. 2012;1:1–9.
68. Zhu T, Liang C, Meng Z, Sun G, Meng Z, Guo S, Zhang R. CottonFGD: an integrated functional genomics database for cotton. BMC Plant Biol. 2017;17(1):101.
69. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 2017;45(W1):W122–9.
70. Ai C, Kong L. CGPS: A machine learning-based approach integrating multiple gene set analysis tools for better prioritization of biologically relevant pathways. J Genet Genomics. 2018;45(9):489–504.
71. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 2017;45(D1):D1040–5.

72. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42(Database issue):D68–73.
73. Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res. 2018;46(W1):W49–54.
74. Killcoyne S, Carter GW, Smith J, Boyle J. Cytoscape: a community-based framework for network modeling. Methods Mol Biol. 2009;563:219–39.

Tables

Table 1 Parental and population statistics for cottonseed oil content trait in 0-153 × sGK9708 population across eight environments

onments	Parents		RIL population					
	Paternal 0-153	Maternal sGK9708	Mean ± SD	Range	Variance	Skew	Kurtosis	CV (%)
r1_14ay	31.64	33.58**	28.49 ± 1.36	24.86~32.00	1.86	0.13	-0.05	4.79
2_14ale	30.20	35.31**	27.38 ± 1.31	22.54~30.92	1.72	-0.11	0.68	4.78
3_14kel	31.32	35.74**	27.27 ± 1.32	24.16~31.59	1.75	0.48	0.26	4.85
4_14shz	30.79	31.82**	27.89 ± 1.25	24.64~31.20	1.55	0.05	-0.07	4.46
r5_15ay	28.49	31.67**	27.53 ± 1.34	23.69~31.05	1.93	0.46	1.19	4.80
6_15ale	28.09	33.25**	25.09 ± 1.34	21.46~29.17	1.79	0.36	0.47	5.34
7_15kel	29.67	35.57**	27.13 ± 1.47	23.97~32.72	2.18	0.80	0.97	5.44
8_15shz	29.50	32.14**	27.26 ± 1.32	24.25~32.10	1.75	0.74	1.50	4.85

**The 0.01 level of significant difference of oil content between two parents.

Env, Environment; ay, Anyang; ale, Alaer; shz, Shihezi; kel, Kuerle; 14ay, year 2014 in Anyang; 15ay, year 2015 in Anyang; SD, standard deviation; CV, Coefficient of variance.

Table 2 Summary of stable QTLs observed for cottonseed oil content trait in the 0-153 × sGK9708 RIL mapping population by QTL.gCIMapping and WinQTLCart.

QTL name	Environment	Chr	Position	Interval markers	LOD	r^2	Additive effect	Reported previously
qOC-chr11-1	14ay 15ale 15shz	At11	22449635~52248998	M7345~M7361	3.19~3.42	4.94~5.68	0.32~0.56	
qOC-chr12-1	14ale 14kel 15kel	At12	9747983~10423316	M7570~M7573	3.05~3.54	6.87~7.35	-0.32~-0.30	
qOC-chr17-3	14ale 14shz 15ay 15kel	Dt04	24137015~37886008	M10128~M10141	3.05~5.96	6.53~14.15	-0.93~-0.46	
qOC-chr22-1	14ale 15ale 15kel	Dt09	7241124~7384663	M10236~M10239	3.36~3.40	6.99~7.61	0.19~0.24	
qOC-chr24-2	14shz 14kel 15kel	Dt11	62294040~64675839	M13116~M13128	3.10~5.84	4.43~13.17	-0.58~-0.32	[8,43]

OC: oil content; *chr*: chromosome; M7345: Marker7345; r^2: the percentage of phenotypic variance explained by QTL.

Figures

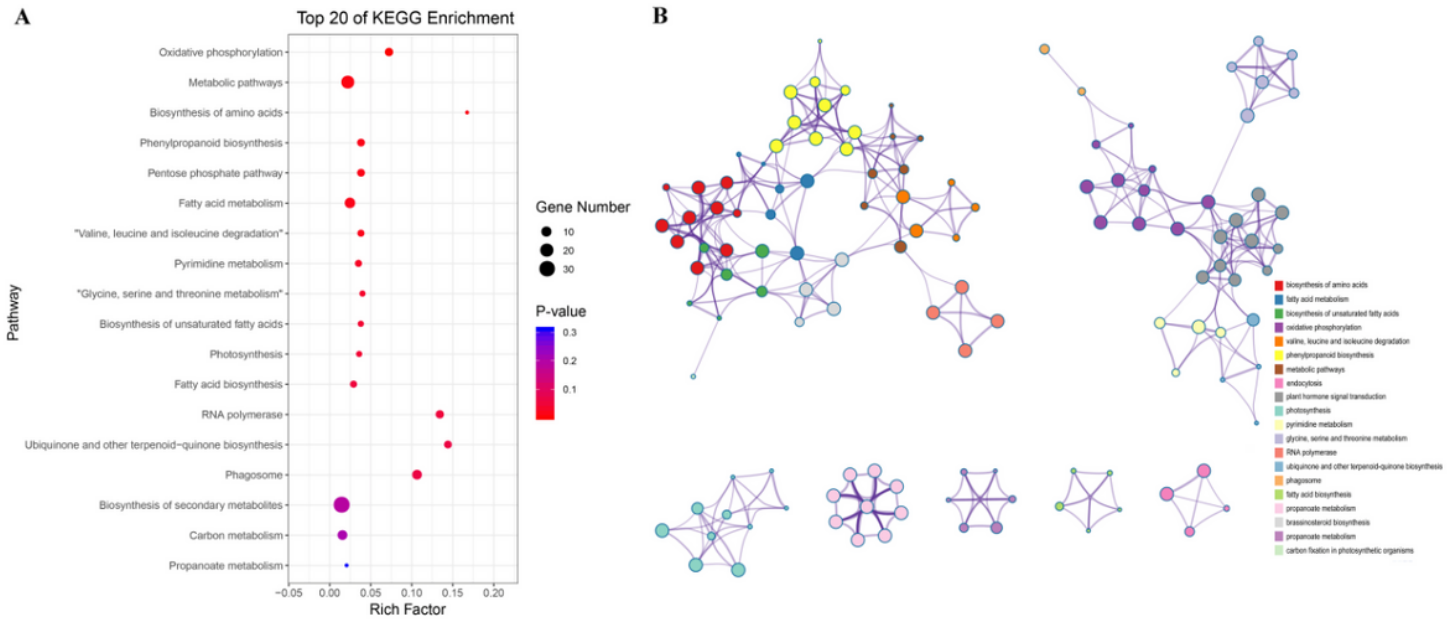

Figure 1

KEGG pathway enrichment analysis of 832 genes around the five stable QTLs. a top 20 enrichment pathways of all genes. Terms with a P-value < 0.05 were considered significant enrichment and selected for further analysis. b the relationships between the top 20 terms. A subset of enriched terms was selected and rendered as a network plot, where terms with a similarity > 0.3 are connected by edges, and where each node represents an enriched term.

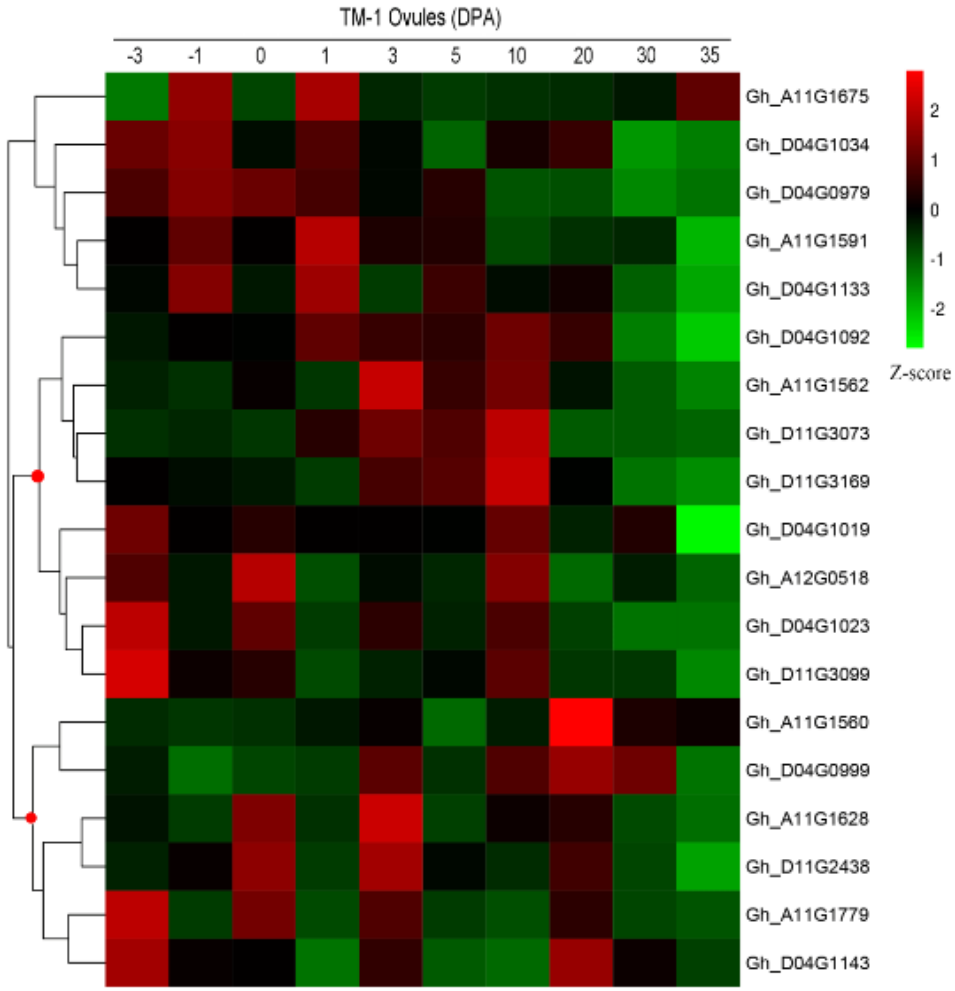

Figure 2

Temporal expression profile of putative genes that are involved in cottonseed oil accumulation in developing ovules by Z-score normalized FPKM.

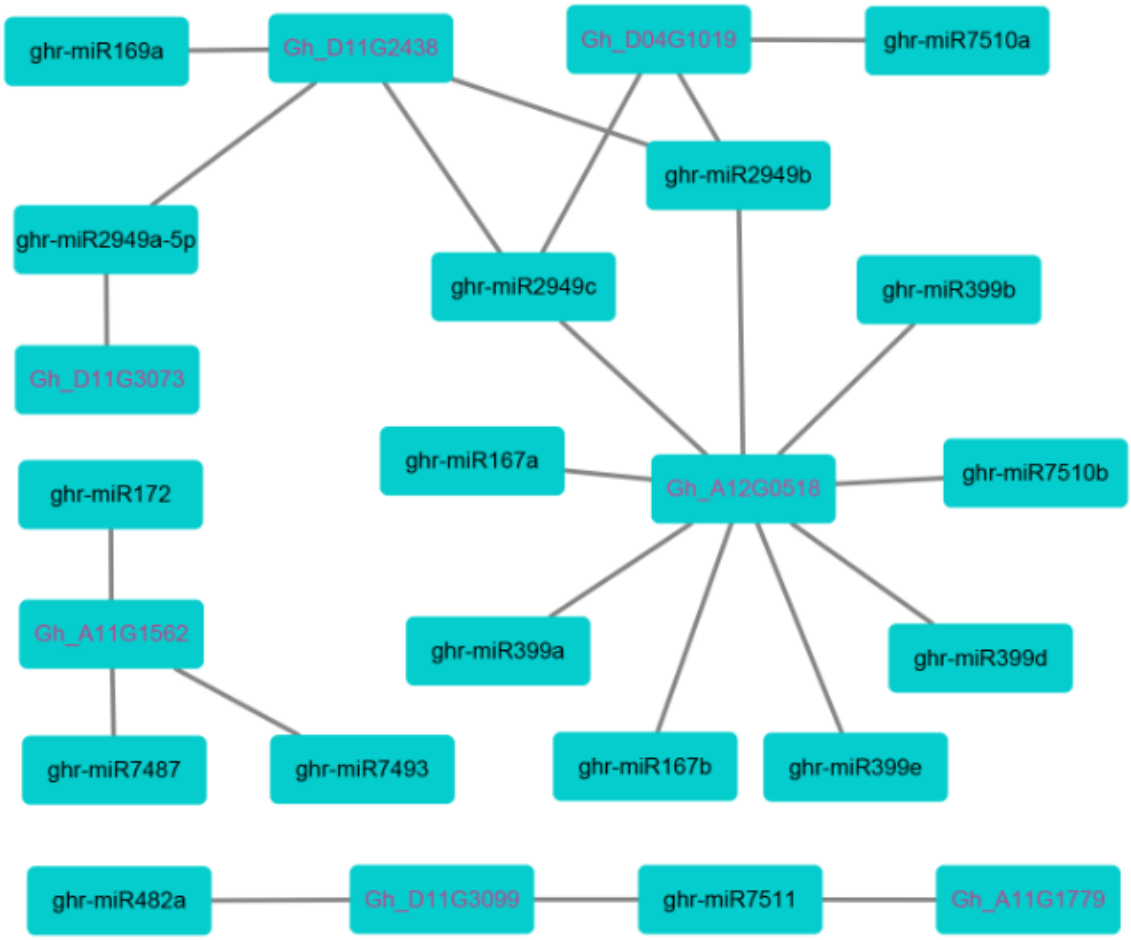

Figure 3

miRNAs-based regulatory network analysis of putative genes involved in cottonseed oil accumulation. This gene network included miRNA-target gene regulatory network and transcription factor information of G. hirsutum L.

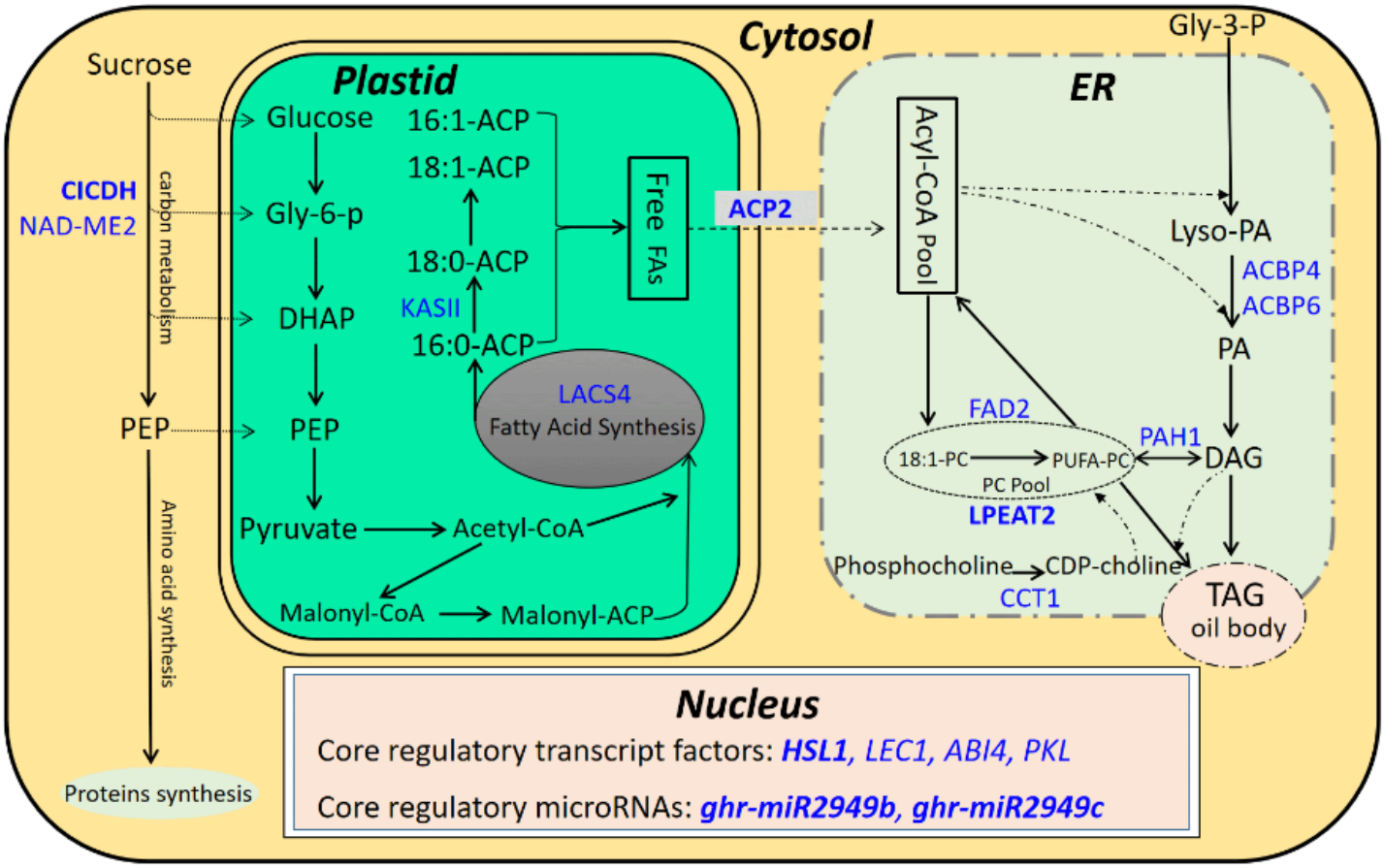

Figure 4

Characterization of the cottonseed oil accumulation model in developing cotton ovules for the regulation of enzymes, TFs, and miRNAs in oil biosynthesis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1ANOVAandbroadsenseheritabilityanalysis.xlsx
- TableS239QTLs.xlsx
- TableS3832genesaroundthefivestableQTLs.xlsx
- TableS4KEGG.xlsx
- TableS5Blastresults.xlsx
- TableS6FPKMvalueof24candidategenes.xlsx
- TableS7Fourtranscriptionfactors.xlsx
- TableS8miRNATargetnetwork.xlsx