

RESEARCH ARTICLE

Open Access



Genome-wide association study reveals the genetic basis of fiber quality traits in upland cotton (*Gossypium hirsutum* L.)

Wei Liu^{1†}, Chengxiang Song^{2†}, Zhongying Ren², Zhiqiang Zhang², Xiaoyu Pei², Yangai Liu², Kunlun He², Fei Zhang², Junjie Zhao², Jie Zhang³, Xingxing Wang^{2,3}, Daigang Yang^{2,3} and Wei Li^{1,2,3*}

Abstract

Background: Fiber quality is an important economic trait of cotton, and its improvement is a major goal of cotton breeding. To better understand the genetic mechanisms responsible for fiber quality traits, we conducted a genome-wide association study to identify and mine fiber-quality-related quantitative trait loci (QTLs) and genes.

Results: In total, 42 single nucleotide polymorphisms (SNPs) and 31 QTLs were identified as being significantly associated with five fiber quality traits. Twenty-five QTLs were identified in previous studies, and six novel QTLs were firstly identified in this study. In the QTL regions, 822 genes were identified and divided into four clusters based on their expression profiles. We also identified two pleiotropic SNPs. The SNP locus i52359Gb was associated with fiber elongation, strength, length and uniformity, while i11316Gh was associated with fiber strength and length. Moreover, these two SNPs were nonsynonymous and located in genes *Gh_D09G2376* and *Gh_D06G1908*, respectively. RT-qPCR analysis revealed that these two genes were preferentially expressed at one or more stages of cotton fiber development, which was consistent with the RNA-seq data. Thus, *Gh_D09G2376* and *Gh_D06G1908* may be involved in fiber developmental processes.

Conclusions: The findings of this study provide insights into the genetic bases of fiber quality traits, and the identified QTLs or genes may be applicable in cotton breeding to improve fiber quality.

Keywords: Upland cotton, Fiber quality, Genome-wide association study, Single nucleotide polymorphism, Quantitative trait locus

Background

Cotton (*Gossypium* spp.) is an important economic crop, and the major source of natural fibers for the textile industry worldwide [1]. Upland cotton (*Gossypium hirsutum* L.), an allotetraploid species, has the advantages of wide adaptability and high production. Consequently, it

is used to produce approximately 95% of the world's cotton fiber [2]. In the past few decades, cotton breeders have been mainly focused on improving cotton yield. At present, with improvements in the quality of consumers' lives and in textile technology, the demand for high-quality fiber is increasing [3, 4]. Thus, improving fiber quality has become a new target of cotton breeding. However, improving the fiber quality of upland cotton by conventional breeding has proven to be inefficient [5]. Therefore, it is important to clarify the genetic bases of upland cotton fiber quality traits, and molecular

* Correspondence: liweili@caas.cn

[†]Wei Liu and Chengxiang Song contributed equally to this work.

¹Collaborative Innovation Center of Henan Grain Crops, Agronomy College, Henan Agricultural University, Zhengzhou 450002, China

²State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

markers will play vital roles in the breeding of high-quality cotton.

The fiber quality traits of cotton mainly include fiber elongation (FE), fiber micronaire (FM), fiber strength (FS), fiber length (FL) and fiber uniformity (FU). These are all complex quantitative traits that are substantially affected by environmental factors, even though they are mainly controlled by genetic factors [5]. Before cotton genomes were published, cotton fiber quality traits were mainly studied using linkage analysis methods with biparental segregating populations. To date, nearly 1000 quantitative trait loci (QTLs) associated with fiber quality traits and distributed across the 26 cotton chromosomes have been reported using this mapping method [6, 7]. These studies provided important information for fiber-related genetics and accelerated the development of cotton breeding for fiber quality.

In recent years, with the release of cotton genome sequences [8–10] and the rapid development of molecular markers [11, 12], a large number of single nucleotide polymorphism (SNP) markers have been identified at the whole-genome level in cotton. Additionally, association mapping has been widely used to discover the genetic bases of complex traits in cotton [13–16]. Compared with traditional linkage mapping, association mapping can reveal the associations between genotypes and phenotypes using natural populations and simultaneously detect many natural allelic variations in a study [17]. Using this mapping method, many QTLs and candidate genes associated with fiber quality have been identified [18]. For example, Sun et al. (2017) performed a genome-wide association study (GWAS) of fiber quality traits using 719 diverse accessions of upland cotton and 10,511 polymorphic SNPs identified using the CottonSNP63K array. In total, they identified 46 significant SNPs associated with five fiber quality traits. Furthermore, a combined GWAS and transcriptome analysis revealed 19 promising genes related to FL and FS [19]. Huang et al. (2017) detected 79 significant SNPs associated with fiber quality traits using a natural population containing 503 *G. hirsutum* accessions through CottonSNP63K genotyping [20]. Moreover, the genes *GhXI-K*, *GhFL1* and *GhFL2* were also identified as being associated with fiber quality traits using the association mapping method [13, 21].

In our previous study, 276 upland cotton accessions were genotyped with the CottonSNP63K array and 10,660 high-quality SNPs were identified [22]. Here, we performed a GWAS to determine the correlations between genetic loci and fiber quality traits using these 276 diverse accessions and 10,660 high-quality SNPs. Moreover, we also mined for candidate genes underlying the fiber quality traits and assessed them by integrating RNA-seq and RT-qPCR analyses. Our results are useful for the breeding of improved fiber quality traits in cotton.

Results

Phenotypic characterization of fiber quality traits

The 276 diverse upland cotton accessions were obtained from different ecological zones, and the best linear unbiased predictions (BLUPs) of the phenotypic data exhibited abundant variation among accessions. FE, FM, FS, FL and FU exhibited values in the ranges of 6.53%–6.82%, 4.24–5.63, 25.08–31.69 cN/tex, 25.17–31.02 mm and 81.55%–85.41%, with average values of 6.70%, 5.05, 28.49 cN/tex, 28.85 mm and 84.35%, respectively (Table 1). All the fiber quality traits were normally distributed (Fig. 1), indicating that these traits were quantitative traits controlled by multiple genes. The coefficients of variation (CVs) of FE, FM, FS, FL and FU were 0.66%, 5.24%, 3.61%, 2.60% and 0.61%, respectively (Table 1).

Correlation analysis of the five fiber quality traits showed that highly significant correlations existed among the five fiber quality traits (Fig. 2). Significant positive correlations were observed among the four traits FE, FS, FL and FU, ranging from 0.51 to 0.84, whereas FM was significantly negatively correlated with FS and FL. In addition, ANOVA revealed the effects of genotype (G), environment (E) and the interaction of genotype and environment (G × E) were significant ($P < 0.001$) on fiber traits (Additional file 1: Table S1). This result suggested that fiber quality traits are affected by both genotype and environment. The broad-sense heritability (H^2) values of FE, FM, FS, FL and FU were 73.31%, 91.38%, 84.78%, 84.54% and 72.06%, respectively (Table 1), suggesting that fiber quality traits are mainly controlled by genetic effects.

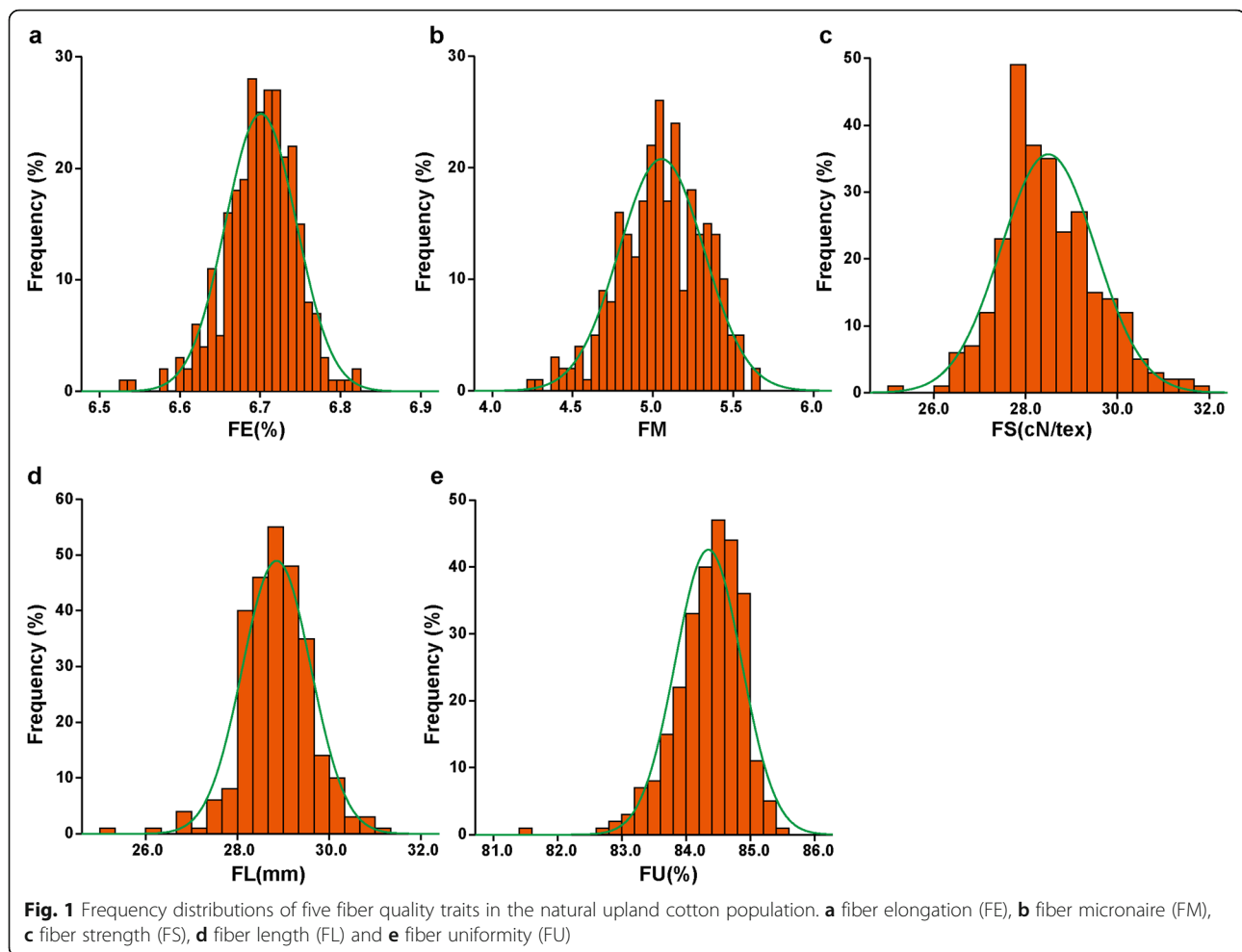
GWAS of fiber quality traits

To identify genetic factors underlying fiber quality traits, we performed a GWAS of the five fiber quality traits, combining 10,660 high-quality SNP markers identified using the CottonSNP63K array and phenotypic data collected from multiple environments. In total, 42 SNPs significantly associated with the five fiber quality traits were detected using the BLUP values (Fig. 3a, Additional file 2: Figure S1 and Additional file 1: Table S2). They were scattered across 13 chromosomes, including

Table 1 Phenotype statistics of fiber quality traits in the accession population

Trait	Min	Max	Mean	SD	CV(%)	H^2 (%)
FE(%)	6.53	6.82	6.70	0.04	0.66	73.31
FM	4.24	5.63	5.05	0.26	5.24	91.38
FS(cN/tex)	25.08	31.69	28.49	1.03	3.61	84.78
FL(mm)	25.17	31.02	28.85	0.75	2.60	84.54
FU(%)	81.55	85.41	84.35	0.52	0.61	72.06

FE Fiber elongation, FM Fiber micronaire, FS Fiber strength, FL Fiber length, FU Fiber uniformity, SD Standard deviation, CV Coefficient of variation, H^2 Broad-sense heritability



At05, At09, At10, At12, Dt01, Dt02, Dt05, Dt06, Dt08, Dt09, Dt10, Dt11 and Dt12. Moreover, the 9 and 33 significant SNPs were located in the At and Dt subgenomes, respectively.

For FE, 10 significant SNPs distributed on chromosomes At09, Dt02, Dt09, Dt11 and Dt12 were identified. The phenotypic variation explained (PVE) by each SNP ranged from 3.00% to 4.41% (Additional file 1: Table S2). Among these SNPs, locus i52264Gb had a positive effect on FE, with the highest $-\log_{10}(P)$ value (4.18), and i39571Gh had a negative effect on FE, with the lowest $-\log_{10}(P)$ value (3.02).

For FM, 11 significant SNPs were detected on chromosomes At10, Dt05, Dt09, Dt10, Dt11 and Dt12, explaining 3.06%–3.98% of the phenotypic variation (Additional file 1: Table S2). Of these SNPs, locus i35898Gh had a negative effect on FM, with the highest $-\log_{10}(P)$ value (3.73), and i09491Gh had a positive effect on FM, with the lowest $-\log_{10}(P)$ value (3.00).

For FS, eight significant SNPs located on chromosomes At12, Dt06, Dt08 and Dt09 were found. They explained 3.33%–4.65% of the phenotypic variation (Additional file 1: Table S2). Locus i52359Gb had a negative effect on FS, with the highest $-\log_{10}(P)$ value (4.03), while i30183Gh had a positive effect on FS, with the lowest $-\log_{10}(P)$ value (3.04).

For FL, six significant SNPs were observed on chromosomes At05, At10, Dt01, Dt06 and Dt09. They explained 3.31%–4.16% of the phenotypic variation (Additional file 1: Table S2). Locus i52359Gb had a negative effect on FL, with the highest $-\log_{10}(P)$ value (3.66), and i33845Gh had a positive effect on FL, with the lowest $-\log_{10}(P)$ value (3.02).

For FU, 11 significant SNPs were identified on chromosomes At10, Dt01, Dt02, Dt05 and Dt09, contributing 2.64%–4.19% of the phenotypic variation (Additional file 1: Table S2). Locus i09849Gh had a negative effect on FU, with the highest $-\log_{10}(P)$ value (4.42), and i65397Gm had a positive effect on FU, with the lowest $-\log_{10}(P)$ value (3.00).

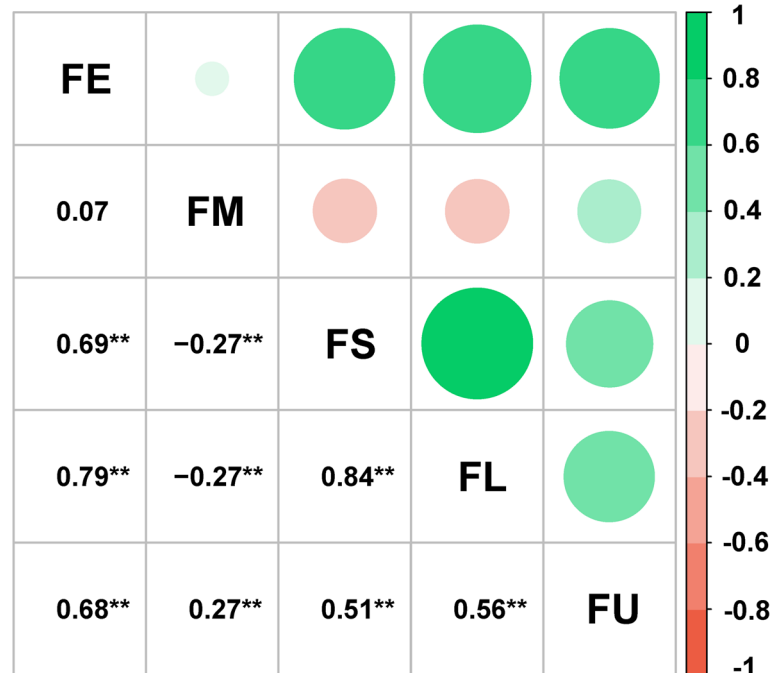


Fig. 2 Correlation analysis of five cotton fiber quality traits. FE, fiber elongation; FM, fiber micronaire; FS, fiber strength; FL, fiber length; FU, fiber uniformity. ** indicates that the correlation was significant at the $P < 0.01$ level

Identification and pleiotropy of QTLs

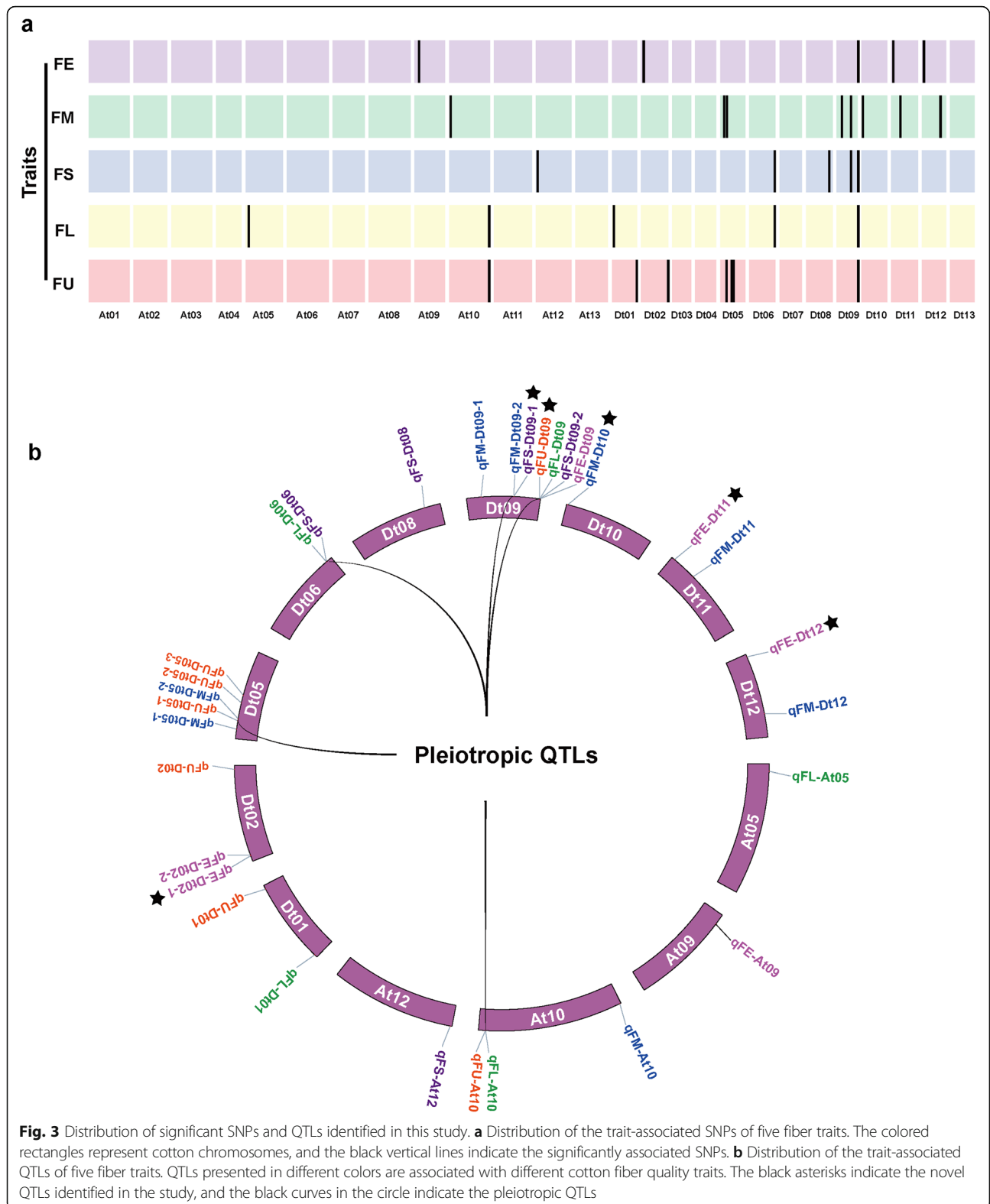
According to the definition of QTL from a previous study [19, 22], 31 QTLs were identified in this study: six for FE, eight for FM, five for FS, five for FL and seven for FU (Fig. 3b). These QTLs were distributed on different chromosomes that included significantly associated SNPs. Among these QTLs, six (*qFE-Dt02-1*, *qFE-Dt11*, *qFE-Dt12*, *qFM-Dt10*, *qFS-Dt09-1* and *qFU-Dt09*) were newly identified, and the remaining QTLs overlapped with previously identified QTLs (Fig. 3b and Additional file 1: Table S3).

Gene linkage and pleiotropic effects are commonly observed between complex agronomic traits [23]. In our study, two significant SNPs (i52359Gb and i11316Gh) exhibited pleiotropy and were associated with four and two fiber traits, respectively (Additional file 1: Table S3). Meanwhile, QTLs containing these significant SNPs were anchored in the same genomic region. For example, there were four QTLs (*qFE-Dt09*, *qFS-Dt09-2*, *qFL-Dt09* and *qFU-Dt09*) on Dt09, which were associated with different fiber traits and simultaneously mapped to the genomic region of Dt09: 50.51–50.91 Mb. On Dt06, two QTLs (*qFS-Dt06* and *qFL-Dt06*) with the same significant SNP were located in the same genomic interval (59.8–60.2 Mb). Moreover, some QTLs overlapped with or were adjacent to other QTLs in terms of physical position, such as *qFL-At10/qFU-At10*, *qFM-Dt05-2/qFU-Dt05-1* and *qFM-Dt09-2/qFS-Dt09-1* (Fig. 3b). These results implied

that these fiber traits might be controlled by a QTL network with multiple effects.

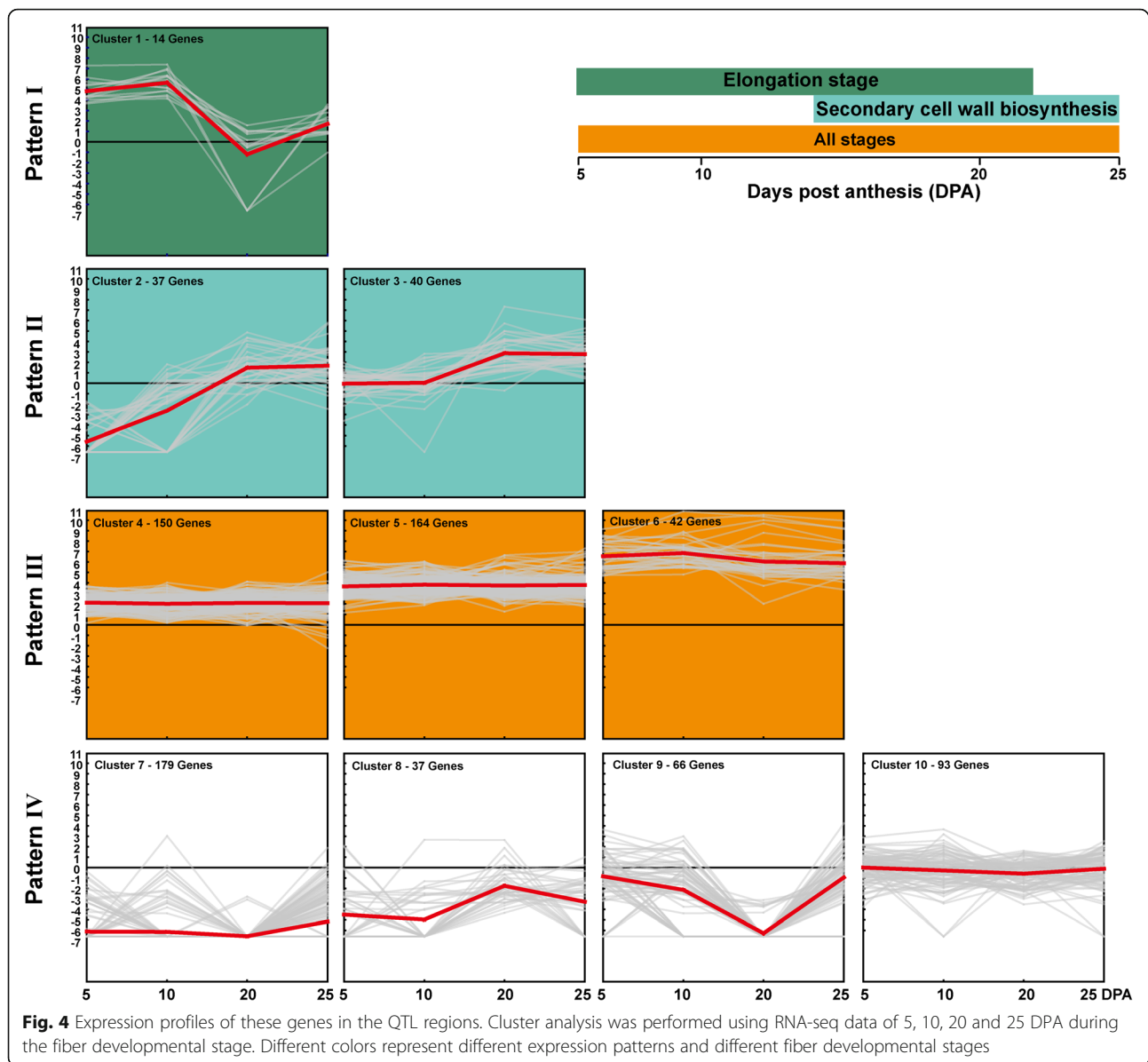
Expression profile analysis of genes in the QTL regions

Based on the *G. hirsutum* TM-1 reference genome sequence [10], 822 genes were detected in these QTL regions (Additional file 1: Table S4). These genes were unevenly distributed across the genome, with 154 and 668 genes in the At and Dt subgenomes, respectively (Additional file 3: Figure S2). Moreover, the number of genes differed among chromosomes. The largest number of genes was located on Dt05 (188 genes), while there were only 14 genes on At09. In addition, we investigated the expression patterns of these genes using transcriptome data on fiber development, which were obtained from the NCBI Sequence Read Archive (SRA) database. The results revealed that these genes could be divided into four patterns, I–IV (Fig. 4). Pattern I included 14 genes highly expressed at the elongation stage for about 5 to 10 days post anthesis (DPA). Pattern II, comprising two clusters (Clusters 2 and 3), contained 77 genes preferentially expressed at the secondary cell wall synthesis stage (20–25 DPA). Pattern III contained 356 genes that maintained high expression levels at all of the fiber developmental stages. Pattern IV included 375 genes with low expression levels at all fiber developmental stages. This finding suggested that these genes could be involved in fiber development and play important roles in different stages of fiber development.



Analysis of pleiotropic SNPs associated with fiber quality
In plants, many traits are controlled by multiple genes, and some genes have pleiotropic effects on

yield traits or other traits [20]. Gene pleiotropy means that a gene determines or influences the formation of multiple traits [24]. In this study, we identified single



genomic regions on both chromosomes Dt09 and Dt06 that exhibited pleiotropic associations with more than one fiber quality trait.

On Dt09, the association signal was located at 49.0–50.7 Mb (Fig. 5). In this candidate region, one SNP was significantly associated with four fiber quality traits (FE, FS, FL and FU) (Fig. 5a) and was located within the gene *Gh_D09G2376*, which encodes a Jumonji N/C and zinc finger domain-containing protein (Fig. 5c). The linkage disequilibrium (LD) block analysis showed that two blocks exist in this region, but the significantly associated SNP did not belong to any block (Fig. 5b). Moreover, i52359Gb is a nonsynonymous SNP that mutated from A to C at the 7215 bp position in the coding region of *Gh_D09G2376*, which resulted in an amino acid

change from aspartic acid (Asp) to alanine (Ala) (Fig. 5c). Interestingly, the two genotypes generated by this SNP locus were associated with the phenotypic performances of these fiber traits. The accessions with the AA genotype showed higher FE, FS, FL and FU values than the accessions with the CC genotype. In particular, FS and FL were significantly different between these two genotypes (Fig. 5d). RT-qPCR analysis revealed that this gene was highly expressed at the fiber elongation stage (10 DPA) (Fig. 5e). Moreover, this gene was significantly higher expression at 10 DPA of Xinluzao30 (high-quality upland cotton cultivar) compared with Sukang191 (low-quality upland cotton cultivar) (Additional file 4: Figure S3a and b). In cotton, the function of the *Gh_D09G2376* gene is unknown. However, its orthologous gene in

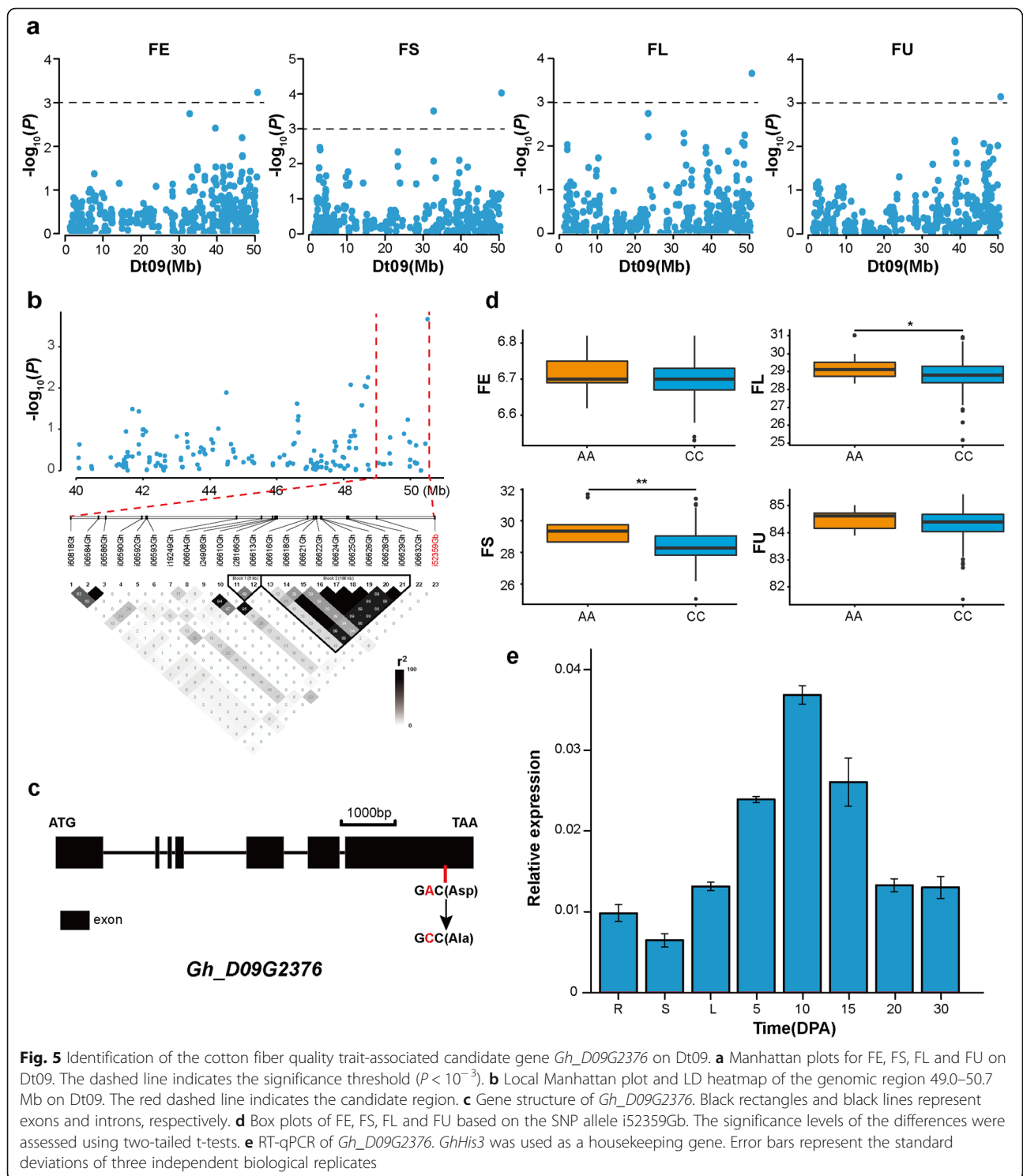


Fig. 5 Identification of the cotton fiber quality trait-associated candidate gene *Gh_D09G2376* on Dt09. **a** Manhattan plots for FE, FS, FL and FU on Dt09. The dashed line indicates the significance threshold ($P < 10^{-3}$). **b** Local Manhattan plot and LD heatmap of the genomic region 49.0–50.7 Mb on Dt09. The red dashed line indicates the candidate region. **c** Gene structure of *Gh_D09G2376*. Black rectangles and black lines represent exons and introns, respectively. **d** Box plots of FE, FS, FL and FU based on the SNP allele i52359Gb. The significance levels of the differences were assessed using two-tailed t-tests. **e** RT-qPCR of *Gh_D09G2376*. *GhHis3* was used as a housekeeping gene. Error bars represent the standard deviations of three independent biological replicates

Arabidopsis, *JMJ12* (*At3g48430*) encodes a Jumonji N/C and zinc finger domain-containing protein and plays an important role in cell elongation (Additional file 5: Figure S4) [25, 26]. Therefore, we hypothesize that this is a pleiotropic gene controlling FE, FS, FL and FU.

In addition, another association signal was associated with FS and FL within 59.8–60.2 Mb on Dt06 (Fig. 6). There were four SNPs in this candidate region, and two of them fell within an LD block (Fig. 6a and b). In this block, only the SNP i11316Gh was significantly

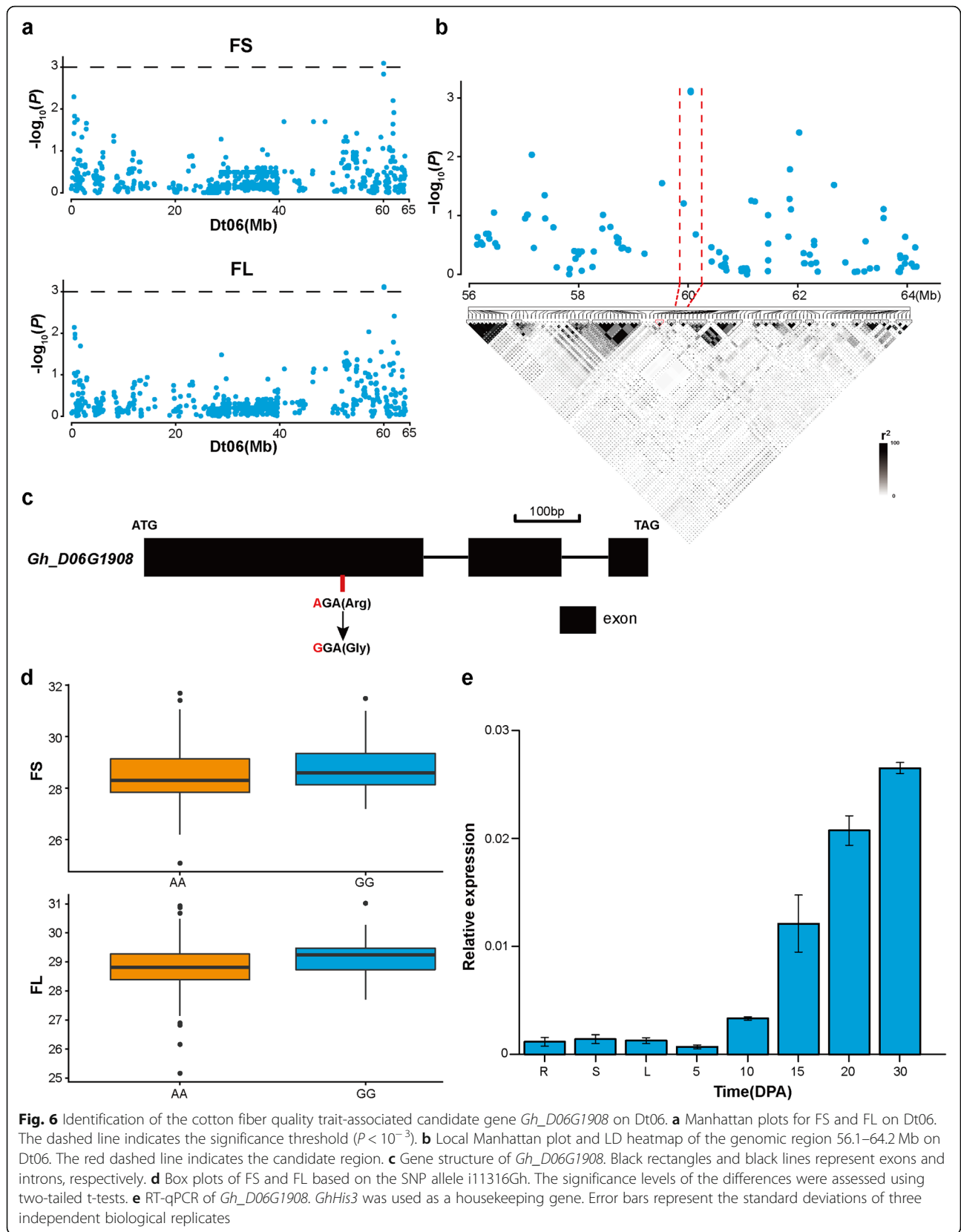


Fig. 6 Identification of the cotton fiber quality trait-associated candidate gene *Gh_D06G1908* on Dt06. **a** Manhattan plots for FS and FL on Dt06. The dashed line indicates the significance threshold ($P < 10^{-3}$). **b** Local Manhattan plot and LD heatmap of the genomic region 56.1–64.2 Mb on Dt06. The red dashed line indicates the candidate region. **c** Gene structure of *Gh_D06G1908*. Black rectangles and black lines represent exons and introns, respectively. **d** Box plots of FS and FL based on the SNP allele i11316Gh. The significance levels of the differences were assessed using two-tailed t-tests. **e** RT-qPCR of *Gh_D06G1908*. *GhHis3* was used as a housekeeping gene. Error bars represent the standard deviations of three independent biological replicates

associated with these traits and was located in the gene *Gh_D06G1908*. This SNP caused an amino acid change from arginine (Arg) to glycine (Gly) at the 310 bp position of this gene (Fig. 6c). And, based on the alleles of this SNP, these accessions were divided into two genotypes, AA and GG. The GG genotype, compared with the AA genotype, had greater FS and longer FL (Fig. 6d). Moreover, the gene was preferentially expressed at the fiber developmental stages, and its expression level increased during fiber development, peaking at 30 DPA (Fig. 6e). We also found that this gene showed higher expression in the high-quality upland cotton cultivar (Xinluzao30) than in the low-quality upland cotton cultivar (Sukang191) at 10, 15 and 20 DPA (Additional file 4: Figure S3a and c). In *Arabidopsis*, *CYSb* (*At3g12490*) is the homologue of *Gh_D06G1908*, and it encodes a protein with cysteine proteinase inhibitor activity (Additional file 6: Figure S5) The overexpression of *CYSb* can stimulate plant growth [27]. These results imply that *Gh_D06G1908* might participate in fiber developmental process.

Discussion

Phenotypic diversity and heritability of fiber quality traits

Genetic diversity, including genotypic and phenotypic diversity, plays a vital role in association mapping [28]. In the present study, the degree of genotypic diversity was similar to that measured in previous studies [19, 20], and it was suitable for a GWAS. Moreover, to increase the reliability of the phenotypic data and reduce environmental effects on the GWAS results, phenotypic data were collected from seven environments over 2 years, and BLUPs for the five fiber quality traits were estimated in this study. Phenotypic statistical results showed that all the traits exhibited phenotypic variation ranging from 0.61%–5.24% (Table 1), which was coincident with the results of previous studies [19, 20, 23].

Heritability is another main factor that influences the accuracy of association mapping [29]. Generally, H^2 is used to judge the degree of stability of inherited traits, and an H^2 value greater than 50% is considered high [30]. The fiber quality traits of upland cotton possess high heritability levels. For example, the H^2 values of these traits reported by Nie et al. (2016), Huang et al. (2017) and Dong et al. (2018) were in the ranges of 86%–93%, 84%–92% and 69.54%–91.05%, respectively [5, 20, 23]. In our research, the H^2 values of these fiber quality traits were also high, ranging from 72.06% (FU) to 91.38% (FM) (Table 1), which was consistent with the results of previous studies [5, 20, 23]. Moreover, ANOVA showed that the variances explained by genotype, environment and their interaction in these tested traits were all significant (Additional file 1: Table S1), suggesting that these traits were influenced by the

environment. However, the genotype effect plays the dominant role. Thus, cotton fiber quality traits are predominately controlled by genetic factors and suitable for association mapping analysis.

Pleiotropic SNPs and QTLs identified in the present study

Fiber quality traits are generally complex quantitative traits that are controlled by a complex network of multiple genes [31]. Some QTLs or genes simultaneously govern multiple fiber quality traits [13, 19]. In the present study, we identified 42 significant SNPs and 31 QTLs associated with these five fiber quality traits using the association mapping method (Additional file 1: Tables S3 and S4, Fig. 3). Interestingly, some of these significant SNPs or QTLs (within the same genomic interval) were associated with more than one trait. The significant SNP i52359Gb was simultaneously associated with FE, FS, FL and FU, and another SNP, i11316Gh, was concurrently associated with FS and FL. Moreover, *qFL-At10*, *qFM-Dt05-2* and *qFM-Dt09-2* overlapped with or were adjacent to *qFU-At10*, *qFU-Dt05-1*, and *qFS-Dt09-1*, respectively. Additionally, correlation analysis of these fiber quality traits showed that FE, FS, FL and FU were significantly positively correlated with each other, and a significant negative correlation was found for FM with both FL and FS. Thus, there were strong correlations between fiber quality traits, which was consistent with the results of previous reports [5, 15, 19]. In addition, correlations between cotton fiber quality traits are generally favorable [32, 33]. Thus, the genes involved in different fiber quality traits can be used to efficiently obtain the desired fiber quality through the improved breeding of cotton fiber.

Comparison of our GWAS results with QTL or GWAS results from previous studies

In recent decades, large numbers of QTLs or GWAS signals associated with fiber quality traits in cotton have been identified using different genetic populations or different mapping methods [18, 34]. Among the 31 QTLs associated with these five fiber quality traits detected in the present study, 25 QTLs fell within or adjacent to QTL intervals or GWAS signals identified in previous studies (Additional file 1: Table S3). For FE, QTLs *qFE-At09*, *qFE-Dt02-2* and *qFE-Dt09* overlapped with DPL0679a [5], NAU3308 – NAU5467 [35] and *qFE23.1* [36, 37], respectively. For FM, six QTLs (*qFM-At10*, *qFM-Dt05-1*, *qFM-Dt05-2*, *qFM-Dt09-2*, *qFM-Dt11* and *qFM-Dt12*) corresponded to ten QTLs reported in previous studies, and some of these QTLs also mapped to regions adjacent to TM33781 [38], *qFM-Chr19-1* [39], NAU1004 [40], qFM23.2 [37] and TM79085 [38]. For FS, the QTLs *qFS-At12*, *qFS-Dt06* and *qFS-Dt08* were associated with *qFS-A12-1* [41], i38606Gh [19] and

D08_54727428 [42], respectively. Moreover, *qFS-Dt09-2* was adjacent to *qFS23.2* [37]. For FL, the QTLs *qFL-At05*, *qFL-At10* and *qFL-Dt01* overlapped with *qFL-c10-1* [43], TM10319 and TM47849 [38], respectively, and QTLs *qFL-Dt06* and *qFL-Dt09* were close to GH185 [44] and TM72969 [38], respectively. For FU, the QTLs *qFU-At10*, *qFU-Dt02*, *qFU-Dt05-1*, *qFU-Dt05-2* and *qFU-Dt05-3* overlapped with *qFU-D5-2* [45], *qFU14.2* [36], HAU1384 [46], TM57244 and TM36632 [38], respectively, and one QTL (*qFU-Dt01*) was located close to the GWAS signal at D01:60914905 [13]. Thus, the GWAS results in the current study appear reliable, and these stably inherited QTLs/genes, which were detected simultaneously by different segregating populations having different genetic backgrounds or by different mapping methods, have great application potential in future breeding programs focused on improving cotton fiber quality. Moreover, we found that more QTLs were located in the Dt subgenome than in the At subgenome, which is consistent with the previous finding that the chromosomes in the Dt subgenome contained 25% more fiber quality QTLs than those in the At subgenome, and the data support the suspicion that the Dt subgenome plays a more important role in fiber-regulatory mechanisms [21, 30, 47, 48].

Potential candidate fiber quality genes

To date, many candidate genes related to fiber developmental stages have been identified using different methods, such as *GhHDD1* and *GhHOX3* by homologous gene cloning [49, 50], *GhMML3_A12* and *GhMML4_D12* by map-based cloning [51, 52] and *GhFL1* and *GhFL2* by GWAS [21]. Moreover, the development of cotton fiber is a complex and dynamic process, involving the initiation, elongation, secondary cell wall synthesis and maturation stages. Correspondingly, the regulatory mechanisms of fiber development are also complex, involving conserved transcription factors, phytohormones, epigenetic modifications and metabolic pathways [31].

In this study, 822 genes were identified in QTL regions (Additional file 1: Table S4). The gene expression profile analysis showed that many genes were preferentially expressed at specific or during all the fiber developmental stages. Among these genes, there were two candidate genes, *Gh_D09G2376* and *Gh_D06G1908* (Figs. 5 and 6). Both genes contain a significantly associated nonsynonymous SNP that causes an amino acid change. Based on the two alleles of these nonsynonymous SNPs, the accessions were classified into two types, carrying AA/CC and AA/GG alleles at i52359Gb and i11316Gh, respectively. Moreover, the gene expression analysis revealed that these two genes exhibited higher expression levels during fiber developmental stages than in other tissues, indicating that they may be involved in fiber development.

We focused on homologous annotation information on *Gh_D09G2376* and *Gh_D06G1908*. The *JMJ12* gene, a homologue of *Gh_D09G2376* in *Arabidopsis*, encodes a Jumonji N/C domain-containing protein. In *Arabidopsis*, *BES1* recruits *JMJ12* to active brassinosteroid (BR) responsive genes, and mutations in the *JMJ12* gene lead to impaired cell elongation [25, 26]. In cotton, BR is a critical regulator of fiber elongation [53], and *BES1* genes have influence on fiber development [54]. *Gh_D06G1908* is a homologue of *Arabidopsis* *CYSb* gene, which encodes a protein that inhibits the catalytic activity of proteinases. In *Arabidopsis*, *CYSb*-overexpression transgenic lines exhibit greater fresh weights than wild-type plants [27]. Therefore, we speculated that these two genes may underlie fiber development in cotton. Functional studies of these candidate genes are needed to further validate their roles in cotton fiber development.

Conclusions

In this study, an association mapping method was used to explore the genetic architecture of fiber quality traits in upland cotton. A total of 42 SNPs and 31 QTLs significantly associated with these fiber quality traits were identified. In total, 822 genes were located in these QTLs. Furthermore, *Gh_D09G2376* and *Gh_D06G1908* were detected as candidate genes controlling fiber quality traits as assessed by RNA-seq and RT-qPCR analyses. In summary, the SNPs, QTLs and candidate genes identified in this study can be used in marker-assisted selection (MAS) of fiber quality traits and to increase our understanding of the molecular mechanisms responsible for fiber quality.

Methods

Plant materials and field experiments

A set of 276 upland cotton accessions were used for the GWAS and detailed information regarding these accessions has been published previously [22]. A randomized complete block design with two replicates was conducted and these accessions were planted in eight environments, including Anyang in Henan Province (in 2016 and 2017), Jingzhou in Hubei Province (in 2016 and 2017), Alaer in Xinjiang Province (in 2016 and 2017), Jiujiang in Jiangxi Province (in 2016) and Huanggang in Hubei Province (in 2017). In all environments, each plot was 6.0 m length with 0.8 m between rows, containing 20–25 plants. The field management followed local agronomic practices.

Phenotypic data collection and statistical analysis

A total of 25 naturally open bolls were harvested from each plot by hand. After ginning, 10–15 g fiber from each sample was sent to the Cotton Fiber Quality Inspection and Testing Center of the Ministry of

Agriculture, Anyang, China, and the fiber quality traits of each accession were measured by the HVI1000 automatic fiber testing system (<http://www.uster.cn/en/instruments/fiber-testing-1/uster-hvi-3/>), including FE (%), FM, FS (cN/tex), FL (upper-half mean FL, mm) and FU (%).

To reduce environment-related errors, the BLUPs of the fiber quality traits were calculated in R software using the lme4 package [55]. The BLUP values were used for subsequent association mapping. Statistical analyses, Pearson's linear correlation coefficients between different fiber quality traits, and an analysis of variance (ANOVA) were also implemented in R software [56]. Moreover, the H^2 values of each trait were calculated according to the previously reported method [22].

Association mapping of fiber quality traits

SNP genotyping was performed using the CottonSNP63K array, and 10,660 high-quality SNPs were identified and employed in the association mapping. Population structure, relative kinship and LD analyses of this population have been reported previously [22]. Here, a GWAS was conducted using the Genome Association and Prediction Integrated Tool (GAPIT) with a mixed linear model (MLM) (PCA + K) [57]. To obtain more significantly associated SNP, a threshold of $P = 1.0 \times 10^{-3}$ was chosen. Manhattan plots were drawn using qqman package in R software [58]. The LD heatmaps near peak SNPs were produced using Haploview 4.2 software [59].

Identification of QTLs and gene expression analysis

In accordance with a previously reported method [19, 22], we selected 200-kb regions upstream and downstream of the significant trait-associated SNPs as QTLs based on the *G. hirsutum* TM-1 reference genome sequence [10]. The co-location analysis of our GWAS results and previously reported results was implemented by the following steps described in previous studies [60]: (1) the previously reported QTLs and GWAS signals were obtained from the cotton QTL database (CottonQTLdb, <http://www2.cottonqtl.org/>) [6], and the primer sequences of these markers were selected from the Cottongen database (<https://www.cottongen.org/>) [61]; (2) the physical positions of these QTLs were obtained using electronic PCR (e-PCR) based on the *G. hirsutum* TM-1 reference genome sequence [10]; and (3) these previous QTLs were compared with the QTLs identified in this study on a physical map.

The genes in the QTL regions were mined from the gene annotations in the *G. hirsutum* TM-1 reference genome [10]. To investigate the expression profiles of these genes, RNA-seq data of *G. hirsutum* TM-1 tissues were obtained from the NCBI Sequence Read Archive (SRA) database under accession PRJNA248163 [10].

And, RNA-seq data were processed using TopHat and Cufflinks software [62], and normalized fragments per kilobase per million mapped read (FPKM) values were used to indicate the gene expression levels.

RNA extraction and RT-qPCR verification

Root, stem, leaf and fiber samples at 5, 10, 15, 20 and 30 DPA obtained from *G. hirsutum* TM-1, and fiber samples at 10, 15 and 20 DPA obtained from Xinlumian30 (high-quality upland cotton cultivar) and Sukang191 (low-quality upland cotton cultivar) were used for total RNA isolation. Total RNA extraction, cDNA synthesis and RT-qPCR were performed following the previously study [22]. Three independent biological replicates were performed for each sample. *GhHis3* was chosen as the internal reference gene. Relative gene expression values were calculated using the comparative Ct method [63]. All primers used in this study are listed in Additional file 1: Table S5.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12870-020-02611-0>.

Additional file 1: Table S1. Analysis of variance (ANOVA) of the cotton fiber quality traits. **Table S2.** Significant SNPs detected for cotton fiber quality traits by a GWAS. **Table S3.** Comparisons of the QTLs identified in this study with those identified in previous studies. **Table S4.** Annotation information of these genes in the QTL regions. **Table S5.** Primer sequences for RT-qPCR in this study.

Additional file 2: Figure S1. Manhattan plots and quantile-quantile plots for FE, FM, FS, FL and FU. The dashed horizontal line indicates the significance threshold ($P < 10^{-3}$).

Additional file 3: Figure S2. Distribution of genes in these QTL regions.

Additional file 4: Figure S3. Comparison of expression levels of *Gh_D09G2376* and *Gh_D06G1908* between Xinlumian30 and Sukang191. **a** Box plots for fiber quality traits of Xinlumian30 and Sukang191. **b** Expression of *Gh_D09G2376* in Xinlumian30 and Sukang191 by RT-qPCR. **c** Expression of *Gh_D06G1908* in Xinlumian30 and Sukang191 by RT-qPCR. *GhHis3* was used as a housekeeping gene. Error bars represent the standard deviations of three independent biological replicates. ** indicates the significance level at 0.01 by using two-tailed t-test method.

Additional file 5: Figure S4. Homology analysis of *Gh_D09G2376*. **a** Phylogenetic tree of *Gh_D09G2376* and *Arabidopsis JMJ* gene family. **b** Protein structure of *Gh_D09G2376* and *At3g48430/JMJ12*.

Additional file 6: Figure S5. Homology analysis of *Gh_D06G1908*. **a** Phylogenetic tree of *Gh_D06G1908* and *Arabidopsis CYS* gene family. **b** Protein structure of *Gh_D06G1908* and *At3g12490/CYSb*.

Abbreviations

FE: Fiber elongation; FM: Fiber micronaire; FS: Fiber strength; FL: Fiber length; FU: Fiber uniformity; GWAS: Genome-wide association study; QTLs: Quantitative trait loci; SNP: Single nucleotide polymorphism; PVE: Phenotypic variation explained; BLUPs: Best linear unbiased predictions; H^2 : Broad-sense heritability; DPA: Days post anthesis; LD: Linkage disequilibrium

Acknowledgements

We acknowledge Peng Huo (Zhengzhou Research Center, Institute of Cotton Research of CAAS, Zhengzhou, China) for technical assistance.

Authors' contributions

WL (Wei Li), DGY and XXW conceived and designed the research. WL (Wei Liu), CXS, ZYR, ZQZ, JJZ and JZ performed the experiments. XYP, YGL, KLH and FZ prepared the materials. WL (Wei Liu) and CXS analyzed the data and wrote the paper. WL (Wei Li), DGY and XXW revised the manuscript. WL (Wei Li) guided the experiments and supervised the research. All authors have read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (31901580), and the Key Project of Science and Technology of Henan Province of China (192102110032 and 182102110306). These funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The raw RNA-seq data are available from NCBI Sequence Read Archive (SRA) database under accession PRJNA248163. All data generated and analyzed during this study are included in this published article and its supplementary information files. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable requests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Collaborative Innovation Center of Henan Grain Crops, Agronomy College, Henan Agricultural University, Zhengzhou 450002, China. ²State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China. ³Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou 450001, China.

Received: 10 February 2020 Accepted: 18 August 2020

Published online: 27 August 2020

References

- Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD, et al. Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol.* 2007;145(4):1303–10.
- Zhang HB, Li Y, Wang B, Chee PW. Recent advances in cotton genomics. *Int J Plant Genomics.* 2008;2008:742304.
- Kohel RJ, Yu J, Park Y-H, Lazo GR. Molecular mapping and characterization of traits controlling fiber quality in cotton. *Euphytica.* 2001;121(2):163–72.
- Su J, Li L, Pang C, Wei H, Wang C, Song M, Wang H, Zhao S, Zhang C, Mao G, et al. Two genomic regions associated with fiber quality traits in Chinese upland cotton under apparent breeding selection. *Sci Rep.* 2016;6:38496.
- Nie X, Huang C, You C, Li W, Zhao W, Shen C, Zhang B, Wang H, Yan Z, Dai B, et al. Genome-wide SSR-based association mapping for fiber quality in nation-wide upland cotton inbred cultivars in China. *BMC Genomics.* 2016;17:352.
- 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* x *G. barbadense* populations. *Mol Genet and Genomics.* 2015;290(4):1615–25.
- 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.
- Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet.* 2014;46(6):567–72.
- Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature.* 2012;492(7429):423–7.
- Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotech.* 2015;33(5):531–7.
- Hulse-Kemp AM, Lemm J, Plieske J, Ashrafi H, Buyyarapu R, Fang DD, Frelichowski J, Giband M, Hague S, Hinze LL, 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.
- 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(1):654.
- Fang L, Wang Q, Hu Y, Jia Y, Chen J, Liu B, Zhang Z, Guan X, Chen S, Zhou B, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet.* 2017;49(7):1089–98.
- Li T, Ma X, Li N, Zhou L, Liu Z, Han H, Gui Y, Bao Y, Chen J, Dai X. Genome-wide association study discovered candidate genes of *Verticillium* wilt resistance in upland cotton (*Gossypium hirsutum* L.). *Plant Biotechnol J.* 2017;15(12):1520–32.
- Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z, Shen C, Li J, Zhang L, et al. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat Genet.* 2017;49(4):579–87.
- Sun Z, Li H, Zhang Y, Li Z, Ke H, Wu L, Zhang G, Wang X, Ma Z. Identification of SNPs and candidate genes associated with salt tolerance at the seedling stage in cotton (*Gossypium hirsutum* L.). *Front Plant Sci.* 2018;9:1011.
- Huang X, Han B. Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol.* 2014;65:531–51.
- Yu JZ, Gervers KA. Genomic analysis of marker-associated fiber development genes in upland cotton (*Gossypium hirsutum* L.). *Euphytica.* 2019;215(4):74.
- Sun Z, Wang X, Liu Z, Gu Q, Zhang Y, Li Z, Ke H, Yang J, Wu J, Wu L, et al. Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in *Gossypium hirsutum* L. *Plant Biotechnol J.* 2017;15(8):982–96.
- Huang C, Nie X, Shen C, You C, Li W, Zhao W, Zhang X, Lin Z. 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.
- Ma Z, He S, Wang X, Sun J, Zhang Y, Zhang G, Wu L, Li Z, Liu Z, Sun G, et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat Genet.* 2018;50(6):803–13.
- Song C, Li W, Pei X, Liu Y, Ren Z, He K, Zhang F, Sun K, Zhou X, Ma X, et al. Dissection of the genetic variation and candidate genes of lint percentage by a genome-wide association study in upland cotton. *Theor Appl Genet.* 2019;132(7):1991–2002.
- Dong C, Wang J, Yu Y, Ju L, Zhou X, Ma X, Mei G, Han Z, Si Z, Li B, et al. Identifying functional genes influencing *Gossypium hirsutum* fiber quality. *Front Plant Sci.* 2018;9:1968.
- Yang C, Li C, Wang Q, Chung D, Zhao H. Implications of pleiotropy: challenges and opportunities for mining big data in biomedicine. *Front Genet.* 2015;6:229.
- Yu X, Li L, Li L, Guo M, Chory J, Yin Y. Modulation of brassinosteroid-regulated gene expression by Jumonji domain-containing proteins *ELF6* and *REF6* in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2008;105(21):7618–23.
- Lu F, Cui X, Zhang S, Jenuwein T, Cao X. *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. *Nat Genet.* 2011;43(7):715–9.
- Zhang X, Liu S, Takano T. Two cysteine proteinase inhibitors from *Arabidopsis thaliana*, *AtCYSa* and *AtCYSb*, increasing the salt, drought, oxidation and cold tolerance. *Plant Mol Biol.* 2008;68(1–2):131–43.
- Li C, Wang Y, Ai N, Li Y, Song J. A genome-wide association study of early-maturation traits in upland cotton based on the CottonSNP80K array. *J Integr Plant Biol.* 2018;60(10):970–85.
- Bernardo R. What proportion of declared QTL in plants are false? *Theor Appl Genet.* 2004;109(2):419–24.
- Zhang C, Li L, Liu Q, Gu L, Huang J, Wei H, Wang H, Yu S. 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.* 2019;10:53.
- Wang Z, Yang Z, Li F. Updates on molecular mechanisms in the development of branched trichome in *Arabidopsis* and nonbranched in cotton. *Plant Biotechnol J.* 2019;17(9):1706–22.

32. Smith CW, Coyle GG. Association of fiber quality parameters and within-boll yield components in upland cotton. *Crop Sci.* 1997;37:1775–9.
33. McCall LL, Verhalea LM, McNew RW. Multidirectional selection for fiber strength in upland cotton. *Crop Sci.* 1986;26:744–50.
34. Said JI, Song M, Wang H, Lin Z, Zhang X, Fang DD, Zhang J. A comparative meta-analysis of QTL between intraspecific *Gossypium hirsutum* and interspecific *G. hirsutum* × *G. barbadense* populations. *Mol Gen Genomics.* 2015;290(3):1003–25.
35. Wang J, Guo W, Zhang T. QTL mapping for fiber quality properties in cotton cultivar Yumian 1. *Acta Agron Sin.* 2007;33(12):1915–21.
36. Tan Z, Zhang Z, Sun X, Li Q, Sun Y, Yang P, Wang W, Liu X, Chen C, Liu D, et al. Genetic map construction and fiber quality QTL mapping using the CottonSNP80K array in upland cotton. *Front Plant Sci.* 2018;9:225.
37. Liu X, Teng Z, Wang J, Wu T, Zhang Z, Deng X, Fang X, Tan Z, Ali I, Liu D, et al. Enriching an intraspecific genetic map and identifying QTL for fiber quality and yield component traits across multiple environments in upland cotton (*Gossypium hirsutum* L.). *Mol Gen Genomics.* 2017;292(6):1281–306.
38. Li C, Fu Y, Sun R, Wang Y, Wang Q. Single-locus and multi-locus genome-wide association studies in the genetic dissection of fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *front. Plant Sci.* 2018;9:1083.
39. Shang L, Wang Y, Wang X, Liu F, Abduweli A, Cai S, Li Y, Ma L, Wang K, Hua J. Genetic analysis and QTL detection on fiber traits using two recombinant inbred lines and their backcross populations in upland cotton. G3 (Bethesda). 2016;6(9):2717–24.
40. Cai C, Ye W, Zhang T, Guo W. Association analysis of fiber quality traits and exploration of elite alleles in upland cotton cultivars/accessions (*Gossypium hirsutum* L.). *J Integr Plant Biol.* 2014;56(1):51–62.
41. Si Z, Chen H, Zhu X, Cao Z, Zhang T. Genetic dissection of lint yield and fiber quality traits of *G. hirsutum* in *G. barbadense* background. *Mol Breed.* 2017;37(1):9.
42. Su J, Ma Q, Li M, Hao F, Wang C. Multi-locus genome-wide association studies of fiber-quality related traits in Chinese early-maturity upland cotton. *Front Plant Sci.* 2018;9:1169.
43. Wang H, Huang C, Zhao W, Dai B, Shen C, Zhang B, Li D, Lin Z. Identification of QTL for fiber quality and yield traits using two immortalized backcross populations in upland cotton. *PLoS One.* 2016;11(12):e0166970.
44. Nie X, Tu J, Wang B, Zhou X, Lin Z. A BIL population derived from *G. hirsutum* and *G. barbadense* provides a resource for cotton genetics and breeding. *PLoS one.* 2015;10(10):e0141064.
45. Hu W, Zhang X, Zhang T, Guo W. Molecular tagging and source analysis of QTL for elite quality in upland cotton. *Acta Agron Sin.* 2008;34(4):578–86.
46. Dong CG, Wang J, Yu Y, Li BC, Chen QJ. Association mapping and favourable QTL alleles for fibre quality traits in upland cotton (*Gossypium hirsutum* L.). *J Genet.* 2018;97(1):e1–e12.
47. Rong J, Feltus FA, Waghmare VN, Pierce GJ, Chee PW, Draye X, Saranga Y, Wright RJ, Wilkins TA, May OL, et al. Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. *Genetics.* 2007;176(4):2577–88.
48. Fang DD, Jenkins JN, Deng DD, McCarty JC, Li P, Wu J. Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in upland cotton (*Gossypium hirsutum* L.). *BMC Genomics.* 2014; 15:397.
49. Walford SA, Wu Y, Llewellyn DJ, Dennis ES. Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, *GhHD-1*. *Plant J.* 2012;71(3):464–78.
50. Shan CM, Shanguan XX, Zhao B, Zhang XF, Chao LM, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ, et al. Control of cotton fibre elongation by a homeodomain transcription factor *GhHox3*. *Nat Commun.* 2014;5:5519.
51. Wan Q, Guan X, Yang N, Wu H, Pan M, Liu B, Fang L, Yang S, Hu Y, Ye W, et al. Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. *New Phytol.* 2016;210(4):1298–310.
52. Wu H, Tian Y, Wan Q, Fang L, Guan X, Chen J, Hu Y, Ye W, Zhang H, Guo W, et al. Genetics and evolution of *MIXTA* genes regulating cotton lint fiber development. *New Phytol.* 2018;217(2):883–95.
53. Yang Z, Zhang C, Yang X, Liu K, Wu Z, Zhang X, Zheng W, Xun Q, Liu C, Lu L, et al. *PAG1*, a cotton brassinosteroid catabolism gene, modulates fiber elongation. *New Phytol.* 2014;203(2):437–48.
54. Liu Z, Qanmber G, Lu L, Qin W, Liu J, Li J, Ma S, Yang Z, Yang Z. Genome-wide analysis of *BEST1* genes in *Gossypium* revealed their evolutionary conserved roles in brassinosteroid signaling. *Sci China Life Sci.* 2018;61(12): 1566–82.
55. Bates D, Machler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;67(1):1–48.
56. Team RDC. R : a language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. *Computing.* 2014;14: 12–21.
57. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. GAPIT: genome association and prediction integrated tool. *Bioinformatics.* 2012;28(18):2397–9.
58. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv.* 2014;(5):005165. <https://doi.org/10.1101/005165>.
59. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2):263–5.
60. Su J, Fan S, Li L, Wei H, Wang C, Wang H, Song M, Zhang C, Gu L, Zhao S, et al. Detection of favorable QTL alleles and candidate genes for lint percentage by GWAS in Chinese upland cotton. *Front Plant Sci.* 2016;7:1576.
61. Yu J, Jung S, Cheng CH, Ficklin SP, Lee T, Zheng P, Jones D, Percy RG, Main D. CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic Acids Res.* 2014;42:D1229–36.
62. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc.* 2012;7(3):562–78.
63. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc.* 2008;3(6):1101–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

