Journal of Advanced Research 65 (2024) 33–46

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Genetic linkage analysis of stable QTLs in *Gossypium hirsutum* RIL population revealed function of *GhCesA4* in fiber development

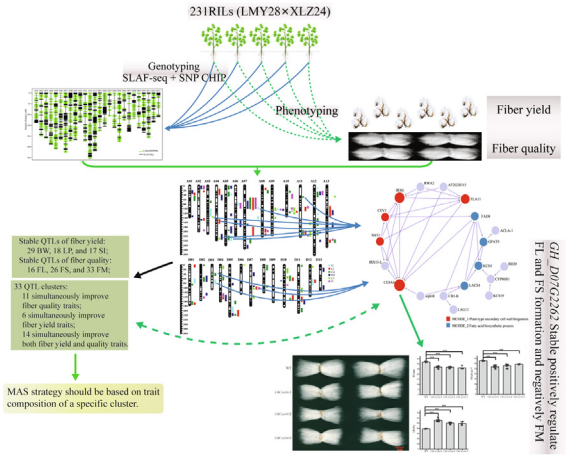
Ruìxián Liú [a,b], Xiānghuī Xiāo [a,b,c], Jw Gōng [a,d], Jùnwén Lǐ [a,d], Hàoliàng Yán [a], Qún Gě [a,d], Quánwěi Lú [c], Péngtāo Lǐ [c], Jìngtāo Pān [a], Hǎihóng Shāng [a,d], Yùzhēn Shí [a], Qúanjiā Chén [b,*], Yǒulù Yuán [a,b,d,*], Wànkuí Gǒng [a,*]

[a] National Key Laboratory of Cotton Bio-breeding and Integrated Utilization, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China
[b] Engineering Research Centre of Cotton, Ministry of Education, College of Agriculture, Xinjiang Agricultural University, 311 Nongda East Road, Urumqi 830052, Xinjiang, China
[c] College of Biotechnology and Food Engineering, Anyang Institute of Technology, Anyang 455000, Henan, China
[d] Zhengzhou Research Base, National Key Laboratory of Cotton Bio-breeding and Integrated Utilization, Zhengzhou University, Zhengzhou 450001, Henan, China

HIGHLIGHTS

- Genome-wide mining of QTLs of fiber yield and fiber quality traits were beneficial to better understanding their genome level distributions.
- A total of 64 stable QTLs of fiber yield and 75 stable QTLs of fiber quality were identified, which formed 33 clusters, indicating that the correlations between fiber yield and quality traits have complex genetic components.
- These QTLs formed expression network, via their candidate genes, to orchestrate fiber development and the formation fiber yield and quality.
- The candidate gene, GH_D07G2262, an upland cotton cellulose synthase 4 (*GhCesA 4*) gene had pleiotropic functions during cotton fiber development, which positively regulates fiber length and strength, and negatively lint percentage.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 31 May 2023
Revised 27 August 2023
Accepted 2 December 2023
Available online 6 December 2023

Keywords:
Gossypium hirsutum
High-density genetic map (HDGM)
Fiber quality and yield trait
Quantitative trait loci (QTLs), *GhCesA4*

ABSTRACT

Introduction: Upland cotton is an important allotetrapolyploid crop providing natural fibers for textile industry. Under the present high-level breeding and production conditions, further simultaneous improvement of fiber quality and yield is facing unprecedented challenges due to their complex negative correlations.

Objectives: The study was to adequately identify quantitative trait loci (QTLs) and dissect how they orchestrate the formation of fiber quality and yield.

Methods: A high-density genetic map (HDGM) based on an intraspecific recombinant inbred line (RIL) population consisting of 231 individuals was used to identify QTLs and QTL clusters of fiber quality and yield traits. The weighted gene correlation network analysis (WGCNA) package in R software was utilized to identify WGCNA network and hub genes related to fiber development. Gene functions were

* Corresponding authors.
E-mail addresses: chqjia@126.com (Q. Chén), yuanyoulu@caas.cn (Y. Yuán), gongwankui@caas.cn (W. Gǒng).

https://doi.org/10.1016/j.jare.2023.12.005

verified via virus-induced gene silencing (VIGS) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 strategies.

Results: An HDGM consisting of 8045 markers was constructed spanning 4943.01 cM of cotton genome. A total of 295 QTLs were identified based on multi-environmental phenotypes. Among 139 stable QTLs, including 35 newly identified ones, seventy five were of fiber quality and 64 yield traits. A total of 33 QTL clusters harboring 74 QTLs were identified. Eleven candidate hub genes were identified via WGCNA using genes in all stable QTLs and QTL clusters. The relative expression profiles of these hub genes revealed their correlations with fiber development. VIGS and CRISPR/Cas9 edition revealed that the hub gene *cellulose synthase 4* (*GhCesA4*, *GH_D07G2262*) positively regulate fiber length and fiber strength formation and negatively lint percentage.

Conclusion: Multiple analyses demonstrate that the hub genes harbored in the QTLs orchestrate the fiber development. The hub gene *GhCesA4* has opposite pleiotropic effects in regulating trait formation of fiber quality and yield. The results facilitate understanding the genetic basis of negative correlation between cotton fiber quality and yield.

Introduction

Allotetraploid cotton, which produces a large amount of natural fiber for textile industry, is one of the economic crops widely planted in the world [1,2]. It includes two cultivated species, *Gossypium hirsutum*, which shares 90 % of total fiber production, features broader adaptability, high yield potential, and fibers of attractive quality, and *G. barbadense*, which has exceptionally high-quality fibers, occupying less than 10 % of fiber yield in production [3]. Thus, developing high-yield and high-quality fiber of *G. hirsutum* cultivars plays a critical role in cotton production. However cotton fiber quality and yield traits are genetically complex, which are controlled by multi-genes. Previous studies reveal that fiber quality and yield are negatively correlated [4,5]. During breeding improvement practices, this negative correlation between traits will lead to a negative selection effect (trade-off effect) of a trait when the improvement of another trait is imposed [6]. Under current agricultural conditions, both cotton breeding and production in major cotton producing countries have reached a high level of practice. These factors render it a challenge for the breeders to further improve cotton fiber quality and yield simultaneously. Quantitative trait loci (QTLs) mapping of these complex quantitative traits, which has been applied to the genetic architecture dissection of crops, was proven effective in marker assisted selection (MAS) in cotton breeding [7]. Following the improvement of sequencing technology and the reduction in cost, high quality drafts of cotton genomes, which includes *G. raimondii*, *G. arboretum* (two diploid species), and *G. hirsutum*, *G. barbadense* (two allotetraploid species), had been assembled, which could further accelerate cotton biotechnological improvement [3,8–14]. In addition, large collections of cotton accessions had been genotyped by high-throughput re-sequencing strategies or by CottonSNP63K or CottonSNP80K arrays [15,16]. The important QTLs and candidate genes of fiber quality and yield traits were thus identified by GWAS analysis of natural populations or by linkage analysis based on high-density genetic map (HDGM) of segregating populations using the low-cost and large-scale single nucleotide polymorphism (SNP) genotyping strategies [15–20].

Improving and producing high quality fibers have always been the primary tasks of cotton breeders and other participants in cotton breeding and production system [2,20]. Based on the morphological characteristics of developing fibers, previous studies had assorted cotton fiber development into four stages, which includes initiation (0–3 days post anthesis, DPA), elongation (3–20 DPA), secondary cell wall (SCW) biosynthesis (15–40 DPA), and maturation (40–50 DPA) [21,22]. A large number of genes are thought to be involved in the complex process of cotton fiber development and formation, and some of these genes may also affect genetic diversity in fiber quality across cultivars and species in *Gossypium* [2,3,23]. Recent studies revealed that complex multi-group signal pathways, including plant hormones [24], plant transcription factor family [25], lipids [26], chitinase-like (CTL) protein [27], α-expansin [28], aquaporins [29], ROS signal transduction [30], may involve in fiber development.

In this study, a population consisting of 231 recombinant inbred lines (231 RILs) was developed from a cross between Lumianyan 28 (LMY28) of high yield cultivar and Xinluzao 24 (XLZ24) a high fiber quality cultivar. The fiber quality traits, which include fiber length (FL), fiber strength (FS) and fiber micronaire (FM), and fiber yield traits, which include boll weight (BW) and lint percentage (LP), of the RILs were investigated across nine environments, while seed index (SI) seven environments. After the population was genotyped by SLAF-seq strategy, together with the previous genotyping result of it by CottonSNP80K array [17], an HDGM was constructed aiming to identify QTLs of fiber quality and fiber yield traits. Eleven candidate genes were identified through weighted gene correlation network analysis (WGCNA), which was conducted using the genes in all stable QTLs and QTL clusters. The relative expression profiles of these candidate genes in RNA-seq data and qRT-PCR validation during fiber development revealed that they played orchestrating roles in the formation of fiber quality and yield traits. Thus, the mechanism of how these QTLs regulate the formation of fiber yield and quality, as well as their integrating and restricting relationships were analyzed. Silencing via virus-induced gene silencing (VIGS) strategy and editing via clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system of a hub gene *cellulose synthase 4* (*GhCesA4*, *GH_D07G2262*) revealed that it has a pleiotropic effect and is involved in the formation of LP, as well as FL and FS during cotton fiber development. The results will be of great significance for candidate gene cloning and functional analysis, as well as for dissecting how candidate genes coordinate cotton fiber development through interactions.

Materials and methods

Plant materials and phenotype evaluation

The experimental materials included two parental *G. hirsutum* cultivars, LMY28 and XLZ24, as well as a RIL population consisting of 231 RILs derived from a cross of LMY28 × XLZ24, which was described in a previous study [17]. Briefly the cross was made in Anyang in 2008 summer. F_2 harvested 237 individual plants, and $F_{2:3}$ harvested 231 lines and from thence, they were proliferated via self-pollinating. From F_2 to $F_{2:6}$, they were alternatively planted in the growing seasons in Sanya, Hainan and Anyang, Henan respectively. At $F_{2:6}$, a single plant was selected from each line to

compose $F_{6:7}$ generation. $F_{6:8}$ and later generations were considered as a RIL population.

From 2013 to 2016, the field phenotypes of FL, FS, FM, BW, and LP were evaluated under nine environments, which included four in Anyang (13AY, 14AY, 15AY, and 16AY), two in Linqing (13LQ and 14LQ), and three in Quzhou (13QZ), Kuerle (14KEL), and Alaer (15ALE). SI was evaluated in seven environments, same as abovementioned except 14AY and 14LQ. The field experiment design and management, and phenotype evaluation were as described in previous study [17]. Briefly, thirty naturally opened bolls from each line were hand harvested and the phenotype evaluation of all the target traits was based on these bolls. The phenotypic evaluation procedures briefly involved air-drying, weighing seed cotton, ginning, and weighing lint and cottonseeds. BW refers to the average single boll weight of the 30 bolls, SI the weight of 100 seeds, and LP the percentage of lint in seed cotton weight. FL, FS and FM were measured using HFT9000 (Premier Evolvics Pvt. Ltd., India) instruments with HVICC Calibration.

DNA extraction and library construction and genotyping

Total genomic DNA samples of XLZ24, LMY28, and 231 RILs were extracted using MiniBEST Plant Genomic DNA Extraction kit (Code: 9768, Takara, Beijing). Two strategies were applied to genotype the experimental materials, SLAF-seq strategy and CHIP strategy. The CHIP strategy was performed with cottonSNP80K array, which consisted of 77,774 SNPs [16,17]. The genotyping procedure was initially delineated in a previous study [17], briefly the genotyping was performed following the protocols provided in the manufacturer's instruction (Illumina Inc., San Diego, CA, United States) [15] and the raw data were reanalyzed following the procedure described in a previous report [16]. The SLAF-seq strategy was performed following the procedures described in previous studies [31,32] on the Illumina HiSeq platform. Briefly, the reference genome of *G. hirsutum* (TM-1) [3] was used *in silico* to simulate the digestion results to select appropriate restriction endonucleases or their combination to generate optimum SLAF marker distribution across *G. hirsutum* genome. A combination of two endonucleases, *Hae*III and *Ssp*I (New England Biolabs, NEB, USA), was selected to digest the genomic DNA of the experimental materials to construct Illumina libraries. SLAF-seq was performed on the Illumina HiSeq platform. The identification and genotyping of SLAF markers were based on Sun et al. [31] and Ali et al. [33]. Raw data were filtered according to the criteria of minimum read depth less than 10 fold in parental lines and less than 1 fold in RILs to obtain clean reads. The clean reads were aligned to TM-1 reference genome [3] by BWA software [34]. After SNPs and Indels were detected using software GATK (Genome Analysis Toolkit) [35]. They were further filtered to generate SLAF markers according to the following standards: 1) SNPs and Indels that had more than 30 % missing data and 2) segregation distortion that reached $p < 0.005$ significant level in RILs were removed.

HDGM construction and QTL mapping

The HDGM was constructed by Lep-Map3 software [36] using Kosambi mapping function [37]. QTLs of the target traits were identified using the composite interval mapping (CIM) method through the Windows QTL Cartographer 2.5 software [38]. A QTL was declared to exist if where the logarithm of odds (LOD) reached the threshold of $p \leq 0.05$ significance level after 1000 permutations [17]. For a target trait, QTLs identified in different environments with fully or partially overlapping confidence intervals were considered the same QTL. QTLs that were identified across at least two environments were regarded as stable QTLs [17]. If a QTL has positive additive effect, it means that the favorable allele of that QTL is contributed by XLZ24, while negative additive effect by LMY28. MapChart 2.2 [39] was utilized to display the genetic linkage map and the QTL positions on the map.

To compare QTLs of current study with those of previous studies, the physical interval of the target QTLs reported in previous studies were realigned to the TM-1 reference genome [3,40]. If the physical marker interval of a QTL partially or fully overlapped with that of a reported QTL in previous studies, the QTL was regarded as a common one. Previous studies of both linkage analysis of segregating populations and GWAS of natural populations were used to compare the QTL results of current study.

Gene identification and annotation

The physical interval of a stable QTL or QTL cluster was determined based on the physical position of all markers of that QTL or QTL cluster on the TM-1 reference genome [3]. All genes in the physical intervals of stable QTLs and QTL clusters were fetched from the CottonFGD database (https://cottonfgd.org/). All gene annotation information was retrieved from Cotton Omics Database (https://cotton.zju.edu.cn/).

RNA-Seq dataset collection

RNA-Seq data of developing fiber samples of sGK156 (P1), 901–001 (P2), MBZ70-053 (L1), and MBZ70-236 (L2), which were sampled on 5, 10, 15, 20, 25, 30 days post anthesis (DPA), had been reported previously [41]. L1 and L2 are two lines of RIL population developed from a cross of P1 × P2. The fiber quality, especially FL and FS, of P2 and L1 were better than that of P1 and L2.

Identification of co-expression modules during fiber development

Weighted gene correlation network analysis (WGCNA) package in R software [42] was utilized to identify the gene co-expression modules relating to fiber development. Initially, all genes within the physical intervals of stable QTLs and QTL clusters were filtrated to obtain their expression profiles according to the criterion of average FPKM ≥ 0.5 for further analyses. Then the modular scale independence and mean connectivity with varying power values were analyzed. When R^2 value reached 0.9, the optimal soft threshold was determined for module analysis, followed by setting the optimal soft threshold, minimum module size, and merge cut height to 8, 30, and 0.25, respectively in stepwise. The modules, which had a highly significant correlation to a fiber development stage, were selected to perform intra-modular analysis to identify both gene significance (GS) and module membership (MM) in the modules of interest. GS represents the significance of gene expression correlation with the fiber development stage, while MM represents the correlation between the eigengene of each module and the module itself. Subsequently, a scatter plot was created to depict the correlation between GS and MM in the given modules.

Enrichment analysis of protein–protein interaction (PPI) network and screen of candidate genes

To identify candidate genes for fiber development, the following steps were performed: 1) the modules, which had a correlation coefficient > 0.80 at significant level of p_value < 0.01, were selected; 2) The genes, which had GS $\geq$ 0.56 and MM $\geq$ 0.93 in the significant modules, were selected; 3) Analyses of gene ontology (GO) and PPI network were performed with Metascape (https://metascape.org/).

Analysis of real-time quantitative polymerase chain reaction (qRT-PCR)

Cotton developing fibers of XLZ24 and LMY28 were sampled from 5, 10, 15, 20, 25, 30 DPA ovules with tweezers on ice. Then, total RNA samples were isolated from aforementioned fiber samples with the RNAprep pure plant plus kit (for samples abundance of polysaccharides & polyphenolics) (DP441, TianGen Biotech, Beijing, China). cDNAs were synthesized utilizing 1 μg total RNA samples and the HiScript III RT superMix for qPCR (+gDNA wiper) reverse transcription (R323-01AA, Vazyme, Biotech, Nanjing, China). The qRT-PCR analyses were performed on an applied biosystems 7500 fast real-time PCR system (ABI) utilizing the chamQ universal SYBR qPCR master mix (Q711-02-AA, Vazyme, Biotech, Nanjing, China). An internal control (*actin* gene) was utilized to normalize the variance of samples. The relative expression level was determined utilizing the $2^{-\Delta\Delta CT}$ method [43], and the analysis was performed on three independent biological replications. All primer sequences are presented in Supplementary Table S1.

VIGS and CRISPR/Cas9 verification of GH_D07G2262

VIGS experiment was conducted basically following protocols described in previous study [44]. A 300-bp fragment of CDs of *GH_D07G2262*, spanning from +2662 to +2961 bp from the start codon, was cloned into the *pCLCrV-A* vector using restriction endonucleases *Spe*I and *Asc*I. The vector *pCLCrV-GH_D07G2262* and the helper vector *pCLCrV-B* were transformed into LBA4404, an *Agrobacterium tumefaciens* strain. The transformed *Agrobacterium* was propagated and reactivated in YEB medium, and was collected in infection buffer, which contained 10 mM 2-(N-morpholine)-ethanesulphonic acid (MES), 10 mM $MgCl_2$ and 0.2 mM acetosyringone, pH 5.8, When the cotyledons of cotton seedlings of TM-1 were fully unfolded, *Agrobacterium* was injected into the two cotyledons. The seedlings injected with *pCLCrV-CHLI500* were used as positive control. Then, the injected seedlings were kept in darkness for 24–36 h, and were grown in greenhouse (25 ℃, 16-h light/8-h dark). At boll setting stage, newly unfolded leaves were sampled to evaluate *GH_D07G2262* expression level. Naturally opened bolls were sampled to evaluated FM and FL phenotypes.

Two single-guide RNA (sgRNA) sequences were designed from the sequence of *GH_D07G2262* and cloned into CRISPR/Cas9 vector using the primer sequences presented in Supplementary Table S1. The CRISPR/Cas9 vectors were then transformed into GV3101, an *A. tumefaciens* strain. The cotton transformation was conducted following previous descriptions [45], and the resulting putative transgenic plants were removed to a greenhouse with a 16-hour light/8-hour dark cycle at 30℃. Following selection based on kanamycin test and transgene PCR amplification test, the genomic DNA and total RNA of developing fibers of positive plants were extracted. The editing target site was cloned and sequenced, and *GH_D07G2262* cDNA of developing fiber was synthesized and cloned and sequenced to confirm the CRISPR/Cas9 system editing result. The fiber quality phenotypes were evaluated in T_2 generation of the edited transgenic cotton plants.

Paraffin cross-sectioning of natural mature fibers

To observe the secondary cell wall of fiber cells, dried mature fibers were treated and observed following the protocol [46] with some modifications. Briefly, the fiber samples were fixed by formaldehyde-acetic acid–ethanol (FAA) for 24 h, and then placed in a dehydration chamber for dehydration treatment with a series of different concentrations of 75 %, 85 %, 90 %, 95 %, 100 %, and 100 % ethanol, with each step treated for 4 h, 2 h, 1.5 h, 1.5 h, 30 min-1 h, 30 min-1 h, respectively, and then washed three times with ethanol + xylene (1:1), xylene, and xylene for 10–20 min in each step. After the pretreatment, the samples were embedded in paraffin three times in step-wise for 1–2 h in each step. After the paraffin solidified, the excessive paraffin was removed and trimmed. The cross sections of the samples were sliced to a thickness of 4 μm, dried, and stored at room temperature. The cross-sections of fiber cells were stained with toluidine blue for 2–5 min, and finally observed under a microscope Nikon Eclipse E100 (Nikon, Japan) and scanned using Pannoramic MIDI Ⅱ (3DHISTECH, Japan). Cell wall thickness of cotton fibers was analyzed with SlideViewer software provided by the manufacturer along with Pannoramic MIDI.

Results and analysis

Phenotypic assessment of traits of fiber quality and yield

The correlation analyses of the traits of fiber quality and yield were preliminary reported in a previous study (Table S2) [17]. The correlations analyses showed that the trait pairs of FL-FS, FS-SI and FL-SI had significant or extremely significant positive correlations across all environments; FS-FM and FS-LP had extremely significant negative correlations; FM-LP, BW-SI and FM-BW had extremely significant positive correlations in most environments; LP-SI, LP-BW and FM-SI had significant or extremely significant negative correlations in most environments.

Construction of HDGM

An HDGM containing 8045 markers, including 2934 SLAF-markers and 5111 CHIP-markers (CottonSNP80K array), was constructed, which spans 4943.01 cM with an average marker interval of 0.61 cM (Fig. 1a; Table 1; Table S3). A_t sub-genome consisted of 5929 markers, including 2336 SLAF-markers and 3593 CHIP-markers, which spans 2904.16 cM with an average marker interval of 0.49 cM. D_t sub-genome consisted of 2116 markers, including 598 SLAF-markers and 1518 CHIP-markers, which covers 2038.85 cM with an average marker interval of 0.96 cM. The longest chromosome was A12, which spanned 303.00 cM and consisted of 758 markers with an average interval of 0.40 cM. The shortest chromosome was D03, which spanned 105.41 cM and contained 27 markers with an average interval of 3.90 cM. The chromosome that contained the largest number of markers was A13, which contained 1006 markers with coverage of 267.58 cM, and an average interval of 0.27 cM. The chromosome that contained the least number of markers was D03, which contained 27 markers with coverage of 105.41 cM and an average interval of 3.90 cM. The largest linkage gap was 41.02 cM on D07. The number of linkage gaps ($\geq$20 cM) on D_t sub-genome was more than that on A_t sub-genome. Collinearity analysis showed that the absolute value of the spearman coefficient of 26 chromosomes was at least 0.83, suggesting that the linkage map and the cotton reference genome (TM-1) had relatively high collinearity (Fig. 1b; Table 1).

QTLs and QTL clusters

A total of 295 QTLs of traits of fiber quality (140 QTLs) and yield (155 QTLs) (Table S4) were identified on 26 chromosomes, explaining 3.38–19.41 % of phenotype variance (PV). Of them, 156 QTLs had positive additive effect on PV and 139 QTLs negative additive effects, indicating that both XLZ24 and LMY28 contributed favorable alleles to increase the trait PV of fiber quality and yield respectively. A total of 139 stable QTLs, including 75 fiber quality QTLs

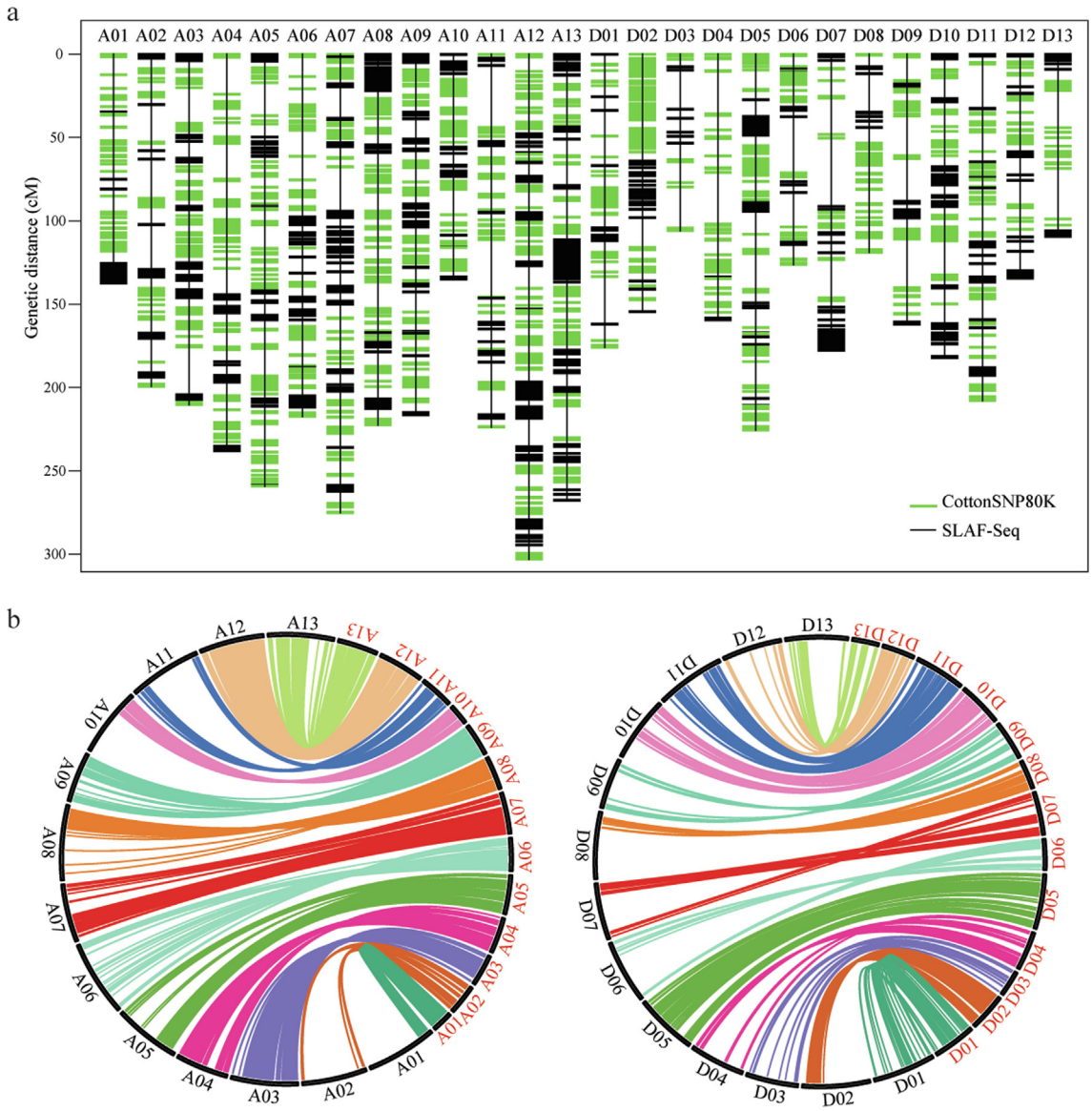


Fig. 1. High density genetic-linkage map (HDGM) construction and evaluation. a. diagram of the map. b. collinearity between the genetic map (black font) and the physical map (red font) of A_t and D_t sub-genomes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(16 of FL, 26 of FS, and 33 of FM) and 64 yield QTLs (29 of BW, 18 of LP, and 17 of SI), were identified simultaneously in at least two environments on 26 chromosomes, which could explain 3.38–13.18 % of PV (Table S5; Fig. 2). Two chromosomes A07 and A12 harbored the largest number of stable QTLs, twelve and eleven respectively, which could explain 3.38–11.42 % of PV. Among these stable QTLs, 69 had positive additive effect, indicating that XLZ24 contributed favorable alleles to improve fiber quality and yield; and 70 had negative additive effect indicating that LMY28 contributed favorable alleles to improve fiber quality and yield. In addition, a total of 201 QTLs, including 92 fiber quality QTLs and 109 yield QTLs, were identified in A_t sub-genome, which was more than twice the number of 94 QTLs identified in the D_t sub-genome, including 48 fiber quality and 46 yield QTLs. Similarly, 88 stable QTLs, including 46 fiber quality traits and 42 yield traits, in A_t sub-genome were also nearly twice as many as 51 stable QTLs, including 29 fiber quality QTLs and 22 yield QTLs, identified in the D_t sub-genome.

Among all QTLs identified in current study, 214 QTLs were in 74 QTL clusters, which were distributed on 25 chromosomes (excluding D09), 51 of the clusters were in A_t sub-genome and 23 of them in D_t sub-genome (Table S6). A07 harbored six QTL clusters, which was the largest number of QTL clusters in a chromosome, followed by A02, A04, A08, A12, and D10, each of which harbored five QTL clusters. Thirty-three of the 74 clusters (Table S7) harbored at least two stable QTLs of different traits, which were regarded as QTL cluster hotspots and were distributed on 20 chromosomes (excluding A03, D03, D06, D08-D09, and D13). Four of the thirty-three QTL clusters exclusively harbored stable QTLs of yield trait (BW, LP, and SI) on A02, A04, A10, and A13. Six of the thirty-three QTL clusters exclusively harbored stable QTLs for fiber quality (FL, FS, and FM) traits on A05, A08, D01, D10, and D12. Twenty-three of the thirty-three QTL clusters harbored stable QTLs of both fiber quality (FL or FS or FM) and yield (BW or LP or SI) traits on A01-A02, A05-A13, D02, D04-D05, D07, and D10-D11. Nine of the twenty-three QTL clusters harbored at least three stable QTLs. The cluster on

Table 1
Detailed information of HDGM of LMY28 × XLZ24 RIL population.

Chr	Genetic length (cM)	No. makers			Average marker interval (cM)	Collinearity coefficient	No. > 20 cM gaps	Largest gap (cM)
		Total	SLAF-seq	Cotton80K array				
A01	137.43	307	78	229	0.45	0.96	–	10.77
A02	199.20	100	31	69	1.99	−0.91	3	26.75
A03	210.24	516	210	306	0.41	0.86	2	28.78
A04	237.87	264	36	228	0.90	0.83	2	22.12
A05	259.19	484	139	345	0.54	−0.85	–	19.81
A06	217.32	263	65	198	0.83	0.93	–	15.64
A07	275.13	532	305	227	0.52	−0.89	1	26.47
A08	222.41	849	303	546	0.26	0.91	–	15.88
A09	216.33	410	143	267	0.53	0.92	–	12.14
A10	134.89	199	88	111	0.68	0.95	–	15.17
A11	223.58	241	38	203	0.93	−0.88	2	37.24
A12	303.00	758	247	511	0.40	−0.86	–	17.09
A13	267.58	1006	653	353	0.27	−0.93	–	19.05
A_t	2904.16	5929	2336	3593	0.49	–	10	–
D01	175.87	137	13	124	1.28	0.93	1	28.49
D02	154.40	365	55	310	0.42	0.87	–	17.82
D03	105.41	27	8	19	3.90	−0.95	2	24.79
D04	159.41	110	15	95	1.45	0.83	2	30.90
D05	225.53	315	71	244	0.72	0.96	1	21.34
D06	126.18	99	13	86	1.27	0.96	2	33.79
D07	177.62	157	103	54	1.13	0.83	3	41.02
D08	118.99	115	13	102	1.03	0.90	1	23.17
D09	161.93	142	46	96	1.14	0.94	3	29.38
D10	182.07	253	148	105	0.72	−0.89	1	20.57
D11	207.80	231	62	169	0.90	−0.94	1	26.75
D12	134.39	87	29	58	1.54	0.87	–	14.69
D13	109.27	78	22	56	1.40	0.88	2	30.29
D_t	2038.85	2116	598	1518	0.96	–	19	–
Total	4943.01	8045	2934	5111	0.61	–	29	–

A07, Clu-A07-1, harbored three stable QTLs of fiber quality (FL or FS or FM) and two stable QTLs of yield trait (BW and SI).

Qtls common to those of previous studies

Upon comparing the stable QTLs identified in the current study with those from previous studies, the results revealed that 93 stable QTLs were common to previous ones, and that 46 stable QTLs were newly identified QTLs. In the 93 stable QTLs, 39 ones, including seven of FL, nine of FS, sixteen of FM, two of BW, two of LP, and three of SI, were common to those in previous linkage analysis of segregating populations (Table S8 with references); 69 ones, including seven of FL, fourteen of FS, nine of FM, thirteen of BW, fifteen of LP, and eleven of SI, were common to those in previous GWAS of natural populations (Table S9 with references). Seventeen of them were common to those of both linkage analysis and GWAS.

In addition, comparing with expression QTL (eQTL) hotspots of fiber development in previous reports [47] revealed that 39 stable QTLs overlap with the eQTL hotspots (Table S10), of which 26 are the abovementioned common QTLs. Five of them, qFL-D03-1, qFL-D05-3, qFS-D10-2, qFM-D03-1, and qBW-A05-1, were simultaneously identified in both linkage analysis and GWAS.

Construction and identification of the gene co-expression modules of stable QTLs and clusters in fiber development

To dissect the potential acting mechanism of the functional genes during the trait formation of cotton fiber quality and yield, a total of 15,167 genes were fetched from the physical region of confidence interval of the stable QTLs and QTL clusters in TM-1 reference genome [3]. Filtration of these genes under the criterion of average FPKM $\geq$ 0.5 resulted in 6275 effectively expressed genes (EEGs) during fiber development. WGCNA of these EEGs generated 14 distinct gene co-expression modules (Fig. 3a; Fig. S1). The high correlation between the module eigengenes and sample of different fiber developing stage indicated that the module of MEpurple displayed a significant positive correlation with 10 DPA, MEskyblue with 15 DPA, MEgreen with 25 DPA, and MEblack with 30 DPA (Fig. 3a). The scatter plots of GS versus MM of above selected modules were presented in Fig. 3b. At the condition of threshold of GS $\geq$ 0.56 and MM $\geq$ 0.92, a total of 650 genes were screened from the above modules with significant correlation with fiber developing stages.

Screening for candidate genes

GO enrichment analysis of the 650 genes revealed that plant-type secondary cell wall biogenesis (GO:0009834), fatty acid biosynthetic process (GO:0006633), and very long-chain fatty acid biosynthetic process (GO:0042761) were enriched (Fig. 3c). PPI analysis of these 650 genes identified 9 hub nodes in the network, of which five were in plant-type secondary cell wall biogenesis and four were in fatty acid biosynthetic process (Fig. 3d). These nine hub nodes corresponded to eleven *G. hirsutum* homologous genes, which were regarded as putative candidate genes orchestrating the fiber development in this study (Table 2). The qRT-PCR result of these eleven candidate genes in XLZ24 and LMY28 showed that except for *GH_A03G0921*, all other genes had differential expression at least at an important stage during fiber development (Fig. 4).

Silencing verification of GH_D07G2262 function in cotton

GH_D07G2262 gene was identified in the QTL cluster Clu-D07-2, which consists of multiple QTLs of traits of both fiber quality and yield. Its physical interval was from 53,125,583 bp to 54,911,218 bp, harbored 134 genes. *GH_D07G2262* (*CesA4*) gene belongs to cellulose synthase gene superfamily in *Gossypium* species (Fig. S2). To investigate the function *GH_D07G2262* gene

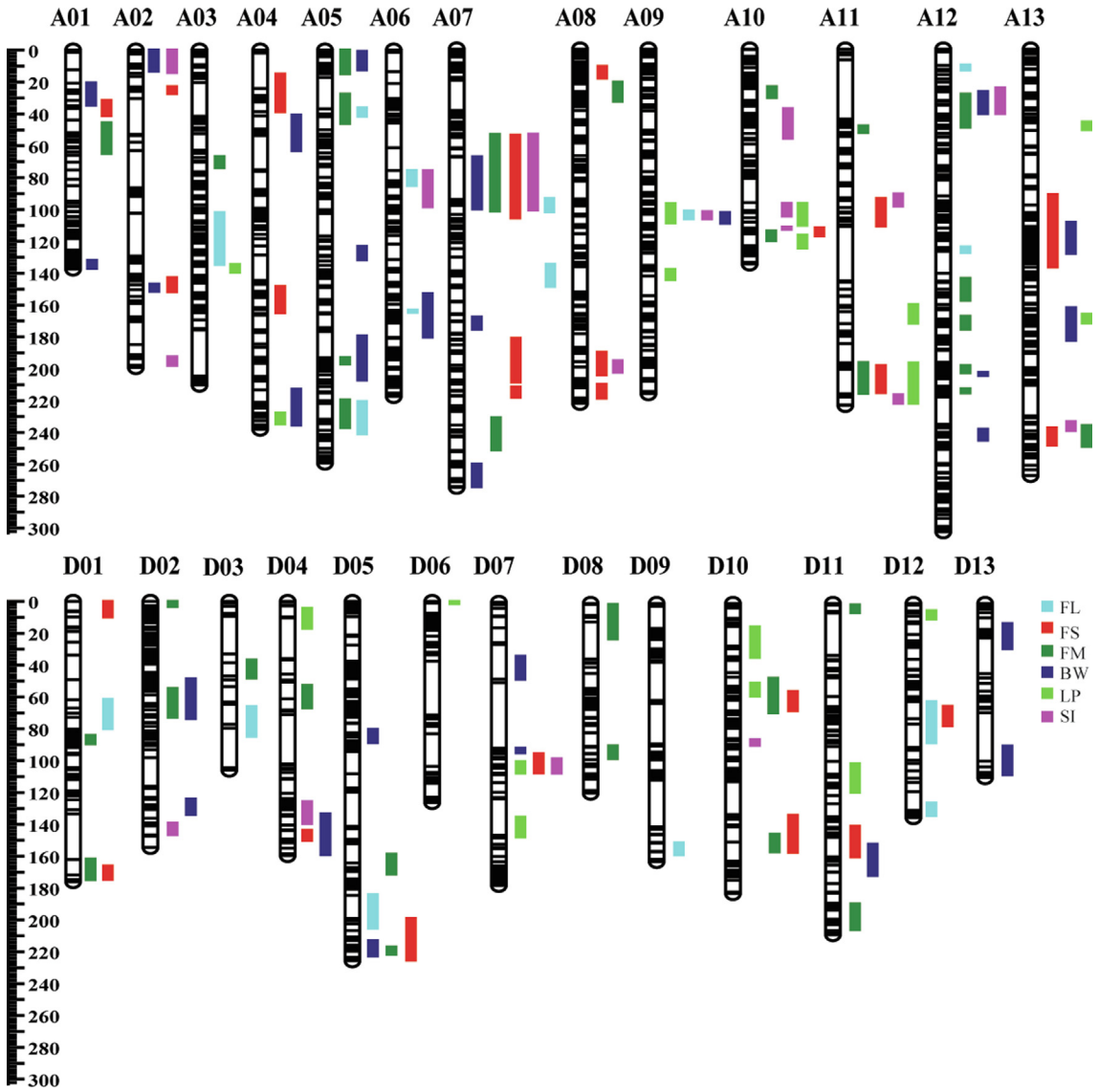


Fig. 2. Identification and distribution of stable QTLs of fiber quality and yield traits on 26 chromosomes.

in cotton fiber development, both VIGS and CRISPR/Cas9 strategies were employed to silence it using upland cultivars TM-1 and Jin668 respectively (Fig. 5). Analysis of the reverse transcription quantitative polymerase chain reaction (RT-qPCR) data revealed that the expression of *GH_D07G2262* was significantly lower in VIGS plants than in control TM-1 plants (Fig. 5a). When cotton bolls naturally mature, they were harvested and FL of the bolls was measured (Fig. 5b-c). The results revealed that average FL of silenced plants were 27.06 mm and 28.18 mm, significantly shorter than that of the control plant (30.79 mm), suggesting that silencing of *GH_D07G2262* significantly inhibited cotton fiber elongation. The CRISPR/Cas9 editing obtained three mutants, all of which were detected the expected mutations guided by sgRNA at the target location of *GH_D07G2262* (Fig. 5d). The observed results revealed that the FL and FS of all T_1 transgenic knockout plants were significantly shorter and weaker than those of control Jin668 plants (Table S11). The FL and FS of knockout plants were 25.00–27.90 mm and 27.20–29.70 cN·tex^{-1}, significantly shorter and weaker than those of WT, which were 28.21 mm (FL) and 30.40 cN·tex^{-1} (FS), respectively. In T_2 homologous *GH_D07G2262*-knockout lines, Both FL and FS were significantly lower and weaker than those of WT lines (Fig. 5e and 5f). Observation of cross-sections of fiber cells revealed that the average cell wall thickness WT, and three edited lines, *GhCesA4*-1, *GhCesA4*-2, and *GhCesA4*-3 was 4.43, 3.26, 3.53, and 3.13 μm, respectively. Statistical analysis revealed that CRISPR/Cas9 editing of *GH_D07G2262* significantly reduced fiber cell wall thickness (Fig. 5g and 5 h). Interestingly, LP of knockout lines were significantly increased than that of WT (Fig. 5f), suggesting that *GH_D07G2262* may have pleiotropic functions during cotton fiber development.

Discussion

HDGM construction and QTL identification

High-throughput sequencing and genotyping strategies are an effective approach to evaluate the genetic diversities or to dissect the genetic makeup of important traits of a given population [16]. In cotton, most of QTLs of fiber quality and yield traits were detected through analyses of linkage and GWAS based on segregation populations [5,6,48,49] and nature accessions [47,50–52],

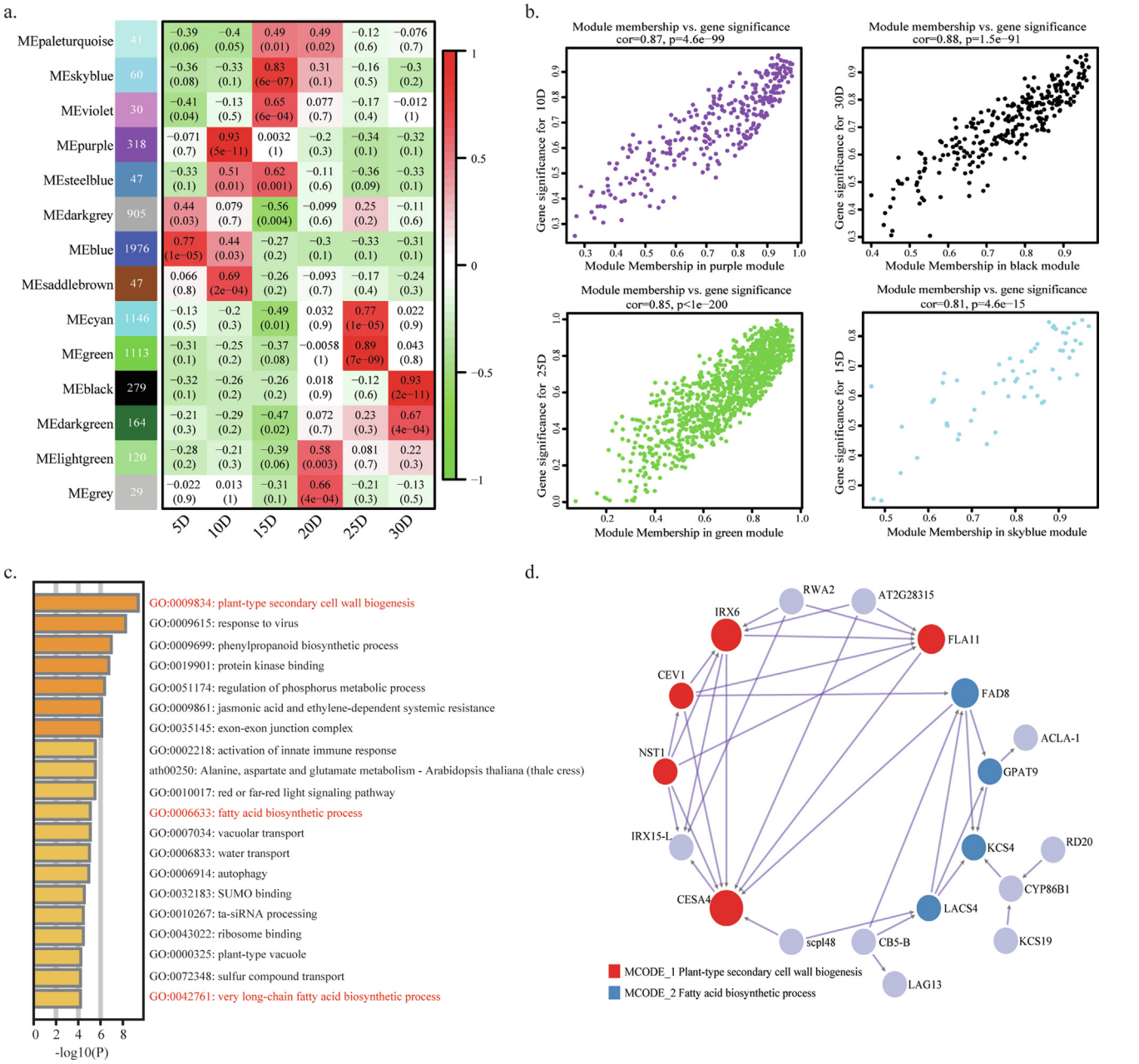


Fig. 3. WGCNA of genes in stable QTL and QTL clusters. a. The heat map of correlations between the modules and sample of different fiber developmental stages. Modules are displayed on the left, and each column corresponds to a fiber developmental stage. The number in each cell represents the pearson correlation coefficient, the *p*_value of the corresponding module-trait is exhibited in parentheses. The color of cells indicates the degree of correlation; b. Scatterplot of correlations between module membership (MM) and gene significance (GS) of the cells of MEpurple/MEgreen/MEblack/MEskyblue showing high correlation between module and fiber developmental stage. The correlation coefficient between MM and GS and *p*_value is listed on top of each scatterplot. One dot represents one gene in the plot. c. The enrichment analysis of 650 genes (GS $\geq$ 0.56 and MM $\geq$ 0.92) in above four modules. d. Protein-protein interaction network and MCODE (molecular complex detection) components of GO terms (GO:0009834: plant-type secondary cell wall biogenesis GO:0006633: fatty acid biosynthetic process).

respectively. To fully identify the QTLs of traits relating to fiber quality and yield harbored in the RILs of the present study, an HDGM was constructed incorporating previous genotyping results through CottonSNP80K array strategy [17] and molecular markers de novo developed through SLAF-Seq strategy. Comparison of the two maps revealed that the HDGM of current study had better coverage of and a much more improved collinearity to the physical reference genome map of upland cotton than the previous one [17] (Fig. 1). Compared with the previous QTL identification results in this population, the present study detected a total of 139 stable QTLs, of which 62 were detected *de novo* and the remaining 77 were duplicated with the previous results of CottonSNP80K array strategy [17]. Of these 77 QTLs, 39 were also stable ones while 38 were detected in a single environment in previous results. These findings confirmed that the improved HDGM has higher value and reliability in linkage analysis and QTL identification.

In reviewing the previous reported literatures, it was observed that 272 of the total 295 QTLs identified in this population, including 130 of the 139 stable ones, were common to those in previous reports and nine stable QTLs were unique loci in this population [17]. When comparing the current results of this population with previous analyses of both linkage and associations, it was observed

Table 2
Eleven candidate genes associated with fiber development.

MCODE	Modules	Gene IDs		Abbr.	Gene annotation in *A. thaliana*
		G. hirsutum	*A. thaliana*		
Plant-type secondary cell wall biogenesis	Skyblue	*GH_A11G3703*	*AT5G03170*	*FLA11*	FASCICLIN-like arabinogalactan-protein 11
		GH_D03G1117	*AT5G15630*	*IRX6*	COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family
		GH_A11G1050	*AT2G46770*	*NST1*	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
	Green	*GH_D03G0812*	*AT5G05170*	*CEV1*	Cellulose synthase family protein
		GH_A07G2317	*AT5G44030*	*CesA4*	cellulose synthase A4
		GH_D07G2262	*AT5G44030*	*CesA4*	cellulose synthase A4
Fatty acid biosynthetic process		*GH_A04G1111*	*AT5G05580*	*FAD8*	fatty acid desaturase 8
		GH_A03G0921	*AT1G19440*	*KCS4*	3-ketoacyl-CoA synthase 4
		GH_A05G1932	*AT1G19440*	*KCS4*	3-ketoacyl-CoA synthase 4
		GH_D03G0938	*AT4G23850*	*LACS4*	AMP-dependent synthetase and ligase family protein
	Black	*GH_A04G1519*	*AT5G60620*	*GPAT9*	glycerol-3-phosphate acyltransferase 9

Genetic linkage analysis of stable QTLs in *Gossypium hirsutum* RIL population revealed function of *GhCesA4* in fiber development.

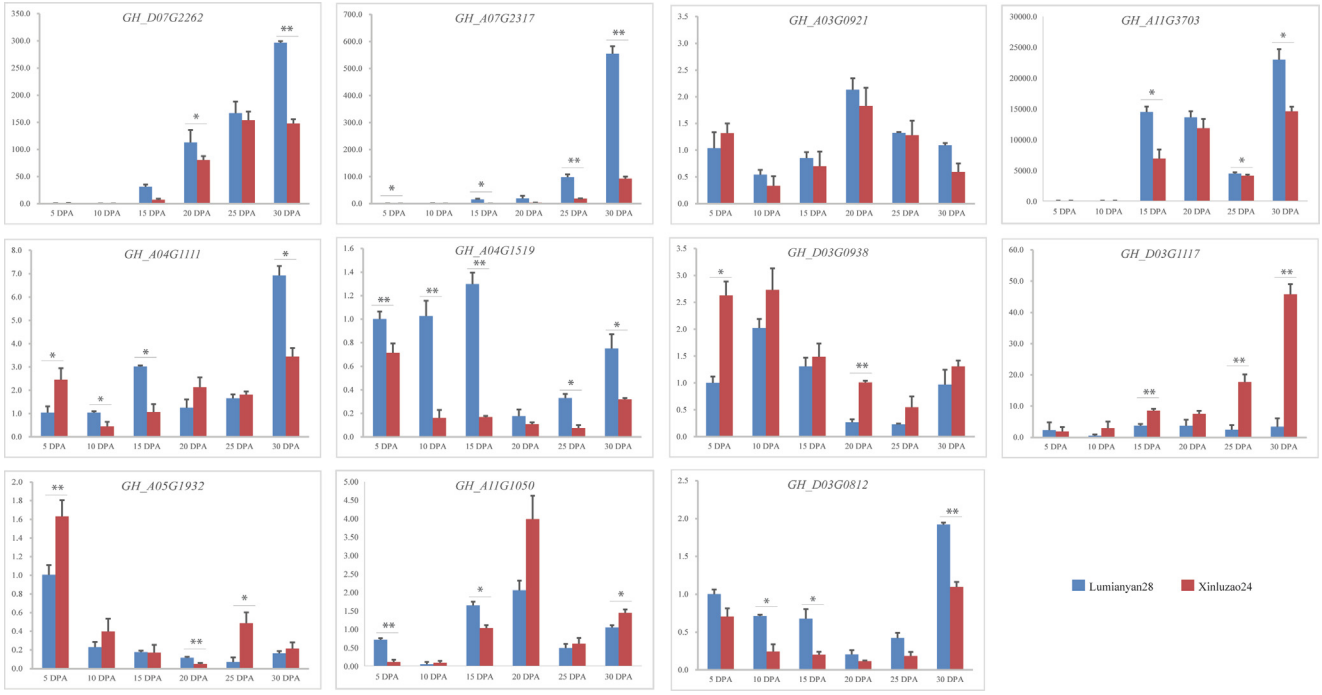


Fig. 4. Quantitative RT-PCR of the expression of 11 candidate genes during fiber development. Fiber sample at 5 DPA of L28 was used as control for all tests.

that 39 stable QTLs, including 32 of fiber quality traits and seven of yield traits, were common to previous linkage analysis of segregating populations (Table S8), and 69 stable QTLs, including 30 of fiber quality traits and 39 of yield traits, were found in GWAS of natural populations (Table S9). Seventeen of the above stable QTLs were simultaneously identified both in segregating and natural populations. An eQTL mapping study [47] revealed that 39 stable QTLs of traits relating fiber quality and yield of present study were expressed, 26 of which were also identified in previous segregation or nature populations. These QTLs common to previous studies suggest the reliability of QTL identification of current study and that they may play important roles in regulating the trait formation of fiber quality or yield. The QTLs specific in the population of current study may suggest new insight in the genetic architecture that underlies the trait formation of fiber quality and yield. These results may also imply a complex genetic mechanism regulating fiber development. One QTL cluster Clu-A07-1, for example, which harbored a stable QTL of FL, FS, FM, BW, and SI [6,53], was identified two candidate genes *Gh_A07G1749* (leucine-rich repeat protein kinase family protein; LRR RLK) and *GH_A07G2179* (*GhSI7*; transcriptional regulator STERILE APETALA) by fine mapping in *G. hirsutum* intraspecific F_2 population [53,54].

QTL effect direction of each trait in QTL clusters and simultaneous improvement of the clustered traits

The pairwise correlations between traits in multi-trait studies were frequently analyzed in previous segregation population reports [6,17]. The correlation analysis (Table S2) indicated that the trait pairs of FL-FS, SI-FL/FS, FM-LP/BW, and BW-SI were positively correlated at significant or extremely significant level and that the trait pairs of FM-FL/FS, LP-FL/FS, FM-SI, BW-LP and SI-LP were negatively correlated at significant or extremely significant level in most of the experimental environments, while the trait pair of FS-BW showed extremely significant positive correlations in some environments and extremely significant negative correlations in other environments. The current study summarized 74 QTL clusters among multi-traits based on the confidence intervals and

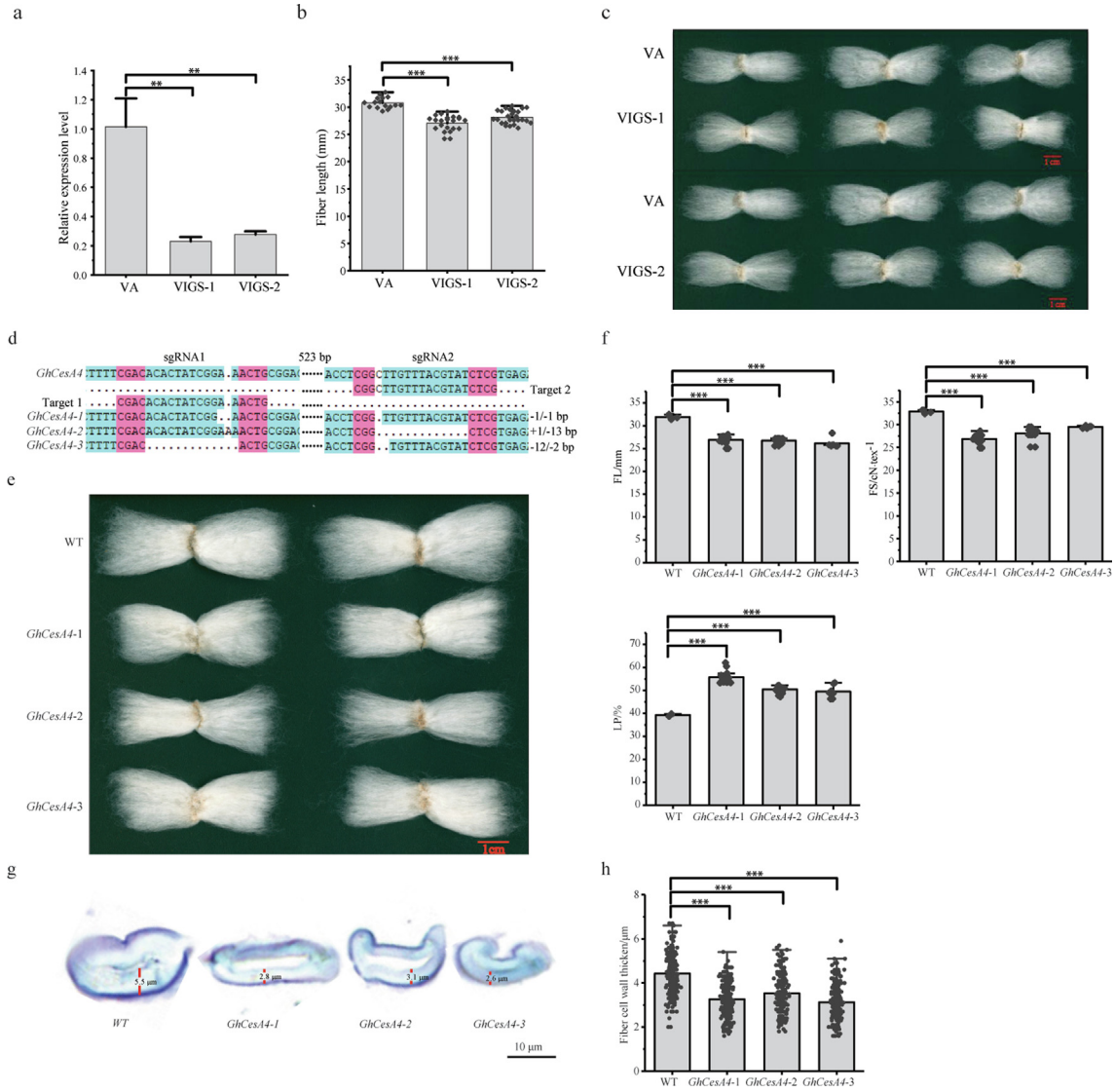


Fig. 5. *GH_D07G2262* positively regulates cotton fiber elongation but negatively regulates lint percentage in knockout experiments. a. The relative expression level of *GH_D07G2262* in control, VIGS-1 and VIGS-2 plants. b. Fiber length (FL, mm) of control and VIGS plants. c. Mature fiber of control and VIGS plants. d. Sanger sequencing-based detection of *GH_D07G2262*-knockout plants obtained after CRISPR/Cas9 gene editing. Nucleotide deletions at the target site are represented by dots. e. Phenotypes of fiber samples of *GH_D07G2262*-knockout and wild type (WT) plants. f. Analysis of variance of FL, fiber strength (FS, cN·tex^{-1}), and lint percentage (LP, %) of WT and CRISPR/Cas9 edited lines. g. Cell wall thickness of mature fibers of the WT and CRISPR/Cas9 edited lines. h. Statistical analysis of the mean values of cell wall thickness of mature fibers. Error bars indicate the standard errors of biological replicates. *** *P_value* < 0.001.

their corresponding physical intervals of these multi-trait QTLs (Fig. 2). In linkage analysis of a segregation population, the effect direction of a QTL represents which of the two parents of the population provides the favorable allele of the locus for the trait. The QTL effect direction of the traits in a QTL cluster corresponds to the phenotypic correlation direction between traits. In the current study, 33 of 74 QTL clusters harbored at least two stable QTLs (Tables S6, S7). Among the positively correlated trait pairs in each of the clusters, the favorable alleles at this locus for these traits are all provided by the same parent, such as FL-FS, FL/FS-SI, FM-LP/BW (Tables S6, S7). These positively correlated traits usually share same additive effect direction, which infers that they can be improved synchronously via simply selecting the common markers in the clusters of these traits through MAS. However, balancing of the yield and quality was always a problem in crop genetic improvement. Between the negatively correlated fiber quality and yield traits, it is always difficult to further improve one without allowing a certain trade-off of the other [6,55–57]. In the present study, the results revealed that 25 of the total 74 QTL clusters harbored two stable QTLs, in which 14 clusters had stable QTL pairs with opposite additive effects, while nine clusters harbored at least three stable QTLs (≥3), in which there were 6 stable QTL pairs, namely, LP-FL/FS, FM-FL/FS, and SI-FM/LP, had opposite additive effects. These cluster loci are the keys to resolve the trade-off effects between the negatively correlated traits. Some studies reckoned that the balancing of complex trade-off was caused by gene pleiotropy and linkage drags [55]. A recent study suggested that the QTL clusters formed by the negatively correlated traits and that in the clusters, the alleles favorable for the trait development contributed by different parents, can explain the trade-off effect [6], which makes it difficult to achieve simultaneous improvement of both traits [55–57]. However, there are currently two possible gene acting mechanisms involved in the trade-off effect, namely linkage drags and gene pleiotropy. In such case, it is necessary to first conduct further genetic dissect of the cluster to clarify which of the two mechanisms is at work, and then use corresponding strategies

to improve the two traits based on the mechanism at work. For linkage drags, via finely mapping the linked causal genes and then recombining them through conventional breeding strategies, the problem of linkage drags can be solved. While for gene pleiotropy, it can be solved by editing *cis*-regulatory regions (CRRs) via a tiling-deletion-based CRISPR/Cas9 screening strategy [55]. Researchers have found that variations on CRRs can regulate the gene expression at different levels [58–61] and thus regulating the quantitative traits. Accumulating the amount of genetic variations in CRRs and screening the favorable variations at different levels would achieve the optimum effects of overcoming trade-off among quantitative traits. However, it still remains unclear which of the two mechanisms, linkage drag or pleiotropic genes, is at work in these clusters, and further dissection is needed to clarify this issue.

GhCesA4 may regulate cotton fiber development

Previous study indicated that plant homomeric cellulose synthase (CesA) trimers comprising *CesA4*/*7*/*8* are in charge of the biosynthesis of the secondary cell walls (Fig. 6) [62]. These homomeric CesA trimers may form complexes (CSCs) to perform their function. After assembled in the Golgi apparatus, CSCs are subsequently transported to the plasma membrane (PM), where they catalyze cellulose synthesis [63] and assemble them into microfibrils of the secondary cell wall (SCW) [64]. Integrity of CSCs is very important for cellulose biosynthesis, mutation of any member of the complexes may leads to rapid CesA proteasome degradation [65]. Knock-out of any monomer of *GhCesA 4*, *7*, and *8* will result in the loss-of-function of the cellulose synthase supercomplex (CSS) and the production of fiber cells without SCW thickening [64]. In rice, different CSCs comprising different *OsCesA* members may be responsible for biosynthesis of primary cell wall or SCW [66]. The heterotrimeric complex comprising OsCesA4/7/9 is responsible for synthesizing rice SCW cellulose [67,68]. While in *Arabidopsis*, *AtCesA4*, *7*, *8* are highly co-expressed in regulating cellulose synthesis of SCW [69,70]. Existing studies have elaborated the mechanisms of regulating CSCs and the action modes of CSCs in the synthesis of cellulose in SCWs. Some results demonstrate that the wall thickening activator NACs could differentially binds at the promoter region of *CesA4* [69]. GhKNL1, a class Ⅱ KNOX protein, which is reckoned to be a transcription factor regulating plant SCW formation, could directly bind to the promoter elements of *GhCesA4-2*/*4–4*/*8–2* and *GhMYB46* to catalyze SCW cellulose synthesis during cotton fiber development [71]. A study identified a

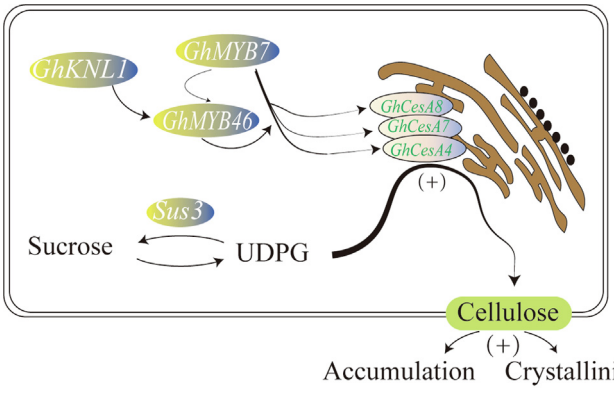


Fig. 6. The regulatory schematic model of *CesAs* in the synthesis and accumulation of cellulose in secondary cell walls during fiber development. The cellulose synthase complexes (CSCs) formed from homomeric CesA trimers are assembled in the Golgi apparatus; subsequently they are transported to the plasma membrane, where they catalyze cellulose synthesis. The assembly of celluloses into microfibrils of the secondary cell wall is affected by both cellulose accumulation and crystallinity.

rice mutant of *OsCesA4*, *fc17*, which significantly reduces cellulose crystallinity, and thus enhances biomass saccharification and lodging resistance [72] (Fig. 6). Another mutant *Bc19* of *OsCesA4* is regarded as a semi-dominant brittle allele, which affects cellulose synthesis in a dosage-dependent manner [73]. *AtCesA8* promoter can drive overexpression of *OsSUS3* to regulate carbon partitioning between cellulose and other polysaccharide biosynthesis [74]. Increased wall thickness may coincide with reduced cellulose crystallinity [74]. Meanwhile, a sucrose-free environment accelerated the relocation of PM-localized GFP–CesA3 towards the periphery of the Golgi apparatus through the clathrin-rich *trans*-Golgi network [75], mediated by clathrin-dependent mechanisms.

Cotton fiber is a highly specialized single cell that develops from the epidermal layer of the ovule after fully elongated and its secondary cell wall thickened, which assembles a large quantity of cellulose [64]. A recent study identified homozygous *GhCesA4* (A_t: *GH_A08G0515*, D_t: *GH_D08G0525*) mutants via CRISPR/Cas9 editing, which displayed significant reduction in both fiber yield and FL (mid panel) in comparison with those of control plants (Fig. S2) [64]. During cotton fiber SCW synthesis, *GhCesA4*, *7*, and *8* assemble into heteromers in 36-*mer*-like CSS [64], which infers that the pathway and modulation of cellulose synthesis of SCW in developing cotton fiber cells may adopt similar mechanisms. In this study, we identified two new *GhCesA4* members, *GH_A07G2317* and *GH_D07G2262*, within the physical intervals of Clu-A07-1 and Clu-D07-2 respectively, which was identified via genetic linkage analysis (Fig. 2). Silencing *GH_D07G2262* via both VIGS technology of TM-1 and CRISPR/Cas9 editing of Jin668 resulted shorter FL and higher LP than those of wild control (Fig. 5). In a previous study, it was observed that when the homologous pair of *GhCesA4*, *GH_A08G0515* and *GH_D08G0525*, were edited, the secondary cell wall of fibers was damaged [64]. Here we demonstrated that editing of *GH_D07G2262* reduced the secondary cell wall thickness of cotton fiber, implying that the homologous pair of *GhCesA4* might additively co-regulate secondary cell wall thickening during fiber development.

We also identified some other candidate genes in the stable QTLs or clusters, including *FLA11* (*fasciclin*-like Arabinogalactan protein 11, *GH_A11G3703* in Clu-A11-3), *IRX6* (*irregular xylem6*, *GH_D03G1117* in Clu-D03-1), *CEV1* (*constitutive expression of the vegetative storage protein1*, *GH_D03G0812* in Clu-D03-1), and *NST1* (*NAC secondary wall thickening promoting factor1*, *GH_A11G1050* in qFM-A11-1), which are correlated to plant-type SCW biogenesis, form hub nodes in their protein interaction network (Fig. 3d). How these hub node genes orchestrate the cellulose biosynthesizes and the development of fiber quality is still open to elucidation, however, previous studies revealed the involvement of them in cellulose biosynthesis in SCW. The *cev1* is testified as an allele of *AtCesA3* and mutations of *cev1* allele lead to less cellulose deposition in its roots than wild-type roots [76]. *AtCesA3* needs to be assembled into functional CSCs to catalyze cellulose in plant SCWs [77]. Arabinogalactan proteins (AGPs) are a class of glycoproteins that are present throughout the plant kingdom and play a crucial role in various aspects of plant development [78]. Fasciclin-like AGPs (FLAs) are an important class of AGPs [79], which maintain conservative fasciclin functional domains, suggesting their functional conservations in fundamental aspects of embryogenesis and seed development [80], including fiber initiation and elongation [78]. The expression of *FLA11* and *FLA12* is SCW-specific and highly correlated to cellulose content and microfibril angle [81].

The *irx6* mutant was identified to be *COBL4* gene defect, which exhibits reduced cellulose content [82], In willow, *SxCOBL4* is involved in cellulose microfibril deposition [83], it may regulate the structure of cell walls via enhancing cellulose content and preserving cellulose crystallinity [84], which infers that fiber quality is

dependent to not only cellulose deposition but also cellulose crystallinity in cotton. The current study also identified that Fatty acids (FAs) biosynthesis pathway plays a vital role in orchestrating cotton fiber development, of which the hub node genes *included long-chain acyl-coA synthetase4* (*LACS4*, *GH_D03G0938* in Clu-D03-1), *fatty acid denaturase8* (*FAD8*, *GH_A04G1111* in Clu-A04-3), *ketoacyl-CoA synthase4* (*KCS4*, *GH_A03G0921* in Clu-A03-1), and *glycerol-3-phosphate acyltransferase9* (*GPAT9*, *GH_A04G1519* in Clu-A04-5) (Fig. 3). The Acyl-CoA synthetase (ACS) family plays a crucial role in both FAs synthesis and degradation, and long-chain ACSs (LACSs) comprise a small subgroup of this family [85]. Ketoacyl-CoA synthase (KCS) is a rate-limiting enzyme required for the synthesis of very-long-chain fatty acids (VLCFA), and several members of this enzyme family have been found to participate in VLCFA synthesis [86–88]. Modulating biosynthesis of VLCFA biosynthesis can affect fiber elongation [86–89].

These findings may provide new insights in dissecting candidate gene functions in modulating SCW cellulose biosynthesis during fiber development. Further studies and verification of their specific functions will have great interests and significance in understanding their orchestrating mechanism of fiber quality formation.

Ethical statement

This article does not contain any studies with human or animal materials.

CRediT authorship contribution statement

Ruìxián Liú: Investigation, Software. **Xiànghuī Xiāo:** Investigation, Methodology, Visualization. **Jw Gōng:** Investigation, Formal analysis, Visualization. **Jùnwén Lǐ:** Investigation. **Hàoliàng Yán:** Investigation, Formal analysis. **Qún Gě:** . **Quánwěi Lú:** . **Péngtāo Lǐ:** . **Jīngtāo Pān:** . **Hǎihóng Shāng:** . **Yùzhēn Shí:** Methodology. **Qúanjiā Chén:** Supervision, Investigation, Methodology. **Yǒulù Yuán:** Supervision, Funding acquisition, Resources. **Wànkuí Gǒng:** Supervision.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

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), China Agriculture Research System—Cotton (CARS-15-02), the National Agricultural Science and Technology Innovation Project for Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2016-ICR), Central Public-interest Scientific Institution Basal Research Fund (CN) (1610162023013), Project of Kashgar Regional Science and Technology Plan (KS2023003), and the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2021D01B114), Talent Development Program for Innovation and Expansion in Xinjiang for Youlu Yuan Group. The authors acknowledge Sci Lang for its linguistic assistance during the preparation of this manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2023.12.005.

References

[1] Ma Z, Zhang Y, Wu L, Zhang G, Sun Z, Li Z, et al. High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement. Nat Genet 2021;53(9):1385–91.
[2] 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(5):516–24.
[3] 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(4):739–48.
[4] Paterson AH, Saranga Y, Menz M, Jiang CX, Wright RJ. QTL analysis of genotype x environment interactions affecting cotton fiber quality. Theor Appl Genet 2003;106(3):384–96.
[5] Gu Q, Ke H, Liu Z, Lv X, Sun Z, Zhang M, et al. A high-density genetic map and multiple environmental tests reveal novel quantitative trait loci and candidate genes for fibre quality and yield in cotton. Theor Appl Genet 2020;133(12):3395–408.
[6] 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(1):239–53.
[7] Huang C, Nie X, Shen C, You C, Li W, Zhao W, et al. Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. Plant Biotechnol J 2017;15(11):1374–86.
[8] Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, et al. Toward sequencing cotton (*Gossypium*) genomes. Plant Physiol 2007;145(4):1303–10.
[9] Wang K, Wang Z, Li F, Ye W, Wang J, Song G, et al. The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet 2012;44(10):1098–103.
[10] 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(7429):423–7.
[11] Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet 2014;46(6):567–72.
[12] 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(5):524–30.
[13] 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(5):531–7.
[14] Wang M, Tu L, Yuan D, Zhu De, Shen C, Li J, et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat Genet 2019;51(2):224–9.
[15] Hulse-Kemp AM, Lemm J, Plieske J, Ashrafi H, Buyyarapu R, Fang DD, et al. Development of a 63K SNP array for cotton and high-density mapping of intraspecific and interspecific populations of *Gossypium* spp. G3 (Bethesda) 2015;5(6):1187–209.
[16] Cai C, Zhu G, Zhang T, Guo W. High-density 80 K SNP array is a powerful tool for genotyping *G. hirsutum accessions* and genome analysis. BMC Genomics 2017;18:654.
[17] Liu R, Gong J, Xiao X, Zhang Z, Li J, Liu A, et al. GWAS analysis and QTL identification of fiber quality traits and yield components in upland cotton using enriched high-density SNP markers. Front Plant Sci 2018;9:1067.
[18] Tan Z, Zhang Z, Sun X, Li Q, Sun Y, Yang P, et al. Genetic map construction and fiber quality QTL mapping using the CottonSNP80K array in upland cotton. Front Plant Sci 2018;9:225.
[19] Zhang K, Kuraparthy V, Fang H, Zhu L, Sood S, Jones DC. High-density linkage map construction and QTL analyses for fiber quality, yield and morphological traits using CottonSNP63K array in upland cotton (*Gossypium hirsutum* L.). BMC Genomics 2019;20:889.
[20] Zhang Z, Ge Q, Liu A, Li J, Gong J, Shang H, et al. Construction of a high-density genetic map and its application to QTL identification for fiber strength in upland cotton. Crop Sci 2017;57(2):774–88.
[21] Haigler CH, Betancur L, Stiff MR, Tuttle JR. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Front Plant Sci 2012;3:104.
[22] Mei H, Qi B, Han Z, Zhao T, Guo M, Han J, et al. Subgenome bias and temporal postponement of gene expression contributes to the distinctions of fiber quality in *Gossypium* species. Front Plant Sci 2021;12:819679.
[23] Jan M, Liu Z, Guo C, Sun X. Molecular regulation of cotton fiber development: a review. Int J Mol Sci 2022;23(9):5004.
[24] Xiao G, Zhao P, Zhang Y. A pivotal role of hormones in regulating cotton fiber development. Front Plant Sci 2019;10:87.
[25] Sun W, Gao Z, Wang J, Huang Y, Chen Y, Li J, et al. Cotton fiber elongation requires the transcription factor *GhMYB212* to regulate sucrose transportation into expanding fibers. New Phytol 2019;222(2):864–81.

[26] Shi Z, Chen X, Xue H, Jia T, Meng F, Liu Y, et al. *GhBZR3* suppresses cotton fiber elongation by inhibiting very-long-chain fatty acid biosynthesis. Plant J 2022;111(3):785–99.
[27] Zhang D, Hrmova M, Wan C-H, Wu C, Balzen J, Cai W, et al. Members of a new group of chitinase-like genes are expressed preferentially in cotton cells with secondary walls. Plant Mol Biol 2004;54(3):353–72.
[28] Li Y, Tu L, Pettolino FA, Ji S, Hao J, Yuan D, et al. *GbEXPATR*, a species-specific expansin, enhances cotton fibre elongation through cell wall restructuring. Plant Biotechnol J 2016;14(3):951–63.
[29] Naoumkina M, Thyssen GN, Fang DD. RNA-seq analysis of short fiber mutants Ligon-lintless-1 (Li 1) and -2 (Li 2) revealed important role of aquaporins in cotton (*Gossypium hirsutum* L.) fiber elongation. BMC Plant Biol 2015;15:65.
[30] Tang W, Tu L, Yang X, Tan J, Deng F, Hao J, et al. The calcium sensor *GhCaM7* promotes cotton fiber elongation by modulating reactive oxygen species (ROS) production. New Phytol 2014;202(2):509–20.
[31] Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, et al. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PLoS ONE 2013;8(3):e58700.
[32] Shen C, Jin X, Zhu D, Lin Z. Uncovering SNP and indel variations of tetraploid cottons by SLAF-seq. BMC Genomics 2017;18(1):247.
[33] Ali I, Teng Z, Bai Y, Yang Q, Hao Y, Hou J, et al. A high density SLAF-SNP genetic map and QTL detection for fibre quality traits in *Gossypium hirsutum*. BMC Genomics 2018;19(1):879.
[34] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60.
[35] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20(9):1297–303.
[36] Rastas P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. Bioinformatics 2017;33(23):3726–32.
[37] Kosambi DD. The estimation of map distance from recombination values. Ann Eugen 1944;12:172–5.
[38] Wang S, Basten C, Zeng Z. Windows QTL Cartographer v2.5. 2012.
[39] Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 2002;1:77–8.
[40] Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL NCBI BLAST a better web interface. Nucleic Acids Res 2008;36:W5-9.
[41] Jiang X, Gong J, Zhang J, Zhang Z, Shi Y, Li J, et al. Quantitative trait loci and transcriptome analysis reveal genetic basis of fiber quality traits in CCRI70 RIL population of *Gossypium hirsutum*. Front Plant Sci 2021;12:753755.
[42] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf 2008;9:559.
[43] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25(4):402–8.
[44] Gu Z, Huang C, Li F, Zhou X. A versatile system for functional analysis of genes and microRNAs in cotton. Plant Biotechnol J 2014;12(5):638–49.
[45] Manghwar H, Lindsey K, Zhang X, Jin S. CRISPR/Cas system: recent advances and future prospects for genome editing. Trends Plant Sci 2019;24(12):1102–25.
[46] Zhou L, Zhu T, Han S, Li S, Liu Y, Lin T, et al. Changes in the Histology of walnut (*Juglans regia* L.) infected with *Phomopsis capsici* and transcriptome and metabolome analysis. Int J Mol Sci 2023;24(5):4879.
[47] Li Z, Wang P, You C, Yu J, Zhang X, Yan F, et al. Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton. New Phytol 2020;226(6):1738–52.
[48] Keerio AA, Shen C, Nie Y, Ahmed MM, Zhang X, Lin Z. QTL mapping for fiber quality and yield traits based on introgression lines derived from *Gossypium hirsutum* × *G. tomentosu*m. Int J Mol Sci 2018;19(1):243.
[49] Yang P, Sun X, Liu X, Wang W, Hao Y, Chen L, et al. Identification of candidate genes for lint percentage and fiber quality through QTL mapping and transcriptome analysis in an allotetraploid interspecific cotton CSSLs population. Front Plant Sci 2022;13:882051.
[50] Wang F, Zhang J, Chen Y, Zhang C, Gong J, Song Z, et al. Identification of candidate genes for key fibre-related QTLs and derivation of favourable alleles in *Gossypium hirsutum* recombinant inbred lines with *G. barbadense* introgressions. Plant Biotechnol J 2020;18(3):707–20.
[51] Wang M, Qi Z, Thyssen GN, Naoumkina M, Jenkins JN, McCarty JC, et al. Genomic interrogation of a MAGIC population highlights genetic factors controlling fiber quality traits in cotton. Commun Biol 2022;5(1):60.
[52] Liu W, Song C, Ren Z, Zhang Z, Pei X, Liu Y, et al. Genome-wide association study reveals the genetic basis of fiber quality traits in upland cotton (*Gossypium hirsutum* L.). BMC Plant Biol 2020;20:395.
[53] Liu X, Hou J, Chen Li, Li Q, Fang X, Wang J, et al. Natural variation of *GhSI7* increases seed index in cotton. Theor Appl Genet 2022;135(10):3661–72.
[54] Fang X, Liu X, Wang X, Wang W, Liu D, Zhang J, et al. Fine-mapping qFS07.1 controlling fiber strength in upland cotton (*Gossypium hirsutum* L.). Theor Appl Genet 2017;130(4):795–806.
[55] Song X, Meng X, Guo H, Cheng Q, Jing Y, Chen M, et al. Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. Nat Biotechnol 2022;40(9):1403–11.
[56] Takatsuji H. Regulating tradeoffs to improve rice production. Front Plant Sci 2017;8:171.
[57] Wang L, Wang D, Yang Z, Jiang S, Qu J, He W, et al. Roles of FERONIA-like receptor genes in regulating grain size and quality in rice. Sci China Life Sci 2021;64(2):294–310.
[58] Liu L, Gallagher J, Arevalo ED, Chen R, Skopelitis T, Wu Q, et al. Enhancing grain-yield-related traits by CRISPR-Cas9 promoter editing of maize CLE genes. Nat Plants 2021;7(3):287–94.
[59] Hendelman A, Zebell S, Rodriguez-Leal D, Dukler N, Robitaille G, Wu X, et al. Conserved pleiotropy of an ancient plant homeobox gene uncovered by cis-regulatory dissection. Cell 2021;184(7):1724–1739.e16.
[60] Li X, Xie Y, Zhu Q, Liu YG. Targeted genome editing in genes and cis-regulatory regions improves qualitative and quantitative traits in crops. Mol Plant 2017;10(11):1368–70.
[61] Zhang S, Yu H, Wang K, Zheng Z, Liu L, Xu M, et al. Detection of major loci associated with the variation of 18 important agronomic traits between *Solanum pimpinellifolium* and cultivated tomatoes. Plant J 2018;95(2):312–23.
[62] Du J, Vandavasi VG, Molloy KR, Yang H, Massenburg LN, Singh A, et al. Evidence for plant-conserved region mediated trimeric CESAs in plant cellulose synthase complexes. Biomacromolecules 2022;23(9):3663–77.
[63] Polko JK, Kieber JJ. The regulation of cellulose biosynthesis in plants. Plant Cell 2019;31(2):282–96.
[64] Wen X, Zhai Y, Zhang Li, Chen Y, Zhu Z, Chen G, et al. Molecular studies of cellulose synthase supercomplex from cotton fiber reveal its unique biochemical properties. Sci China Life Sci 2022;65(9):1776–93.
[65] Li F, Xie G, Huang J, Zhang R, Li Y, Zhang M, et al. *OsCESA9* conserved-site mutation leads to largely enhanced plant lodging resistance and biomass enzymatic saccharification by reducing cellulose DP and crystallinity in rice. Plant Biotechnol J 2017;15(9):1093–104.
[66] Wang Y, Fan C, Hu H, Li Y, Sun D, Wang Y, et al. Genetic modification of plant cell walls to enhance biomass yield and biofuel production in bioenergy crops. Biotechnol Adv 2016;34(5):997–1017.
[67] Zhang J, Liu Z, Sakamoto S, Mitsuda N, Ren A, Persson S, et al. ETHYLENE RESPONSE FACTOR 34 promotes secondary cell wall thickening and strength of rice peduncles. Plant Physiol 2022;190(3):1806–20.
[68] Tanaka K, Murata K, Yamazaki M, Onosato K, Miyao A, Hirochika H. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. Plant Physiol 2003;133:73–83.
[69] Olins JR, Lin L, Lee SJ, Trabucco GM, MacKinnon KJM, Hazen SP. Secondary wall regulating NACs differentially bind at the promoter at a CELLULOSE SYNTHASE A4 Cis-eQTL. Front Plant Sci 2018;9:1985.
[70] Guo B, Huang X, Qi J, Sun H, Lv C, Wang F, et al. *Brittle culm 3*, encoding a cellulose synthase subunit 5, is required for cell wall biosynthesis in barley (*Hordeum vulgare* L.). Front Plant Sci 2022;13:989406.
[71] Wang Y, Li Y, Gong S-Y, Qin L-X, Nie X-Y, Liu D, et al. *GhKNL1* controls fiber elongation and secondary cell wall synthesis by repressing its downstream genes in cotton (*Gossypium hirsutum*). J Integr Plant Biol 2022;64(1):39–55.
[72] Li F, Liu S, Xu H, Xu Q. A novel FC17/CESA4 mutation causes increased biomass saccharification and lodging resistance by remodeling cell wall in rice. Biotechnol Biofuels 2018;11:298.
[73] Ma X, Li C, Huang R, Zhang K, Wang Q, Fu C, et al. Rice *Brittle Culm19* encoding cellulose synthase subunit CESA4 causes dominant brittle phenotype but has no distinct influence on growth and grain yield. Rice (N Y) 2021;14:95.
[74] Fan C, Feng S, Huang J, Wang Y, Wu L, Li X, et al. *AtCesA8*-driven *OsSUS3* expression leads to largely enhanced biomass saccharification and lodging resistance by distinctively altering lignocellulose features in rice. Biotechnol Biofuels 2017;10:221.
[75] Fujimoto M, Suda Y, Vernhettes S, Nakano A, Ueda T. Phosphatidylinositol 3-kinase and 4-kinase have distinct roles in intracellular trafficking of cellulose synthase complexes in *Arabidopsis thaliana*. Plant Cell Physiol 2015;56(2):287–98.
[76] Ellis C, Karafyllidis I, Wasternack C, Turner JG. The Arabidopsis mutant cev1 links cell wall signaling to jasmonate and ethylene responses. Plant Cell 2002;14(7):1557–66.
[77] Qiao Z, Lampugnani ER, Yan X-F, Khan GA, Saw WG, Hannah P, et al. Structure of *Arabidopsis* CESA3 catalytic domain with its substrate UDP-glucose provides insight into the mechanism of cellulose synthesis. Proc Natl Acad Sci U S A 2021;118(11):e2024015118.
[78] Huang GQ, Gong SY, Xu WL, Li W, Li P, Zhang CJ, et al. A fasciclin-like arabinogalactan protein, *GhFLA1*, is involved in fiber initiation and elongation of cotton. Plant Physiol 2013;161(3):1278–90.
[79] Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR. A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. Plant Physiol 2010;153(2):485–513.
[80] Costa M, Pereira AM, Pinto SC, Silva J, Pereira LG, Coimbra S. In silico and expression analyses of fasciclin-like arabinogalactan proteins reveal functional conservation during embryo and seed development. Plant Reprod 2019;32(4):353–70.
[81] Ma Y, MacMillan CP, de Vries L, Mansfield SD, Hao P, Ratcliffe J, et al. FLA11 and FLA12 glycoproteins fine-tune stem secondary wall properties in response to mechanical stresses. New Phytol 2022;233(4):1750–67.
[82] Sato K, Suzuki R, Nishikubo N, Takenouchi S, Ito S, Nakano Y, et al. Isolation of a novel cell wall architecture mutant of rice with defective *Arabidopsis COBL4* ortholog BC_1 required for regulated deposition of secondary cell wall components. Planta 2010;232(1):257–70.

[83] Gritsch C, Wan Y, Mitchell RA, Shewry PR, Hanley SJ, Karp A. G-fibre cell wall development in willow stems during tension wood induction. J Exp Bot 2015;66(20):6447–59.
[84] Sun X, Xiong H, Jiang C, Zhang D, Yang Z, Huang Y, et al. Natural variation of *DROT1* confers drought adaptation in upland rice. Nat Commun 2022;13:4265.
[85] Zhao H, Kosma DK, Lü S. Functional role of long-chain acyl-CoA synthetases in plant development and stress responses. Front Plant Sci 2021;12:640996.
[86] Qin YM, Zhu YX. How cotton fibers elongate: a tale of linear cell-growth mode. Curr Opin Plant Biol 2011;14(1):106–11.
[87] Qin YM, Hu CY, Pang Y, Kastaniotis AJ, Hiltunen JK, Zhu YX. Saturated very-long-chain fatty acids promote cotton fiber and Arabidopsis cell elongation by activating ethylene biosynthesis. Plant Cell 2007;19(11):3692–704.
[88] Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, et al. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. Plant Cell 2006;18(3):651–64.
[89] Yang Z, Liu Z, Ge X, Lu L, Qin W, Qanmber G, et al. Brassinosteroids regulate cotton fiber elongation by modulating very-long-chain fatty acid biosynthesis. Plant Cell 2023;35:2114–31.