



## Linkage and association analyses reveal that hub genes in energy-flow and lipid biosynthesis pathways form a cluster in upland cotton



Juwu Gong<sup>a,b,e,1</sup>, Yan Peng<sup>c,1</sup>, Jiwen Yu<sup>a,e,1</sup>, Wenfeng Pei<sup>a</sup>, Zhen Zhang<sup>a</sup>, Daoran Fan<sup>a</sup>, Linjie Liu<sup>a,b</sup>, Xianghui Xiao<sup>a,b</sup>, Ruixian Liu<sup>a,b</sup>, Quanwei Lu<sup>d</sup>, Pengtao Li<sup>d</sup>, Haihong Shang<sup>a,e</sup>, Yuzhen Shi<sup>a,e</sup>, Junwen Li<sup>a,e</sup>, Qun Ge<sup>a</sup>, Aiyong Liu<sup>a</sup>, Xiaoying Deng<sup>a</sup>, Senmiao Fan<sup>a</sup>, Jingtao Pan<sup>a</sup>, Quanjia Chen<sup>b,\*</sup>, Youlu Yuan<sup>a,b,e,\*</sup>, Wankui Gong<sup>a,\*</sup>

<sup>a</sup>State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China

<sup>b</sup>Engineering Research Centre of Cotton, Ministry of Education, College of Agriculture, Xinjiang Agricultural University, 311 Nongda East Road, Urumqi 830052, Xinjiang, China

<sup>c</sup>Third Division of the Xinjiang Production and Construction Corps Agricultural Research Institute, Tumushuke, Xinjiang 843900, China

<sup>d</sup>College of Biotechnology and Food Engineering, Anyang Institute of Technology, Anyang, China

<sup>e</sup>School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, Henan, China

### ARTICLE INFO

#### Article history:

Received 5 October 2021

Received in revised form 11 April 2022

Accepted 11 April 2022

Available online 15 April 2022

#### Keywords:

Hub gene

Gene cluster

Interaction network

Kernel oil content

Quantitative trait loci

Upland cotton

### ABSTRACT

Upland cotton is an important allotetraploid crop that provides both natural fiber for the textile industry and edible vegetable oil for the food or feed industry. To better understand the genetic mechanism that regulates the biosynthesis of storage oil in cottonseed, we identified the genes harbored in the major quantitative trait loci/nucleotides (QTLs/QTNs) of kernel oil content (KOC) in cottonseed via both multiple linkage analyses and genome-wide association studies (GWAS). In ‘CCRI70’ RILs, six stable QTLs were simultaneously identified by linkage analysis of CHIP and SLAF-seq strategies. In ‘0-153’ RILs, eight stable QTLs were detected by consensus linkage analysis integrating multiple strategies. In the natural panel, thirteen and eight loci were associated across multiple environments with two algorithms of GWAS. Within the confidence interval of a major common QTL on chromosome 3, six genes were identified as participating in the interaction network highly correlated with cottonseed KOC. Further observations of gene differential expression showed that four of the genes, *LtnD*, *PGK*, *LPLAT1*, and *PAH2*, formed hub genes and two of them, *FER* and *RAV1*, formed the key genes in the interaction network. Sequence variations in the coding regions of *LtnD*, *FER*, *PGK*, *LPLAT1*, and *PAH2* genes may support their regulatory effects on oil accumulation in mature cottonseed. Taken together, clustering of the hub genes in the lipid biosynthesis interaction network provides new insights to understanding the mechanism of fatty acid biosynthesis and TAG assembly and to further genetic improvement projects for the KOC in cottonseeds.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The growth of the world's population and the improvement in living standards have greatly increased the global demand for vegetable oil [1]. As an important economic crop, upland cotton

( $2n = 4x = 26$ , AADD) is not only the main natural fiber source for the textile industry [66, 79] but also the fifth largest vegetable oil source for the food and feed industry [11]. Cottonseed oil contains a high proportion of unsaturated fatty acids (FAs), including linoleic acid (18-carbon chain with two double bonds; C18:2) and oleic acid (C18:1) [46]. Due to the beneficial effects of unsaturated FAs on human health and the significant advantages of low cost and stable flavor compared with rapeseed oil or olive oil, cottonseed oil is becoming increasingly attractive in various food processing techniques [70]. The carbon chain length of 99% FAs of cottonseed oil is between C16 and C18, similar to that of fossil diesel, which is usually between C15 and C18. In addition, the biodiesel converted from cottonseed oil is oxygen-rich but sulfur-free, which makes the biodiesel burn more thoroughly and causes less

\* Corresponding authors.

E-mail addresses: [gongjuwu@caas.cn](mailto:gongjuwu@caas.cn) (J. Gong), [pengyanyunfei@163.com](mailto:pengyanyunfei@163.com) (Y. Peng), [yujuwen@caas.cn](mailto:yujuwen@caas.cn) (J. Yu), [peiwenfeng@caas.cn](mailto:peiwenfeng@caas.cn) (W. Pei), [zhangzhen@caas.cn](mailto:zhangzhen@caas.cn) (Z. Zhang), [82101192184@caas.cn](mailto:82101192184@caas.cn) (D. Fan), [liulinjie258@163.com](mailto:liulinjie258@163.com) (L. Liu), [xiaoxianghui4953@163.com](mailto:xiaoxianghui4953@163.com) (X. Xiao), [1120405872@qq.com](mailto:1120405872@qq.com) (R. Liu), [13707667581@163.com](mailto:13707667581@163.com) (Q. Lu), [lipengtao1056@126.com](mailto:lipengtao1056@126.com) (P. Li), [shanghaihong@caas.cn](mailto:shanghaihong@caas.cn) (H. Shang), [shiyuzhen@caas.cn](mailto:shiyuzhen@caas.cn) (Y. Shi), [lijunwen@caas.cn](mailto:lijunwen@caas.cn) (J. Li), [gequn@caas.cn](mailto:gequn@caas.cn) (Q. Ge), [liuayiyong@caas.cn](mailto:liuayiyong@caas.cn) (A. Liu), [dengxiaoying@caas.cn](mailto:dengxiaoying@caas.cn) (X. Deng), [fansenmiao@caas.cn](mailto:fansenmiao@caas.cn) (S. Fan), [panjingtao@caas.cn](mailto:panjingtao@caas.cn) (J. Pan), [chqjia@126.com](mailto:chqjia@126.com) (Q. Chen), [yuanyoulu@caas.cn](mailto:yuanyoulu@caas.cn) (Y. Yuan), [gongwankui@caas.cn](mailto:gongwankui@caas.cn) (W. Gong).

<sup>1</sup> These authors contributed equally to this work.

<https://doi.org/10.1016/j.csbj.2022.04.012>

2001-0370/© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

environmental pollution. These facts make cottonseed oil an ideal raw biofuel material, showing promising application prospects in the production of biodiesels [20]. Cottonseed oil can also be used to synthesize lubricants, inks, paints, soaps, and other chemical products. However, in cotton-related breeding and research projects, breeders put little effort into improving seed composition and quality, let alone cottonseed oil content and composition [80]. Therefore, on the premise of improving fiber yield and quality, it is necessary to improve the yield and quality of cottonseed to meet the high nutritional and industrial demand for cottonseed derivatives. Mapping the quantitative trait loci/nucleotides (QTLs/QTNs) of kernel oil content (KOC) in cottonseed and mining its candidate genes are effective ways to dissect the regulatory mechanism of cottonseed oil biosynthesis and to improve the KOC of cottonseed through corresponding breeding strategies. Cotton growers may also benefit from these joint improvement programs for both fiber and cottonseed qualities.

Vegetable oil exists in the form of triacylglycerol (TAG), and its biosynthesis is a complex process of co-expression and regulation of several enzymes and genes. It involves FA biosynthesis and TAG assembly [8], a continuous procedure of incorporating acyl CoA FAs from the cytosol into a glycerol-3-phosphate (G3P) backbone in the endoplasmic reticulum (ER) [82]. G3P is produced by the reduction of dihydroxyacetone phosphate (DHAP), an intermediate of glycolysis, by glycerophosphate dehydrogenase [105]. At present, a large number of genes have been cloned to dissect gene expression regulation and mechanisms in the process of oil biosynthesis [131,133]. WRKYs are a large family of transcription factors that are mainly involved in plant growth, development, and response to stress [10]. In *A. thaliana*, WRKY6 influences seed oil accumulation and FA composition in developing *Arabidopsis* seeds [90]. *GLUCOSE-6-PHOSPHATE/PHOSPHATE TRANSLOCATOR1 (GPT1)*, the key target gene of WRKY2/WRKY34, drives Glc6P from the cytosol into the plastid in the late stage of pollen development, which is an important step in lipid biogenesis in pollen [12]. In *Gossypium* species, WRKY12 has 198 potential target genes, and their differential expression may alter the FA biosynthetic pathway [39]. The serine/threonine/tyrosine kinase (STYK) gene phosphorylates oil body proteins, which in turn affect seed oil content. The seeds of the *styk* mutant of *Arachis hypogaea* accumulated 10–12% less TAG compared to the wild type, validating the role of STYK phosphorylation in lipid accumulation [78]. The peanut  $\beta$ -ketoacyl-CoA synthase (*KCS*) genes, *AhKCS1* and *AhKCS28*, are involved in the regulation of very long-chain FA (VLCFA) content in peanut seeds. Overexpression of *AhKCS1* or *AhKCS28* in *A. thaliana* can increase VLCFA content, especially the content of very long-chain saturated FAs (VLCSFAs) [38]. The hydroxysteroid dehydrogenase (*HSD1*) gene, which has two copies, *At5g50600* and *At5g50700*, in *Arabidopsis*, is specifically and highly induced in the oil-accumulating tissues of maturing seeds [7]. The Phospholipid:diacylglycerol acyltransferase (*PDAT*) gene increases seed oil content in rice [41], soybean [63], *A. thaliana* [68], and upland cotton [123]. In *Brassica napus*, the genes *POCKET* and *BnPMT6s* negatively regulate seed oil content [97].

Among the four domesticated cotton species, upland cotton accounts for 90–95% of the total cotton fiber production due to its wide adaptability and attractive yield [47,66]. The commercially planted upland cotton cultivars are the result of long-term selection and domestication [33], in which cottonseed oil content varies greatly from 12.60 to 39.33% of the total kernel weight [33,103,130]. To better understand the genetic mechanism of oil accumulation, some early studies have tentatively tackled QTL identification in cottonseed oil [85,120]. Many QTLs have thus been identified using intraspecific populations of *G. hirsutum* [112], interspecific populations between *G. hirsutum* and *G. barbadense* [120,134], and, more recently, natural accessions in

genome-wide association studies (GWAS) [22,130]. The rapid development of genome sequencing technologies has provided a new platform for QTL mapping based on linkage and association analyses of upland cotton and gene function studies [26,37,52,107,125]. The lysophosphatidic acid acyltransferase (*LPAAT*) gene has been reported to be involved in oil biosynthesis. Enhancing the expression of an *LPAAT* gene, *At-Gh13LPAAT5*, significantly increases the yield of total TAG and other FAs [108]. The FA desaturase (*FAD2*) gene encodes seed-specific desaturase, which catalyzes the conversion of oleic acid to linoleic acid in cotton seed and is the main factor determining the content of essential polyunsaturated FAs (with multiple unsaturated bonds) in seed oil [21]. There are at least two copies of this gene in tetraploid *Gossypium* species, and knockouts of *GhFAD2* genes with the CRISPR/Cas9 system have generated high-KOC materials [11]. The transcription factor WRINKLED 1 (*GhWRI1*) participates in the upregulation of FA biosynthesis during seed development. Overexpression of *GhWRI1* has been observed to increase KOC and seed weight in transgenic *A. thaliana* and upland cotton [131]. *GhCPS1*, *GhCPS2*, and *GhCPPS3* are cyclopropane synthase homologues in cotton, and the expression of *GhCPS1* and *GhCPS2* correlates with cyclic FA distribution [121]. *GbSWEET* and *GbACBP6* from *G. barbadense* were also demonstrated to have the potential for improving the KOC in cottonseed [133]. These studies are sporadic, aiming at isolated single genes; however, there are few reports on the regulatory network of oil synthesis in cottonseed, which is critical to dissect the genetic basis of biosynthesis of storage lipids. The current study combined multiple linkage and association strategies for major QTL identification and candidate gene verification of KOC in cottonseed across two recombinant inbred line (RIL) populations and a panel of 207 natural accessions to dissect the genetic regulatory mechanism of KOC accumulation in cottonseed. The hub genes were identified through expression network analysis of the candidate genes harbored in stable QTL intervals. Two hub genes, *LPLAT1* and *PAH2*, coding lysophospholipid acyltransferase [106] and  $Mg^{2+}$ -dependent phosphatidic acid phosphohydrolases [25], respectively, which coordinate the key steps of the Kennedy pathway and the Lands cycle for TAG assembly, were identified to colocalize in a cluster, determining the accumulation of oil in cells. These results provide a better understanding of the molecular regulatory mechanism in the metabolic pathways of lipid biosynthesis and oil accumulation during cottonseed development. This study also identified elite germplasm for future improvement of KOC in cottonseed in pragmatic breeding projects.

## 1. Materials and methods

### 1.1. Plant materials

The experimental materials included two RIL populations for linkage analyses and a panel of 219 natural accessions (natural panel) for association analysis.

One RIL population consisted of 250 individual lines derived from an elite hybrid cultivar 'CCRI70', a commercial hybrid approved by the State Crop Variety Approval Committee (Approval certificate 2008011; CCRI70 RIL) [122], and parental lines 'sGK156' (female) and '901-001' (male). In 2017, the  $F_{5:6}$  generation of the 250 CCRI70 lines was reached and thereafter regarded as RILs. The development of the population has been detailed in previous studies [19,32]. From 2015 to 2017, the phenotypic performance of the CCRI70 RIL population was evaluated at multiple locations. The experiments were performed using a randomized block design, and the experimental locations and layouts are detailed in Table S1.

The other RIL population consisted of 196 individual lines derived from a cross between parental lines 'sGK9708' and '0-153' (0–153 RIL). The construction procedure and characteristics of the parental lines have been described elsewhere [95,128]. The phenotypic performance of KOC was evaluated across eight environments from 2014 to 2015 in Anyang, Alaer, Kuerle, and Shihezi. Briefly, the field design and layout were conducted in a completely randomized block design with two replicates in each experimental location, which were detailed in a previous report [128].

The natural panel was collected from a pool of commercial planting cultivars and breeding nurseries, which represented the cotton production levels of the main cotton planting regions. The panel also contains a collection of advanced lines and germplasm in current breeding projects. The phenotypic performance of the natural panel was evaluated in four environments, namely Anyang in 2017 and Alaer, Anyang, and Wuhan in 2019. The experiments were performed in a randomized block design, which is detailed in Table S1. Field management for both RIL populations and the natural panel followed local practices in cotton production using 'CCRI60,' a commercial cultivar, as the control.

## 2. Trait phenotypic evaluation

In October, 30 normal cotton bolls in each line or accession were sampled. The cotton lint fiber and cottonseed were separated with a roller gin. The broken and immature seeds were discarded manually, and 50 healthy seeds were randomly sampled from each line or accession. The husks of the sampled seeds were removed manually. Then, the KOC in cottonseed was measured with three repeats using NMI20 Analyst (Shanghai, China).

## 3. Data statistical analyses

The best linear unbiased estimates (BLUE) of the KOC of the three populations and the KOC broad-sense heritability ( $H^2$ ) in different environments were estimated using QTL IciMapping 4.2 software [73]. Statistical analysis of the phenotypic performance and BLUE was performed with Minitab 17 statistical software. The Chart.Correlation function in the Performance Analytics Package of R was applied to create the correlation graphs and figures.

## 4. QTL identification

Two linkage maps were used to detect the KOC QTLs of the CCRI70 RILs. One was constructed with the Intl Cotton SNP Consortium\_80k CHIP containing 8292 SNP markers, with a total coverage of 4876.7 cM genetic distance and an average marker interval of 1.08 cM [136]. The other map was constructed with SLAF-seq technology, containing 24,425 SNPs with a total coverage of 4850.47 cM genetic distance and an average marker interval of 0.73 cM (Table S2) (Jiang et al., unpublished data). The linkage map of 0–153 RILs was a consensus map integrating multiple strategies, which has been detailed in a previous report (Table S2) [128]. The markers and genetic linkage map were re-analyzed using reference genome data of TM-1 (v2.1) [37].

QTLs were detected using the composite interval mapping (CIM) method of Windows QTL Cartographer v2.5 software [9,110]. In the parameter setting, a likelihood ratio (LR) value of 11.5, which is equivalent to an LOD value of 2.5, was taken as the threshold of QTL existence, that is, when the LOD of the marker interval was greater than or equal to 2.5, a QTL existed at the peak of the LOD value in that interval. The scanning of the whole genetic linkage map for QTL existence was performed with 1000 permutation tests at a step length of 0.5 cM with  $p < 0.05$ . The nomenclature for QTL followed Sun et al. [95], and QTLs that were

consistently identified in at least three environments were regarded as stable. Positive additive effect indicated that the favorable allele was from parent 901–001 (CCRI70 RIL) or 0–153 (0–153 RIL), and negative additive effect indicated that the favorable allele was from parent sGK156 (CCRI70 RIL) or sGK9708 (0–153 RIL).

Stable QTLs were compared with the Cotton QTLdb database (<https://www.cottonqtl.org>) or previous GWAS [23,69,120,134] to determine whether they were identified *de novo* or previously published. If the markers in the confidence interval of QTLs in the current study shared the same or overlapped physical location with those in QTLs in the Cotton QTLdb database or if the physical position of GWAS QTL markers in previous studies rested on confidence intervals of the current study, then QTLs in the current study were regarded as previously identified; otherwise, they were regarded as new QTLs.

## 5. GWAS of the natural panel

The natural panel was genotyped using the SLAF-seq technology described previously [128]. Briefly, 3,464,561 SNPs were obtained. The reads containing these SNPs were aligned to the reference genome [37] using BWA software [53]. SAMtools [54] and GATK [72] software were used to call SNPs and short indel variants. After filtering at the integrity level of  $\text{no} < 0.85$  and a minimum allele frequency of 0.05, a total of 32,994 SNP markers were obtained to genotype the natural panel (Table S2). The GLM and MLM models [127] of the genetic association and prediction integrated tool (GAPIT package) [59] in R software were used to calculate the significance of QTNs. The signal threshold of  $-\log_{10}(P) > 4$  was determined to declare the existence of a significant association site.

## 6. Acquisition of candidate genes in QTLs and QTNs

Cotton genome data and genome-wide functional annotation files were downloaded from the cotton genome database (CottonFGD, <https://cottonfgd.org/>) [135]. QTLs and QTNs of the KOC were aligned to the physical map. The candidate genes corresponding to the upland cotton genome (TM-1 v2.1) [37] were searched based on genome coordinates or genetic markers. Genes within the range of 100 kb flanking QTNs and within the confidence marker interval of QTLs were regarded as candidate genes of the KOC in cottonseed.

## 7. Prediction, annotation, and expression profiling of candidate genes

All candidate genes and their corresponding protein sequences were blasted with the *Arabidopsis* annotation database (<https://www.arabidopsis.org/>). Candidate genes in QTL regions were analyzed for Gene ontology annotation (GO) using agriGO 2.0 software [99]. Metabolic pathway enrichment analysis of the Kyoto Encyclopedia of Genes and Genomes (KEGG) was performed using KOBAS 3.0 [117]. Transcription factors (TFs) were obtained through annotation analysis from the PlantTFDB v5.0 database [104].

Cotton expression profile data were downloaded from the database (<https://cotton.zju.edu.cn/>). The database contains FPKM values of developing ovules at 0, 1, 3, 5, 10, and 20 days post anthesis (dpa) of 'TM-1,' a standard upland cotton cultivar with a low KOC of 30.82%, and 'Hai7124,' a cultivar with a high KOC of 35.31%. The TM-1 database was used to analyze the expression profiles of candidate genes in KOC QTLs, and the Hai7124 database was used to compare the differential expression of these genes during cotton ovule development between high and low KOC materials.

### 8. Construction of a regulatory network of lipid synthesis-related genes

The interaction between proteins was explored in the Metascape database [132] and rendered as a network plot, where terms with a similarity > 0.3 were connected by edges. The terms with the best *p*-value were selected from each cluster based on the constraint that the terms selected did not exceed 15 per cluster and that the total number of terms did not exceed 250. Using all interactions (details) in STRING 10.0 (<https://string-db.org/cgi/input.pl>), the synthesized network exclusively kept the subset of proteins that physically interacted with at least one other member on the list. If the network contained 3–500 proteins, then the molecular complexity detection (MCODE) algorithm [40,55] was used to identify the network components with dense node–node connections. For each MCODE component, pathway and process enrichment analyses were performed independently. The three best scoring items of the *p*-value were reserved to describe the function of the corresponding component. Finally, a regulatory network was constructed.

Cytoscape 3.9.0 [45] was used to visualize the regulatory network, where each node represented an enriched term, and then, the network was pigmented by its cluster ID and then by its *p*-value.

### 9. Results and analysis

#### 9.1. Statistical analysis of KOC phenotypic data in CCRI70 RIL and natural populations

In the 14 environment phenotypic evaluations of CCRI70 RILs and their parental lines, sGK156 and 901–001, the KOC of 901–001 was significantly higher than that of sGK156 (Table 1 and Fig. 1A). The phenotypic distributions of KOC of the CCRI70 RILs in the Cotton Regions of the Yellow River Basin (namely 14AY, 15AY, 16AY, 17AY, 15LQ, and 16LQ), of Northwest Inland (namely 16ALE, 16KEL, 17KEL, 16SHZ, and 17SHZ), and of the Yangtze River

Basin (namely 16CD, 17CD, and 17DT) were 19.45–27.24%, 23.48–41.17%, and 21.47–40.24%, respectively, which exceeded the values of their two parents, indicating that both parents of CCRI70 RILs had alleles that favored KOC accumulation in cottonseed. The results also showed that the average KOC of CCRI70 RILs was between those of the two parents and that they varied significantly across different environments.

The KOC of the accessions of the natural panel across four environments was distributed within the range of 24.27–38.79%, with an average of 30.16% (Table 1). Skewness and kurtosis showed that the phenotypes of both RILs and the natural panel followed a normal distribution and the typical characteristics of quantitative traits. Significant differences in KOC across different environments were also observed in both populations. The continuous variations in KOC in both the RILs and the natural panel indicated that it might be controlled by multiple genes.

### 10. Broad-sense heritability and correlations of KOC across different environments

The effect of genotype (G), environment (E), and genotype by environment (G × E) of KOC reached a significant level (*p* < 0.01) (Table S3). The average broad-sense heritability of the KOC of CCRI70 RILs and the natural panel was 90.51% and 92.30%, respectively, based on the above variance estimations. The KOC of upland cotton seed had strong heredity, showing little response to environmental variations. The stability and consistency of KOC in different environments provided a good basis for further genetic analysis.

Correlation analysis showed a highly significant positive correlation across different environments, with correlation coefficients of 0.82–0.95 and 0.82–0.97 in CCRI70 RILs and the natural panel, respectively. Correlation analysis between 14 environments and BLUE values (Fig. 1B, 1C) showed a highly significant positive correlation, indicating that the environmental data were consistent and accurate.

**Table 1**  
Phenotype analysis of CCRI70 RILs and natural panel across multiple environments.

Population	Env <sup>a</sup>	Parents <sup>b</sup>	Population	N	Mean ± SD <sup>c</sup>	Range	Skewness	Kurtosis	CV <sup>d</sup> (%)
CCRI70 RILs	15AY	21.73	P2	250	26.21 ± 3.87	19.45–33.31	−0.05	−1.54	14.78
	16AY	25.04	P1	250	27.86 ± 3.15	21.78–34.23	−0.15	−1.34	11.31
	17AY	27.45	P2	250	31.74 ± 3.09	24.29–37.24	−0.23	−1.35	9.74
	15LQ	26.56	P1	250	29.48 ± 3.21	23.38–34.40	−0.15	−1.49	10.89
	16LQ	26.45	P2	250	29.68 ± 2.98	23.04–35.50	−0.21	−1.23	10.03
	15ALE	25.76	P1	250	29.05 ± 2.95	23.48–34.78	−0.08	−1.23	10.16
	16ALE	26.29	P2	250	28.62 ± 2.35	23.75–32.98	−0.02	−1.22	8.19
	16KEL	26.23	P1	250	29.33 ± 2.71	23.97–34.62	−0.28	−1.37	9.24
	17KEL	29.16	P2	250	34.74 ± 2.61	29.23–41.17	−0.19	−0.99	7.51
	16SHZ	29.19	P1	250	30.71 ± 2.39	25.60–35.41	−0.13	−1.04	7.76
	17SHZ	30.52	P2	250	34.39 ± 2.90	28.23–40.69	−0.08	−1.08	8.42
	16CD	25.75	P1	250	28.31 ± 2.95	21.47–33.49	−0.06	−1.29	10.43
	17CD	26.66	P2	250	31.98 ± 2.98	25.55–37.26	−0.22	−1.27	9.32
	17DT	31.75	P1	250	35.02 ± 2.63	28.38–40.24	−0.09	−1.08	7.50
	BLUE <sup>CCRI70</sup>			250	30.37 ± 2.69	24.90–34.86	−0.18	−1.50	8.86
	GWAS	17AY		193	31.60 ± 2.41	26.27–38.72	1.06	1.14	7.61
		19AY		219	29.35 ± 3.03	24.27–37.32	1.01	0.10	10.33
19ALE			219	30.03 ± 2.83	24.69–38.79	1.24	1.16	9.42	
19WH			219	29.67 ± 2.63	24.50–36.76	0.81	−0.03	8.85	
BLUE <sup>GWAS</sup>			219	29.69 ± 2.70	24.86–37.06	1.11	0.39	9.08	

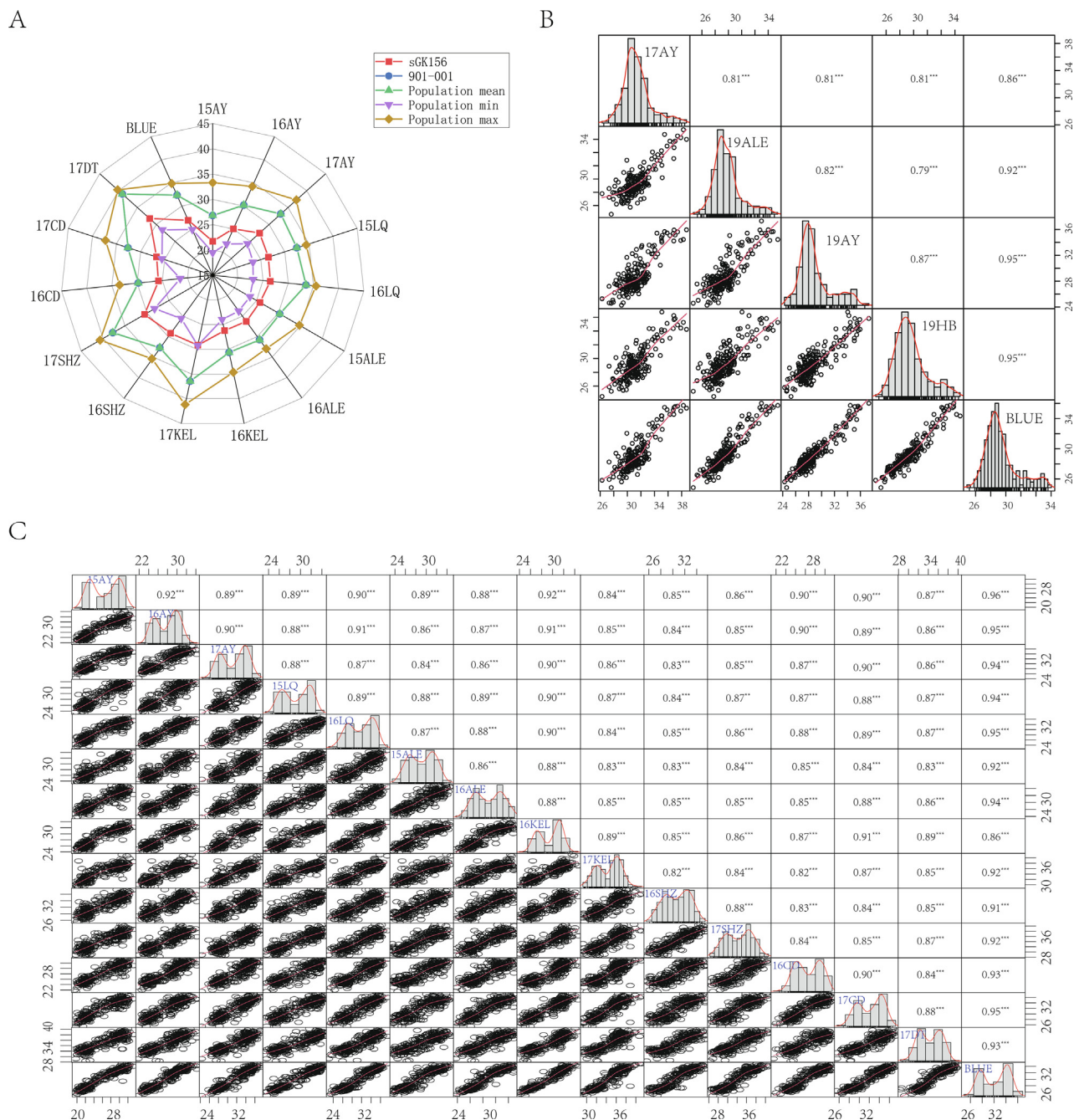
<sup>a</sup> Env., 15AY: 2015 Anyang, 16AY: 2016 Anyang, 17AY: 2017 Anyang, 15LQ: 2015 Linqing, 16LQ: 2016 Linqing, 15ALE: 2015 Alaer, 16ALE: 2016 Alaer, 16KEL: 2016 Kuerle, 17KEL: 2017 Kuerle, 16SHZ: 2016 Shihezi, 17SHZ: 2017 Shihezi, 16CD: 2016 Changde, 17CD: 2017 Changde, 17DT: 2017 Dangtu, 19ALE: 2019 Alaer, 19AY: 2019 Anyang, 19WH: 2019 Wuhan, BLUE<sup>CCRI70</sup> and BLUE<sup>GWAS</sup>: best linear unbiased estimates of CCRI70 RILs and GWAS, respectively.

<sup>b</sup> Parents, P1: sGK156, P2: 901–001.

<sup>c</sup> SD, standard deviation.

<sup>d</sup> CV, coefficient of variation.

\*\* indicates that the differences between sGK156 (P1) and 901–001 (P2) reached a 0.01 significance level.



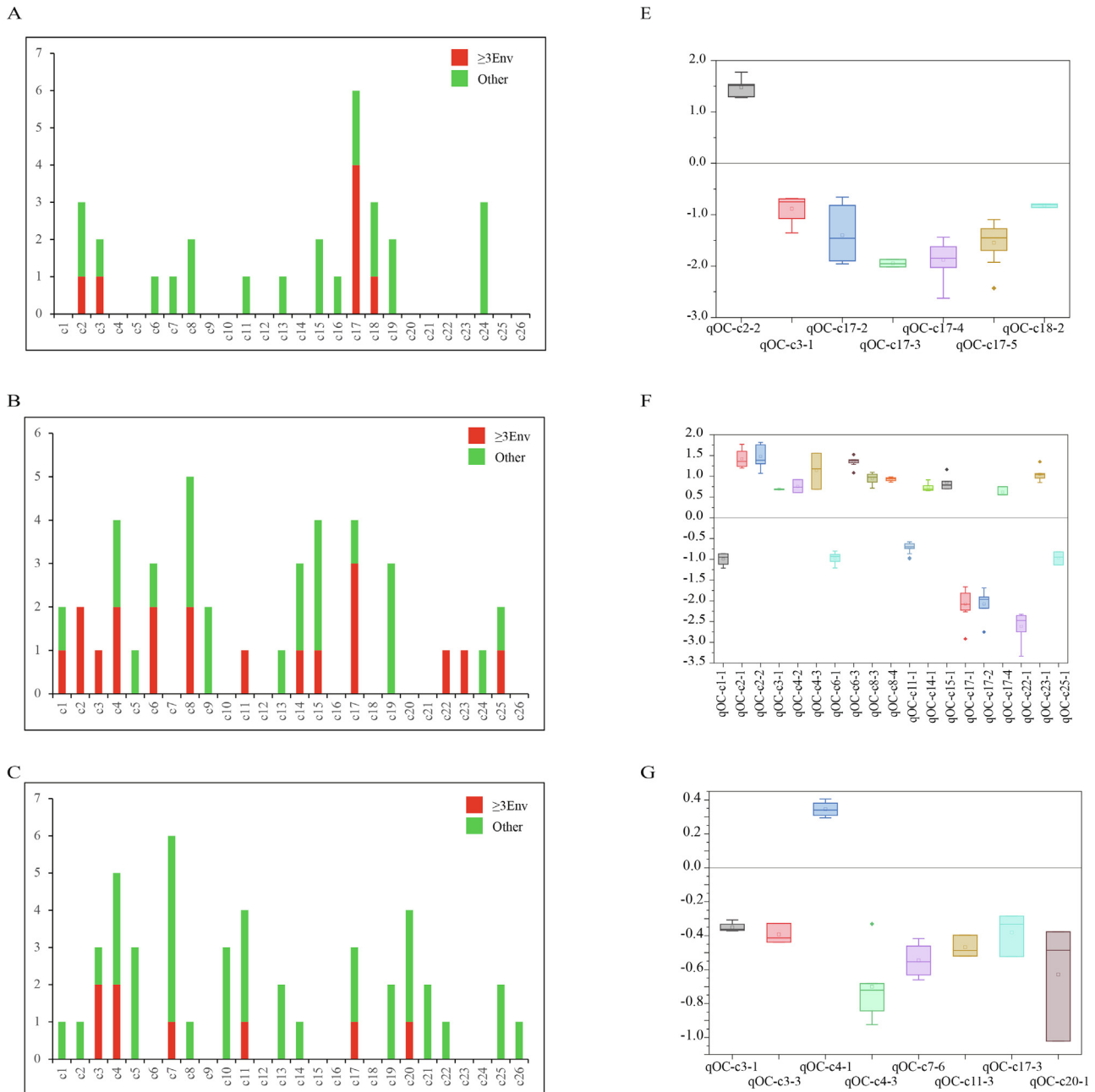
**Fig. 1.** Phenotypic statistics of kernel oil content (KOC) of CCRI70 RILs and their parental lines, and correlations of the natural panel and CCRI70 RILs across different environments. A. Phenotypic performance of KOC of CCRI70 RILs (including highest, lowest, and average values) and their parental lines. B. Correlation analysis of the natural panel across five environments. C. Correlation analysis of CCRI70 RILs across 14 environments and their BLUE values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 11. QTL analysis of KOC in multiple RIL populations and GWAS of the natural panel

**CCRI70 RILs** Based on the genetic linkage map constructed by the Intl Cotton SNP Consortium\_80k CHIP [136], 28 QTLs for KOC in cottonseed were detected by the CIM method using Windows QTL Cartographer v2.5 software (Table S4, Figure S1A). The number and distribution of QTLs throughout the genome are detailed in Fig. 2A. The LOD values of these QTLs ranged from 2.51 to 29.13, and their contribution rate to phenotypic values ranged from 2.58% to 63.07%. Seven of these QTLs were detected in at least three environments (Table 2), mainly distributed on chromosome 2 (c2),

c3, c17, and c18, with additive effects from – 2.63 to 1.78 (Fig. 2D). Chromosome c17 (D03) harbored the most stable KOC QTLs (4), one of which, qOC<sup>chip</sup>-c17-5, was detected simultaneously in 14 environments. Its contribution rate to the phenotypic value ranged from 7.43% to 36.92%.

Based on the genetic linkage map constructed by SLAF-seq technology, 41 QTLs for the KOC in cottonseed were identified by the CIM method (Table S5, Figure S1B). The number and distribution of QTLs throughout the genome are detailed in Fig. 2B. The LOD value of these QTLs ranged from 2.50 to 43.61, and the contribution rate ranged from 2.01 to 72.80%. Nineteen of these QTLs were detected in at least three environments, mainly distributed on c1,



**Fig. 2.** Distribution of quantitative trait loci (QTLs) across the genome of cotton and the additive effects of stable QTLs. A. Distribution of QTLs of CCRI70 RILs identified by the CHIP strategy. B. Distribution of QTLs of CCRI70 RILs identified by the SLAF-seq strategy. C. Distribution of QTLs of 0–153 RILs. D. Additive effects of stable QTLs of CCRI70 RILs identified by the CHIP strategy. E. Additive effects of stable QTLs of CCRI70 RILs identified by the SLAF-seq strategy. F. Additive effects of stable QTLs of 0–153 RILs. The red bars indicate the number of QTLs that are stably expressed in at least three environments, and the green bars indicate the number of QTLs that are only expressed in one or two environments. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

c2, c3, c4, c6, c8, c11, c14, c15, c17, c22, c23, and c25, with additive effects from  $-3.33$  to  $1.82$  (Fig. 2E). A major QTL,  $qOC^{slaf}\text{-c17-1}$ , on c17 (D03) was detected simultaneously in all 14 environments, with a LOD value of  $25.4\text{--}43.61$  and a contribution rate to the phenotype of  $30.64\text{--}54.03\%$ . Its negative additive effect, which ranged from  $-2.92$  to  $-1.66$ , indicated that its additive allele, which favors phenotypic formation of KOC, was obtained from parent 901–001. In summary, the above two linkage maps simultaneously identified two DNA fragments on c2 (A02) and c17 (D03). In the c2 fragment, the Intl Cotton SNP Consortium\_80k CHIP strategy identified locus  $qOC^{chip}\text{-c2-2}$ , and the SLAF-seq strategy identified two loci,  $qOC^{slaf}\text{-c2-1}$  and  $qOC^{slaf}\text{-c2-2}$ . In the c17 fragment, the Intl

Cotton SNP Consortium\_80k CHIP strategy identified four loci,  $qOC^{chip}\text{-c17-2}$ ,  $qOC^{chip}\text{-c17-3}$ ,  $qOC^{chip}\text{-c17-4}$ , and  $qOC^{chip}\text{-c17-5}$ , and the SLAF-seq strategy identified one locus,  $qOC^{slaf}\text{-c17-2}$ . Previous studies have isolated genes controlling the KOC in cottonseed on these chromosome loci [130,133].

**0–153 RILs** Based on the genetic linkage map [128], which was constructed by the integration of SSRs, SLAF-SNPs, and chip-SNPs using TM-1 v2.1 as a reference genome [37], 44 QTLs for KOC in cottonseed were obtained by the CIM method (Table S6, Figure S1C). The distribution of these QTLs across the genome is detailed in Fig. 2C. These QTLs have LOD values of  $2.1\text{--}5.9$ , contributing  $4.87\text{--}13.02\%$  of KOC phenotypic variance. Eight of these

**Table 2**  
Stable QTLs detected in CCRI70 and 0–153 RIL populations.

QTL name <sup>a</sup>	Chr	Env <sup>b</sup>	Position(cM)	LOD	Additive	R <sup>2</sup>	Physical start	Physical end	References
qOC <sup>slaf</sup> -c1-1	A01	15LQ,17SHZ,17DT,BLUE <sup>CCRI70</sup>	86.81–86.81	2.59–3.38	–1.21–0.86	4.95–6.81	12,460,819	13,709,725	[134]
qOC <sup>chip</sup> -c2-2	A02	17AY,16LQ,16KEL,17DT,16AY	147.51–151.61	3.02–4.30	1.28–1.78	5.33–8.16	1,049,090	1,483,801	
qOC <sup>slaf</sup> -c2-1	A02	15LQ,17KEL,17SHZ,16CD	105.21–112.51	2.70–3.11	1.20–1.77	4.88–6.05	1,190,248	1,269,625	
qOC <sup>slaf</sup> -c2-2	A02	16ALE,16SHZ,17SHZ,16CD,17DT,16LQ,BLUE <sup>CCRI70</sup>	121.71–124.81	3.01–3.53	1.07–1.82	5.11–6.25	1,069,052	1,238,794	
qOC <sup>slaf</sup> -c3-1	A03	16CD,17CD,16KEL	6.01–7.01	2.67–2.77	0.68–0.70	4.52–5.49	1,547,094	1,800,523	
qOC <sup>0-153</sup> -c3-1	A03	14AY,15SHZ,14ALE,BLUE <sup>0-153</sup>	3.81–7.81	2.50–3.40	–0.37–0.31	6.66–7.77	835,242	2,119,737	
qOC <sup>chip</sup> -c3-1	A03	15AY,16ALE,17KEL,16SHZ	74.01–76.11	2.57–3.29	–1.35–0.68	4.45–5.80	22,751,743	57,120,804	[134][61]
qOC <sup>0-153</sup> -c3-2	A03	14ALE, BLUE <sup>0-153</sup>	85.21–86.81	2.50–3.20	0.32–0.42	5.32–6.44	16,858,604	92,580,366	[134]
qOC <sup>0-153</sup> -c3-3	A03	14ALE,15ALE,BLUE <sup>0-153</sup>	115.11–115.11	2.70–3.30	–0.44–0.33	5.72–7.01	102,477,680	104,626,789	
qOC <sup>slaf</sup> -c4-2	A04	17AY,15LQ,17KEL	90.91–90.91	2.53–3.93	0.61–0.92	4.12–6.22	70,463,958	79,831,756	
qOC <sup>slaf</sup> -c4-3	A04	16ALE,16KEL,17CD	204.81–204.81	2.89–4.21	0.69–1.56	4.82–6.94	1,216,715	2,315,633	
qOC <sup>0-153</sup> -c4-1	A04	15KEL,14SHZ,15SHZ,BLUE <sup>0-153</sup>	0.01–0.01	2.50–3.70	0.29–0.41	5.36–7.30	84,518,722	86,498,365	
qOC <sup>0-153</sup> -c4-2	A04	14ALE,14SHZ	31.81–32.61	3.20–3.40	–0.90–0.69	6.52–7.54	80,133,670	80,594,623	
qOC <sup>0-153</sup> -c4-3	A04	14AY,15AY,15ALE,14SHZ,BLUE <sup>0-153</sup>	35.61–40.61	2.90–5.90	–0.92–0.33	5.99–11.84	79,085,783	80,126,560	
qOC <sup>0-153</sup> -c5-1	A05	15KEL,BLUE <sup>0-153</sup>	14.31–14.31	2.50–2.60	0.26–0.46	5.14–5.51	107,249,639	109,379,072	
qOC <sup>slaf</sup> -c6-1	A06	16ALE,16SHZ,17SHZ,17CD,BLUE <sup>CCRI70</sup>	85.11–85.41	2.69–3.26	–1.21–0.80	4.50–5.73	120,173,225	121,010,649	
qOC <sup>slaf</sup> -c6-3	A06	16AY,17AY,15LQ,15ALE,16ALE,16SHZ,17SHZ,BLUE <sup>CCRI70</sup>	107.71–107.71	2.88–5.4	1.08–1.52	4.99–9.04	119,831,704	120,002,055	
qOC <sup>0-153</sup> -c7-1	A07	14AY,14SHZ	0.01–0.01	3.80–3.90	–0.47–0.46	7.78–7.95	93,488,530	94,679,900	
qOC <sup>0-153</sup> -c7-2	A07	14ALE,14SHZ	16.11–17.11	4.00–5.70	0.56–0.61	8.11–13.02	90,291,121	93,073,159	[69]
qOC <sup>0-153</sup> -c7-5	A07	BLUE,15KEL	67.71–69.41	2.80–3.50	0.54–0.83	5.87–7.57	25,450,547	26,472,025	
qOC <sup>0-153</sup> -c7-6	A07	15KEL,15SHZ,BLUE,14SHZ	80.41–81.71	2.8–3.4	–0.66–0.42	5.76–6.83	21,941,174	25,186,324	
qOC <sup>slaf</sup> -c8-3	A08	16AY,17AY,16ALE,17KEL,16KEL,16SHZ,17SHZ,16CD,17CD,17DT,BLUE <sup>CCRI70</sup>	177.41–184.31	2.96–4.75	0.71–1.10	4.88–7.98	1,777,312	2,135,693	
qOC <sup>slaf</sup> -c8-4	A08	16AY,17AY,16LQ,BLUE <sup>CCRI70</sup>	190.71–190.71	2.60–3.39	0.86–0.98	4.33–5.69	1,572,717	1,768,288	
qOC <sup>0-153</sup> -c10-2	A10	14ALE,BLUE <sup>0-153</sup>	109.21–111.71	2.90–3.30	–0.43–0.32	6.09–8.26	73,178,541	94,556,940	
qOC <sup>slaf</sup> -c11-1	A11	15AY,16AY,15LQ,16LQ,15ALE,16ALE,17KEL,17KEL,16SHZ,17SHZ,16CD,17DT,BLUE <sup>CCRI70</sup>	122.11–128.51	2.75–4.44	–0.99–0.58	4.49–7.41	23,689,874	58,859,230	
qOC <sup>0-153</sup> -c11-2	A11	14KEL,15AY	18.11–19.61	2.70–2.70	0.57–0.87	5.62–5.83	110,990,183	113,750,731	
qOC <sup>0-153</sup> -c11-3	A11	15KEL,15ALE,BLUE <sup>0-153</sup>	58.21–60.21	2.30–2.90	–0.52–0.40	5.25–6.01	59,121,526	59,583,114	
qOC <sup>slaf</sup> -c14-1	D02	15LQ,16CD,BLUE,16ALE,17SHZ	137.31–143.51	2.69–3.78	0.65–0.91	4.33–6.23	60,566,070	61,911,566	
qOC <sup>slaf</sup> -c15-1	D01	15LQ,15ALE,17KEL,16SHZ,17DT	94.51–95.11	2.84–4.77	0.70–1.16	4.75–7.90	10,767,132	11,338,952	
qOC <sup>slaf</sup> -c17-1	D03	15AY,16AY,17AY,15LQ,16LQ,15ALE,16ALE,17KEL,16KEL,16SHZ,17SHZ,16CD,17CD,17DT,BLUE <sup>CCRI70</sup>	43.81–45.61	25.4–43.61	–2.92–1.66	30.64–54.03	42,061,076	43,311,612	[130][23]
qOC <sup>slaf</sup> -c17-2	D03	15AY,17AY,15LQ,15ALE,16CD,17DT,BLUE <sup>CCRI70</sup>	50.51–51.51	27.00–37.80	–2.75–1.69	35.95–49.09	31,778,268	42,054,654	[23] [130,134]
qOC <sup>slaf</sup> -c17-4	D03	17AY,15ALE,16CD	140.91–144.11	2.55–3.54	0.55–0.75	2.01–3.25	2,476,696	2,492,953	
qOC <sup>chip</sup> -c17-2	D03	15AY,16AY,15LQ,16LQ,15ALE,17KEL,6KEL,17SHZ,16CD,17CD,17DT	40.31–42.91	2.52–15.49	–1.96–0.66	2.86–24.18	39,490,039	39,800,209	[130,134]
qOC <sup>chip</sup> -c17-3	D03	16LQ,15ALE,BLUE <sup>CCRI70</sup>	48.51–48.51	21.80–25.28	–2.01–1.87	33.36–37.28	35,884,335	39,490,039	[130,134]
qOC <sup>chip</sup> -c17-4	D03	15AY,16AY,17AY,15LQ,16ALE,17KEL,16KEL,16SHZ,17SHZ,16CD,17CD,17DT	51.11–54.71	18.29–29.13	–2.63–1.44	24.56–41.05	34,643,835	35,888,354	[130]
qOC <sup>chip</sup> -c17-5	D03	15AY,16AY,17AY,15LQ,16LQ,15ALE,16ALE,16KEL,16SHZ,17SHZ,16CD,17CD,17DT,BLUE <sup>CCRI70</sup>	58.21–60.61	5.97–25.28	–2.43–1.09	7.43–36.92	32,858,408	34,643,835	[130]
qOC <sup>0-153</sup> -c17-3	D03	15ALE,14SHZ,BLUE <sup>0-153</sup>	116.41–117.21	2.10–4.10	–0.52–0.28	4.92–9.54	608,489	1,048,342	
qOC <sup>slaf</sup> -c22-1	D04	15AY,16AY,15LQ,15ALE,16KEL,17CD,BLUE <sup>CCRI70</sup>	144.91–148.91	4.50–11.58	–3.33–2.32	60.70–72.8	313,345	54,237,568	
qOC <sup>0-153</sup> -c19-1	D05	15SHZ,14AY	40.11–41.91	2.40–2.60	0.38–0.40	5.16–5.67	24,566,964	28,945,620	[69]
qOC <sup>0-153</sup> -c19-2	D05	15KEL,15SHZ	47.91–47.91	2.50–2.60	0.37–0.42	5.48–5.74	16,717,496	23,760,060	[60,84]
qOC <sup>slaf</sup> -c25-1	D06	15AY,17SHZ,17CD	33.11–33.11	2.50–2.92	–1.14–0.82	4.55–5.57	6,710,763	6,853,575	
qOC <sup>0-153</sup> -c25-1	D06	14SHZ,15KEL	37.31–39.71	2.50–2.50	0.51–0.52	5.38–5.45	27,591,300	43,038,644	
qOC <sup>0-153</sup> -c25-2	D06	15ALE,15KEL	58.91–58.91	3.10–3.20	–0.57–0.53	7.42–7.56	6,133,046	8,278,619	

(continued on next page)

Table 2 (continued)

QTL name <sup>a</sup>	Chr	Env <sup>b</sup>	Position(cM)	LOD	Additive	R <sup>2</sup>	Physical start	Physical end	References
qOC <sup>slaf</sup> -c23-1	D09	16AY,17AY,17SHZ,17CD,BLUE <sup>CCRI70</sup>	26.81–26.81	2.59–5.17	0.85–1.35	4.69–8.86	48,167,983	48,221,325	
qOC <sup>0-153</sup> -c20-1	D10	14KEL,BLUE,15ALE	10.11–10.61	2.80–4.80	-1.02–0.38	6.05–8.02	61,328,789	62,929,532	[69,130]
qOC <sup>0-153</sup> -c20-3	D10	15KEL,15SHZ	75.21–77.31	2.80–2.90	0.55–0.57	6.06–6.38	6,479,656	8,103,827	
qOC <sup>chip</sup> -c18-2	D13	17KEL,17CD,17DT	74.71–75.31	2.72–2.90	-0.86–0.79	4.77–5.21	31,737,192	45,707,863	[61]

<sup>a</sup> In the name of a QTL, <sup>chip</sup> indicates that the QTL is identified on the genetic linkage map of CCRI70 by CHIP strategy, <sup>slaf</sup> indicates that the QTL is identified on the genetic linkage map of CCRI70 by SLAF-seq strategy, and <sup>0-153</sup> indicates that the QTL is identified on the genetic linkage map of 0–153.

<sup>b</sup> Env, 14ALE: 2014 Alaer, 14AY: 2014 Anyang, 14KEL: 2014 Kuerle, 14SHZ: 2014 Shihezi, 15ALE: 2015 Anyang, 15KEL: 2015 Kuerle, 15LQ: 2015 Linqing, 15SHZ: 2015 Shihezi, 16ALE: 2016 Alaer, 16AY: 2016 Anyang, 16CD: 2016 Changde, 16KEL: 2016 Kuerle, 16LQ: 2016 Linqing, 16SHZ: 2016 Shihezi, 17AY: 2017 Anyang, 17CD: 2017 Dangtu, 17KEL: 2017 Kuerle, 17SHZ: 2017 Shihezi, BLUE<sup>CCRI70</sup> and BLUE<sup>0-153</sup>: best linear unbiased estimates of CCRI70 RILs and 0–153 RILs, respectively.

QTLs were detected in at least three environments (Table 2), mainly distributed on chromosomes c3, c4, c7, c11, c17, and c20 (Figure S1C). The distribution of additive effects is shown in Fig. 2F.

**GWAS of the natural panel** GWAS of the natural panel was performed with both GLM and MLM algorithms (Figures 3 and S2). The evaluation of phenotypic data and SNP filtering exhibited their reliabilities (Figure S2A, S2B, S2C, S2D, S2E). GWAS was carried out under the condition of K = 4, under which the population was in the most appropriate structure (Figure S2F). The quantile–quantile plot (QQ plot) showed that there was a significant correlation between the KOC phenotype and genotypes of the natural panel (Fig. 3B and 3D). The results finally associated 34 and 22 loci with the KOC in cottonseed, among which 13 and 8 loci, respectively, were detected in at least two environments (Fig. 3A and 3C, Table S7), distributed mainly on chromosomes A1, A3, A5, A6, A7, A10, A11, D5, D11, and D13 (Table 3). Among these loci, seven were simultaneously associated with KOC in both models, and all of them were on c18 (D13). These significant SNP markers explained 9.38–18.33% of the phenotypic variance, with genetic effects of - 7.33 to 6.9. These results suggest that the chromosome regions where the significant SNPs were located may harbor candidate genes regulating KOC accumulation in cottonseed.

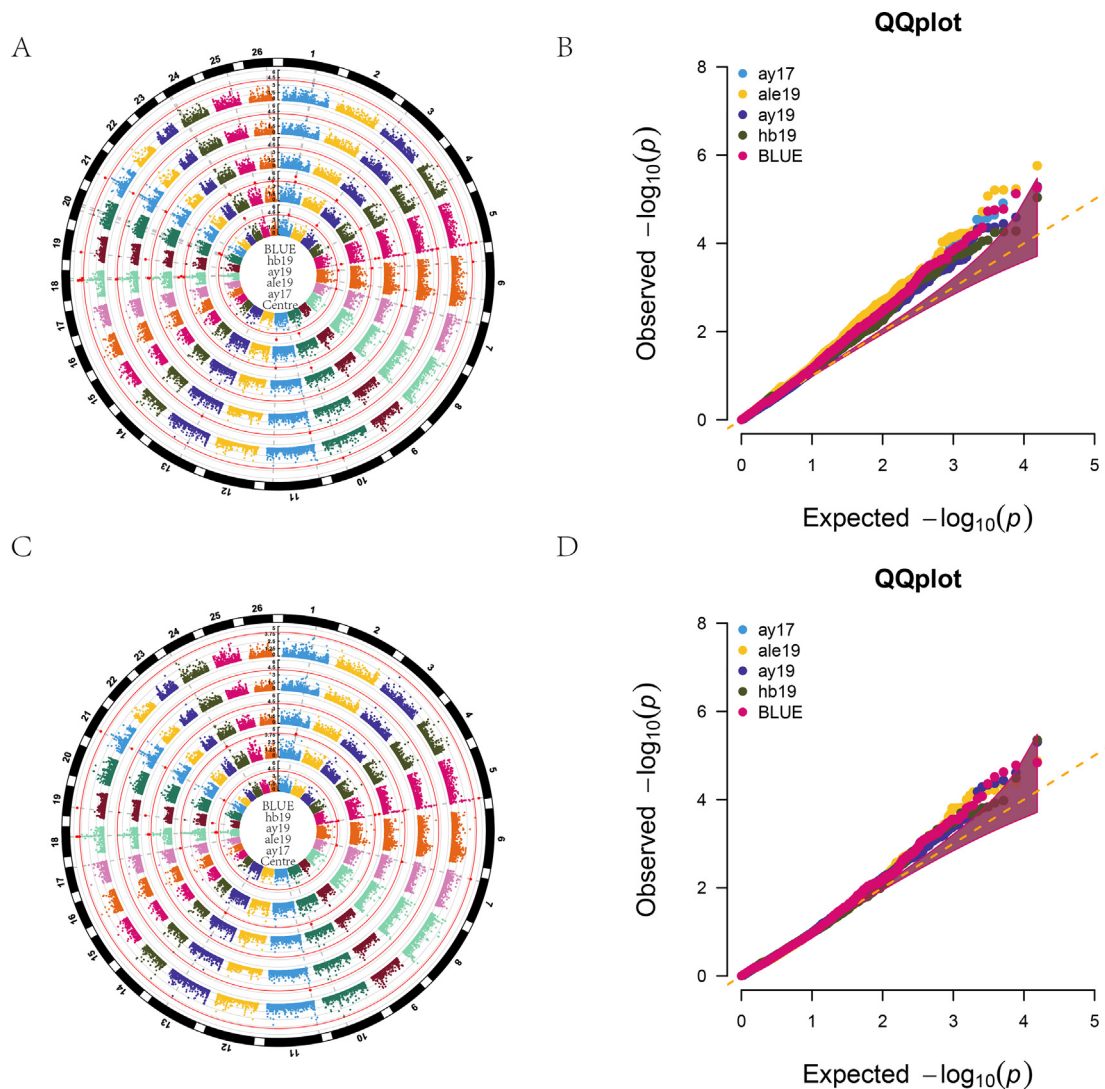
Combining the results of linkage analyses of CCRI70 and 0–153 RILs and association study of the natural panel, 11 QTLs mapped by SLAF-seq and 10 QTLs mapped by chip technology in CCRI70 RILs, 10 QTLs mapped in 0–153 RILs, and 8 QTNs mapped by GWAS could be mapped simultaneously on at least two linkage maps (CCRI70 RIL) or at least in two populations (CCRI70 and 0–153 RILs and the natural panel) (Table S8). The chromosome regions of these QTLs or QTNs are important research objects for candidate gene mapping and metabolic pathway analysis in this study.

## 12. Candidate genes in stable QTLs and their GO enrichment and KEGG analysis

The SNPs at both ends of the QTL confidence interval were mapped back to the upland cotton genome (TM-1\_V2.1) [37] based on their sequence information. Based on the gene annotation information of the physical interval in the cotton genome, a total of 832 and 3177 genes were obtained in the confidence interval stable QTLs identified on the two genetic linkage maps of the Intl Cotton SNP Consortium\_80k chip and SLAF-seq technique, respectively. Using the identification results of the two linkage maps in CCRI70 RILs, 3763 genes were identified, and 246 genes were identified simultaneously on both maps (Table 4, Fig. 4A). In 0–153 RILs, 3736 genes were identified in the confidence interval of the stable QTLs based on linkage [128] (Table 4). In the natural panel, 205 genes were associated with KOC (Table 4). Based on the results of linkage analyses, 7126 genes were identified in the two linkage populations (CCRI70 and 0–153 RILs), among which 373 genes were identified simultaneously in both populations (Table 4, Fig. 4B). Finally, 7196 candidate genes were detected through both linkage and association analyses (Table S9, Fig. 4B) on 20 chromosomes of the *G. hirsutum* genome (Table 4). Among the genes identified in the three populations, 754 candidate genes were detected simultaneously on the two genetic linkage maps of CCRI70 RILs or in each pair of the populations (Table 4, Fig. 4A and 4B).

GO analysis of the 754 candidate genes showed that these candidate genes were mainly enriched in the categories of biological process, cellular component, and molecular function (Table S10, Fig. 4C). Biological process mainly included cellular process, metabolic process, and single-organism process. Cellular components mainly included the cell, cell part, and organelle, and molecular functions mainly included binding and catalytic activity. Thirteen genes, *GH\_A03G0921*, *GH\_A03G0964*, *GH\_A03G0975*, *GH\_A03G0976*,





**Fig. 3.** Manhattan map and the quantile–quantile (QQ) chart in the genome-wide association studies (GWAS) of the natural panel. A. Manhattan map of GLM model. B. QQ chart of GLM model. C. Manhattan map of MLM model. D. QQ chart of MLM model.

*GH\_A03G0979*, *GH\_A03G1006*, *GH\_A03G1056*, *GH\_A03G1070*, *GH\_A03G1100*, *GH\_A10G1508*, *GH\_A11G2175*, *GH\_A11G2207*, and *GH\_D03G0938*, were identified to participate in the FA metabolic process, VLCFA metabolic process, and FA elongase complex (Table S11, Fig. 4D), which were regarded to be involved in important pathways of lipid biosynthesis [25,42,56].

KEGG analysis of these candidate genes showed that they mainly play important roles in the pathways of metabolism, genetic information processing, and cellular processes (Table S12). A total of 32 genes were involved in the metabolic processes of pathways related to lipid biosynthesis (Fig. 4E, Tables S12 and S13). Among them, *GH\_A04G1201* was identified as being involved in the metabolic process of FA degradation (ko00071), biosynthesis of secondary metabolites (ko01110), FA metabolism (ko01212), metabolic pathways (ko01100), and alpha-linolenic acid metabolism (ko00592). *GH\_A03G0956* was involved in the biosynthesis of secondary metabolites (ko01110), metabolic pathways (ko01100), alpha-linolenic acid metabolism (ko00592), and glycerophospholipid metabolism (ko00564). *GH\_A11G2246* was involved in the biosynthesis of secondary metabolites (ko01110), metabolic pathways (ko01100), starch and sucrose metabolism (ko00500), and cyanoo-amino acid metabolism (ko00460). *GH\_A03G0949* is involved in the

biosynthesis of secondary metabolites (ko01110), metabolic pathways (ko01100), and glycerophospholipid metabolism (ko00564). *GH\_D13G1396* was involved in the biosynthesis of secondary metabolites (ko01110), metabolic pathways (ko01100), and lipoic acid metabolism (ko00785). Three genes, *GH\_D03G0980*, *GH\_A03G1138*, and *GH\_D13G1354*, were involved in the biosynthesis of secondary metabolites (ko01110), metabolic pathways (ko01100), and starch and sucrose metabolism (ko00500). *GH\_D13G1401* was involved in metabolic pathways (ko01100) and lipoic acid metabolism (ko00785). Both *GH\_D03G0934* and *GH\_D03G0968* were involved in metabolic pathways (ko01100) and starch and sucrose metabolism (ko00500) (Fig. 4F).

At present, a large number of achievements have been made in research relating to *Arabidopsis* lipid metabolic pathways [8,42,44,48,106]. A database (<https://aralip.plantbiology.msu.edu/pathways/pathways>) [48] has collected comparatively complete gene pools relating to lipid metabolic pathways in *Arabidopsis*. Comparing the results of GO term enrichment and KEGG pathway annotation revealed that they shared 338 genes. Referring these common genes to the gene pool relating to oil metabolism in *Arabidopsis*, we found that 28 genes had ascertained functions in synthetic oil metabolism (Fig. 4G, Table S14).

**Table 3**

The significant loci that are associated with KOC in at least two environments using models GLM and MLM in GWAS.

QTN	Chr	Position	Allele	Env*	GLM				MLM				Reference	
					p-value	Maf	R <sup>2</sup>	effect	p-value	Maf	R <sup>2</sup>	effect		
c01_69712855	A01	69,712,855	G/A	19ALE	6.21E-06	0.0097	13.94	-7.33	/	/	/	/	[134]	
					17AY	6.15E-05	0.0108	13.81	-6.12	/	/	/		/
c03_18925549	A03	18,925,549	A/G	19ALE	5.13E-05	0.0097	11.85	4.73	/	/	/	/		
					17AY	2.63E-05	0.0108	14.74	4.58	/	/	/		/
c05_107481128	A05	107,481,128	A/G	19ALE	8.42E-06	0.0242	13.64	3.05	/	/	/	/		
					BLUE <sup>GWAS</sup>	6.84E-05	0.0242	11.50	2.73	/	/	/		/
c05_107481161	A05	107,481,161	A/G	19ALE	1.73E-06	0.0217	15.24	3.43	1.32E-05	0.0217	18.33	3.42		
					17AY	1.23E-05	0.0162	15.58	3.38	1.16E-05	0.0162	16.25		3.69
					19AY	4.29E-05	0.0217	13.10	3.36	5.18E-05	0.0217	12.42		3.49
					BLUE <sup>GWAS</sup>	5.45E-06	0.0217	14.01	3.26	1.45E-05	0.0217	14.35		3.35
c06_17240705	A06	17,240,705	C/G	19WH	5.64E-05	0.0217	9.84	3.03	/	/	/	/		
					19ALE	6.29E-05	0.0097	11.65	4.67	/	/	/		/
					17AY	6.83E-05	0.0108	13.69	4.33	/	/	/		/
					BLUE <sup>GWAS</sup>	6.49E-05	0.0097	11.55	4.68	/	/	/		/
c07_8706328	A07	8,706,328	T/C	19AY	/	/	/	/	6.61E-05	0.0242	12.18	-3.17		
					BLUE <sup>GWAS</sup>	/	/	/	/	8.41E-05	0.0242	12.65	-2.72	
c10_77548836	A10	77,548,836	T/A	19ALE	6.14E-06	0.0072	13.96	6.45	4.10E-05	0.0072	17.28	6.36		
					17AY	2.26E-05	0.0081	14.90	5.70	1.58E-05	0.0081	15.91	6.27	
					BLUE <sup>GWAS</sup>	7.94E-05	0.0072	11.35	5.63	/	/	/	/	
c11_104291533	A11	104,291,533	G/A	19ALE	6.73E-05	0.0048	11.58	6.54	/	/	/	/		
					17AY	6.03E-06	0.0054	16.38	6.96	/	/	/	/	
c18_42111539	D13	42,111,539	A/G	19ALE	5.85E-06	0.0193	14.00	3.70	4.63E-05	0.0193	12.52	4.18		
					19AY	3.60E-05	0.0193	13.27	3.87	/	/	/	/	
					BLUE <sup>GWAS</sup>	7.47E-06	0.0193	13.69	3.67	3.05E-05	0.0193	13.62	3.91	
					19WH	5.40E-05	0.0193	9.89	3.47	/	/	/	/	
c18_42113870	D13	42,113,870	G/A	19AY	4.55E-05	0.0169	13.04	-3.80	4.26E-05	0.0169	12.60	-4.09		
					BLUE <sup>GWAS</sup>	4.29E-05	0.0169	11.95	-3.32	9.63E-05	0.0169	12.52	-3.51	
c18_42507564	D13	42,507,564	A/G	19ALE	5.89E-05	0.0193	11.71	2.80	/	/	/	/		
					19AY	2.56E-05	0.0193	13.60	3.39	2.47E-05	0.0193	13.14	3.69	
					BLUE <sup>GWAS</sup>	1.75E-05	0.0193	12.84	3.02	4.40E-05	0.0193	13.27	3.24	
					19WH	8.98E-05	0.0193	9.38	2.88	/	/	/	/	
c18_46810731	D13	46,810,731	G/A	19ALE	1.90E-05	0.0242	12.82	-2.54	/	/	/	/		
					BLUE <sup>GWAS</sup>	5.51E-05	0.0242	11.71	-2.40	/	/	/	/	
c19_19777635	D5	19,777,635	G/A	19AY	5.13E-06	0.0145	15.19	-6.44	4.95E-06	0.0145	14.73	-6.86		
					BLUE <sup>GWAS</sup>	1.67E-05	0.0145	12.89	-5.28	2.38E-05	0.0145	13.87	-5.69	
c21_22911784	D11	22,911,784	C/A	19AY	4.75E-05	0.0145	13.00	-3.94	3.67E-05	0.0145	12.75	-4.10		
					BLUE <sup>GWAS</sup>	1.87E-05	0.0145	12.77	-3.62	1.68E-05	0.0145	14.21	-3.72	
					19WH	9.11E-06	0.0145	11.69	-3.96	4.47E-06	0.0145	13.07	-4.19	

\* Env., 17AY: 2017 Anyang, 19AY: 2019 Anyang, 19ALE: 2019 Alaer, 19WH: 2019 Wuhan, and BLUE<sup>GWAS</sup>: best linear unbiased estimates of GWAS, respectively.

### 13. Dynamic expression patterns of candidate genes relating to lipid metabolism in developing ovules

Combining analyses of GO, KEGG, and homologues in pathways related to lipid metabolism in *Arabidopsis*, 64 candidate genes, 57 coding genes, and 7 TFs were considered to be related to lipid metabolism in cottonseed. Using the FPKM values of gene expression levels of transcriptome data of TM-1 at different developmental stages of 0, 1, 3, 5, 10, and 20 dpa ovules to validate the dynamic expression patterns of these candidate genes, almost all of them had some specific expression at these stages, except five, which did not have any expression (Table S15).

Clustering analyses of the gene expression pattern, the expression curves across multiple stages and expression tendency sorted these genes into seven expression modules (Figure S4). The lowest gene number in a module had five genes, and the highest had 12 genes. The accumulation pattern of lipid and FA components during the development of cottonseed showed an upward trend from 2 dpa to maturity, and the oil content reached its peak in mature seeds [36,76]. Notably, Cluster 5 included six members: *GH\_A11G2214*, *GH\_D03G0988*, *GH\_D03G0922*, *GH\_A11G2180*, *GH\_D13G1385*, and *GH\_A04G1183*. The genes had a steady expression before 10 dpa, and then, their expression significantly increased at 20 dpa. Gene annotation revealed that *GH\_A11G2214*, which encodes Probable beta-D-xylosidase 6 BXL6, *GH\_D03G0988*,

which encodes 2-keto-3-deoxy-L-rhamnonate aldolase rhmA, and *GH\_D03G0922*, which encodes arabinogalactan peptide 20 (AGP20), were involved in glycolysis. The genes in Clusters 1 and 2 had high expression at 10 dpa, while the genes in Clusters 3 and 4 had high expression at 5 dpa.

### 14. Hub genes in the interaction network of differentially expressed candidate genes

To further capture the interaction between these differentially expressed genes, the interaction network among proteins of 57 coding genes and 7 TFs was explored using the Metascape database [132]. The network construction revealed that 24 hub genes formed interaction nodes that had at least three directions of gene interaction (Fig. 5). The network recognized three core MCODE components, which contained 13 hub or key genes and 5 TFs, including RAV1 (*GH\_A03G1130*), ERF1-3 (*GH\_D03G1065*), FERONIA (FER, a receptor-like kinase) (*GH\_A03G1011*), BHLH93 (*GH\_A11G2218*), and MYB28 (*GH\_D03G1077*), which were involved in the interaction of genes in the core MCODE components (Fig. 5 and Table S16). The genes in MCODE 1, which mainly participated in the carbon metabolic process, included *atpB* (*GH\_A03G1009*), *MENB* (*GH\_D03G1063*), *PGK* (coding phosphoglycerate kinase) (*GH\_A03G0979*), *SHM1* (*GH\_A11G2207*), and *LtnD* (*GH\_A03G1070*). The genes in MCODE 2, which participated in the coenzyme meta-

**Table 4**

Number of candidate genes identified from KOC QTLs that are identified simultaneously by at least two strategies or two populations.

Chr	CCRI70 RILs		0–153 RILs	Natural panel	Total	
	SLAF	CHIP			Common	Total
A01	46	/	/	2	/	48
A02	21**	33**	/	/	21 <sup>a</sup>	33
A03	17**	240**	892**	7**	17 <sup>b</sup> + 240 <sup>c</sup> + 7 <sup>e</sup>	892
A04	277**	/	253**	/	26 <sup>b</sup>	504
A05	/	/	119**	13**	13 <sup>e</sup>	119
A06	41	/	/	3	/	44
A07	/	/	351	12	/	363
A08	43	/	/	/	/	43
A10	/	/	374**	9**	9 <sup>e</sup>	374
A11	415**	/	265**	9	81 <sup>b</sup>	608
D01	41	/	/	/	/	41
D02	47	/	/	/	/	47
D03	396**	225**	56	/	225 <sup>a</sup>	452
D04	1818	/	/	/	/	1818
D05	/	/	887**	21**	21 <sup>e</sup>	887
D06	9**	/	344**	/	9 <sup>b</sup>	344
D09	6	/	/	/	/	6
D10	/	/	195	/	/	195
D11	/	/	/	3	/	3
D13	/	334**	/	126**	85 <sup>d</sup>	375
Total genes	3177	832	3736	205	754	7196

/ indicates that no common QTLs were identified.

\*\* indicates that the genes are identified in the marker intervals of common QTLs.

<sup>a</sup> indicates that the genes are identified in the common QTLs of CCRI70\_Chip and CCRI70\_SLAF.<sup>b</sup> indicates that the genes are identified in the common QTLs of CCRI70\_SLAF and 0–153 RILs.<sup>c</sup> indicates that the genes are identified in the common QTLs of CCRI70\_Chip and 0–153 RILs.<sup>d</sup> indicates that the genes are identified in the common QTLs of CCRI70\_Chip and the natural panel.<sup>e</sup> indicates that the genes are identified in the common QTLs of 0–153 RILs and the natural panel.

bolic process, cofactor metabolic process, and small molecule biosynthetic process, included *EMB8* (*GH\_D03G0974*), *GDSL-like Lipase* (*GH\_D13G1338*), *MJE7.13* (*GH\_A10G1509*), *SPCC663.09c* (*GH\_A03G0993*), and *D27* (*GH\_D03G0969*). The genes in MCODE 3, which participated in the lipid biosynthetic process and cellular lipid metabolic process, included *LACS4* (*GH\_D03G0938*), *PAH2* (*GH\_A03G0964*), and *LPLAT1* (*GH\_A03G1006*) (Fig. 5, Table S16).

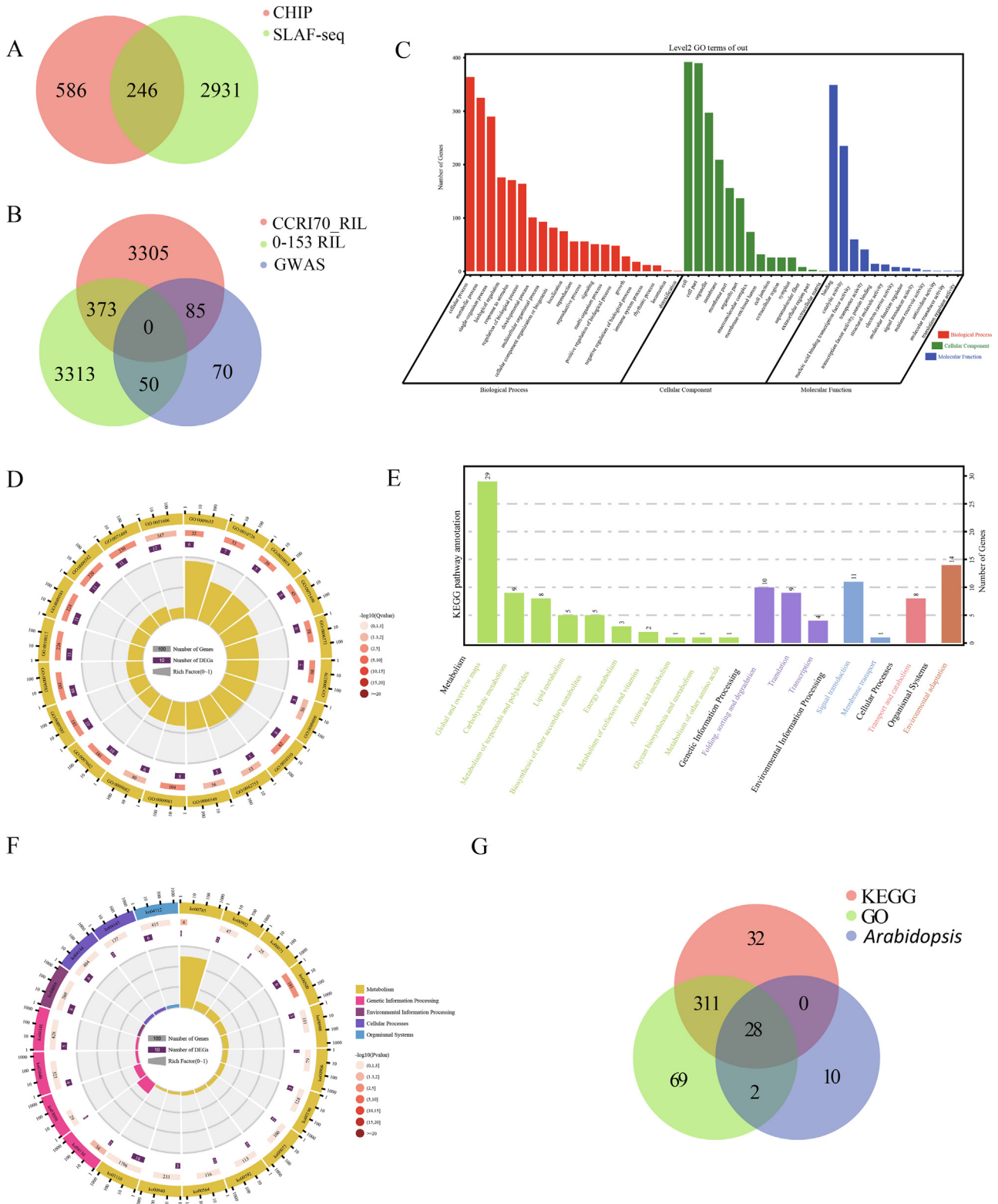
### 15. Expression verification of hub genes in different KOC germplasm

To further clarify the expression differences of the genes in relation to lipid biosynthesis between materials of high and low KOC, 17 genes recognized in the network were further analyzed for their expression differences during ovule development between Hai7124 (high KOC of 35.31%) and TM-1 (low KOC of 30.82%) (Figure S5A). The genes included 13 hub or key genes of core MCODE components (Figures S5B, S5C, S5D, S5E, S5F, S5H, S5I, S5J, S5L, S5M, S5N, S5O, S5Q) and 4 TF genes (Figures S5G, S5K, S5P, S5R). According to the FPKM values of the transcriptome data of 0, 1, 3, 5, 10, and 20 dpa ovules of Hai7124 and TM-1, *GH\_A02G0124* (*FPA*), *GH\_A03G0949* (*Atlg22950*), *GH\_D03G0961* (*ARF9*), and *GH\_D03G1065* (*ERF1-3*) (Figure S5Q) had a higher FPKM expression in Hai7124 than in TM-1 almost across all six ovule development stages (Figures S5B, S5C, S5P, and S5R), while *GH\_A03G1070* (*LtnD*) and *GH\_A10G1508* (*DIR1*) had a higher expression in TM-1 than in Hai7124 (Figures S5I, S5L, Table S17). *GH\_A03G0964* (*PAH2*), *GH\_A03G1075* (*NMT1*), *GH\_D03G0956* (*P14KG2*), and *GH\_D03G0974* (*EMB8*) had a higher expression in 5 and 10 dpa ovules in TM-1 than in Hai7124 (Figures S5D, S5J, S5O, and S5Q). *GH\_A03G0991* (*NTF4*), *GH\_A10G1509* (*MJE7.13*), and *GH\_D03G0938* (*LACS4*) had a higher expression in 10 and 20 dpa ovules in Hai7124 than in TM-1, while *GH\_A03G1006* (*LPLAT1*) had a higher expression in TM-1 than in Hai7124 in these stages (Figures S5E, S5M, S5N,

and S5F). *GH\_A03G1056* (*ABCG11*) had a high expression in 3 dpa ovules in TM-1, and *GH\_A03G1130* (*RAV1*) had a high expression exclusively before 3 dpa in Hai7124 (Figures S5H and S5K).

### 16. Identification of functional gene clusters in carbon metabolism and the Kennedy pathway

Based on the gene expression profiles and differences in developing ovules and the interaction network of the differentially expressed candidate genes, to identify causal genes within the QTL intervals, we examined the genetic variations of the core region of these hub genes. First, we mapped the hub genes back to the QTLs and aligned them to the physical location of the reference genome of *G. hirsutum* [37]. Five of the hub genes, *GH\_A03G0964* (*PAH2*), *GH\_A03G0979* (*PGK*), *GH\_A03G1006* (*LPLAT1*), *GH\_A03G1011* (*FER*), and *GH\_A03G1070* (*LtnD*), were identified in the common region of a QTL identified in both CCRI70 and 0–153 RILs ( $qOC^{chip-c3-1}$  and  $qOC^{0-153-c3-2}$ ), in which the association analysis of the natural panel identified 20 haplotypes (Fig. 6A, 6B, and 6C). Analysis of these genes in the currently available *Gossypium* genome assembly revealed that the clustering of these genes varied slightly [37,107] (Fig. 6D). Genetic variation analysis of these genes showed that all five genes had SNP differences between sGK156 and 901–001, the two parental lines of CCRI70 RILs (Fig. 6E). Three of these can also be verified in the natural panel (Fig. 6F). These genes have a dynamic expression profiling across different organs and tissues, especially the developmental stages of the ovule and fiber (Fig. 6G). Functional analysis indicated that *GH\_A03G0964* (*PAH2*) and *GH\_A03G1006* (*LPLAT1*) coordinated the key steps of the Kennedy pathway and Lands cycle, respectively, and *GH\_A03G0979* (*PGK*), *GH\_A03G1011* (*FER*), and *GH\_A03G1070* (*LtnD*) were responsible for catalyzing the steps in carbon metabolism that provide carbon sources and energy-flow for FA biosynthesis and TAG assembly (Fig. 7).



**Fig. 4.** Bioinformatics studies of the candidate genes harbored in stable QTLs. A. The gene numbers harbored in the stable QTLs in the CHIP and SLAF-seq strategies of CCRI70 RILs. B. Comparison of the genes in the stable QTLs of CCRI70 RILs, 0–153 RILs, and the natural panel. C. gene ontology annotation (GO) term enrichment analysis of the candidate genes. D. The first 20 significantly enriched GO terms. E. The primary and secondary level annotations of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of the candidate genes. F. The first 20 significantly enriched KEGG pathways. G. The number of genes related to oil metabolism in GO term enrichment and KEGG pathway annotation after referring to the genes related to oil metabolism in *Arabidopsis*.

## 17. Discussion

### 17.1. Influencing factors of the KOC in cottonseed

Oil is an important storage substance in plant seeds, and its accumulation is closely correlated with seed development [36,133]. Previously, research on KOC emphasized the effect of the environment on oil accumulation in mature cottonseed [77]. An early study found that the KOC in cottonseed varied according to the locations where the seed samples were collected [102]. These results suggest that environmental impact accounts for a considerable proportion of the total variations in KOC. More recently, multiple environmental data of upland cotton cultivars have shown that the genotype, environment, and their interaction impose significant impacts on the KOC in cottonseed [31]. All genetic effects, including additive, dominance, dominance  $\times$  environment, and additive–dominance epistasis  $\times$  environment effects, were involved in the genetic architecture of the KOC in cottonseed. However, it has been demonstrated that the dominance effect contributes the most to the total KOC heritability [23]. Maternal or cytoplasmic effects may also contribute to KOC variations to some extent [62].

In the current study, phenotypic evaluations were performed in a multi-year multi-location style (Table 1). The phenotypic variations of KOC ranged from 19.45% to 41.17% in CCRI70 RILs and from 24.27% to 38.72% in the natural panel accessions. Both showed a high broad-sense heritability of 86.23–98.81% and 90.13–96.12%, respectively, which was consistent with the results of previous research [70,76]. The high heritability of oil content ensures the effect of the genotype of each line (RIL) or accession (natural panel) on phenotypic performance and the reliability of QTLs mapped on this basis. Significant differences between years or between locations indicate that the environmental effect and the interaction effect between environment and genotype also contribute greatly to phenotypic performance [31]. The variation between different years was greater than that between different locations. It can be reckoned that the optimum collection of parental germplasm, targeting selection, and multi-environment/location evaluations are the key elements to guide breeders to obtain high oil content progeny materials.

### 18. Linkage and association analysis of the KOC loci of cottonseed

Previous studies have also tackled the KOC in cottonseed, mainly using genetic linkage analysis [61,112] or association studies of natural populations [22,69]. GWAS mapping utilizes abundant genetic variations in natural populations, which bestows its advantages of high mapping accuracy. However, linkage disequilibrium makes inefficient mapping unable to detect loci of low abundance, yielding false-positive association results [4]. Genetic linkage mapping can identify low-abundance alleles with low marker coverage in the genome. However, the low mapping resolution renders the mapped fragments large and imprecise, which affects the research expectations. The genetic segregation of the target traits involves only differences between parents [67,114,128]. At present, strategies combining linkage and association analyses have been successfully applied in wheat [64,88], maize [65,67], rice [43], and oil rape (*Brassica napus* L) [100]. In cotton, based on the GWAS of a panel of 355 natural accessions and linkage analysis of an F<sub>2</sub> population derived from two extreme materials in fiber length, a major QTL, which explains >10% of the phenotypic variance, on D03 was simultaneously identified in both strategies. A further study revealed that a candidate gene (*Gh\_D03G1338*) in

the QTL region may be responsible for determining cotton fiber elongation by regulating the synthesis of sucrose [124]. Wen et al. [114] combining linkage and association analyses of brown fiber accessions, fine-mapped the brown fiber region, *Lc1*, and dissected it into two loci, qBFA07-1 and qBF-A07-2. They further identified *Gh\_A07G2341* and *Gh\_A07G0100* as the causal genes of these two loci, respectively. The current study also combined both strategies of multiple linkage and association analyses. The linkage maps genotyping CCRI70 RILs with the Intl Cotton SNP Consortium\_80k chip and SLAF-seq technique identified 7 and 19 stable QTLs, respectively, of which six were simultaneously identified by both strategies, namely qOC<sup>chip</sup>-c2-2 and qOC<sup>slaf</sup>-c2-1, qOC<sup>chip</sup>-c2-2 and qOC<sup>slaf</sup>-c2-2, qOC<sup>slaf</sup>-c17-2 and qOC<sup>chip</sup>-c17-2, qOC<sup>slaf</sup>-c17-2 and qOC<sup>ccri70</sup>-c17-3, qOC<sup>slaf</sup>-c17-2 and qOC<sup>ccri70</sup>-c17-4, and qOC<sup>slaf</sup>-c17-2 and qOC<sup>ccri70</sup>-c17-5 (Table 1). Notably, qOC<sup>chip</sup>-c3-1 and qOC<sup>0-153</sup>-c3-2 shared identical confidence marker intervals in which a locus, c03\_18925549, was also significantly associated with KOC in the GWAS of the natural panel. In the confidence marker interval of qOC<sup>chip</sup>-c18-2, three loci, c18\_42111539, c18\_42113870, and c18\_42507564, were significantly associated with KOC in the GWAS of the natural panel. Linkage analysis of 0–153 RILs identified 21 stable QTLs for KOC in cottonseed (Table 1), among which, qOC<sup>0-153</sup>-c3-1, qOC<sup>0-153</sup>-c4-3 (together with qOC<sup>0-153</sup>-c4-2), qOC<sup>0-153</sup>-c11-3, and qOC<sup>0-153</sup>-c25-2 shared identical marker intervals of qOC<sup>slaf</sup>-c3-1, qOC<sup>slaf</sup>-c4-2, qOC<sup>slaf</sup>-c11-1, and qOC<sup>slaf</sup>-c25-1 in CCRI70 RILs, respectively. In the confidence marker interval of qOC<sup>0-153</sup>-c5-1, two loci, c05\_107481128 and c05\_107481161, were significantly associated with KOC in the GWAS of the natural panel. qOC<sup>0-153</sup>-c10-2 and qOC<sup>0-153</sup>-c19-2 each harbored a locus, c10\_77548836 and c19\_19777635, respectively, that was significantly associated with the GWAS of the natural panel. The mutual confirmation of the mapping results between different linkage maps (within CCRI70 RIL), across different populations (between CCRI70 and 0–153 RILs) and between linkage and association analyses, ensures the accuracy of the major QTLs. The major QTLs and QTNs mapped by linkage and association analyses were consistent, which could significantly improve the precision and accuracy of the loci [98]. The mapping results provide a good opportunity to further identify the hub genes in the lipid biosynthesis pathways to dissect the genetic mechanism of KOC accumulation during ovule development through interaction analysis of the candidate genes harbored in these loci.

To better understand the QTL/QTN identification results, markers in the QTL interval were aligned with the physical map of TM-1 (v2.1) [37] to compare with those reported in previous studies based on their respective physical confidence intervals. The stable QTLs that were obtained through linkage analysis of CCRI70 RILs (qOC<sup>slaf</sup>-c1-1, qOC<sup>slaf</sup>-c3-1, qOC<sup>slaf</sup>-c17-1, qOC<sup>slaf</sup>-c17-2, qOC<sup>slaf</sup>-c17-4, qOC<sup>chip</sup>-c17-2, qOC<sup>chip</sup>-c17-3, qOC<sup>chip</sup>-c17-4, qOC<sup>chip</sup>-c17-5, qOC<sup>chip</sup>-c18-2, and qOC<sup>slaf</sup>-c22-1) were identified in regions already reported in previous studies (Table 2) [23,61,130,134]. The remaining 17 QTLs were identified *de novo* in the current study. The stable QTLs obtained through linkage analysis of 0–153 RILs (qOC<sup>0-153</sup>-c3-2, qOC<sup>0-153</sup>-c5-1, qOC<sup>0-153</sup>-c7-2, qOC<sup>0-153</sup>-c10-2, qOC<sup>0-153</sup>-c19-1, qOC<sup>0-153</sup>-c19-2, and qOC<sup>0-153</sup>-c20-1) were identified in the regions already reported in previous studies (Table 2) [60,69,84,134]. The remaining 16 QTLs were identified *de novo* in the current study. In the natural panel, 14 stable significant loci were associated with the KOC in cottonseed (Table 1). The locus on chromosome A03 (c03\_18925549) overlapped with that reported by Zhu et al. [134]. The loci on D13 (c18\_42111539, c18\_42113870, and c18\_42507564) overlapped with the regions reported in several previous studies [61]. The locus of c19\_19777635 also had overlapping regions in previous results [60,69,84]. The remaining seven loci, c01\_69712855,

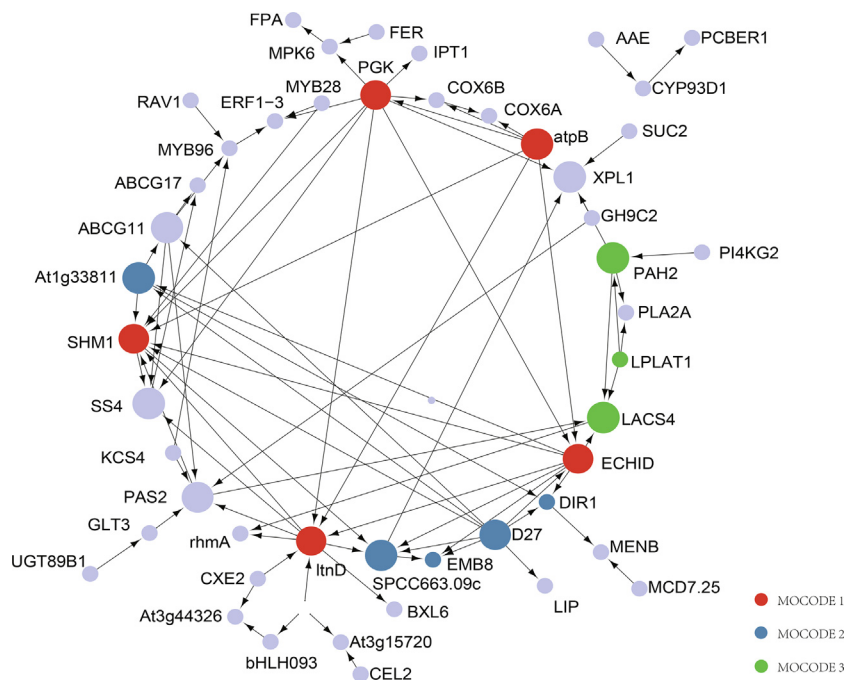


Fig. 5. Analysis of the protein interaction network of candidate genes under the molecular complexity detection (MCODE) algorithm.

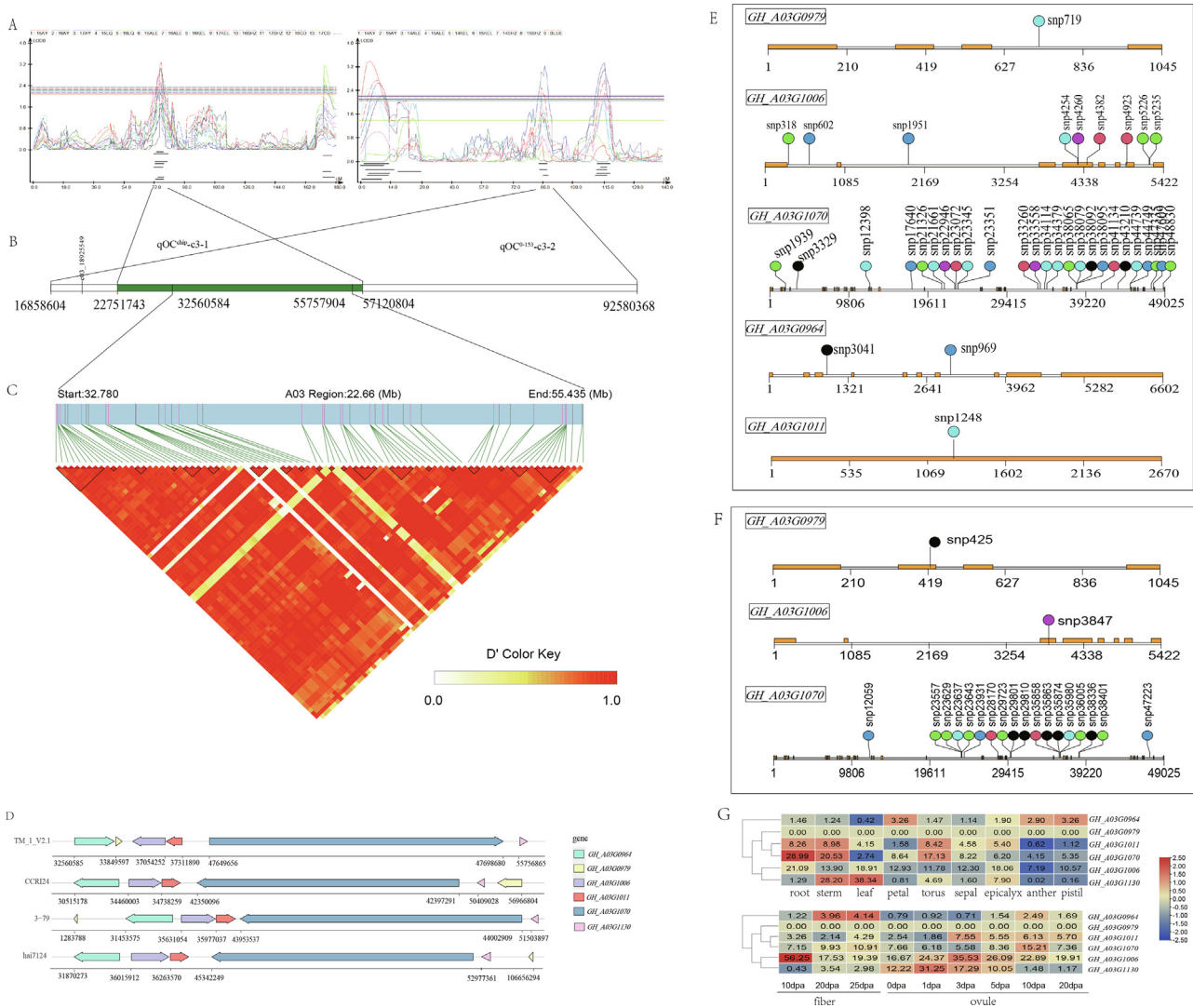
c05\_107481161, c06\_17240705, c07\_8706328, c11\_104291533, c18\_46810731, and c21\_22911784, were identified *de novo* in the current study.

### 19. Clustering and regulation of hub genes in carbon metabolism and oil synthesis pathways

The main component of vegetable oil is TAG, which is synthesized by assembling three FA molecules onto a glycerol skeleton [57]. The process of oil anabolism is complex and involves several pathways, including FA biosynthesis, FA desaturation, and glycerophospholipid metabolism [115]. The FA biosynthesis pathway is the most clearly understood pathway in oil synthesis. The essential first step in the *de novo* biosynthesis of FAs is the catalyzation of acetyl-CoA to produce malonyl-CoA by ACCase. Specific overexpression of heteromeric GhACCase in upland cotton can effectively increase KOC in mature cottonseed [16]. The activation of FAs through the formation of ATP-dependent thioester-linked FA CoA conjugates is required for further TAG synthesis [93]. Hexose and triose are the main carbohydrates for the anabolism of plant seed oil. They are transmitted to plastids to synthesize FA via glycolysis [6]. The FA synthesis pathway in plastids determines the final length and unsaturation level of FA chains in seed oils. Finally, the nascent TAG is transported to and stored in the oil bodies, the main storage organelles of lipids in plant cells [48].

The polyploidy of the *G. hirsutum* genome implies the complexity of gene function in the pathways related to FA biosynthesis and TAG assembly [26,52,125]. Studies have revealed that the genes in oil-related pathways in allotetraploid *G. hirsutum* are tightly co-regulated in response to artificial and natural selection during the domestication process [36]. The interaction network provides a strong indicator of gene function in a biological process [86,113], it may also indicate the expression characteristics of a gene cluster in the metabolic pathway. Here, we demonstrated that the hub genes in the interaction network correlated to the oil biosynthetic pathway in *G. hirsutum* co-localize in a cluster, which is harbored in the confidence marker interval of a major QTL iden-

tified simultaneously in both CCRI70 and 0–153 RILs (qOC<sup>chip</sup>-c3-1 and qOC<sup>0-153</sup>-c3-2) (Fig. 6). In the cluster, *LtnD* [126], *FER* [24,34,49], *PGK* [111,133], and *RAV1* [96,129] are involved in the related pathways of carbon metabolism in plastids, supplying carbon source and energy for FA synthesis, while *PAH2* [6,25] and *LPLAT1* [12,92] coordinate the Kennedy pathway and Lands cycle for TAG assembly [106] (Fig. 7). *LtnD* catalyzes the oxidation of L-threonate to produce 2-oxo-tetronate and synthesizes DHAP, an intermediate of glycolysis [126]. DHAP is then reduced to G3P via the catalysis of G3P dehydrogenase [105]. *PGK* catalyzes the conversion of 1,3-bisphosphoglycerate (1,3-BPG) to G3P in glycolysis [5,56]. Increasing *PGK* activity has been demonstrated to increase the KOC in sunflower seeds [101]. A significant expression difference in *PGK* was also observed in developing ovules between different KOC lines of cotton [133]. *FER* interacts with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) [118], which catalyzes the conversion of GAP to 1,3-BPG, an NADPH-dependent biochemical reaction. *FER* belongs to a large family of plasma membrane-localized receptor-like kinases (RLKs) [27,34]. It has a transmembrane domain [50,51], and its plasma-membrane localization and signaling require direct interaction with the glycosylphosphatidylinositol-anchored protein (GPI-AP) LORELEI-like GPI-AP1 (LLG1) [27,51]. The RAV TFs have two specific DNA-binding domains: one is the AP2/ERF domain about 60–70 amino acids in size that is contained in the AP2/ERF TF superfamily, and the other is the B3 domain about 110 amino acids in size that is contained in another plant-specific B3 TF superfamily, including VP1/ABI3 [74,96]. A recent study revealed that the stability of WRINKLED1, an AP2 TF, enhances oil accumulation in *Arabidopsis* [129], implying a role of the AP2 domain in TAG biosynthesis. The *LPLAT* gene encodes a lysophospholipid acyltransferase (LPLAT) with broad specificity, which participates in the opposite transport of nascent FAs from the chloroplast to the ER through phosphatidylcholine (PC) acyl editing [44] and mediates the transformation of lysophosphatidylethanolamine (1-acyl-*sn*-glycerin-3-phosphoethanolamine or LPE) to phosphatidylethanolamine (1,2-diacyl-*sn*-glycerin-3-phosphoethanolamine or PE). It can also use lysophosphatidylglycerol as a substrate *in vitro*, together with

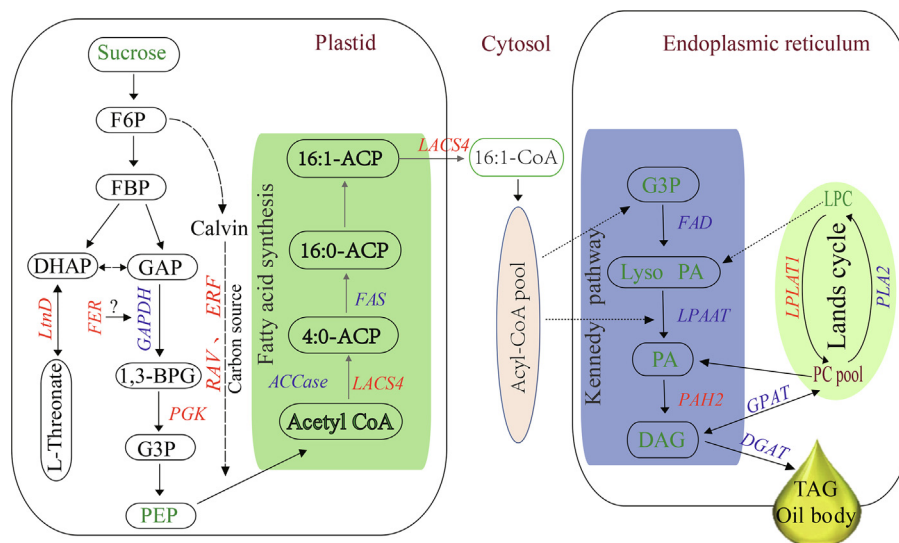


**Fig. 6.** Candidate gene identification in a major quantitative trait locus (QTL) on chromosome 3 (A03). A. The genetic position and significance of the common QTL identified both in CCR170 and 0–153 RILs. B. Physical position of the marker interval of common QTL. C. Haplotypes in the overlapping region of common QTL, where the hub genes were identified. D. Arrangement of the hub genes on the physical maps of the different *Gossypium* genome assemblies. E. SNP differences identified in the hub genes between sGK156 and 901–001, the two parental lines of CCR170 RILs. F. SNPs identified in the hub genes in the natural panel. G. The gene expression levels (FPKM) in  $\log_2(\text{FPKM} + 1)$  in different organs, tissues, and developmental stages of fiber and ovule.

*LPCAT2*, to export FAs synthesized *de novo* from the chloroplast to the PC pool through acyl editing [106]. In upland cotton, a homolog of *LPCAT1* (*Gh\_A03G0735*) was identified in a major QTL for KOC in cottonseed [61]. *PAH1* and *PAH2* act redundantly to inhibit phospholipid biosynthesis in the ER [25]. The phospholipid content in the leaves from *pah1 pah2* double mutants is ~ 1.8-fold more than that from the wild type. The dephosphorylation of phosphatidic acid to diacylglycerol (DAG) is a rate-limiting step in TAG formation [89]. Thus, PAH may control the formation of oil bodies through DAG; a decrease in TAG content is then observed in seeds of the mutants [2]. A homolog of *PAH2* (*Gh\_D05G1934*) in upland cotton was identified in the interval of the major QTL qOC-Dt5-1 [69].

Gene clustering in eukaryotes was first identified in the biosynthesis pathway of secondary metabolite cyclic hydroxamic acids in maize [30]. This phenomenon was also found in major plant species, including oat, rice [81,109,116], *Arabidopsis* [28,29], wheat, and barley [35]. Most of the clustering genes in early studies are correlated with the biosynthesis pathway of secondary metabolites [3], however, an increasing number of studies in recent years have shown that the genes of the primary metabolic pathways of both

fungi and plants may also form clusters [75]. Plants usually produce a large number of compounds that underpin the traits of ecological and agronomic importance to help shape ecological interactions and adaptations to changing environments [86]. Each specific compound is often restricted to specific taxon lineages and tissues or synthesized under certain environmental conditions [86]. Preformed or induced compounds that plants produce in response to biotic or abiotic stress are major secondary metabolites synthesized by plants [28]. Clustering the genes relating to the metabolic pathways of these compounds is conducive to the formation of plant resistance to corresponding biotic or abiotic stress. Examples of these compounds include at least cyclic hydroxamic acids in maize (*Zea mays*) [30], triterpene avenacin in diploid oat (*Avena strigosa*) [14,81], and momilactones in rice [109]. The disruption of any member of the clustered genes may lead not only to the loss of the pathway end-products but also to the accumulation of toxic/bioactive intermediates [14,29], which may further impose strong selection for the maintenance of intact clusters [71,75]. Previous studies have established a correlation between lipid metabolism and plant responses to abiotic stress [18,94]. In the gene cluster of the oil symbiotic pathway in the current study,



**Fig. 7.** A working model of the hub genes *LPLAT1* and *PAH2* in the Kennedy pathway. This model was modified based on Li-Beisson et al. [48], Wang et al. [106], Zhao et al. [131], and Zhu et al. [133].

some genes were also involved in plant responses to stress. *FER* signaling pathways are multifunctional for plant growth and development in response to internal physiological signals [51,119] or for plant defense establishment in responses to biotic or abiotic environmental signals [17,58]. *PGK1* and *PGK3* of upland cotton were targeted by downregulated lncRNAs [91]. *RAV1* also responded to environmental signals, including cold, NaCl, and PEG treatments [83]. When cells are exposed to excess nutrients, FAs are esterified to G3P to form TAG [15,93]. Oil bodies provide an effective mechanism to keep lipids in dynamic turnover equilibrium, sequestering potentially toxic lipids and providing an on-demand source of FAs during periods of increased cellular need in response to ER stress, or lipid sensing in the ER [13,87,111]. The involvement of these clustered genes in plant responses to stress indirectly supports their expression interaction in the pathways of lipid anabolism.

In summary, we identified 16 major QTLs/QTNs for the KOC in cottonseed that were simultaneously detected across multiple environments and populations based on both linkage analysis and GWAS (Table S8). In a short DNA fragment on chromosome 3 (A03) of the confidence interval of a major QTL, *qOC<sup>chip</sup>-c3-1* in CCR170 RILs and *qOC<sup>0-153</sup>-c3-2* in 0–153 RILs harbored a cluster-like hub or key genes in lipid biosynthesis-related pathways (Fig. 6), proposing a working model of their roles in lipid biosynthesis (Fig. 7). These clustered genes consisted of two distinct but interacting groups: those that regulate carbon metabolism and energy-flow, including *LtnD*, *FER*, *PGK*, and *RAV1*, and those that regulate oil synthesis and assembly, including *LPLAT1* and *PAH2*. Upon the plant internal developmental signal of lipid biosynthesis initiation, the former four genes modulated the flow of hexose or triose and energy to FA synthesis, while the latter two genes catalyzed the metabolic interaction of the key steps of the Lands Cycle and Kennedy Pathway in TAG assembly in upland cottonseed. A literature survey indicated that oil metabolism is also involved in the response of plants to abiotic stress, thereby inferring the expression characteristics of these clustered regulatory genes.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

The authors thank supports by the grants from the National Natural Science Foundation of China (32070560,31621005), the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2021D01B114), China Agriculture Research System of MOF and MARA, Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences, Xinjiang Production and Construction Corps Innovation Program for Science and Technology Talents (2020CB005), Central Public-interest Scientific Institution Basal Research Fund (CN) (1610162022002), Scientific and Technological Talents Innovation Project “Corps Cotton Germplasm Resources Research And Innovation Team” (2020CB005). The authors acknowledge SciLang for its linguistic assistance during the preparation of this manuscript.

### Author contributions

Y.Y., J.G. and W.G. conceive the project. Y.P. and J.Y. collect the data. W.P., Z.Z., D.F., L.L., X.X. and R.L. make the field evaluations. Q.L., P.L., H.S., Y.S., J.L. and Q.G. help analyze the data. A.L., X.D., S. F. and J.P. provide water lab techniques, Q.C., Y.Y. and W.G. finalize the whole project analysis. J.G. and W.G. wrote and finalize the manuscript with comments from all authors.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.04.012>.

### References

- [1] Abbassian, A., Cabezas, S.C., Mustafa, S., and Doro, E. (2018). OECD-FAO Agricultural Outlook 2018–2027.
- [2] Adeyo O, Horn PJ, Lee S, Binns DD, Chandras A, Chapman KD, et al. The yeast lipin orthologue Pah1p is important for biogenesis of lipid droplets. *J Cell Biol* 2011;192:1043–55.
- [3] Alam B, Li J, Gē Q, Khan MA, Gōng J, Mehmood S, et al. Endophytic Fungi: From Symbiosis to Secondary Metabolite Communications or Vice Versa? *Front Plant Sci* 2021;12:791033.
- [4] Atwell S, Huang YS, Vilhjalmsón BJ, Willems G, Horton M, Li Y, et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 2010;465:627–31.
- [5] Banks RD, Blake CC, Evans PR, Haser R, Rice DW, Hardy GW, et al. Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-bending enzyme. *Nature* 1979;279:773–7.



- [6] Bates PD, Stymne S, Ohlrogge J. Biochemical pathways in seed oil synthesis. *Curr Opin Plant Biol* 2013;16:358–64.
- [7] Baud S, Dichow NR, Kelemen Z, d'Andrea S, To A, Berger N, et al. Regulation of HSD1 in seeds of *Arabidopsis thaliana*. *Plant Cell Physiol* 2009;50:1463–78.
- [8] Beisson F, Koo AJ, Ruuska S, Schwender J, Pollard M, Thelen JJ, et al. *Arabidopsis* genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol* 2003;132:681–97.
- [9] Bernardo R. Genomewide markers as cofactors for precision mapping of quantitative trait loci. *Theor Appl Genet* 2013;126:999–1009.
- [10] Chen J, Nolan T, Ye H, Zhang M, Tong H, Xin P, et al. *Arabidopsis* WRKY46, WRKY54 and WRKY70 Transcription Factors Are Involved in Brassinosteroid-Regulated Plant Growth and Drought Response. *Plant Cell* 2017;29:1425–39.
- [11] Chen Y, Fu M, Li H, Wang L, Liu R, Liu Z, et al. High-oleic acid content, nontransgenic allotetraploid cotton (*Gossypium hirsutum* L.) generated by knockout of GhFAD2 genes with CRISPR/Cas9 system. *Plant Biotechnol J* 2021;19:424–6.
- [12] Cheung A, Zheng Y, Deng X, Qu A, Zhang M, Tao Y, et al. Regulation of pollen lipid body biogenesis by MAP kinases and downstream WRKY transcription factors in *Arabidopsis*. *PLOS Genet* 2018;14:e1007880.
- [13] Choudhary V, Golani G, Joshi AS, Cottier S, Schneiter R, Prinz WA, et al. Architecture of Lipid Droplets in Endoplasmic Reticulum Is Determined by Phospholipid Intrinsic Curvature. *Curr Biol* 2018;28:915–26.
- [14] Chu HY, Wegel E, Osbourn A. From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. *Plant J* 2011;66:66–79.
- [15] Coleman RA, Lewin TM, Muoio DM. Physiological and nutritional regulation of enzymes of triacylglycerol synthesis. *Annu Rev Nutr* 2000;20:77–103.
- [16] Cui Y, Liu Z, Zhao Y, Wang Y, Huang Y, Li L, et al. Overexpression of Heteromeric GhACCase Subunits Enhanced Oil Accumulation in Upland Cotton. *Plant Mol Biol Rep* 2017;35:287–97.
- [17] De Smet I, Voß U, Jürgens G, Beeckman T. Receptor-like kinases shape the plant. *Nat Cell Biol* 2009;11:1166–73.
- [18] Degenkolbe T, Giavalisco P, Zuther E, Seiwert B, Hinch DK, Willmitzer L. Differential remodeling of the lipidome during cold acclimation in natural accessions of *Arabidopsis thaliana*. *Plant J* 2012;72:972–82.
- [19] Deng X, Gong J, Liu A, Shi Y, Gong W, Ge Q, et al. QTL mapping for fiber quality and yield-related traits across multiple generations in segregating population of CCRI 70. *J Cotton Res* 2019;2:112–21.
- [20] Dhamodaran G, Krishnan R, Pochareddy YK, Pyarelal HM, Sivasubramanian H, Ganeshram AK. A comparative study of combustion, emission, and performance characteristics of rice-bran-, neem-, and cottonseed-oil biodiesels with varying degree of unsaturation. *Fuel* 2017;187:296–305.
- [21] Du C, Chen Y, Wang K, Yang Z, Zhao C, Jia Q, et al. Strong co-suppression impedes an increase in polyunsaturated fatty acids in seeds overexpressing FAD2. *J Exp Bot* 2019;70:985–94.
- [22] Du X, Huang G, He S, Yang Z, Sun G, Ma X, et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat Genet* 2018;50:796–802.
- [23] Du X, Liu S, Sun J, Zhang G, Jia Y, Pan Z, et al. Dissection of complicate genetic architecture and breeding perspective of cottonseed traits by genome-wide association study. *BMC Genomics* 2018;19:451.
- [24] Duan Q, Kita D, Li C, Cheung AY, Wu H-M. FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development. *Proc Natl Acad Sci U S A* 2010;107:17821–6.
- [25] Eastmond PJ, Quettier AL, Kroon JT, Craddock C, Adams N, Slabas AR. PHOSPHATIDIC ACID PHOSPHOHYDROLASE1 and 2 Regulate Phospholipid Synthesis at the Endoplasmic Reticulum in *Arabidopsis*. *Plant Cell* 2011;22:2796–811.
- [26] Fang L, Wang Q, Hu Y, Jia Y, Chen J, Liu B, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet* 2017;49:1089–98.
- [27] Feng W, Kita D, Peaucelle A, Cartwright HN, Doan V, Duan Q, et al. The FERONIA Receptor Kinase Maintains Cell-Wall Integrity during Salt Stress through Ca<sup>2+</sup> Signaling. *Curr Biol* 2018;28:666–75.
- [28] Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H, Osbourn AE. Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci U S A* 2011;108:16116–21.
- [29] Field B, Osbourn AE. Metabolic diversification-independent assembly of operon-like gene clusters in different plants. *Science* 2008;320:543–7.
- [30] Frey M, Chomet P, Glawischnig E, Stettner C, Grün S, Winklmair A, et al. Analysis of a chemical plant defense mechanism in grasses. *Science* 1997;277:696–9.
- [31] Gong J, Kong D, Liu C, Li P, Liu P, Xiao X, et al. Multi-environment evaluations across ecological regions reveal that the kernel oil content of cottonseed is equally determined by genotype and environment. *J Agric Food Chem* 2022;70:2529–44.
- [32] Gong J, Liu A, Li J, Jiang X, Duan L, Ge Q, et al. Major Gene Plus Polygene Genetic Analysis of Lint Percent in Upland Cotton. *Cotton Sci* 2019;31:192–200.
- [33] Grover CE, Gallagher JP, Jareczek JJ, Page JT, Udall JA, Gore MA, et al. Re-evaluating the phylogeny of allopolyploid *Gossypium* L. *Mol Phylogenet Evol* 2015;92:45–52.
- [34] Guo H, Li L, Ye H, Yu X, Algreen A, Yin Y. Three related receptor-like kinases are required for optimal cell elongation in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 2009;106:7648–53.
- [35] Hen-Avivi S, Savin O, Racovita RC, Lee W-S, Adamski NM, Malitsky S, et al. A Metabolic Gene Cluster in the Wheat W1 and the Barley Cer-cqu Loci Determines  $\beta$ -Diketone Biosynthesis and Glauconess. *Plant Cell* 2016;28:1440–60.
- [36] Hu G, Hovav R, Grover CE, Faigenboim-Doron A, Kadmon N, Page JT, et al. Evolutionary Conservation and Divergence of Gene Coexpression Networks in *Gossypium* (Cotton) Seeds. *Genome Biol Evol* 2016;8:3765–83.
- [37] Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet* 2019;51:739–48.
- [38] Huai D, Xue X, Li Y, Wang P, Li J, Yan L, et al. Genome-Wide Identification of Peanut KCS Genes Reveals That AhKCS1 and AhKCS2 Are Involved in Regulating VLCFA Contents in Seeds. *Front Plant Sci* 2020;11:406.
- [39] Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, et al. Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat Genet* 2020;52:516–24.
- [40] Hwang W, Cho Y-R, Zhang A, Ramanathan M. A novel functional module detection algorithm for protein-protein interaction networks. *Algorithm Mol Biol* 2006;1:24.
- [41] Izadi-Darbandi A, Younessi-Hamzekhanlu M, Sticklen M. Metabolically engineered rice biomass and grain using genes associated with lipid pathway show high level of oil content. *Mol Biol Rep* 2020;47:7917–27.
- [42] Jessen D, Roth C, Wiermer M, Fulda M. Two activities of long-chain acyl-coenzyme A synthetase are involved in lipid trafficking between the endoplasmic reticulum and the plastid in *Arabidopsis*. *Plant Physiol* 2015;167:351–66.
- [43] Kang, J.-W., Kabange, N.R., Phyo, Z., Park, S.-Y., Lee, S.-M., Lee, J.-Y., Shin, D., Cho, J.H., Park, D.-S., Ko, J.-M., et al. (2020). Combined Linkage Mapping and Genome-Wide Association Study Identified QTLs Associated with Grain Shape and Weight in Rice (*Oryza sativa* L.). *Agronomy-Basel* 10:1532.
- [44] Karki N, Johnson BS, Bates PD. Metabolically Distinct Pools of Phosphatidylcholine Are Involved in Trafficking of Fatty Acids out of and into the Chloroplast for Membrane Production. *Plant Cell* 2019;31:2768–88.
- [45] Killcoyne S, Carter GW, Smith J, Boyle J. Cytoscape: a community-based framework for network modeling. *Methods Mol Biol* 2009;563:219–39.
- [46] Konukan, D.B., Yilmaztekin, M., Mert, M., Gener, O., and Konukan (2017). Tarm Bilimleri Dergisi Physico-Chemical Characteristic and Fatty Acids Compositions of Cottonseed Oils. *Tarım Bilimleri Dergisi* 23, 253–259.
- [47] Lacape JM, Jacobs J, Arioli T, Derjicker R, Forestier-Chiron N, Llewellyn D, et al. A new interspecific, *Gossypium hirsutum* × *G. barbadense*, RIL population: towards a unified consensus linkage map of tetraploid cotton. *Theor Appl Genet* 2009;119:281–92.
- [48] Li-Beisson Y, Shorosh B, Beisson F, Andersson MX, Arondel V, Bates PD, et al. Acyl-lipid metabolism *Arabidopsis* Book 2013;11:e0161.
- [49] Li C, Liu X, Qiang X, Li X, Li X, Zhu S, et al. EBPI nuclear accumulation negatively feeds back on FERONIA-mediated RALF1 signaling. *PLOS Biol* 2018;16:e2006340.
- [50] Li C, Wu HM, Cheung AY. FERONIA and Her Pals: Functions and Mechanisms. *Plant Physiol* 2016;171:2379–92.
- [51] Li, C., Yeh, F.-L., Cheung, A.Y., Duan, Q., Kita, D., Liu, M.-C., Maman, J., Luu, E.J., Wu, B.W., Gates, L., et al. (2015a). Glycosylphosphatidylinositol-anchored proteins as chaperones and co-receptors for FERONIA receptor kinase signaling in *Arabidopsis*. *eLife* 4:e06587.
- [52] Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol* 2015;33:524–30.
- [53] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- [54] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- [55] Li N, Li J, Desiderio DM, Zhan X. SILAC quantitative proteomics analysis of ivermectin-related proteomic profiling and molecular network alterations in human ovarian cancer cells. *J Mass Spectrom* 2021;56:e4659.
- [56] Li R, Qiu Z, Wang X, Gong P, Xu Q, Yu QB, et al. Pooled CRISPR/Cas9 reveals redundant roles of plastidial phosphoglycerate kinases in carbon fixation and metabolism. *Plant J* 2019;98:1078–89.
- [57] Lin L, Allemekinders H, Dansby A, Campbell L, Durance-Tod S, Berger A, et al. Evidence of health benefits of canola oil. *Nutr Rev* 2013;71:370–85.
- [58] Lindner H, Müller LM, Boisson-Dernier A, Grossniklaus U. CrRLK1L receptor-like kinases: not just another brick in the wall. *Curr Opin Plant Biol* 2012;15:659–69.
- [59] Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics* 2012;28:2397–9.
- [60] Liu D, Liu F, Shan X, Zhang J, Tang S, Fang X, et al. 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.
- [61] Liu HY, Zhang L, Mei L, Quampah A, He QL, Zhang BS, et al. qOil-3, a major QTL identification for oil content in cottonseed across genomes and its candidate gene analysis. *Ind Crop Prod* 2020;145:112070.
- [62] Liu J, Hao W, Liu J, Fan S, Zhao W, Deng L, et al. A Novel Chimeric Mitochondrial Gene Confers Cytoplasmic Effects on Seed Oil Content in Polyploid Rapeseed (*Brassica napus*). *Mol Plant* 2019;12:582–96.

- [63] Liu JY, Zhang YW, Han X, Zuo JF, Zhang Z, Shang H, et al. An evolutionary population structure model reveals pleiotropic effects of GmPDAT for traits related to seed size and oil content in soybean. *J Exp Bot* 2020;71:6988–7002.
- [64] Liu K, Sun X, Ning T, Duan X, Wang Q, Liu T, et al. Genetic dissection of wheat panicle traits using linkage analysis and a genome-wide association study. *Theor Appl Genet* 2018;131:1073–90.
- [65] Liu M, Tan X, Yang Y, Liu P, Zhang X, Zhang Y, et al. Analysis of the genetic architecture of maize kernel size traits by combined linkage and association mapping. *Plant Biotechnol J* 2020;18:207–21.
- [66] Liu R, Xiao X, Gong J, Li J, Zhang Z, Liu A, et al. QTL mapping for plant height and fruit branch number based on RIL population of upland cotton. *J Cotton Res* 2020;3:5.
- [67] Lu Y, Zhang S, Shah T, Xie C, Hao Z, Li X, et al. Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proc Natl Acad Sci U S A* 2010;107:19585–90.
- [68] Lunn D, Le A, Wallis JG, Browse J. Castor LPCAT and PDAT1A Act in Concert to Promote Transacylation of Hydroxy-Fatty Acid onto Triacylglycerol. *Plant Physiol* 2020;184:709–19.
- [69] Ma J, Liu J, Pei W, Ma Q, Wang N, Zhang X, et al. Genome-wide association study of the oil content in upland cotton (*Gossypium hirsutum* L.) and identification of GhPRXR1, a candidate gene for a stable QTLqOC-Dt5-1. *Plant Sci* 2019;286:89–97.
- [70] Ma M, Ren Y, Xie W, Zhou D, Tang S, Kuang M, et al. Physicochemical and functional properties of protein isolate obtained from cottonseed meal. *Food Chem* 2018;240:856–62.
- [71] McGary KL, Slot JC, Rokas A. Physical linkage of metabolic genes in fungi is an adaptation against the accumulation of toxic intermediate compounds. *Proc Natl Acad Sci U S A* 2013;110:11481–6.
- [72] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- [73] Meng L, Li HH, Zhang LY, Wang JK. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J* 2015;3:269–83.
- [74] Nakano T, Suzuki K, Fujimura T, Shinshi H. Genome-Wide Analysis of the ERF Gene Family in Arabidopsis and Rice. *Plant Physiol* 2006;140:411–32.
- [75] Nützmann HW, Scacciocchio C, Osbourn A. Metabolic Gene Clusters in Eukaryotes. *Annu Rev Genet* 2018;52:159–83.
- [76] Overstreet SA. Foreign medical trainees. *J Ky Med Assoc* 1970;68:115–6.
- [77] Pandey SN, Thejappa N. Study on relationship between oil, protein, and gossypol in cottonseed kernels. *J Am Oil Chem Soc* 1975;52:312–5.
- [78] Parthibane V, Iyappan R, Vijayakumar A, Venkateshwari V, Rajasekharan R. Serine/Threonine/Tyrosine Protein Kinase Phosphorylates Oleosin, a Regulator of Lipid Metabolic Functions. *Plant Physiol* 2012;159:95–104.
- [79] Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 2012;492:423–7.
- [80] Pettigrew WT, Dowd MK. Molecular biology and physiology: Interactions between irrigation regimes and varieties result in altered cottonseed composition. *J Cotton Sci* 2012;16:42–52.
- [81] Qi X, Bakht S, Leggett M, Maxwell C, Melton R, Osbourn A. A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc Natl Acad Sci U S A* 2004;101:8233–8.
- [82] Ramli US, Baker DS, Quant PA, Harwood JL. Use of control analysis to study the regulation of plant lipid biosynthesis. *Biochem Soc Trans* 2002;30:1043–6.
- [83] Ren C, Li H, Wang Z, Dai Z, Lecourieux F, Kuang Y, et al. Characterization of Chromatin Accessibility and Gene Expression upon Cold Stress Reveals that the RAV1 Transcription Factor Functions in Cold Response in *Vitis Amurensis*. *Plant Cell Physiol* 2021;62:1615–29.
- [84] 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;290:1615–25.
- [85] Said JI, Lin Z, Zhang X, Song M, Zhang J. A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. *BMC Genomics* 2013;14:776.
- [86] Schläpfer P, Zhang P, Wang C, Kim T, Banf M, Chae L, et al. Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. *Plant Physiol* 2017;173:2041–59.
- [87] Schuck S, Prinz WA, Thorn KS, Voss C, Walter P. Membrane expansion alleviates endoplasmic reticulum stress independently of the unfolded protein response. *J Cell Biol* 2009;187:525–36.
- [88] Shi CN, Zheng YT, Geng JY, Liu CY, Pei H, Ren Y, et al. Identification of herbicide resistance loci using a genome-wide association study and linkage mapping in Chinese common wheat. *Crop J* 2020;8:666–75.
- [89] Skinner JR, Shew TM, Schwartz DM, Tzekov A, Lepus CM, Abumrad NA, et al. Diacylglycerol Enrichment of Endoplasmic Reticulum or Lipid Droplets Recruits Perilipin 3/TIP47 during Lipid Storage and Mobilization. *J Biol Chem* 2009;284:30941–8.
- [90] Song G, Li X, Munir R, Khan AR, Azhar W, Yasin MU, et al. The WRKY6 transcription factor affects seed oil accumulation and alters fatty acid compositions in *Arabidopsis thaliana*. *Physiol Plant* 2020;169:612–24.
- [91] Song Y, Li L, Yang Z, Zhao G, Zhang X, Wang L, et al. Target of Rapamycin (TOR) Regulates the Expression of lncRNAs in Response to Abiotic Stresses in Cotton. *Front Genet* 2019;9:690.
- [92] Ståhl U, Ståhlberg K, Stymne S, Ronne H. A family of eukaryotic lysophospholipid acyltransferases with broad specificity. *FEBS Lett* 2008;582:305–9.
- [93] Stevenson J, Huang EY, Olzmann JA. Endoplasmic Reticulum-Associated Degradation and Lipid Homeostasis. *Annu Rev Nutr* 2016;36:511–42.
- [94] Sui N, Wang Y, Liu S, Yang Z, Wang F, Wan S. Transcriptomic and Physiological Evidence for the Relationship between Unsaturated Fatty Acid and Salt Stress in Peanut. *Front Plant Sci* 2018;9:7.
- [95] Sun F-D, Zhang J-H, Wang S-F, Gong W-K, Shi Y-Z, Liu A-Y, et al. QTL mapping for fiber quality traits across multiple generations and environments in upland cotton. *Mol Breeding* 2012;30:569–82.
- [96] Suzuki M, McCarty DR. Functional symmetry of the B3 network controlling seed development. *Curr Opin Plant Biol* 2008;11:548–53.
- [97] Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, et al. Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. *Mol Plant* 2021;14:470–87.
- [98] Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-García S, et al. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 2011;43:159–62.
- [99] Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res* 2017;45:W122–9.
- [100] Tian W, Wang C, Gao Q, Li L, Luan S. Calcium spikes, waves and oscillations in plant development and biotic interactions. *Nat Plants* 2020;6:750–9.
- [101] Troncoso-Ponce MA, Garces R, Martinez-Force E. Glycolytic enzymatic activities in developing seeds involved in the differences between standard and low oil content sunflowers (*Helianthus annuus* L.). *Plant Physiol Biochem* 2010;48:961–5.
- [102] Turner JH, Ramey HH, Worley S. Influence of Environment on Seed Quality of Four Cotton Cultivars. *Crop Sci* 1976;16:407–9.
- [103] Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthi V. Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). *Theor Appl Genet* 2014;127:283–95.
- [104] van der Helm E, Imamovic L, Hashim Ellabaan MM, van Schaik W, Koza A, Sommer MOA. Rapid resistome mapping using nanopore sequencing. *Nucleic Acids Res* 2017;45:e61.
- [105] 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;5:431–41.
- [106] Wang L, Shen W, Kazachkov M, Chen G, Chen Q, Carlsson AS, et al. Metabolic interactions between the Lands cycle and the Kennedy pathway of glycerolipid synthesis in *Arabidopsis* developing seeds. *Plant Cell* 2012;24:4652–69.
- [107] Wang M, Tu L, Yuan D, Zhu Shen, C., Li, J., Liu, F., Pei, L., Wang, P., Zhao, G., et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat Genet* 2019;51:224–9.
- [108] Wang N, Ma J, Pei W, Wu M, Li H, Li X, et al. 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 Genomics* 2017;18:218.
- [109] Wang Q, Hillwig ML, Peters RJ. CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *Plant J* 2011;65:87–95.
- [110] Wang, S., Basten, C.J., and Zeng, Z.B. (2012b). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC (<https://brwcbportalcosncsu.edu/qtlcart/WQTLCart.htm>).
- [111] Wang S, Idrissi F-Z, Hermansson M, Grippa A, Ejsing CS, Carvalho P. Seipin and the membrane-shaping protein Pex30 cooperate in organelle budding from the endoplasmic reticulum. *Nature Commun* 2018;9:2939.
- [112] Wang W, Sun Y, Yang P, Cai X, Yang L, Ma J, et al. A high density SLAF-seq SNP genetic map and QTL for seed size, oil and protein content in upland cotton. *BMC Genomics* 2019;20:599.
- [113] Wei H, Persson S, Mehta T, Srinivasasainagendra V, Chen L, Page GP, et al. Transcriptional Coordination of the Metabolic Network in *Arabidopsis*. *Plant Physiol* 2006;142:762–74.
- [114] Wen T, Wu M, Shen C, Gao B, Zhu Zhang, X., You, C., and Lin, Z. Linkage and association mapping reveals the genetic basis of brown fibre (*Gossypium hirsutum*). *Plant Biotechnol J* 2018;16:1654–66.
- [115] White SW, Zheng J, Zhang YM, Rock. The structural biology of type II fatty acid biosynthesis. *Annu Rev Biochem* 2005;74:791–831.
- [116] Wilderman PR, Xu M, Jin Y, Coates RM, Peters RJ. Identification of SynPimara-7,15-Diene Synthase Reveals Functional Clustering of Terpene Synthases Involved in Rice Phytoalexin/Allelochemical Biosynthesis. *Plant Physiol* 2004;135:2098–105.
- [117] Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 2011;39:W316–22.
- [118] Yang T, Wang L, Li C, Liu Y, Zhu S, Qi Y, et al. Receptor protein kinase FERONIA controls leaf starch accumulation by interacting with glyceraldehyde-3-phosphate dehydrogenase. *Biochem Biophys Res Commun* 2015;465:77–82.

- [119] Yu F, Li J, Huang Y, Liu L, Li D, Chen L, et al. FERONIA Receptor Kinase Controls Seed Size in *Arabidopsis thaliana*. *Mol Plant* 2014;7:920–2.
- [120] Yu J, Yu S, Fan S, Song M, Zhai H, Li X, et al. 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.
- [121] Yu XH, Rawat R, Shanklin J. Characterization and analysis of the cotton cyclopropane fatty acid synthase family and their contribution to cyclopropane fatty acid synthesis. *BMC Plant Biol* 2011;11:97.
- [122] Yuan Y, Shi Y, Li J, Liu A, Zhai X, Li Y, et al. Transgenic insect-resistant high quality hybrid cotton – CCR1 70. *China Cotton* 2009;36:17.
- [123] Zang X, Geng X, Ma L, Wang N, Pei W, Wu M, et al. A genome-wide analysis of the phospholipid: diacylglycerol acyltransferase gene family in *Gossypium*. *BMC Genomics* 2019;20:402.
- [124] Zhang, C., Li, L., Liu, Q., Gu, L., Huang, J., Wei, H., Wang, H., and Yu, S. (2019). Identification of Loci and Candidate Genes Responsible for Fiber Length in Upland Cotton (*Gossypium hirsutum* L.) via Association Mapping and Linkage Analyses. *Front Plant Sci* 10:53.
- [125] Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol* 2015;33:531–7.
- [126] Zhang XS, Carter MS, Vetting MW, Francisco BS, Zhao SW, Al-Obaidi NF, et al. Assignment of function to a domain of unknown function: DUF1537 is a new kinase family in catabolic pathways for acid sugars. *Proc Natl Acad Sci U S A* 2016;113:E4161–9.
- [127] Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 2010;42:355–60.
- [128] Zhang Z, Li J, Jamshed M, Shi Y, Liu A, Gong J, et al. 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;18:239–53.
- [129] Zhang, Z., Li, J., Muhammad, J., Cai, J., Jia, F., Shi, Y., Gong, J., Shang, H., Liu, A., Chen, T., et al. (2015b). High Resolution Consensus Mapping of Quantitative Trait Loci for Fiber Strength, Length and Micronaire on Chromosome 25 of the Upland Cotton (*Gossypium hirsutum* L.). *Plos One* 10(8): e0135430.
- [130] Zhao W, Kong X, Yang Y, Nie X, Lin Z. Association mapping seed kernel oil content in upland cotton using genome-wide SSRs and SNPs. *Mol Breeding* 2019;39:105.
- [131] Zhao Y, Wang Y, Huang Y, Cui Y, Hua J. Gene network of oil accumulation reveals expression profiles in developing embryos and fatty acid composition in Upland cotton. *J Plant Physiol* 2018;228:101–12.
- [132] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10:1523.
- [133] Zhu D, Le Y, Zhang R, Li X, Lin Z. A global survey of the gene network and key genes for oil accumulation in cultivated tetraploid cottons. *Plant Biotechnol J* 2021;19:1170–82.
- [134] Zhu D, Li X, Wang Z, You C, Nie X, Sun J, et al. Genetic dissection of an allotetraploid interspecific CSSLs guides interspecific genetics and breeding in cotton. *BMC Genomics* 2020;21:431.
- [135] Zhu T, Liang C, Meng Z, Sun G, Meng Z, Guo S, et al. CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol* 2017;17:101.
- [136] Zou XY, Gong JW, Duan L, Jiang X, Zhen Z, Fan SM, et al. High-density genetic map construction and QTL mapping for fiber strength on Chr24 across multiple environments in a CCR170 recombinant inbred lines population. *Euphytica* 2018;214:102.