Genome-Wide Association Study of Micronaire Using a Natural Population of Representative Upland Cotton (Gossypium Hirsutum L.)

Jikun Song
Cotton Research Institute https://orcid.org/0000-0002-2812-0747
Weifeng Pei
Cotton Research Institute
Jianjiang Ma
Cotton Research Institute
Shuxian Yang
Cotton Research Institute
Bing Jia
Xinjiang Agricultural University
Yingying Bian
Cotton Research Institute
Yue Xin
Cotton Research Institute
Luyao Wu
Cotton Research Institute
Xinshan Zang
Cotton Research Institute
Yanying Qu
Xinjiang Agricultural University
Jinfa Zhang
New Mexico State University
Man Wu (wuman2004@163.com)
Cotton Institute of the Chinese Academy of Agricultural Sciences
Jiwen Yu
Cotton Research Institute

Research

Keywords: Upland cotton (Gossypium hirsutum L.), fiber micronaire, GWAS, candidate genes, GhFLA9

Posted Date: December 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-114976/v1

Abstract

Background

Micronaire is a comprehensive index reflecting fineness and maturity of cotton fibers. It is one of the important internal quality indicators of cotton fibers and is closely related to the use value of cotton fibers. Understanding the genetic basis of micronaire is required for the genetic improvement of the trait. However, the underlying genetic architecture at the genomic level is not yet clear. The aim of this study was to use GWAS to dissect the genetic mechanism underlying the micronaire trait in 83 representative upland cotton lines grown in multiple environments.

Results

The GWAS of micronaire was conducted using 83 upland cotton accessions and the Cotton 63K Illumina Infinium SNP array. A total of 11 quantitative trait loci (QTLs) on 10 chromosomes for micronaire were detected. Among these 11 QTLs, 27 genes with specific expression patterns were identified and selected. One novel QTL, *qFM-A12-1*, included 12 significant SNPs, and *GhFLA9* was identified as a candidate gene based on the strong, direct LD relationships between the significantly related SNPs and the gene and a haplotype block analysis. *GhFLA9* was highly expressed during secondary wall thickening at 20-25 days post-anthesis (DPA). Its expression level was significantly higher in low-micronaire line (Msco-12) than in high-micronaire line (Chuangyou-9).

Conclusions

This study provides a genetic reference for the genetic improvement of cotton fiber micronaire (FM) and lays a foundation for the functional verification of *GhFLA9.*

Background

Cotton is an economically important crop and is the most important natural fiber raw material in the textile industry (Yu 2018). Due to its high yield and wide adaptability, *G. hirsutum L.* accounts for more than 95% of cotton production worldwide. The cotton fiber yield is of primary importance for cotton growers, whereas the textile industry demands high fiber quality (Ali et al. 2018). The cotton fiber quality components that are important for spinning primarily include length, strength and micronaire. Micronaire is one of the most important fiber characteristics for international cotton classers and spinners. It is a comprehensive indicator of fiber fineness and maturity, and it is an important characteristic in the spinning process (He 2005). High-micronaire fibers are normally coarse, which is undesirable from the point of view of spinning and yarn evenness. Cotton fibers are the longest and fastest-growing cells in plants. Each cotton fiber is composed of a single cell produced on the surface of the ovule. The fiber development process mainly includes four stages: the initial period of fiber development, elongation, secondary wall thickening and maturity (Pang et al. 2019). Cotton fiber micronaire (FM) is mainly affected by the formation characteristics of the fiber secondary wall (Wu et al. 2020). The thickening of

the cotton fiber secondary wall is a complex physiological process that mainly consists of cellulose synthesis and deposition. It usually starts 16–19 days after flowering and is jointly regulated by the expression of many genes (Yan 2010). Therefore, analyzing and identifying candidate genes that regulate FM at the QTL mapping level has important theoretical value for the molecular breeding of cotton quality and elucidating the genetic mechanism of cotton fiber development.

At present, linkage mapping is usually used to detect QTLs related to FM in specific mapping populations (Said et al. 2013, 2015a, 2015b). Four QTLs for FM were detected traits in A01, A02 and A07 by using a population of 143 recombinant inbred lines (RILs) (Fan et al. 2018). Twenty-two QTLs for FM were detected by using 180 RILs, among which 13 QTLs were detected in two or more environments (Ail et al. 2018). In addition, 27 QTLs for FM were detected by using BC_3F_2, $BC_3F_{2:3}$ and $BC_3F_{2:4}$ populations, among which 11 QTLs were located near the same marker in different populations or near linked markers in the same population (Wang et al. 2017a). However, the populations used for linkage mapping usually exhibit a large positioning interval, limited recombination times and low genotype variation. With the rapid development of genome sequencing technologies, genome-wide association studies (GWAS) have been successfully applied for the genetic dissection of fiber quality traits, including FM. Based on the resequencing of markers, 3 and 533 significant SNPs for FM were identified in groups of 362 and 419 diverse upland cotton accessions, respectively (Wang et al. 2017b; Ma et al. 2018). In addition, 503 upland cotton accessions were individually genotyped using the Cotton 63K Illumina single-nucleotide polymorphism (SNP) array, resulting in the identification of 3 stable QTLs associated with the FM trait, located on A05, D05 and D12. Previous studies have identified many QTLs related to FM, which are mainly distributed on chromosomes A03, A07, A12, D03, D08 and D11. Based on orthologs in other organisms, genes such as *GhADF1*, *GhWLIM1a*, and *GhXTH* have been found to be related to the development of fiber cell walls in upland cotton. Fasciclin-like arabinogalactan protein (Fasciclin-like arabinogalactan protein, FLA) is a type of arabinogalactan protein (arabinogalactan protein, AGP) with one or two Fasciclin domains that is usually reported to be involved in plant cell development (Showalter et al. 2001). Previous studies have shown that the FLA protein may play a role in cell elongation and secondary wall maturation. Huang et al. (2008) isolated and identified 19 *GhFLAs* from cotton, 7 of which were highly expressed during fiber development, but genes regulating FM during fiber secondary wall thickening were not reported. Furthermore, this previous study focused only on certain specific genes, and it is still very valuable to explore the whole genetic architecture of FM in association with QTL mapping. To better understand the genetic variation of FM at a natural population level, a diversity panel consisting of 83 upland cotton accessions was genotyped with the Cotton 63K Illumina Infinium SNP array. Fiber micronaire was measured in five different environments. A GWAS was performed to identify SNP loci or QTL regions associated with FM in upland cotton. Candidate genes controlling FM were further predicted by haplotype block analysis in novel and stable QTL regions. These results will lay a foundation for FM improvement through marker-assisted breeding.

Materials And Methods

Plant materials

A diversity panel consisting of 83 representative upland cotton accessions obtained from cotton germplasm collections housed in our laboratory and the low-temperature germplasm genebank of the Cotton Research Institute, Chinese Academy of Agricultural Sciences (CRI-CAAS), was examined. The detailed information of the 83 upland cotton samples was described in a previous study from our laboratory (Ma et al. 2018).

Field experiments and phenotyping

All 83 upland cotton accessions were planted at Anyang (Ay), Henan, China (36.06°N, 114.49°E), in 2014, 2015 and 2016 (designated 14_Ay, 15_Ay and 16_Ay, respectively) and at Alaer (Ale), Xinjiang, China (40.55°N, 81.28°E), in 2016 (designated 16_Ale). These two sites are representative cotton production locations in the Yellow River Valley and northwest inland area, respectively. The area of Sanya (Sy), Hainan, China (18.41°N, 109.20°E), in 2016 (designated 16_Sy) is another representative cotton production area. The field experiments followed a randomized complete block design with three replications. At Anyang, each accession was grown in a single-row plot with 18–23 plants, with a plot length of 4 m and row spacing of 0.38 m, while at Alaer, each accession was grown in a plot with 30–40 plants in two rows, with a plant spacing of 0.1 m and a row spacing of 0.45 m. In addition, the planting patterns in Sanya were characterized by a plant spacing of 0.11 m, a row spacing of 0.38 m, a plot length of 5 m, and a total of 50 plants. The trial management procedures followed standard breeding field practices. A total of 20 randomly selected fully opened middle cotton bolls were manually harvested from each accession in September in each year. The FM of each harvested sample was evaluated using a High-Volume Instrument (HVI) 900 (Test Center of Cotton Fiber Quality affiliated with the Agriculture Ministry of China, Institute of Cotton Research, Chinese Academy of Agriculture Science, Anyang, Henan, China). Statistical analyses, including the descriptive statistical analysis and correlation analysis of FM in the 83 upland cotton accessions across five environments, were performed using SPSS 24.0 (IBM, New York, USA). The combined broad-sense heritability (H^2) of micronaire in different environments was estimated with QTL IciMapping 4.2.0 (Meng et al. 2015).

QTL mapping for FM

SNP genotyping

Genomic DNA from all of the accessions was extracted from young leaf tissue using a quick cetylrimethyl ammonium bromide (CTAB) method (Zhang and Stewart 2000). The SNP genotyping of the association panel was performed using a 63K Illumina Infinium SNP array-based approach (Hulse-Kemp et al. 2015). A total of 15,369 SNP markers were identified and used for the subsequent analysis. The details were reported previously by Ma et al. (2018).

Identification of SNPs associated with FM

Previous studies have analyzed population structure and linkage disequilibrium (Ma et al. 2018). The best linear unbiased predictions (BLUPs) for FM across the five environments were estimated using R software. The BLUP values for the five environments and the phenotypic values for FM in each environment were used for GWAS. The general linear model (GLM) in TASSEL 5.0 was used for GWAS, and the population structure was considered a fixed effect. The Bonferroni threshold of SNP significance was $P < 6.51 \times 10^{-5}$ ($p = 1/n$, n = the number of SNPs, -log10(1/15,369) ≈ 4.19) (Li et al. 2013; Yang et al. 2014; Liu et al. 2016). Manhattan plots were generated using the R software package "CMplot" (Turner 2014). Based on descriptions from recent studies, stable QTLs were defined as those identified in two or more different environments in the study. The QTLs were named as follows: q + trait abbreviation - chromosome - QTL number (Zhang et al. 2015).

Candidate gene identification

Transcriptome sequencing data (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA490626/) from *G. barbadense* Hai7124 and *G. hirsutum* TM-1 tissues including 0 days post-anthesis (0 DPA), 1 DPA, 3 DPA, 5 DPA, fiber at 10 days post-anthesis (fiber-10 DPA), fiber-20 DPA and fiber-25 DPA were available (Hu et al. 2019). Regarding the screening criteria, the fiber expression levels at 20–25 DPA were higher than those at 0–15 DPA, and the expression levels measured in the fiber at 20–25 DPA were significantly different between the two materials. The genes related to FM were further screened according to gene function annotation. Furthermore, according to the stable *qFM-A12-1* SNPs, R software was used to construct haplotype blocks with reference to the transcriptome database and gene function annotations, and differentially expressed genes were screened in the intervals (Su et al. 2016).

To further verify the expression trends of the candidate genes, qRT-PCR analysis was performed. The total RNA of fibers collected from five developmental time points (5, 10, 15, 20 and 25 DPA) in the high-micronaire line (Msco-12) and the low-micronaire line (Chuangyou-9) was extracted with the FastPure Plant Total RNA Isolation Kit (Polysaccharides & Polyphenolics–rich) (Nanjing Vazyme BIOTECH CO., LTD.). And cDNA was synthesized using a reverse transcription kit (HiScript II Q RT SuperMix for qPCR). Furthermore, the ChamQ Universal SYBR quantitative fluorescent enzyme and the quantitative fluorescent PCR instrument (ABI7500) produced by Eppendorf Germany were used to carry out qRT-PCR verification. The qRT-PCR conditions were as follows: the first step was 94 °C (30 s) for DNA polymerase activation; the second step was 40 cycles of 94 °C (5 s), 58 °C (15 s), and 72 °C (12 s); the third step was to add a default process of melting curve analysis; and the last step was 12 °C (1 min). In this study, Histone3 (AF024716) was used as a housekeeping gene (Tu et al. 2007). The experimental design included 3 technical and 3 biological replicates for each gene, and the $2^{-\Delta\Delta CT}$ method was used to calculate the relative gene expression value (Livak et al. 2001).

Results

Phenotypic variation of FM

Significant variation in FM was observed among the 83 upland cotton accessions, ranging from 2.73 to 6.50, with an average of 4.65 (Table 1 and Fig. S1). In the 14_Ay, 15_Ay, 16_Ay, 16_Sy and 16_Ale environments, the natural population exhibited average FM values of 5.00, 4.76, 4.92, 4.23 and 4.32, respectively. The coefficients of variation (CVs) for FM ranged from 9.09–14.77% in the five environments, indicating that the cotton FM index shows adequate variation in this population (Table 1). As shown in Fig. 1, FM was normally distributed in the five environments. The analysis showed that the FM index in the cotton area of the Yellow River Valley is relatively thicker, while in the northwest inland cotton area, it is relatively finer. Correlation analysis showed that the FM trait exhibited a very significant positive correlation in the five environments, and the correlation coefficient was between 0.4 and 0.7 (Fig. 2). ANOVA indicated that the genotype and environment had significant effects on FM ($P < 0.01$), and the estimated broad-sense heritability was 86.27% (Table 1, Table S1). These results suggest that FM in upland cotton is highly heritable, although the role of environmental factors in fiber development should not be ignored. The results indicated that this population is suitable for the GWAS of the cotton FM trait.

Table 1
Phenotypic statistics of fiber micronaire in upland cotton in five environments

Environment	Min	Max	Mean	SD	CV(%)	Skewness	Kurtosis	H^2(%)
14_Ay	3.46	5.95	5.00	0.45	9.09	-0.80	1.56	86.27%
15_Ay	3.15	5.80	4.76	0.53	11.16	-0.61	0.48	
16_Ay	2.73	5.99	4.92	0.53	10.80	-1.17	2.87	
16_Sy	2.90	6.50	4.23	0.62	14.77	0.29	1.13	
16_Ale	3.08	5.50	4.32	0.60	13.80	-0.14	-0.52	

GWAS

In this study, GWAS were conducted for FM using BLUPs across five environments and one individual environment in the GLM (Q). A threshold of -log10 (P) > 4.19 was considered to indicate significant SNPs. A total of 18, 37, 67, 21 and 7 SNPs were significantly associated with FM in 14_Ay, 15_Ay, 16_Ay, 16_Sy and 16_Ale, respectively (Fig. 3; Fig. S2). Through further analysis of the significantly related SNPs, a total of 55 common SNPs were obtained, among which 16 SNPs could be detected in two environments. These 55 significantly related SNPs were distributed on 10 chromosomes: A10, A12, A13, D03, D06, D08, D09, D10, D11, and D12, among which there were 12 significant SNPs on chromosome A12. According to the specific LD value of each chromosome, the QTL interval is the significant SNP position ± LD (Ma et al. 2018). According to this approach, a total of 11 stable QTLs related to FM were obtained, explaining 18.10–31.64% of the phenotypic variation in FM. Among these SNPs, the phenotypic contribution rates of 11 markers in *qFM-A12-1* were more than 20%. To further verify the reliability of the 11 stable QTLs related to FM obtained in this study, we compared the GWAS results with those of previous association

and linkage studies for FM. We compiled the SSR and SNP markers of the previous QTLs related to FM and aligned the previously studied marker sequences to the *G. hirsutum* TM-1 genome using the local BLAST method (Zhang et al. 2015). Through comparison, it was found that eight of the 11 stable QTLs for FM were reported in previous studies and that 3 QTLs were newly detected in this study (Table 2).

Table 2
Summary of SNPs significantly associated with fiber micronaire in five environments

QTL	Marker	Chr.	Position(bp)	$-\log_{10}(p)$	R^2(%)	Environment	Reported previously
qFM-A10-1	i12339Gh	A10	100,279,301	4.85	20.89	16_Ay	(Jamshed et al. 2016)
	i12239Gh	A10	100,587,583	6.84	29.39	15_Ay	
qFM-A12-1	i61168Gt	A12	67,381,479	4.91	20.9	16_Ay	
	i08143Gh	A12	68,066,658	4.41	18.97	16_Ale	
	i27811Gh	A12	68,205,857	4.22–4.78	21.35–24.04	16_Ay 16_Sy	
	i41990Gh	A12	68,321,416	4.22–4.78	21.35–24.04	16_Ay 16_Sy	
	i47170Gh	A12	68,410,009	4.22–4.78	21.35–24.04	16_Ay 16_Sy	
	i38887Gh	A12	68,472,462	4.49–5.59	19.32-24.00	16_Ay 16_Sy BLUP	
	i21853Gh	A12	68,540,285	4.22–4.78	21.35–24.04	16_Ay 16_Sy	
	i43838Gh	A12	68,551,427	4.22–4.78	21.35–24.04	16_Ay 16_Sy	
	i08151Gh	A12	68,597,376	4.34–4.73	20.34–28.45	16_Ay 16_Sy	
	i32282Gh	A12	68,673,029	4.48–5.27	19.91–23.44	16_Ay 16_Sy BLUP	
	i27971Gh	A12	68,695,670	4.53	22.95	16_Sy	
	i41399Gh	A12	68,715,788	4.53	22.95	16_Sy	
qFM-A13-1	i20979Gh	A13	2,132,189	4.23	28.53	14_Ay	(Wang et al. 2017; Jamshed et 2016)
	i13041Gh	A13	2,704,662	4.29	28.23	16_Ay	
qFM-A13-2	i35441Gh	A13	74,952,854	6.64	31.45	15_Ay	
	i32750Gh	A13	75,105,822	4.39–4.60	22.35-23.00	15_Ay 16_Sy	
	i46646Gh	A13	75,109,913	5.14–5.39	22.12–22.95	15_Ay 16_Sy BLUP	

QTL	Marker	Chr.	Position(bp)	-log_{10}(p)	R^2(%)	Environment	Reported previously
	i41773Gh	A13	75,116,067	5.14–5.39	22.12–22.95	15_Ay 16_Sy BLUP	
	i33120Gh	A13	75,129,388	4.32–4.63	22.25–23.65	15_Ay 16_Sy BLUP	
	i30758Gh	A13	75,130,352	4.21	18.1	16_Ale	
	i35636Gh	A13	75,242,672	5.14–5.39	22.12–22.95	15_Ay 16_Sy	
	i01190Gh	A13	75,398,430	4.26	18.29	16_Ale	
	i42228Gh	A13	75,472,091	4.35	19.55	16_Ale	
	i13652Gh	A13	75,504,093	4.44	19.09	16_Ale	
qFM-D03-1	i03515Gh	D03	42,362,797	4.63	23.13	16_Ay	(Diouf et al. 2018)
	i21465Gh	D03	43,276,036	4.91	20.9	16_Ay	
	i61219Gt	D03	44,406,129	4.39	22.35	15_Ay	
	i63893Gm	D03	44,581,155	4.46–6.29	19.27–26.62	15_Ay BLUP	
	i43459Gh	D03	45,627,279	4.31	18.31	15_Ay	
	i14956Gh	D03	45,689,086	4.31	18.31	15_Ay	
qFM-D06-1	i10547Gh	D06	547,763	4.44–4.81	20.47–25.26	14_Ay 15_Ay	(Chen et al. 2018)
	i19728Gh	D06	689,943	4.76	23.7	15_Ay	
qFM-D08-1	i15175Gh	D08	58,849,763	4.56	30.39	14_Ay	(Wang et al. 2017; Jia et al. 2018)
	i15177Gh	D08	58,905,788	4.56	30.39	14_Ay	
	i04608Gh	D08	61,265,566	4.86	20.71	15_Ay	
	i04610Gh	D08	61,327,863	5.14	25.34	15_Ay	
	i44633Gh	D08	61,354,352	5.14	25.34	15_Ay	
	i47262Gh	D08	61,379,191	4.98	21.45	15_Ay	
qFM-D09-1	i38863Gh	D09	29,968,447	4.91	20.9	16_Ay	(Jia et al. 2018)
	i32012Gh	D09	30,108,133	4.91	20.9	16_Ay	

QTL	Marker	Chr.	Position(bp)	-log_{10}(p)	R^2(%)	Environment	Reported previously
	i45347Gh	D09	30,133,934	4.91	20.9	16_Ay	
	i27999Gh	D09	30,418,262	4.75	23.66	16_Ay	
	i19190Gh	D09	30,866,022	4.43	18.84	15_Ay	
	i06010Gh	D09	30,870,947	4.43	18.84	15_Ay	
	i06011Gh	D09	30,873,661	4.43	18.84	15_Ay	
qFM-D10-1	i11504Gh	D10	877,970	4.58–5.85	21.02–26.08	14Ay 16Sy BLUP	
	i11512Gh	D10	944,253	4.89	21.04	16_Sy	
	i19999Gh	D10	946,181	4.28	21.83	16_Sy	
	i11564Gh	D10	1,938,451	4.30–4.36	18.48–18.72	16_Sy 16Ale	
qFM-D11-1	i00901Gh	D11	5,706,308	6.36	31.64	15_Ay	(Zhang et al. 2016; Jia et al. 2018)
	i50152Gb	D11	5,792,360	4.51	22.88	16_Sy	
qFM-D12-1	i08142Gh	D12	43,263,390	4.51	26.42	14_Ay	(Huang et al. 2017)
	i08144Gh	D12	43,267,732	4.41	28.97	16_Ale	

Identification of candidate genes in 11 stable QTLs

In this study, we used the method of combining the functional annotation of genes in stable QTLs with the analysis of Arabidopsis homologous genes and the cotton fiber transcriptome database to mine candidate genes related to FM (Du et al. 2018). The 11 QTLs contained a total of 1,594 genes, which were compared to the upland cotton genome from Zhejiang University using local BLAST (http://ibi.zju.edu.cn/cotton). The fiber quality of *G. barbadense* is significantly better than that of *G. hirsutum*, and there are significant differences in FM between *G. barbadense* Hai7124 and *G. hirsutum* TM-1. The Hai7124 and TM-1 transcriptome databases were employed to screen the differences in gene expression levels during fiber development (0, 1, 3, 5, 10, 20, and 25 DPA). The fiber expression levels at 20 to 25 DPA were higher than those at 0 to 15 DPA, and the expression levels in the fibers of the two materials were significantly different. A total of 27 genes were initially screened (Table 3, Fig. 4). Through the annotation analysis of the 27 differentially expressed genes, genes related to fiber development that have been previously reported by researchers, such as *GhTUB6* and *GhFLA*, were identified.

Table 3
Candidate genes for micronaire

Gene	Physical position	Homologs in A. thaliana	Functional annotation
Gh_A10G2157	A10:100434822–100435289	*AT3G11110*	RING/U-box superfamily protein
Gh_A10G2169	A10:100571847–100572047	-	-
Gh_A10G2174	A10:100618393–100647910	*AT4G27190*	NB-ARC domain-containing disease resistance protein
Gh_A10G2176	A10:100676093–100688392	-	-
Gh_A12G1287	A12:68269264–68269998	*AT1G03870*	FASCICLIN-like arabinoogalactan 9(FLA9)
Gh_A13G0120	A13:1383789–1385806	*AT1G52340*	NAD(P)-binding Rossmann-fold superfamily protein
Gh_A13G0121	A13:1394586–1395093	*AT2G41430*	dehydration-induced protein (ERD15)
Gh_A13G0133	A13:1519272–1521026	*AT5G17540*	HXXXD-type acyltransferase family protein
Gh_A13G0188	A13:2123316–2124229	*AT2G37590*	DNA binding with one finger 2.4(DOF2.4)
Gh_A13G0220	A13:2631476–2632431	*AT3G01390*	vacuolar membrane ATPase 10(VMA10)
Gh_A13G1555	A13:74370130–74370630	*AT1G09310*	Protein of unknown function, DUF538
Gh_A13G1563	A13:74415625–74417872	*AT1G09380*	nodulin MtN21/EamA-like transporter family protein
Gh_A13G1570	A13:74527856–74528134	-	-
Gh_A13G1573	A13:74594101–74600768	*AT2G24520*	H(+)-ATPase 5(HA5)
Gh_A13G1640	A13:75324608–75326119	*AT5G25820*	Exostosin family protein
Gh_D03G1332	D03:41738284–41744222	*AT5G22920*	CHY-type/CTCHY-type/RING-type Zinc finger protein
Gh_D03G1452	D03:43497008–43499238	*AT5G12250*	beta-6 tubulin(TUB6)

Gene	Physical position	Homologs in A. thaliana	Functional annotation
Gh_D03G1533	D03:44563981–44565739	*AT1G62790*	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
Gh_D06G0032	D06:322107–322808	*AT1G75390*	basic leucine-zipper 44(bZIP44)
Gh_D06G0041	D06:479879–483912	*AT5G42800*	dihydroflavonol 4-reductase
Gh_D09G0646	D09:29655772–29660487	*AT3G17180*	serine carboxypeptidase-like 33(scpl33)
Gh_D10G0078	D10:622188–623235	*AT4G16380*	Heavy metal transport/detoxification superfamily protein
Gh_D11G0597	D11:5211532–5212612	*AT3G51520*	diacylglycerol acyltransferase family
Gh_D11G0639	D11:5582090–5583296	*AT5G13870*	xyloglucan endotransglucosylase/hydrolase 5
Gh_D11G0718	D11:6198554–6200040	*AT3G28150*	TRICHOME BIREFRINGENCE-LIKE 22(TBL22)
Gh_D11G0750	D11:6489218–6489475	*AT5G54100*	SPFH/Band 7/PHB domain-containing membrane-associated protein family
Gh_D12G1409	D12:43357297–43358031	*AT1G03870*	FASCICLIN-like arabinoogalactan 9(FLA9)

Mining of candidate genes in qFM-A12-1

qFM-A12-1 contained the most significant SNPs, and its phenotypic contribution rate was stable and high, but previous studies have not located any genes related to FM at this locus. Therefore, this study focused on the analysis of *qFM-A12-1* to identify candidate genes affecting FM. To further narrow the QTL interval, the haplotype block of *qFM-A12-1* was constructed, and the interval was further narrowed to between i27811Gh and i41399Gh according to the direct strong LD relationships between the significantly related SNPs and the genes (Fig. 5a). The i41990Gh marker in the interval was used to genotype upland cotton. It was found that the accessions carrying the AA (44) allele exhibited a significantly higher micronaire than those with the GG (35) allele (Fig. 5b), indicating that this interval may significantly affect FM; Chuangyou-9 and Msco-12 harbor the AA and GG alleles, respectively. *GhFLA9* is located between the i27811Gh and i41990Gh markers, and its fiber expression level at 20 to 25 DPA is significantly higher than that at 5 to 10 DPA in Hai7124 and TM-1. The highest expression level is observed at 20 DPA, and the fiber expression level of *GhFLA9* at 20 to 25 DPA in Hai7124 is significantly higher than that in TM-1 (Fig. 4). In this study, qRT-PCR was used to determine the expression level of *GhFLA9* in different fiber development stages (5, 10, 15, 20 and 25 DPA) in the high-micronaire genotype

and the low-micronaire genotype. The expression level of *GhFLA9* increased throughout the fiber development period. In addition, the expression levels in fiber at 20 and 25 DPA were significantly different between Msco-12 and Chuangyou-9, with Msco-12 showing significantly higher expression than Chuangyou-9 (Fig. 6). The above data indicate that this gene may be involved in the regulation of FM, acting as a negative regulator.

Discussion

The quality of cotton fiber depends mainly on being “long, strong and fine”, where “fine” refers to micronaire, which has a direct impact on the processing of cotton fiber and the quality of the obtained products. A moderate FM confers cotton fiber with high yarn strength and good spinning quality. Therefore, the improvement of cotton FM has become a hot spot in research on the genetic improvement of cotton fiber quality. In the early stage of the project, the 63K Illumina Infinium SNP array was applied to 83 upland cotton accessions, and 15,369 high-quality SNPs were obtained. The population was used to perform an association analysis of dynamic fiber length and oil traits, and candidate genes were mined separately (Ma et al. 2018; Ma et al. 2019). These studies indicated that this population is suitable for the GWAS of related traits. The FM of the 83 upland cotton accessions that we used in this study showed abundant phenotypic variation and a significant variation range in the 5 environments (Table 1). The H^2 of FM was relatively high (86.27%) in this study and was very similar to the H^2 (76.37-94.00%) values obtained in previous studies (Sun et al. 2010; Ma et al. 2020). The above results indicate that the FM trait is mainly affected by genotype and is suitable for GWAS. Through the GWAS of FM in 5 environments, 11 stable QTLs containing 55 significant SNPs were obtained, among which 4 QTLs were located in the At subgenome and 7 in the Dt subgenome (Table 2). We compared the mapped interval with the previous QTL mapping results of micronaire traits using genetic population linkage analysis or natural population association analysis methods and found that 8 of the 11 QTLs identified in this study were comparable to QTLs identified in previous studies (Jamshed et al. 2016; Zhang et al. 2016; Huang et al. 2017; Wang et al. 2017a; Wang et al. 2017b; Chen et al. 2018; Diouf et al. 2018; Jia et al. 2018). One of these previous studies (Huang et al. 2017) used the Illumina Infinium SNP array for a GWAS of 503 upland cotton accessions and obtained the stable QTL *qGhMV-c26-1*, which coincides with *qFM-D12-1* from the present study, and our interval is relatively small. This result indicates that the GWAS of FM conducted in 83 upland cotton accessions in this study is credible. Our GWAS of the FM trait combined with transcriptome and qRT-PCR data analyses identified genes related to fiber development such as *GhTUB6* and *GhFLA*. *GhTUB6* is predominantly expressed in the period of the rapid elongation of cotton fiber cells and primary cell wall synthesis (He et al. 2008). *GhFLA* genes are not only involved in cotton fiber development but can also promote the expression of primary wall synthesis genes (Huang et al. 2008). The new QTLs found in this study with stability in multiple environments provide references for molecular marker-assisted selection breeding targeting cotton micronaire.

The stable QTL *qFM-A12-1* identified in this study has not been previously reported, and this QTL contains 12 adjacent significant SNPs. Therefore, we focused on the differential expression of genes in

this interval. The interval was further narrowed according to the haplotype block and then based on the transcriptome data. The Fasciclin-like arabinogalactan protein gene *GhFLA9* was screened, which is mainly expressed in the secondary wall thickening stage (20–25 DPA) during fiber development, and its expression trend was negatively correlated with micronaire (Fig. 4). Previous studies have found that this gene shows the highest expression in Kezi cotton fibers at 10 DPA (Huang et al. 2008), which is different from the expression trend observed in the most recent *G. hirsutum* TM-1 transcriptome, indicating that the expression of this gene may differ among different materials. Subsequently, we performed genotyping analysis based on the significant SNP (i41990Gh) and found that the high-micronaire line (Chuangyou-9) and the low-micronaire line (Msco-12) belonged to two genotypes with significant differences. The subsequent qRT-PCR analysis of *GhFLA9* in the fiber tissues of the two materials at five stages (5, 10, 15, 20 and 25 DPA) showed that the expression trend of this gene was consistent with the expression trend in the transcriptome. In the secondary wall thickening stage (20–25 DPA), the expression level in the low-micronaire material was higher than that in the high micronaire material. The above results indicate that the *GhFLA9* expression trend identified in this study is relatively reliable.

Conclusions

GWAS of FM was performed in 5 environments, and 11 QTLs related to FM were detected, among which 8 QTLs colocalized with QTLs identified in previous studies, and 3 QTLs were newly discovered in this study. Twenty-seven candidate genes related to FM were screened from the 11 stable QTLs, and haplotype block analysis and qRT-PCR analysis of the stable QTL *qFM-A12-1* showed that *GhFLA9* may negatively regulate the development of cotton FM.

Abbreviations

FM: Fiber micronaire; DPA: Days post-anthesis; GWAS: Genome-wide association study; LD: Linkage disequilibrium; QTL: Quantitative trait locus; SNP: Single-nucleotide polymorphism

FLA: Fasciclin-like arabinogalactan

Declarations

Ethics approval and consent to participate

Not applicable

Consent to publish

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

This research was financed by National Key Research and Development program of China (grant nos. 2018YFD0101402, 2018YFD0100300 and 2016YFD0101400), and the Natural Science Foundation of Xinjiang Uygur Autonomous Region of China (grant nos. 2020D01A135), and Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences.

Authors' contributions

Yu JW, Wu M conceived and designed the research. Song JK, Pei WF and Ma JJ performed the experiments. Yang SX, Jia B, Bian YY, Xin Y, Wu LY and Zang XS performed the field cultivation of cotton plants. Song JK wrote the paper. Yu JW, Wu M, Zhang JF and Qu YY revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors wish to thank New Mexico Agricultural Experiment Station, New Mexico, USA

References

Ali I, Teng ZH, Bai YT, et al. A high density SLAF-SNP genetic map and QTL detection for fibre quality traits in Gossypium hirsutum. BMC Genomics. 2018;19(1):879.

Chen Y, Liu GD, Ma HH, et al. Identification of Introgressed Alleles Conferring High Fiber Quality Derived From Gossypium barbadense L. in Secondary Mapping Populations of G. hirsutum L. Front Plant Sci. 2018;9:1023.

Diouf L, Magwanga RO, Gong WF, et al. QTL Mapping of Fiber Quality and Yield-Related Traits in an Intra-Specific Upland Cotton Using Genotype by Sequencing (GBS). Int J Mol Sci. 2018;19(2):441.

Du XM, Liu SY, Sun JL, et al. Dissection of complicate genetic architecture and breeding perspective of cottonseed traits by genome-wide association study. BMC Genomics. 2018;19(1):451.

Fan LP, Wang LP, Wang XY, et al. A high-density genetic map of extra-long staple cotton (*Gossypium barbadense*) constructed using genotyping-by-sequencing based single nucleotide polymorphic markers and identification of fiber traits-related QTL in a recombinant inbred line population. BMC Genomics. 2018;19(1):489.

Han LB, Li YB, Wang HY, et al. The dual functions of WLIM1a in cell elongation and secondary wall formation in developing cotton fibers. The Plant Cell. 2013;25(11):4421-4438.

He JX. Micronaire value is an important parameter for cotton quality. Chinese Cotton Processing. 2005; (3):25-26.

He XC, Qin YM, Xu Y, et al. Molecular cloning, expression profiling, and yeast complementation of 19 beta-tubulin cDNAs from developing cotton ovules. J Exp Bot. 2008;59(10):2687-2695.

Hu Y, Chen JD, Fang L, et al. Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet. 2019;51(4):739-748.

Huang C, Nie XH, Shen C, et al. Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. Plant Biotechnol J. 2017;15(11):1374-1386.

Huang GQ, Xu WL, Gong SY, et al. Characterization of 19 novel cotton FLA genes and their expression profiling in fiber development and in response to phytohormones and salt stress. Physiol Plant. 2008;134(2):348-359.

Hulse-Kemp AM, Lemm J, Plieske J, 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.

Jamshed M, Jia F, Gong JW, et al. Identification of stable quantitative trait loci (QTLs) for fiber quality traits across multiple environments in *Gossypium hirsutum* recombinant inbred line population. BMC Genomics. 2016;17:197.

Jia XY, Wang HT, Pang CY, et al. QTL delineation for five fiber quality traits based on an intra-specific *Gossypium hirsutum* L. recombinant inbred line population. Mol Genet Genomics. 2018;293(4):831-843.

Li H, Peng ZY, Yang XH, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet. 2013;45(1):43-50.

Liu S, Fan CC, Li JN, et al. A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*. Theor Appl Genet. 2016;129(6):1203-1215.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 2001;25:402-408.

Ma JJ, Geng YH, Pei WF, et al. Genetic variation of dynamic fiber elongation and developmental quantitative trait locus mapping of fiber length in upland cotton (*Gossypium hirsutum L.*). BMC Genomics. 2018;19(1):882.

Ma JJ, Liu J, Pei WF, et al. Genome-wide association study of the oil content in upland cotton (*Gossypium hirsutum L.*) and identification of GhPRXR1, a candidate gene for a stable QTLqOC-Dt5-1. Plant Sci. 2019;286:89-97.

Ma LL, Su Y, Nie HS, et al. QTL and genetic analysis controlling fiber quality traits using paternal backcross population in upland cotton. J Cotton Res. 2020;3:22.

Ma ZY, He SP, Wang XF, et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat Genet. 2018;50(6):803-813.

Michailidis G, Argiriou A, Darzentas N, et al. Analysis of xyloglucan endotransglycosylase/hydrolase (XTH) genes from allotetraploid (*Gossypium hirsutum*) cotton and its diploid progenitors expressed during fiber elongation. J Plant Physiol. 2009;166(4):403-416.

Pang CY, Wang H, Pang Y, et al. Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and Arabidopsis root hair elongation. Mol Cell Proteomics. 2010;9(9):2019-2033.

Said JI, Lin ZX, Zhang XL, et al. 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.

Said JI, Knapka JA, Song MZ, et al. Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between *Gossypium hirsutum* and *G. hirsutum × G. barbadense* populations. Mol Genet Genomics. 2015;290(4):1615-25.

Said JI, Song MZ, Wang HT, et al. A comparative meta-analysis of QTL between intraspecific *Gossypium hirsutum* and interspecific *G. hirsutum × G. barbadense* populations. Mol Genet Genomics. 2015;290(3):1003-25.

Showalter AM. Arabinogalactan-proteins: structure, expression and function. Cell Mol Life Sci. 2001;58(10):1399-1417.

Su JJ, Fan SL, Li LB, 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.

Sun FD, Li JW, Liu AY, et al. Analysis of genetic variation for fiber quality and yield traits in an recombinant inbred population of *Gossypium hirsutum L*. Cotton Science. 2010;22(04):319-325.

Tu LL, Zhang XL, Liu DQ, et al. Suitable internal control genes for qRT-PCR normalization in cotton fiber development and somatic embryogenesis. Chinese Science Bulletin. 2007;52(22):3110-3117.

Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Biorxiv. 2014:005165.

Wang BH, Zhuang ZM, Zhang ZS, et al. Advanced Backcross QTL Analysis of Fiber Strength and Fineness in a Cross between *Gossypium hirsutum* and *G. mustelinum*. Front Plant Sci. 2017a;8:1848.

Wang MJ, Tu LL, Lin M, et al. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. Nat Genet. 2017b;49(4):579-587.

Wang HY, Wang J, Gao P, et al. Down-regulation of *GhADF1* gene expression affects cotton fibre properties. Plant Biotechnol J. 2009;7(1):13-23.

Wu M, Li LY, Pei WF, et al. Identification of differentially expressed genes in developing cotton fibers between two groups of backcross inbred lines differing in fiber micronaire. Cotton Science. 2020;32(1):52-62.

Yan HC. Analysis of the digital gene expression profiling of cotton fiber development during cell elongation and secondary cell wall biogenesics. MD Dissertation of Graduate School of Huangzhong Agricultural University, 2010. pp 13-14.

Yang N, Lu YL, Yang XH, et al. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLoS Genetics. 2014;10(9):e1004573.

Yu SX. The development of cotton production in the recent hundred years of China. Journal of Agriculture. 2018;8(1):85-91.

Zhang JF, Stewart JW. Economical and rapid method for extracting cotton genomic DNA. J Cotton Sci. 2000;4(3):193-201.

Zhang SW, Zhu XF, Feng LC, et al. Mapping of fiber quality QTLs reveals useful variation and footprints of cotton domestication using introgression lines. Sci Rep. 2016;6:31954.

Zhang TZ, Hu Y, Jiang WK, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum*L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015;33(5):531–537.

Zhang Z, Li JW, Muhammad J, et al. High resolution consensus mapping of quantitative trait loci for fiber strength, length and micronaire on chromosome 25 of the upland cotton (*Gossypium hirsutum* L.). PLoS One. 2015;10(8):e0135430.

Figures

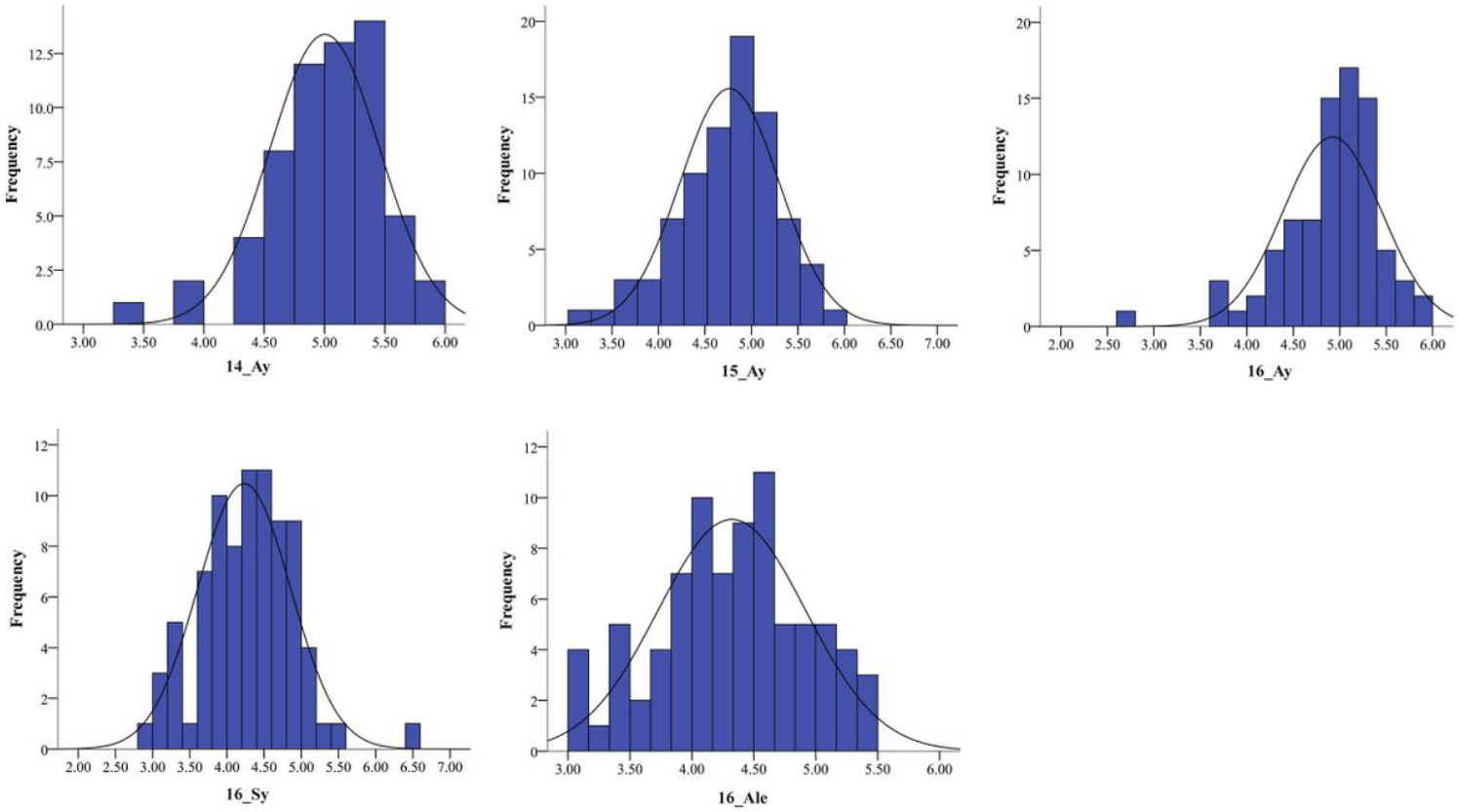

Figure 1

Frequency distribution of fiber micronaire in upland cotton in five environments

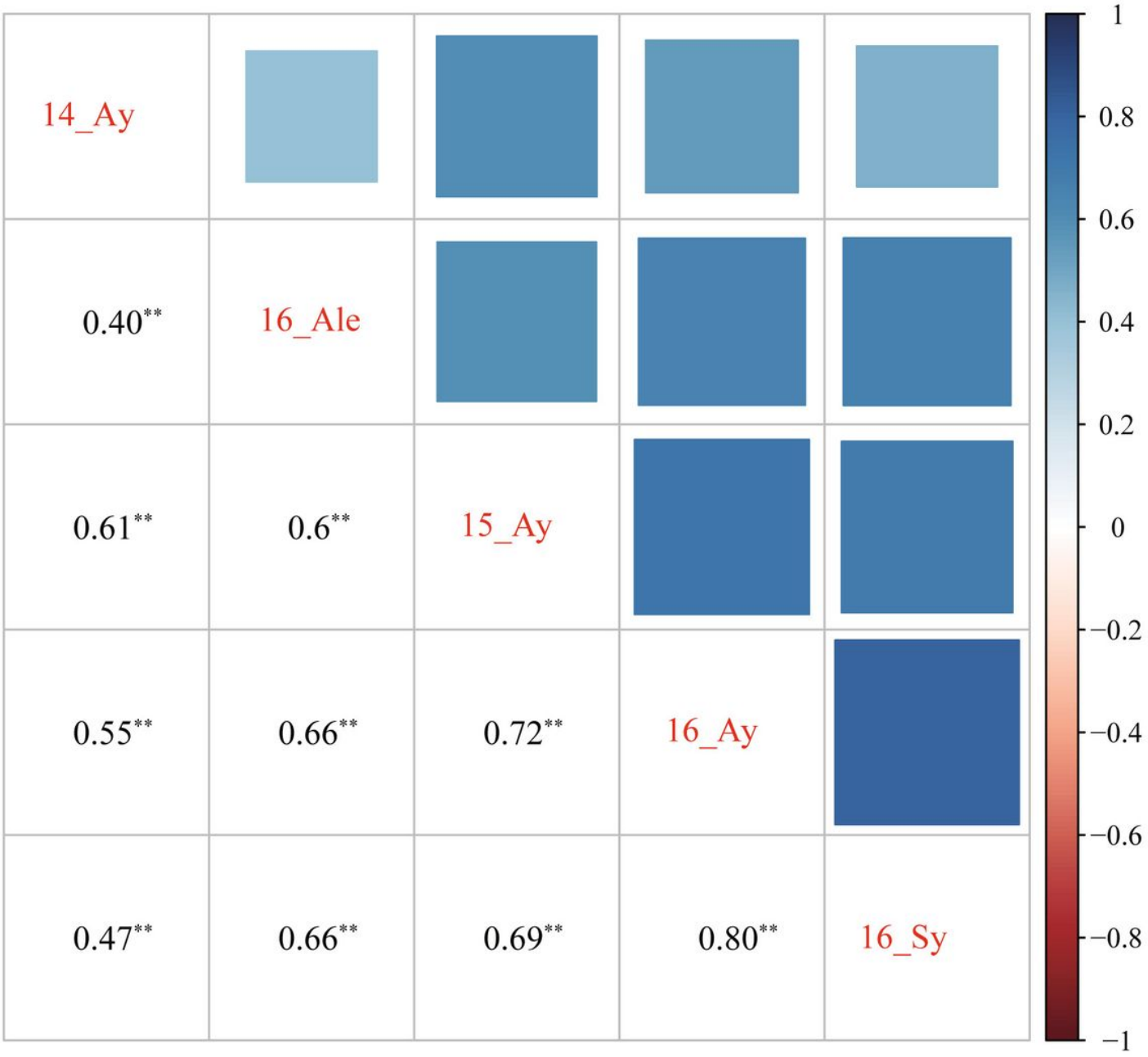

Figure 2

Correlation analysis of fiber micronaire in upland cotton in five environments

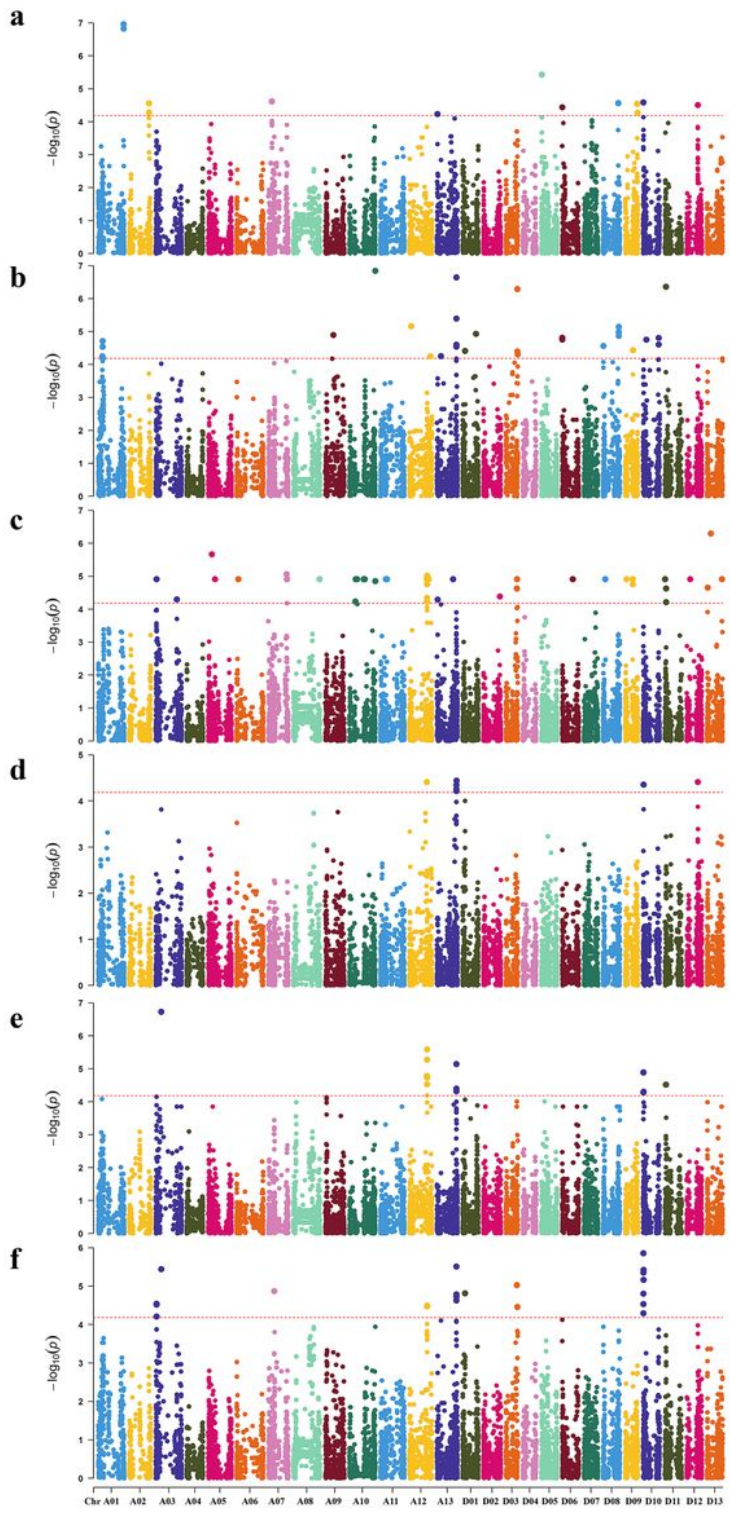

Figure 3

Manhattan plots of fiber micronaire in upland cotton in five environments with BLUP The lowercase letters a, b, and c represent the Manhattan plots of the GLM in Anyang in 2014, 2015 and 2016, respectively; the letters d and e represent the Manhattan plots of the GLM in Alaer and Sanya in 2016, respectively; and f represents the Manhattan plots of the GLM for BLUP.

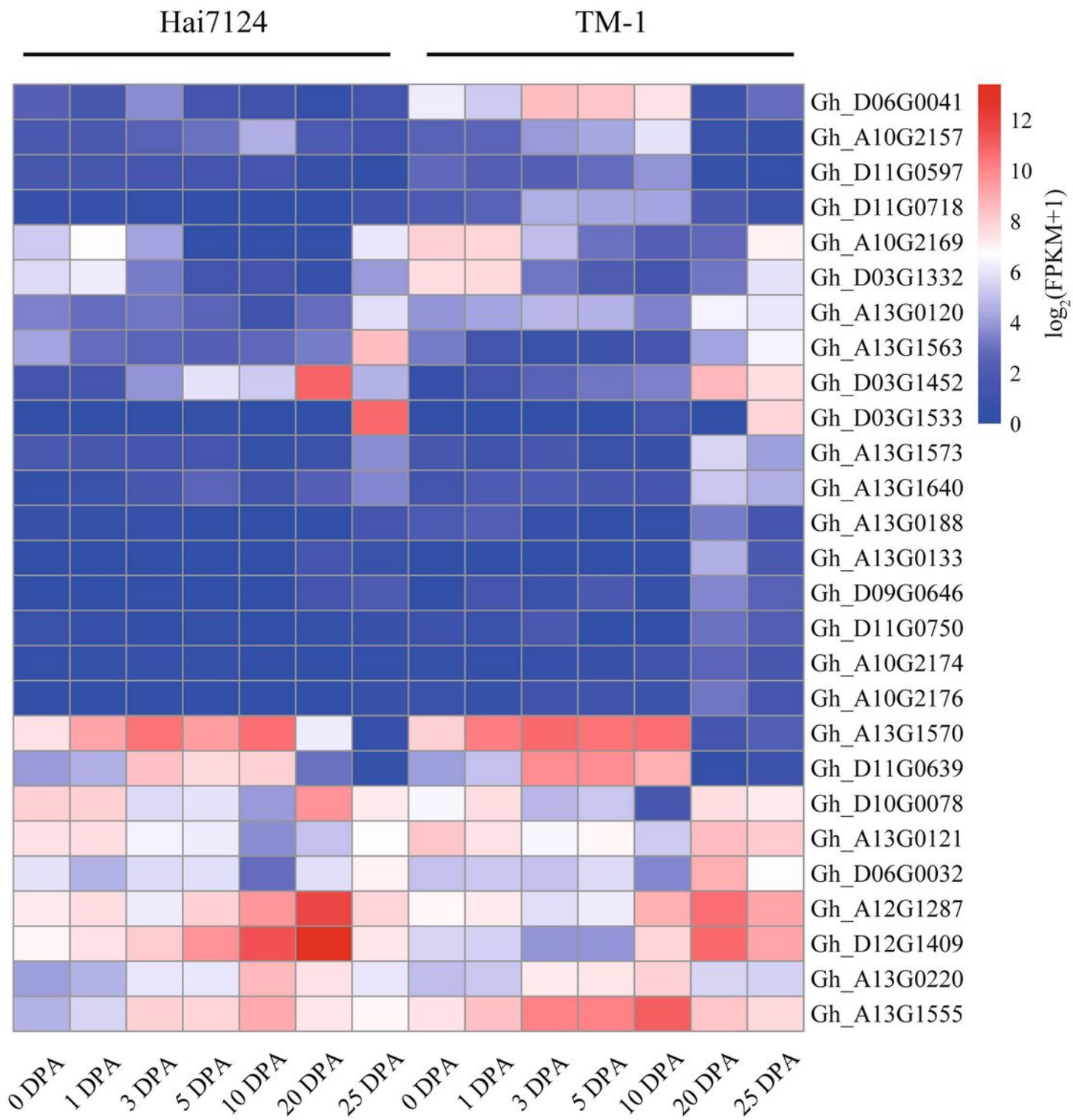

Figure 4

The expression patterns of 27 genes in fibers at seven different developmental time points in Hai7124 and TM-1

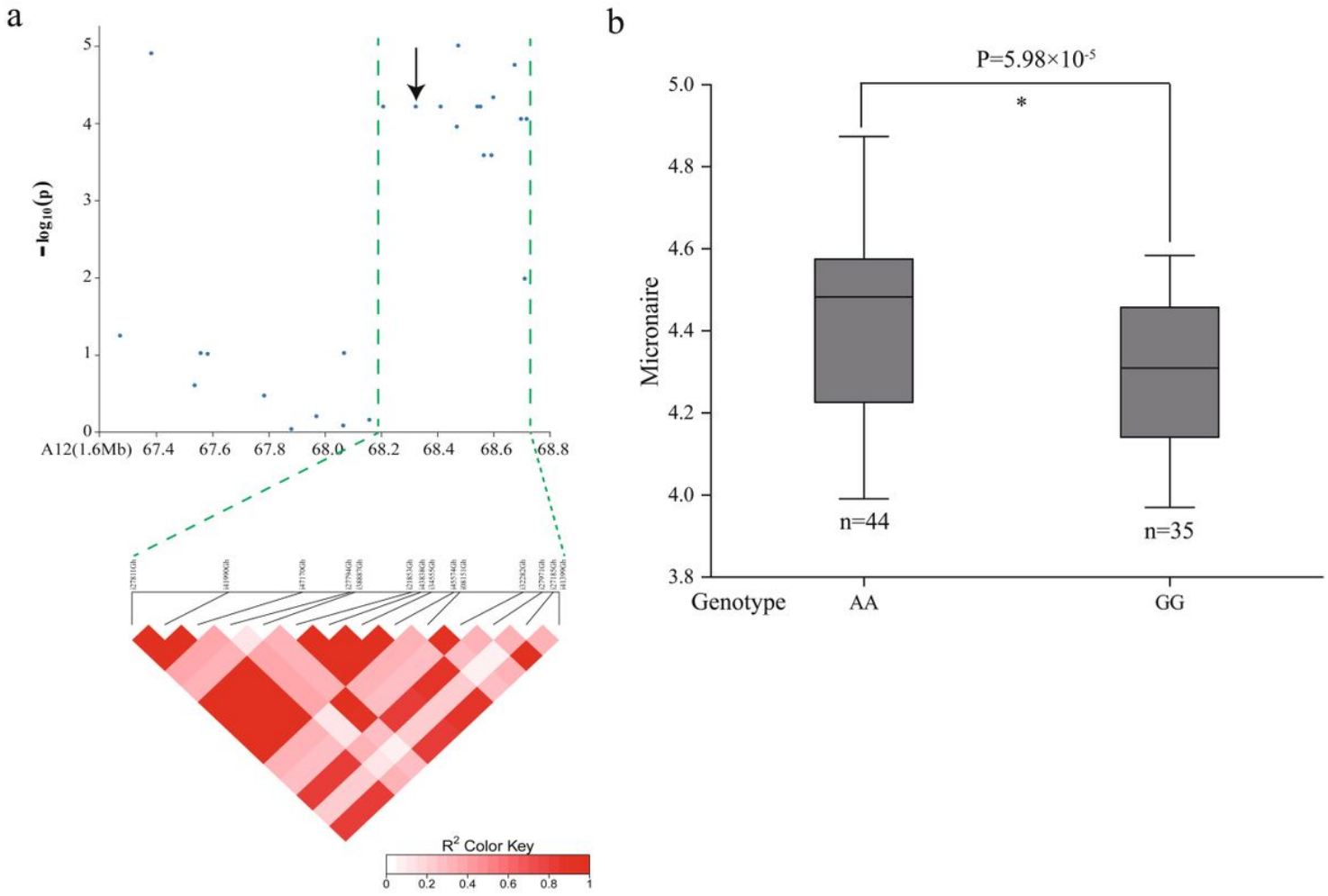

Figure 5

Identification of GhFLA9 on chromosome A12 a: Local Manhattan plot (top) and heat map (bottom) surrounding the peak on chromosome A12. Black arrows indicate the positions of the significant SNP i41990Gh. b: Box plot for micronaire based on the two haplotypes of i41990Gh. n indicates the number of accessions with the same genotype. * indicates significant differences at p = 0.05.

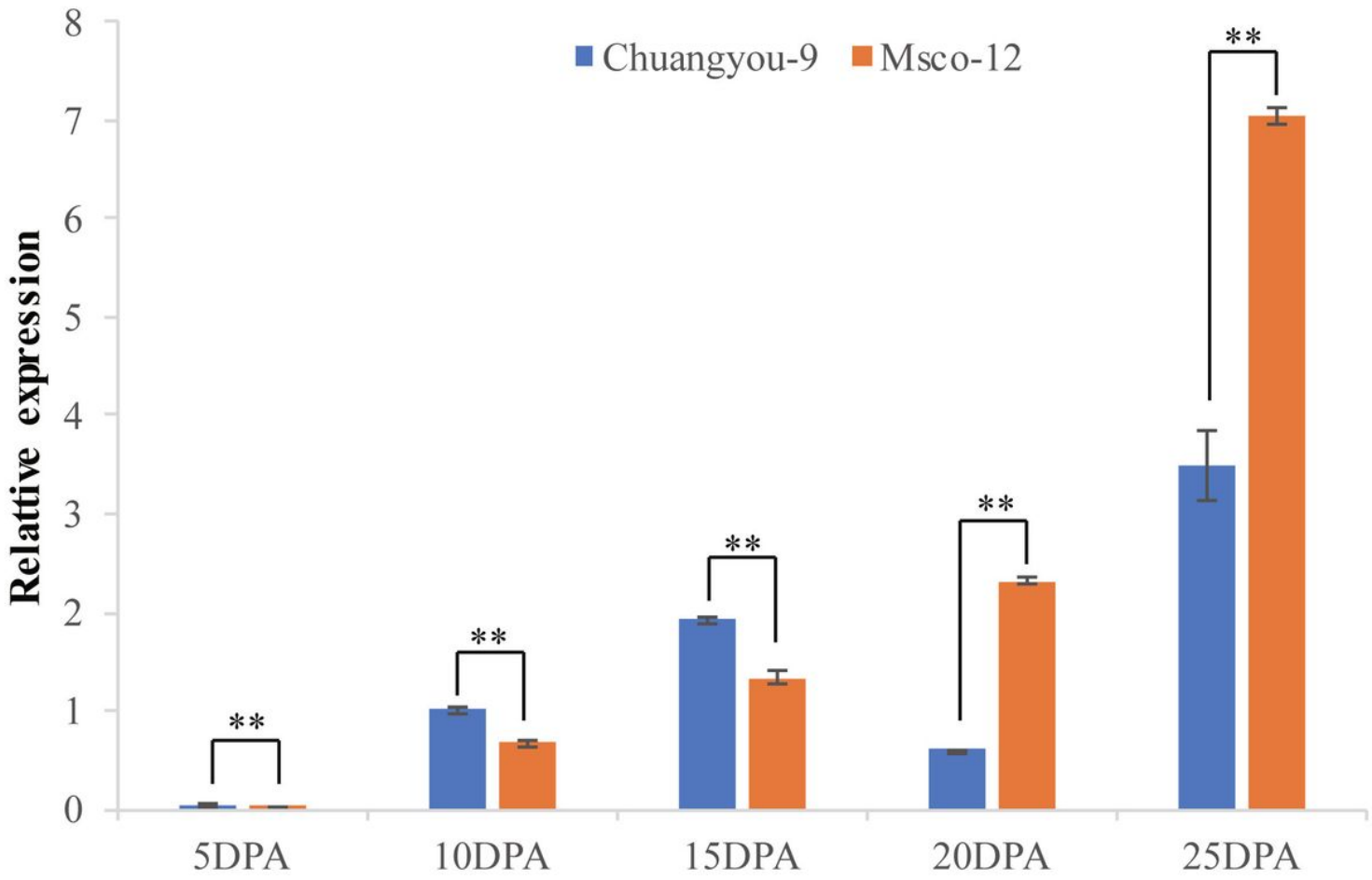

Figure 6

qRT-PCR analysis of GhFLA9 during fiber development Expression levels of GhFLA9 at five developmental stages of fiber. ** indicates significant differences at p = 0.01.

Supplementary Files

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

- AdditionalFile1TableS1.xlsx
- AdditionalFile2Fig.S1.pdf
- AdditionalFile2Fig.S2.pdf