Theoretical and Applied Genetics (2022) 135:4483–4494
https://doi.org/10.1007/s00122-022-04233-6

ORIGINAL ARTICLE

Fine-mapping and candidate gene analysis of *qFL-c10-1* controlling fiber length in upland cotton (*Gossypium hirsutum* L.)

Ruiting Zhang[1] · Chao Shen[2] · De Zhu[1] · Yu Le[1] · Nian Wang[1] · Yuanxue Li[1] · Xianlong Zhang[1] · Zhongxu Lin[1]

Received: 23 June 2022 / Accepted: 1 October 2022 / Published online: 11 October 2022

Abstract

Key message **A fiber length QTL, *qFL-c10-1*, was fine-mapped to a 96.5-kb region containing one gene that has not been characterized in plants.**

Abstract Fiber length is an important component of cotton fiber quality, which is associated with other quality properties such as fiber strength, fiber maturity, and fineness. In our previous studies, a stable QTL *qFL-c10-1* controlling fiber length had been identified on chromosome A10 in an upland cotton recombinant inbred line (RIL) population from a cross between Jimian5 and DH962. To fine-map *qFL-c10-1*, an F_2 population with 1081 individual plants from a cross between a recombinant line DJ61 and Jimian5 was established. Using linkage analysis and progeny recombination experiment, *qFL-c10-1* was mapped into a 96.5-kb genomic region that just contained one proper transcript *Ghir_A10G022020* (described as *GhFL10*), an undescribed gene in plants. One 214-bp deletion was identified in the promoter region of DJ61 compared with Jimian5. Quantitative real-time PCR (qRT-PCR) and comparative analysis of parental sequences suggested that *GhFL10* was the most promising candidate gene for *qFL-c10-1*. According to RNA-seq, yeast two-hybrid assay and bimolecular fluorescence complementation (BiFC), we speculate that *GhFL10* interacts with NF-YA transcription factors to negatively regulate fiber elongation.

Introduction

Cotton is one of the most crucial cash crops in the world. Upland cotton (*Gossypium hirsutum* L.) dominates more than 95% of the annual cotton production (Chen et al. 2007). For a long time, breeders focused on increasing the yield of cotton. In recent years, with the continuous innovations for efficiency in the textile technology, improving cotton fiber quality has become extremely important. Therefore, it is necessary to effectively improve upland cotton fiber quality to meet the demands of modern textile industry. Generally, cotton fiber quality consists of five component traits, namely fiber length, fiber strength, micronaire value, fiber elongation, and fiber uniformity. Among them, fiber length is one of the most important aspects of fiber quality, which is the initial property used to evaluate the quality of cotton fiber. Therefore, fine-mapping stable quantitative trait loci (QTL) for fiber length and understanding the complex genetic basis

Communicated by Tianzhen Zhang.

✉ Zhongxu Lin
linzhongxu@mail.hzau.edu.cn

Ruiting Zhang
ruiting_z@outlook.com

Chao Shen
383895867@qq.com

De Zhu
zhude@caas.cn

Yu Le
1453762887@qq.com

Nian Wang
627808558@qq.com

Yuanxue Li
834954226@qq.com

Xianlong Zhang
xlzhang@mail.hzau.edu.cn

[1] National Key Laboratory of Crop Genetic Improvement, College of Plant Science & Technology, Huazhong Agricultural University, Wuhan 430070, Hubei, China

[2] College of Biological and Food Engineering, Guangdong University of Petrochemical Technology, Maoming 525000, Guangdong, China

of fiber elongation have become one of the important goals of current cotton biology.

Fiber length is a genetically complex quantitative trait, controlled by multiple genes and easily affected by environment factors. QTL mapping with molecular markers provides a powerful approach to dissect the molecular mechanism underlying complex fiber quality traits (Ijaz et al. 2019). Based on data from Cottongen website (https://www.cottongen.org/), 705 QTLs for fiber length, 617 QTLs for fiber elongation, 134 QTLs for fiber fineness, and 908 QTLs for fiber strength have been mapped with intraspecific and interspecific populations (Yu et al.2021). However, to our knowledge, few genes controlling the fiber traits have been fine-mapped in cotton up to now. Xu et al. (2017) fine-mapped a fiber length QTL *qFL-chr1*, and two differential expression genes, *GOBAR07705* and *GOBAR25992*, were supported as candidates for *qFL-chr1*. Fang et al. (2017) fine-mapped *qFS07.1*, a QTL for fiber strength on chromosome 7, into a 62.6-kb genomic region that contained four annotated genes; qRT-PCR and comparative analysis of candidate genes revealed a leucine-rich repeat protein kinase (*LRR RLK*) family protein to be a promising candidate gene for *qFS07.1*. Feng et al. (2020) fine-mapped a fiber strength QTL *qFS-Chr.D02* into to a 550.66-kb region on chromosome D02; qRT-PCR and SNP analysis revealed that *GH_D02G2269* and *GH_D02G2289* were the most likely candidate genes. Obviously, these findings are not enough to support our knowledge of the molecular basis determining cotton fiber quality, especially fiber length.

Due to its agronomic importance, cotton fiber length has always been the focus of cotton research. A number of factors affecting the development of cotton fiber have been identified, such as Ca^{2+} (Tang et al. 2014), ascorbate peroxidase (Qin et al. 2008), phytohormone (Hu et al. 2016; Yang et al. 2014), sucrose (Sun et al. 2019), and cellulose (Huang et al. 2021; Zhang et al. 2021). Several studies have uncovered the critical role of transcription factors in controlling fiber elongation. For example, silencing *GhMYB25* in cotton resulted in short fibers, while the overexpression lines increased fibers initiation and leaf trichome number (Machado et al. 2009). The cotton homeodomain leucine zipper (HD-ZIP) TF, *GhHOX3*, acted as a core regulator of fiber elongation (Shan et al. 2014). *GhTCP4* interacted antagonistically with *GhHOX3* to constrain fiber elongation growth; the dynamics of *GhHOX3* expression and the miR319-targeted *TCP* expression patterns modulated the cotton fiber transition from cell elongation to cell wall thickening (Cao et al. 2020). *GhFSN1* is a NAC transcription factor and is specifically expressed in secondary cell wall (SCW) thickening fibers, which activated or repressed numerous fiber SCW-related genes to regulate fiber length (Zhang et al. 2018). Basic helix–loop–helix/helix–loop–helix (*bHLH*/*HLH*) transcription factor GhFP2 interacted with a bHLH protein GhACE1 and positively regulated fiber elongation (Lu et al. 2022). Cotton fiber is the unique tissue of cotton, whether other genes contributing to fiber development remain little understood.

In our previous studies, we constructed F_2, recombinant inbred line (RIL) and immortalized backcross populations by using Jimian5 and DH962 as parents, and mapped QTLs for yield and fiber quality. A stable QTL, *qFL-c10-1,* was detected in multiple environments and populations, explaining 5.79–37.09% of the phenotypic variation (Lin et al. 2009; Wang et al. 2015a, b, 2016). In order to fine-map and clone the candidate gene of the *qFL-c10-1* locus, a large F_2 population was developed. We applied whole-genome resequencing, KASP markers development, and rapid amplification of cDNA ends (RACE)-polymerase chain reaction (PCR) to successfully clone the candidate gene *GhFL10*. We also performed RNA-seq and protein interaction experiment to explore the putative function of *GhFL10*. This study discovered an uncharacterized gene related to fiber length, which will provide new insights of fiber development and will facilitate the improvement of cotton fiber length.

Materials and methods

Plant materials and fiber quality measurement

By using molecular markers to screen the RIL populations in the background of Jimian5, a RIL DJ61 with markedly increased fiber length and fiber strength compared to Jimian5 was selected. In DJ61, the QTL *qFL-c10-1* allele is derived from DH962, whereas the other regions with alleles affecting fiber quality are derived from Jimian5. To fine-map *qFL-c10-1*, we crossed Jimian5 and DJ61, and the resultant F_1 plants were self-pollinated to construct the F_2 mapping population. The F_2 population including 1081 individuals was grown in the summer of 2017 at Huazhong Agricultural University, Wuhan, Hubei, China. F_3/F_4 recombination individuals derived from F_2 plants were planted in the summer of 2018 and 2019 at Huazhong Agricultural University.

Mature fibers were collected from the field-grown plants of two parents, F_2 individuals and recombination individuals. Fiber samples (> 8 g) were collected to evaluate fiber quality using a high volume instrument (HVI) method (HFT9000, Premier, India).

For measurements of length of immature fibers at different development stages (8, 10, 12, 15, 18, 20, and 20 day post-anthesis (DPA)), bolls were collected simultaneously from the same positions on parental plants. The method of measuring immature fiber length was as described previously (Tang et al. 2014).

Whole-genome resequencing and marker development

Genomic DNA of two parents was extracted via the modified CTAB method (Paterson et al. 1993) and was sequenced on the Illumina HiSeq platform with 30 × coverage each. To redo SNP calling, all the clean sequencing reads were mapped to the *G. hirsutum* reference TM-1 genome (Wang et al. 2019) using BWA software version 0.7.10 (McKenna et al. 2010), and single nucleotide polymorphisms (SNPs) and insertion/deletions (Indels) were detected using the GATK software (Li et al. 2009).

The primers for the Indel markers and KASP markers for SNPs were designed using Primer Premier 5.0 software (Premier Biosoft, San Francisco, CA, USA). Each KASP marker contained two allele-specific forward primers and one common reverse primer. Simple sequence repeat (SSR) loci were scanned by the MicroSAtellite (MISA) identification tool from (http://pgrc.ipk-gatersleben.de/misa/) based on the sequence of the *G. hirsutum* (Beier et al. 2017). In addition, some SSR and Indel markers previously developed in our laboratory were used. All primer sequences are listed in Table S1.

Genomic DNA extraction and marker analysis

The fresh young leaves of individual plants were collected into 2 ml centrifuge tube and then were grounded in ice box. The CTAB method was used to extract total DNA from the two parental lines and individuals of the F_2 populations (Paterson et al. 1993). The PCR program for the initial denaturing step was at 95 ℃ for 5 min, followed by 34 cycles for 30 s at 95 ℃, 30 s at 55 ℃, 30 s at 72 ℃, with a final extension at 72 ℃ for 5 min. The Indel and SSR markers were genotyped using 8% (w/v) nondenaturing gels and viewed by silver staining, and the SNPs were genotyped using the KASP Master mix (LCG, UK) on ABI 7500 instrument.

Fine mapping of *qFL-c10-1*

All SSR and Indel polymorphic markers between the parents were selected, and 1081 F_2 individuals were genotyped using polymorphic markers to construct linkage maps and to narrow down candidate regions using the Inclusive Composite Interval Mapping (ICIM) module of QTL IciMapping 4.1 software (IciMapping V4.1; http://www.isbreeding.net). The fiber length of F_2 individuals was used for QTL detection under a LOD score of 2.5 (Wang et al. 2015a, b). The Kosambi's mapping function was used to count genetic distances (cM).

The progeny of recombinant individuals screened from F_2 population was used for fine mapping of the identified QTL. The newly developed KASP markers were used to screen recombinant individuals to further narrow the mapping interval. The fiber length of the homozygous recombinant was analyzed using Student's *t* test and was compared with that of Jimian5 at a significance level of $P < 0.05$.

Quantitative real-time PCR

Total RNA was extracted from the developing fibers (8, 10, 12, 15, and 18 DPAs) using a plant RNA purification kit (Tiangen, Beijing, China). The first strand of complementary DNA (cDNA) was reverse transcribed by M-MLV reverse transcriptase (Biotech Co. Ltd, Promega, Beijing). The ABI Prism 7500 system was used to evaluate the expression levels of target genes by qRT-PCR. Each qRT-PCR program was performed in total volume of 20 µl schema containing 1 × SYBR Green Master Mix. The relative quantitative method was used to evaluate quantitative variation. The qRT-PCR program was directed at 95 ℃ for 2 min, followed by heating for 15 s at 95 °C and 35 s at 60 °C for 40 cycles. A minimum of three biological replicates was tested for each sample. The relative expression was determined by the $2^{-\Delta\Delta CT}$ method with *GhUBQ7* (DQ116441) as the internal control. Gene-specific primers used in this analysis are listed in Table S1.

Sequence analysis of candidate genes

Whole coding sequence (CDS) of *GhFL10* was cloned by the gene-specific primers. The primers (Table S1) were designed based on the TM-1 genome (Wang et al. 2019). The Phanta Super-Fidelity DNA Polymerase (Vazyme Biotech, Nanjing, China) was used for PCR amplification. The PCR product was detected by 1% agarose gel electrophoresis, and the target band was recovered and purified by a Gel Extraction Kit (Majorbio, Shanghai, China). The sequence with full-length ORF was cloned into pGEM-T Easy vector (Biotech Co. Ltd, Promega, Beijing), and the vector was transformed into *E. coli* (Top10). The inserts were sequenced using vector M13 primers (5'-CCCAGTCACGACGTTGTAAAACG-3', 5'-GCGGATAACAATTTCACACAGGA-3').

To confirm the cDNA of *GhFL10*, RACE was performed using the SMARTer RACE 5'/3' Kit on total RNA from 8 DPA fibers of Jimian5 according to the manufacturer's instructions (Takara). Specific primers (Table S1) were used to amplify the CDS of *GhFL10* from the prepared cDNA of fibers of Jimian5 and DJ61. All fragments were sequenced by Beijing TSINGKE Biological Technology Co., Ltd. DNAMAN software (Lynnon Biosoft, San Ramon, CA, USA) was used for multiple sequence alignments. Primers for these experiments are listed in Table S1.

Subcellular localization

The CDS of *GhFL10* from DJ61 was cloned into the N-terminal fusion green fluorescent protein (GFP) vector pMDC43. This construct was introduced into *Agrobacterium tumefaciens* strain GV3101 and then was infiltrated into the leaves of *Nicotiana benthamiana* plants. The GFP was detected after 48 h of infiltration using an Olympus FV1200 confocal microscope. The primer information is shown in Table S1.

Transcriptome analysis

Total RNA extraction was performed using the RNAprep Pure Plant Kit (Tiangen, Beijing, China) and was sequenced with the Illumina HiSeq 2000 system (paired-end 150 bp). Three biological replications were sequenced for each parental line. The clean RNA-seq reads of Jimian5 and DJ61 were used for alignment to the *G. hirsutum* reference genome (Wang et al. 2019) using HISAT 2.0 software (http://ccb.jhu.edu/software.shtml). Fragments per kilobase of exon per million mapped reads (FPKM) value were calculated by StringTie for gene expression levels (Pertea et al. 2016), and only criterion of FPKM ≥ 1 in each library was considered as expressed genes. The differentially expressed genes (DEGs) were identified using an R package of DESeq (Anders and Huber 2010). Gene Ontology (GO) function annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation were carried out using the online data of the new reference genome project (https://www.cottongen.org/data/download/genome_tetraploid/AD1) updated by our laboratory.

Yeast two-hybrid and BiFC

The full-length cDNA segment of *GhFL10* was amplified from DJ61 and was cloned into the pGBKT7 vector for transcriptional activation activity analysis. Full-length cDNAs of the candidate interacting proteins were amplified and cloned into the pGBKT7 or pGADT7 vector for further verification. Cotransformed yeast clones were serially diluted (1:10) and then were spotted on SD-Leu-Trp medium and SD-Leu-Trp-Ade-His medium for growth. For BiFC assays, the full-length cDNA of *GhFL10* from DJ61 was cloned into N-terminal fusion yellow fluorescent protein (YFP) vector pCAMBIA1301, and NF-YA transcription factors were cloned into the C-terminal fusion YFP vector pCAMBIA1301. The plasmids were co-expressed in *N. benthamiana* leaf cells by Agrobacterium-mediated infiltration. A confocal microscope (Olympus FV1200) was used to observe the fluorescence approximately 60 h later. All primer sequences are listed in Table S1.

Results

Phenotypic characterization of Jimian5 and DJ61

Fiber length and fiber strength play important roles in fiber quality. We investigated and compared these two traits in Jimian5 and DJ61. The average fiber length was 31.84 ± 0.56 mm and the average fiber strength was 34.17 ± 1.06 cN/tex in DJ61, which were significantly higher ($P < 0.01$) than those in Jimian5 (27.58 ± 0.55 mm and 27.93 ± 0.56 cN/tex, respectively) (Fig. 1a, b, c). These results indicated that the recombinant segment from DH962 in DJ61 contributed to the increase in fiber length and fiber strength. The fiber length was measured in Jimian5 and DJ61 at different time points of developing fiber (8, 10, 12, 15, 18, 20, 22, and 25 DPAs) (Fig. 1d, e), which showed that the fiber length was increasing at 8–18 DPAs and reached maximum at 18 DPA.

Fine mapping of *qFL-c10-1*

To narrow down the *qFL-c10-1* region and to identify the genes controlling fiber length in DJ61, a large F_2 population with 1081 individuals was planted in 2017. In order to develop more markers, the whole-genome resequencing data of two parents were aligned to the reference genome TM-1, and 43 SNPs and 33 Indels were detected in the initial mapping interval. A total of 33 Indel markers were developed, and four Indels (A10-9, A10-25, A10-29, and A10-32) showed clear and easily identifiable polymorphisms between the parental lines. Using the four newly developed Indel markers and six previously linked markers in the QTL, an updated linkage map for *qFL-c10-1* was constructed, and the QTL explained 16.05% of the phenotypic variation (Fig. 2a). Genetic analysis of the F_2 population suggested that *qFL-c10-1* was a semidominant allele from DJ61 (Fig. S1a, b). The fiber length difference of three genotypes was detected by analysis of variance, which showed that there were significant differences among them and large variances due to sample size (Fig. S1c).

Following a modified progeny testing strategy, three SSR and seven Indel markers were used for recombinant screening from all the F_2 individuals, and 4 recombinants events between the flanking markers HAU-J5638 and A10-32 were obtained. The homozygous recombinants were planted in 2019; combined genotype and phenotype analysis, the *qFL-c10-1* locus was narrowed down to a region between A10-9 and A10-29. A total of 43 SNPs were identified from the whole-genome resequencing of two parents, and 9 SNPs between A10-9 and 2385-2 were selected to developed KASP markers. Two polymorphic markers, named SNP13 and SNP294, were obtained, which made a distinction of alleles in 4 recombinant

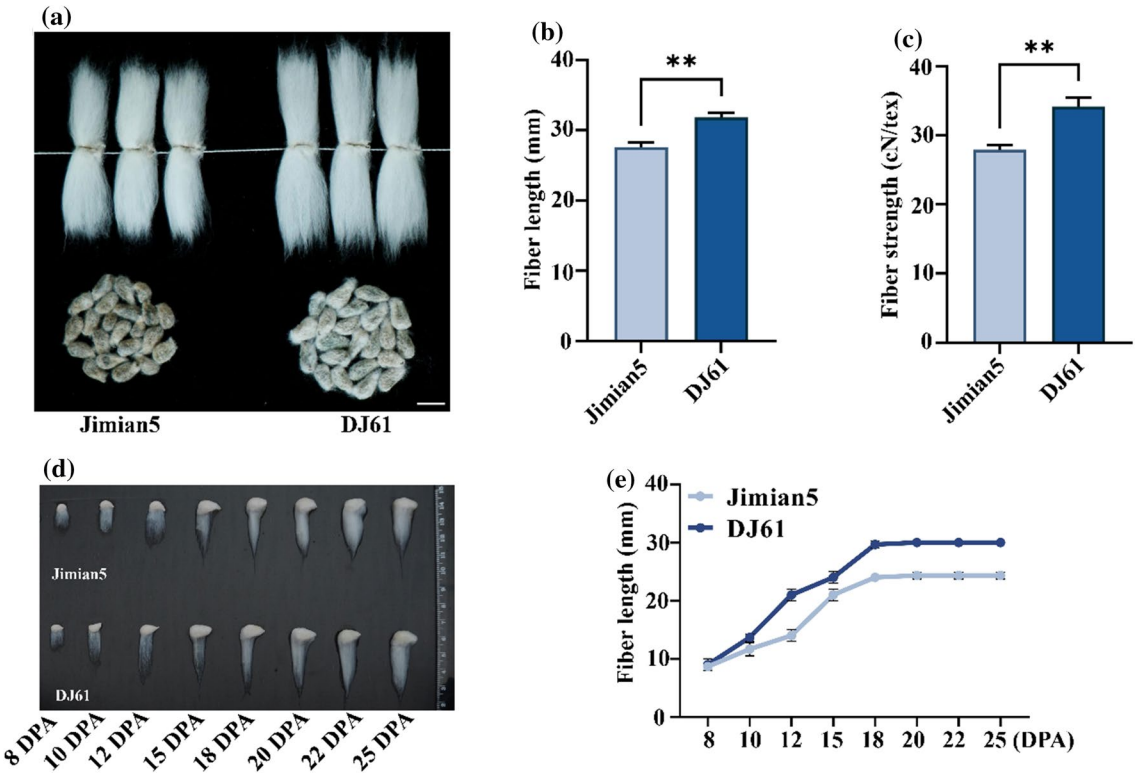

Fig. 1 Fiber phenotypes between Jimian5 and DJ61. **a** Fiber length and fuzz of seeds between Jimian5 and DJ61 at mature stage. Scale bar, 1 cm. **b** Fiber length, **c** fiber strength, and **d** fiber length between Jimian5 and DJ61 at different fiber developing stages. Scale bar, 1 cm. **e** Fiber length of Jimian5 and DJ61 at different stages. Data in **b**, **c**, **e** are shown as means ± SD. Student's *t* test was employed for the evaluation of *P* value (**$P \leq 0.01$)

events. Comparing the genotype and phenotype of the homozygous recombinants (C1-C4) and Jimian5, C1 and C2 recombinants did not show statistically significant difference, while C3 and C4 recombinants showed statistically significant difference. Therefore, the *qFL-c10-1* was delimited to an interval flanked by markers SNP-294 and A10-29 (Fig. 2b). However, the markers corresponding region in the reference genomes was confused, because we found that annotation genes were quite different in the five versions of published genomes of TM-1 (Chen et al. 2020; Hu et al. 2019; Li et al. 2015; Wang et al. 2019; Yang et al. 2019) (Table S2). A further RACE experiment was performed to investigate the real transcripts for all the annotation genes. Sequencing the clones of the 5'/3'RACE amplicons revealed that only the *Ghir_A10G022020* transcript in TM-1 (HAU) was real. The *qFL-c10-1* corresponds to a 96.5-kb genomic fragment from 110,417,074 to 110,513,590 bp based on the TM-1 reference genome sequence (https://cottonfgd.org/, HAU). Therefore, *Ghir_A10G022020* (named as *GhFL10*) was designated as the most likely gene for *qFL-c10-1*.

Variations in *GhFL10*

According to the annotation information from the TM-1 reference genome, *GhFL10* was composed of one exon of 2052-bp without introns and was predicted to encode a protein of unknown function in plants. To define causal variants, we sequenced the entire 2682-bp gene region of *GhFL10* between Jimian5 and DJ61 using gene-specific PCR, including the 581-bp promoter region, the 2052-bp open reading frame (ORF), and the 49-bp 3'-untranslated region (3' UTR). Several variants were identified, including SNPs and Indels in the 2682-bp region (Fig. 3a). There is one synonymous mutation (C 1917 T) and seven nonsynonymous ones (C 113 T, T 161 C, G 579 T, C 1015 A, G 1052 A, A 1126 G, TT 1723 GA) in the coding sequence. Besides these minor variants, one conspicuous 214-bp deletion was present in the longer fiber parent DJ61 at positions -534 and -314 in the promoter region (Fig. 3a, b, and Fig. S2). Transcription factor (TF) binding motif predictions showed that the 214-bp deletion resulted in the deletion of binding sites for MYB, HD-ZIP, bHLH, and C2H2 TF families (Table S3). We next checked the expression pattern of *GhFL10* in developing fibers between Jimian5 and DJ61, and significant differences were observed in fiber elongation period (5 DPA-18 DPA)

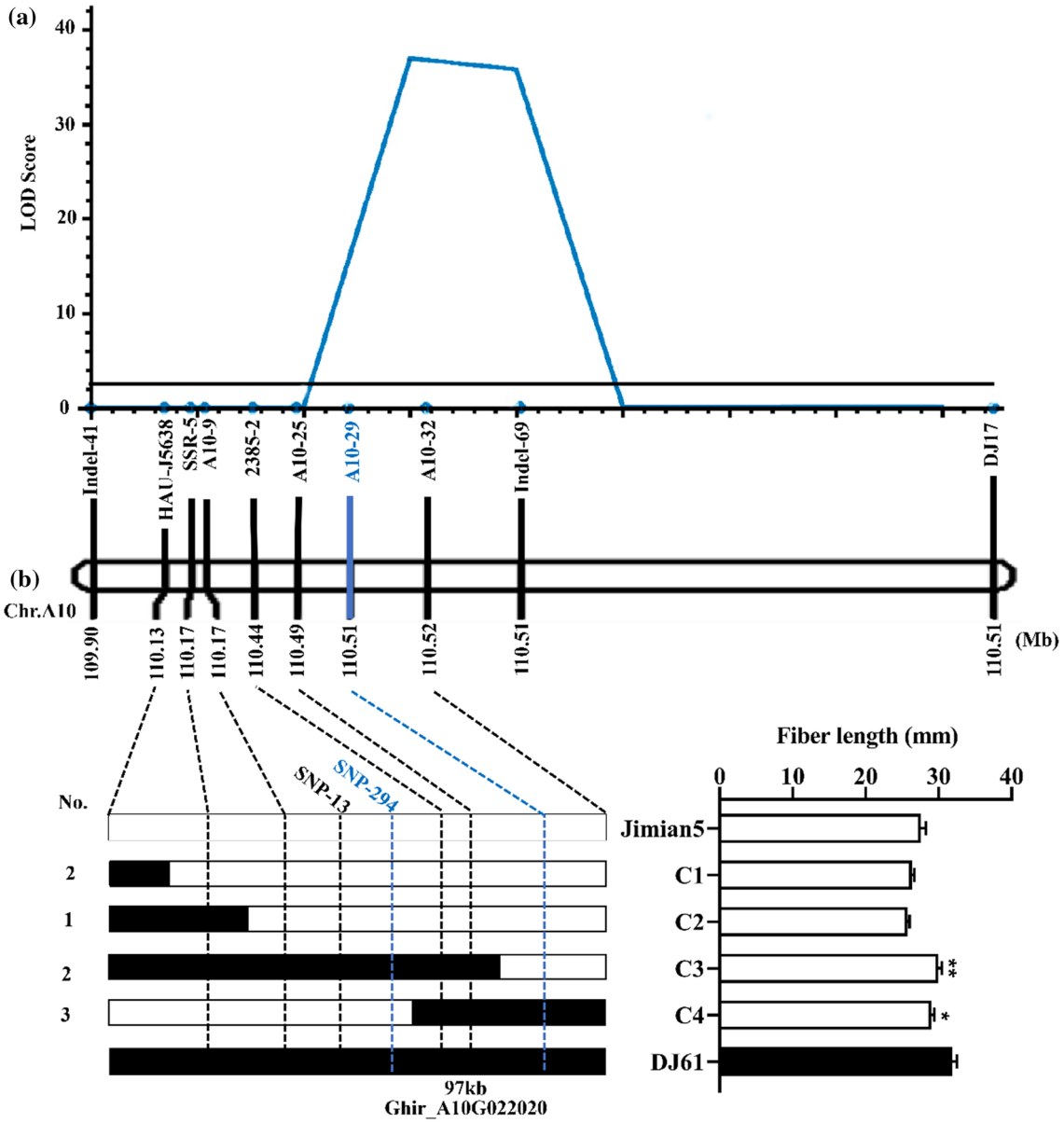

Fig. 2 Fine mapping of *qFL-c10-1*. **a** Fine mapping based on 1081 F_2 individual plants in 2017. **b** *qFL-c10-1* was mapped to 96.5-kb region between markers SNP294 and A10-29 using homozygous recombination plants. The numbers on the left indicate the number of recombinants. Black filled and open bars represent substitution segments homozygous for the DJ61 and Jimian5 alleles, respectively. Fiber length is shown for recombinant plants (C1–C4) and parents. Data are shown as means ± SD, $*P \leq 0.05$, and $**P \leq 0.01$

with higher expression level in Jimian5 than in DJ61 during 5–12 DPAs (Fig. 3c). Thus, *GhFL10* was the candidate gene of *qFL-c10-1* affecting fiber length.

GhFL10 subcellular localization and transcriptional activity analysis

We searched for homologous proteins of GhFL10 in the National Center for Biotechnology Information (NCBI) database and found that the closest ortholog in Arabidopsis is AT3G27500 and shares 30% amino acid identity with GhFL10 (Fig. S3). AT3G27500 has been described as Cysteine/Histidine-rich C1 domain family protein and has not been functionally characterized. To further determine the type of this gene, transient expression vectors were constructed by fusing a GFP with *GhFL10* under the control of the CaMV35S promoter and were introduced into *N. benthamiana* leaves. Fluorescence signals were found to be localized both on the plasma membrane and in nuclear (Fig. 4a), indicating that GhFL10 functioned both on plasma membrane and in nuclear. A transcriptional activation assay showed that the GhFL10 and DNA

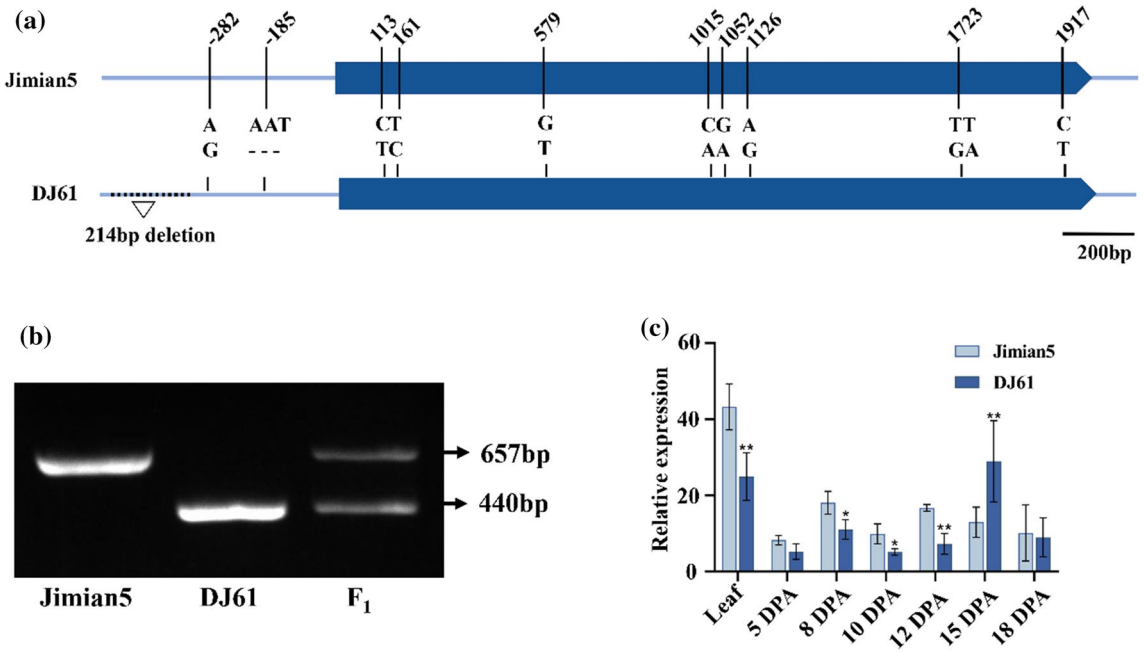

Fig. 3 Sequence variations and differential expression of *GhFL10* **a** Allelic variations in the candidate gene *GhFL10*. **b** Large fragment differences were detected for the 214-bp deletion in the *GhFL10* promoter region. **c** Expression of *GhFL10* in developing fibers of Jimian5 and DJ61. *GhUBQ7* was used as internal control gene. Data are shown as means ± SD, $*P \leq 0.05$, and $**P \leq 0.01$

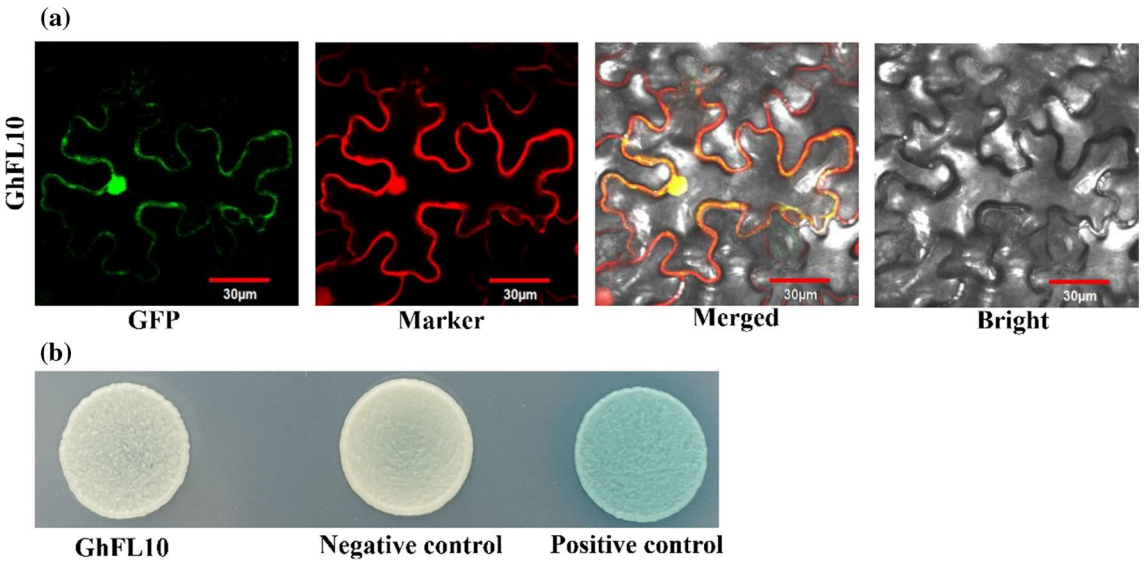

Fig. 4 Subcellular localization and transcriptional activity of GhFL10. **a** Subcellular localization of GhFL10 in *N. benthamiana* leaves. **b** Transcriptional activation activity analysis of GhFL10 in yeast

binding domain (BD) fusion protein in yeast did not activate the expression of the reporter gene (Fig. 4b), implying that GhFL10 has no transcriptional activity, and is not a transcription factor.

GhFL10 is predicted to be interacted with NF-YA transcription factors

Considering that fiber rapidly elongated at 8–18 DPAs, and *GhFL10* was differentially expressed in 8 DPA, we conducted RNA-seq using 8 DPA developing fibers from Jimian5 and DJ61 in order to identify the regulatory networks

in which *GhFL10* was involved. In total, we detected 366 DEGs (fold change > 2, $P \leq 0.01$) with 169 upregulated and 197 downregulated ones (Fig. 5a). Gene Ontology (GO) cluster analysis showed that the DEGs regulated by *GhFL10* were classified in diverse molecular functions, cellular components, and biological processes (Fig. S4), suggesting that *GhFL10* was involved in a complex network regulating fiber elongation. As the major variant of 214-bp deletion was in the promoter region that was related to binding of TFs, we analyzed the differentially expressed TFs in the RNA-seq data, and 14 TFs were found, including the nuclear transcription factor Y subunit (NF-YA), WRKY transcription factor, and ethylene-responsive transcription factor (Fig. 5b).

Because *GhFL10* is an unknown functional protein, there is no guidance for us to follow to study its function. To identify the putative interacting proteins of GhFL10, we used the full-length coding sequences of *GhFL10* from DJ61 as bait to perform a yeast two-hybrid screening. Interestingly, we obtained three candidate interacting proteins that all belong to the NF-YA transcription factors (Fig. 5c); among them, *Ghir_A11G034270* and *Ghir_D11G034910* were found to be differentially expressed in the transcriptome profiling. To verify the interaction in vivo, the interactions between GhFL10 and NF-YA transcription factors were further confirmed by BiFC analysis in *N. benthamiana*. The interaction of GhFL10-nYFP with Ghir_D08G016110-cYFP and GhFL10-nYFP with Ghir_D11G034910-cYFP elicited a nuclear YFP signal, which showed that GhFL10 could direct interact with Ghir_D08G016110 and Ghir_D11G034910 protein, respectively, in vivo (Fig. 5d). Next, we measured the expression levels of *Ghir_D08G016110* and *Ghir_D11G034910* using qRT-PCR between Jimian5 and DJ61. The qRT-PCR results showed that the relative expression of *Ghir_D08G016110* in the short fiber material Jimian5 was lower than in the longer fiber material DJ61 at 10 and 12 DPA, the period of cotton fiber elongating rapidly (Fig. 5e). The expression of *Ghir_D11G034910* was consistent with the results in RNA-seq; in addition, its expression level at all fiber elongation periods in Jimian5, with the exception of 18 DPA, was significantly lower than in DJ61. The expression pattern of *Ghir_D11G034910* was opposite to that of *GhFL10* (Fig. 5f). These results suggested that *GhFL10* may modulate fiber development via interacting with NF-YA transcription factors.

Discussion

Fiber length, controlled by quantitative trait loci (QTL), is an important component of fiber quality and is a key target trait in cotton breeding. However, identification of the genetic components underlying fiber length is challenging due to the complexity of the cotton genome and the difficulty of cloning fiber length QTLs. According to Cottongen website data (https://www.cottongen.org/), a total of 705 genes/QTLs for fiber length have been identified and are distributed across nearly all the 26 chromosomes in cotton (Yu et al.2021). However, none of the QTLs were fine-mapped on chromosome A10. In this study, *qFL-c10-1* was mapped on chromosome A10 flanking by markers SNP294 and A10-29 (Fig. 2b). However, the complex gene annotation in different versions of upland cotton genome challenged us to find the exact gene. Fortunately, we found the real transcript of *Ghir_A10G022020* by using RACE through multiplex PCR amplification and screening. Although the reference genome facilitates the gene cloning, when we clone genes from cotton in a region with a complex genome structure, it is not enough rely on the reference genome sequence. The 214-bp Indel in the promoter of *GhFL10* gene between the two parental lines resulted in TFs binding sites deletion for MYB, HD-ZIP, bHLH, and C2H2 TF families in the longer fiber parent DJ61 (Fig. S2, Table S3). The TFs of MYB, HD-ZIP, bHLH have been reported to control fiber elongation in cotton (Lu et al. 2022; Machado et al. 2009; Shan et al. 2014), and C2H2 zinc-finger transcription factor is responsible for trichome formation in tomato (Chang et al. 2018). The Indel-derived 2385-2 marker (Fig. 3b) will be useful in marker-assisted selection for using *GhFL10* allele in cotton breeding.

The process of cotton fiber cell development can be divided into four distinct but overlapping periods: initiation, elongation, secondary cell wall thickening, and maturation. Fiber initiation starts from 3 DPA, enters the extreme elongation (2 mm/day) phase immediately, and lasts until 20 DPA. The length of elongation period determines the final length of fiber cells. Secondary cell wall thickening stage initiates at 16 DPA via deposition of nearly pure cellulose, which lasts until 40 DPA, followed by the dehydration and maturation of cotton fibers (Haigler et al. 2012). In our study, the expression of *GhFL10* in Jimian5 was higher than in DJ61 at 5–12 DPAs and decreased at 15 DPA with the fiber elongation (Fig. 3c), indicating that *GhFL10* is a negative regulator in fiber elongation. The 15DPA is the overlapping stage and turning point from fiber elongation to the secondary cell wall thickening. We speculated that the significant upregulation of *GhFL10* at 15 DPA in DJ61 should be involved in the cell wall biosynthesis because the fiber strength of DJ61 is better than that of Jimian5. Through homologous searching, we found that *GhFL10* has not been characterized in plants, and we preliminarily classified *GhFL10* as a functional gene by subcellular localization and transcriptional activity analysis (Fig. 4a, b). Cotton fiber is a specialized and elongated single cell that is derived from the seed coat, and the unknown *GhFL10* may have unique functions in cotton fiber elongation.

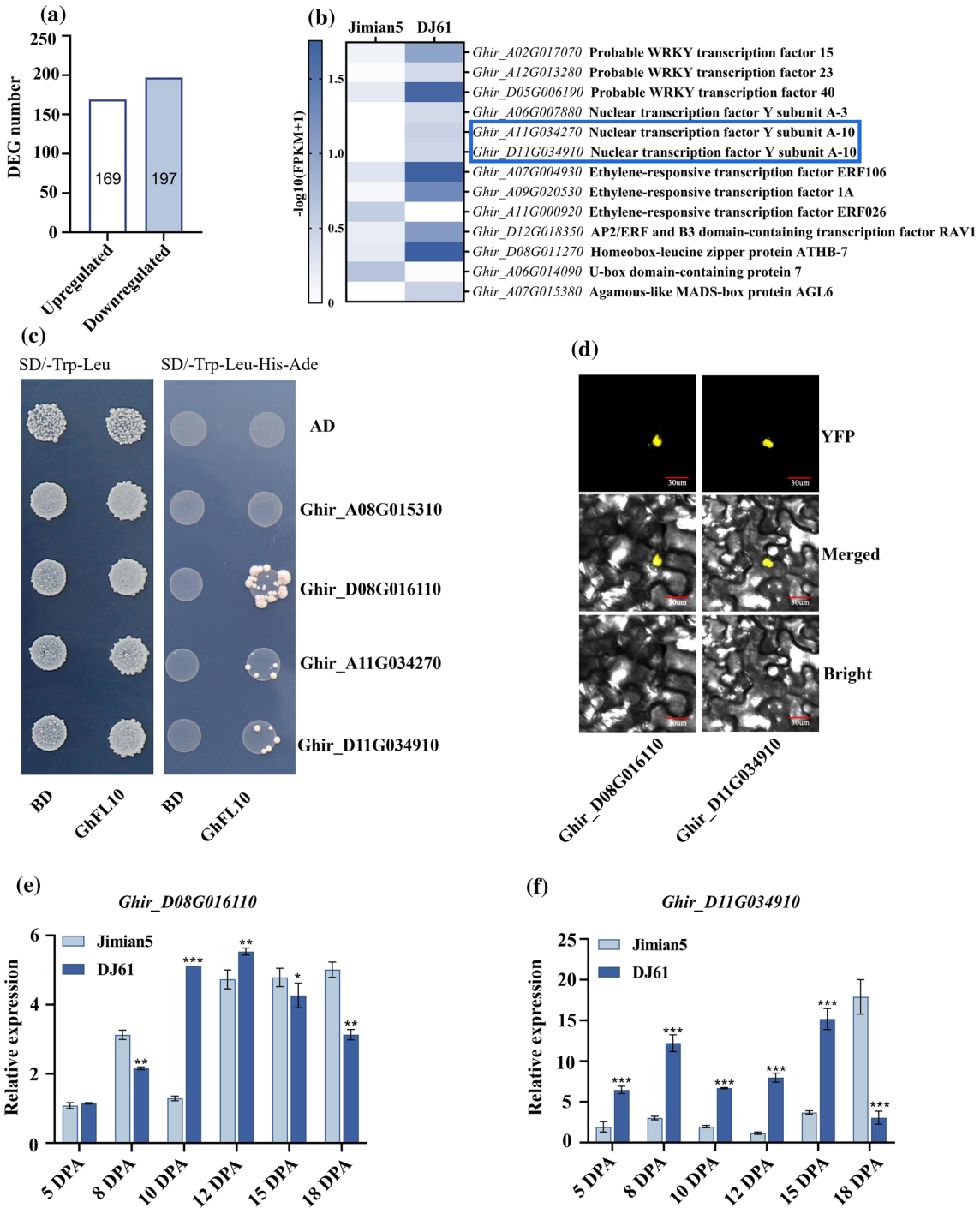

Fig. 5 The interaction between GhFL10 and NF-YA transcription factors. **a** The differentially expressed genes (DEGs) between Jimian5 and DJ61 were identified using a significant cutoff of $P < 0.01$ and a fold change > 2. **b** The transcript differences of 14 transcription factors between Jimian5 and DJ61 based on -log10 (FPKM + 1), where FPKM is fragments per kilobase of transcript per million mapped reads. **c** Yeast two-hybrid assays show the interaction of GhFL10 and NF-YA transcription factors. Yeast cells were plated on SD-Trp-Leu and SD-Trp-Leu-His-Ade media. **d** BiFC assays between GhFL10-nYFP and NF-YA-cYFP. **e**, **f** Expression of *Ghir_D08G016110* and *Ghir_D11G034910* in Jimian5 and DJ61, respectively. *GhUBQ7* was used as internal control gene. Data are shown as means ± SD. $*P \leq 0.05$, $**P \leq 0.01$, and $***P \leq 0.001$

Fiber length is regulated by multiple genes and is associated with complex regulatory networks. TFs play essential regulatory roles during plant growth and development (Mitsuda and Ohme-Takagi 2009), and in recent years, an increasing number of TFs have been reported to function in fiber development of cotton (Cao et al. 2020; Lu et al. 2022; Machado et al. 2009; Shan et al. 2014; Wang et al. 2021b; Zhang et al. 2018). In this study, transcriptome analysis revealed that the expression levels of 14 TFs (encoding WRKY, NF-YA, and AP2/ERF TFs) during cotton fiber elongation were prominently altered in DJ61 fibers compared to Jimian5 (Fig. 5b). WRKY transcription factors play a critical role in the regulation of fiber development in cotton. The cotton WRKY TF, *GhWRKY16*, promoted fiber initiation and elongation by directly regulating the expression levels of downstream target genes (such as *GhMYB25*, *GhHOX3*, *GhMYB109*, and *GhCesA6D_D11*) (Wang et al. 2021b). In cotton, *GhERF4L* and *GhERF54L* acted as key players in salt tolerance (Long et al. 2019). NF-YA silenced plants of cotton indicated that *GhNF-YA10* and *GhNF-YA23* participated in the regulation of response to salt stress (Zhang et al. 2020). It will be promising to explore the functions of these transcription factors in cotton fiber development.

To try to discover the putative function of GhFL10, we used yeast two-hybrid assay and BiFC assay to detect protein-protein interactions and found that GhFL10 directly interacted with NF-YA transcription factor Ghir_D08G016110 and Ghir_D11G034910 protein, respectively (Fig. 5c, d). The expression pattern of NF-YA TFs was opposite to that of *GhFL10* during fast elongation stage (8–12 DPA) (Fig. 5e, f), which demonstrated that *GhFL10* might regulate cotton fiber elongation through antagonizing NF-YA transcription factors. NF-Y is a heterotrimeric complex that contains at least three distinct subunits: NF-YA, NF-YB, and NF-YC. The NF-YA subunit provides specific sequence that is required to specifically binding the CCAAT box in the promoter regions of target genes (Nardini et al. 2013). NF-YA proteins are a class of plant-specific TFs that regulate various plant developmental and abiotic stress tolerance (Myers and Holt III 2018; Petroni et al. 2012). In tomato, the NF-Y complexes are composed of NF-YB8a/8b/8c and NF-YC1a/1b/1d/9, and NF-YA1b/9 played important roles in flavonoid biosynthesis (Wang et al. 2021a). *GmNFYA* acted as an upstream activator of both *GmZF392* and *GmZF351* to activate seed oil biosynthesis in soybean (Lu et al. 2021). *OsNF-YAs* physically interacted with JA signaling transcription factors OsMYC2/3 to inhibit plant antiviral defense by repressing JA pathways in rice (Tan et al. 2022). However, no NF-YA TFs have been reported in cotton fiber development to date, and unraveling their exact mechanism will be a challenge in the future.

In summary, we identified a functional unknown gene, *GhFL10*, which expresses predominately during fiber elongation stage. *GhFL10* may be involved in the interactions with NF-YA transcription factors to regulate cotton fiber elongation. This newly discovered gene will not only accelerate molecular breeding of cotton fiber length, but also will provide a novel route for elucidating the molecular mechanism of fiber elongation regulation in cotton.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-022-04233-6.

Author contribution statement ZX Lin conceived and designed the project. RT Zhang performed most of the experiments. D Zhu, C Shen, and N Wang analyzed the experimental data. Y Le and YX Li conducted a part of experiments. We thank the high-performance computing center at National Key Laboratory of Crop Genetic Improvement in Huazhong Agricultural University. RT Zhang prepared the draft of manuscript. XL Zhang and ZX Lin revised the manuscript.

Funding This work was supported by the National Natural Science Foundation of China (31871669), the Guangdong Basic and Applied Basic Research Foundation (2019A1515110288), and the Projects of Talents Recruitment of Guangdong University of Petrochemical Technology (2019rc112).

Data availability The re_sequencing data and RNA-Seq data in this manuscript have been deposited in NCBI Sequence Read Archive under accession number PRJNA848262 and PRJNA848260.

Declarations

Conflicts of interest The authors declare no conflict of interest.

References

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33:2583–2585

Cao JF, Zhao B, Huang CC, Chen ZW, Zhao T, Liu HR, Hu GJ, Shangguan XX, Shan CM, Wang LJ (2020) The miR319-targeted *GhTCP4* promotes the transition from cell elongation to wall thickening in cotton fiber. Mol Plant 13:1063–1077

Chang J, Yu T, Yang Q, Li C, Xiong C, Gao S, Xie Q, Zheng F, Li H, Tian Z (2018) *Hair*, encoding a single C2H2 zinc-finger protein, regulates multicellular trichome formation in tomato. Plant J 96:90–102

Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD (2007) Toward sequencing cotton (*Gossypium*) genomes. Plant Physiol 145:1303–1310

Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J (2020) Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. Nat Genet 52:525–533

Fang X, Liu X, Wang X, Wang W, Liu D, Zhang J, Liu D, Teng Z, Tan Z, Liu F (2017) Fine-mapping *qFS07.1* controlling fiber strength in upland cotton (*Gossypium hirsutum L.*). Theor Appl Genet 130:795–806

Feng L, Zhou C, Su Q, Xu M, Yue H, Zhang S, Zhou B (2020) Fine-mapping and candidate gene analysis of *qFS-Chr.D02*, a QTL for fibre strength introgressed from a semi-wild cotton into Gossypium hirsutum. Plant Sci 297:110524

Haigler CH, Betancur L, Stiff MR, Tuttle JR (2012) Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Front Plant Sci 3:104

Hu H, He X, Tu L, Zhu L, Zhu S, Ge Z, Zhang X (2016) *GhJAZ2* negatively regulates cotton fiber initiation by interacting with the R2R3-MYB transcription factor *GhMYB 25-like*. Plant J 88:921–935

Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J (2019) *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet 51:739–748

Huang J, Chen F, Guo Y, Gan X, Yang M, Zeng W, Persson S, Li J, Xu W (2021) *GhMYB7* promotes secondary wall cellulose deposition in cotton fibres by regulating *GhCesA* gene expression through three distinct cis-elements. New Phytol 232:1718–1737

Ijaz B, Zhao N, Kong J, Hua J (2019) Fiber quality improvement in upland cotton (*Gossypium hirsutum* L.): quantitative trait loci mapping and marker assisted selection application. Front Plant Sci 10:1585

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol 33:524–530

Lin Z, Zhang Y, Zhang X, Guo X (2009) A high-density integrative linkage map for *Gossypium hirsutum*. Euphytica 166:35–45

Long L, Yang WW, Liao P, Guo YW, Kumar A, Gao W (2019) Transcriptome analysis reveals differentially expressed ERF transcription factors associated with salt response in cotton. Plant Sci 281:72–81

Lu L, Wei W, Li QT, Bian XH, Lu X, Hu Y, Cheng T, Wang ZY, Jin M, Tao JJ, Yin CC, He SJ, Man WQ, Li W, Lai YC, Zhang WK, Chen SY, Zhang JS (2021) A transcriptional regulatory module controls lipid accumulation in soybean. New Phytol 231:661–678

Lu R, Li Y, Zhang J, Wang Y, Zhang J, Li Y, Zheng Y, Li XB (2022) The bHLH/HLH transcription factors *GhFP2* and *GhACE1* antagonistically regulate fiber elongation in cotton. Plant Physiol 189(2):628–643

Machado A, Wu Y, Yang Y, Llewellyn DJ, Dennis ES (2009) The MYB transcription factor *GhMYB25* regulates early fiber and trichome development. Plant J 59:52–62

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

Mitsuda N, Ohme-Takagi M (2009) Functional analysis of transcription factors in Arabidopsis. Plant cell physiol 50:1232–1248Myers ZA, Holt III BF (2018) NUCLEAR FACTOR-Y: still complex after all these years? Curr Opin Plant Biol 45:96–102

Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, Vonrhein C, Moras D, Romier C, Bolognesi M (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell 152:132–143

Paterson AH, Brubaker CL, Wendel JF (1993) A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Rep 11:122–127

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650–1667

Petroni K, Kumimoto RW, Gnesutta N, Calvenzani V, Fornari M, Tonelli C, Holt BF III, Mantovani R (2012) The promiscuous life of plant nuclear factory transcription factors. Plant Cell 24:4777–4792

Qin YM, Hu CY, Zhu YX (2008) The ascorbate peroxidase regulated by H2O2 and ethylene is involved in cotton fiber cell elongation by modulating ROS homeostasis. Plant Signal Behav 3:194–196

Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ (2014) Control of cotton fibre elongation by a homeodomain transcription factor *GhHOX3*. Nat Commun 5:1–9

Sun W, Gao Z, Wang J, Huang Y, Chen Y, Li J, Lv M, Wang J, Luo M, Zuo K (2019) Cotton fiber elongation requires the transcription factor *GhMYB212* to regulate sucrose transportation into expanding fibers. New Phytol 222:864–881

Tan X, Zhang H, Yang Z, Wei Z, Li Y, Chen J, Sun Z (2022) NF-YA transcription factors suppress jasmonic acid-mediated antiviral defense and facilitate viral infection in rice. PLoS Pathog 18:e1010548

Tang W, Tu L, Yang X, Tan J, Deng F, Hao J, Guo K, Lindsey K, Zhang X (2014) The calcium sensor *GhCaM7* promotes cotton fiber elongation by modulating reactive oxygen species (ROS) production. New Phytol 202:509–520

Wang H, Huang C, Guo H, Li X, Zhao W, Dai B, Yan Z, Lin Z (2015a) QTL mapping for fiber and yield traits in upland cotton under multiple environments. PLoS ONE 10:e0130742

Wang J, Li H, Zhang L, Meng L (2015b) QTL IciMapping: Integrated software for building genetic linkage maps and mapping quantitative trait genes. Crop J 3:269–283

Wang H, Huang C, Zhao W, Dai B, Shen C, Zhang B, Li D, Lin Z (2016) Identification of QTL for fiber quality and yield traits using two immortalized backcross populations in upland cotton. PLoS ONE 11:e0166970

Wang M, Tu L, Yuan D, Zhu D, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G (2019) Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat Genet 51:224–229

Wang J, Li G, Li C, Zhang C, Cui L, Ai G, Wang X, Zheng F, Zhang D, Larkin RM, Ye Z, Zhang J (2021a) NF-Y plays essential roles in flavonoid biosynthesis by modulating histone modifications in tomato. New Phytol 229:3237–3252

Wang NN, Li Y, Chen YH, Lu R, Zhou L, Wang Y, Zheng Y, Li XB (2021b) Phosphorylation of *WRKY16* by *MPK3-1* is essential for its transcriptional activity during fiber initiation and elongation in cotton (*Gossypium hirsutum*). Plant Cell 33:2736–2752

Xu P, Gao J, Cao Z, Chee PW, Guo Q, Xu Z, Paterson AH, Zhang X, Shen X (2017) Fine mapping and candidate gene analysis of *qFL-chr1*, a fiber length QTL in cotton. Theor Appl Genet 130:1309–1319

Yang Z, Zhang C, Yang X, Liu K, Wu Z, Zhang X, Zheng W, Xun Q, Liu C, Lu L (2014) *PAG1*, a cotton brassinosteroid catabolism gene, modulates fiber elongation. New Phytol 203:437–448

Yang Z, Ge X, Yang Z, Qin W, Sun G, Wang Z, Li Z, Liu J, Wu J, Wang Y (2019) Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. Nat Commun 10:1–13

Yu J, Jung S, Cheng CH, Lee T, Zheng P, Buble K, Crabb J, Humann J, Hough H, Jones D, Campbell JT, Udall J, Main D (2021) CottonGen: The community database for cotton genomics, genetics, and breeding research. Plants 10(12):2805

Zhang J, Huang GQ, Zou D, Yan JQ, Li Y, Hu S, Li XB (2018) The cotton (*Gossypium hirsutum*) NAC transcription factor (*FSN1*) as a positive regulator participates in controlling secondary cell wall biosynthesis and modification of fibers. New Phytol 217:625–640

Zhang Q, Zhang J, Wei H, Fu X, Ma L, Lu J, Wang H, Yu S (2020) Genome-wide identification of NF-YA gene family in cotton and the positive role of *GhNF-YA10* and *GhNF-YA23* in salt tolerance. Int J Biol Macromol 165:2103–2115

Zhang X, Xue Y, Guan Z, Zhou C, Nie Y, Men S, Wang Q, Shen C, Zhang D, Jin S (2021) Structural insights into homotrimeric assembly of cellulose synthase CesA7 from *Gossypium hirsutum*. Plant Biotechnol J 19:1579–1587