Journal of Cotton Research

RESEARCH Open Access

Genome-wide association study of micronaire using a natural population of representative upland cotton (*Gossypium hirsutum* L.)

SONG Jikun[1,2], PEI Wenfeng[1,2], MA Jianjiang[1], YANG Shuxian[1], JIA Bing[2], BIAN Yingying[1], XIN Yue[1], WU Luyao[1], ZANG Xinshan[1], QU Yanying[2], ZHANG Jinfa[3], WU Man[1*] and YU Jiwen[1*]

Abstract

Background: Micronaire is a comprehensive index reflecting the fineness and maturity of cotton fiber. Micronaire is one of the important internal quality indicators of the cotton fiber and is closely related to the value of the cotton fiber. Understanding the genetic basis of micronaire is required for the genetic improvement of the trait. However, the genetic architecture of micronaire at the genomic level is unclear. The present genome-wide association study (GWAS) aimed to identify the genetic mechanism of the micronaire trait in 83 representative upland cotton lines grown in multiple environments.

Results: GWAS of micronaire used 83 upland cotton accessions assayed by a Cotton 63 K Illumina Infinium single nucleotide polymorphism (SNP) array. A total of 11 quantitative trait loci (QTLs) for micronaire were detected on 10 chromosomes. These 11 QTLs included 27 identified genes with specific expression patterns. A novel QTL, *qFM-A12–1*, included 12 significant SNPs, and *GhFLA9* was identified as a candidate gene based on haplotype block analysis and on strong and direct linkage disequilibrium between the significantly related SNPs and gene. *GhFLA9* was expressed at a high level during secondary wall thickening at 20~25 days post-anthesis. The expression level of *GhFLA9* was significantly higher in the low micronaire line (Msco-12) than that in the high micronaire line (Chuangyou-9).

Conclusions: This study provides a genetic reference for genetic improvement of cotton fiber micronaire and a foundation for verification of the functions of *GhFLA9*.

Keywords: Upland cotton (*Gossypium hirsutum* L.), Fiber micronaire, GWAS, Candidate genes, *GhFLA9*

Background

Cotton is an economically important crop and the most important source of the natural fiber for the textile industry (Yu 2018). *Gossypium hirsutum* L. is characterized by high yield and wide adaptability, accounting 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 important fiber quality properties for spinning primarily include length, strength and micronaire. Micronaire is one of the most important fiber characteristics for international cotton classers and spinners. Micronaire is a comprehensive indicator of fiber fineness and maturity and an important characteristic for 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 known plants. Each cotton fiber is composed of

* Correspondence: wuman2004@163.com; yujw666@hotmail.com
[1]State Key Laboratory of Cotton Biology, Ministry of Agriculture and Rural Affairs/Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China
Full list of author information is available at the end of the article

a single cell produced on the surface of an ovule. Fiber development mainly includes four stages: initial development, elongation, secondary wall thickening, and maturity (Pang et al. 2010). Cotton fiber micronaire (FM) is mainly influenced by the formation of the secondary wall of the fiber (Wu et al. 2020). Thickening of the secondary wall of the cotton fiber is a complex physiological process that mainly involves the synthesis and deposition of cellulose, usually starts 16~19 days after flowering and is jointly regulated by the expression of many genes (Yan 2010). Therefore, the analysis and identification of the candidate genes that regulate FM at the QTL mapping level have important theoretical value for molecular breeding of cotton quality and identification of genetic mechanisms of cotton fiber development.

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 in the A01, A02, and A07 chromosomes using a population of 143 recombinant inbred lines (RILs) (Fan et al. 2018). Twenty-two QTLs for FM were detected using 180 RILs, and 13 of these QTLs were detected in two or more environments (Ali et al. 2018). Additionally, 27 QTLs for FM were detected using BC_3F_2, $BC_3F_{2:3}$, and $BC_3F_{2:4}$ populations, including 11 QTLs that were located near the same marker in various populations or near the linked markers in the same population (Wang et al. 2017a). However, the populations used for linkage mapping are usually characterized by a large positioning interval, limited recombination times, and low genotype variation. The rapid development of genome sequencing resulted in successful applications of genome-wide association studies (GWAS) for genetic dissection of the fiber quality traits, including FM. Resequencing of the markers identified 3 and 533 significant SNPs for FM in the groups of 362 and 419 diverse upland cotton accessions, respectively (Ma et al. 2018b; Wang et al. 2017b). Additionally, 503 upland cotton accessions were individually genotyped using a Cotton 63 K Illumina SNP array, resulting in the identification of 3 stable QTLs associated with the FM trait located on the A05, D05 and D12 chromosomes. Previous studies identified many QTLs for FM mainly distributed on the A03, A07, A12, D03, D08, and D11 chromosomes. Comparison with the orthologs in other organisms demonstrated that various genes, such as *GhADF1*, *GhWLIM1a*, and *GhXTH*, are related to the development of the fiber cell walls in upland cotton (Han et al. 2013; Michailidis et al. 2009; Wang et al. 2009). Fasciclin-like arabinogalactan protein (FLA) is an arabinogalactan protein (AGP) with one or two fasciclin domains involved in plant cell development (Showalter 2001). FLA 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 expressed at a high level during fiber development; however, genes regulating FM during thickening of the secondary wall of the fiber were not reported. Furthermore, the study focused only on certain specific genes; however, exploration of the whole genetic architecture of FM with QTL mapping is very important. A diversity panel consisting of 83 upland cotton accessions was genotyped using a Cotton 63 K Illumina Infinium SNP array to investigate genetic variation of FM at a natural population level. Fiber micronaire was measured in five environments. GWAS was performed to identify the SNP loci or QTL regions associated with FM in upland cotton. Candidate genes controlling FM were predicted by haplotype block analysis in the novel and stable QTL regions. These results provide a foundation for FM improvement by marker-assisted breeding.

Materials and methods

Plant materials

The present study investigated a diversity panel of 83 representative upland cotton accessions obtained from cotton germplasm collections housed in our laboratory and the low-temperature germplasm gene bank of the Institute of Cotton Research, Chinese Academy of Agricultural Sciences (ICR-CAAS). The detailed information about 83 upland cotton samples has been presented in a previous study from our laboratory (Ma et al. 2018a, b).

Field experiments and phenotyping

All of the 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 Aral (Ale), Xinjiang, China (40.55°N, 81.28°E), in 2016 (designated 16_Ale). These two sites were used as representative cotton production locations in the Yellow River Valley region and northwest inland region, respectively. They were planted at Sanya (Sy), Hainan, China (18.41°N, 109.20°E), another representative cotton production area, in 2016 (designated 16_Sy). 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 a row spacing of 0.38 m; at Aral, 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. The planting at Sanya had a plant spacing of 0.11 m, a row spacing of 0.38 m, and a plot length of 5 m with a total of 50 plants. The trial management procedures followed standard breeding field practices. A total of 20 naturally opening bolls in the middle of plants were manually harvested from each accession in September at Anyang and Aral. The FM of each harvested sample was evaluated using a high-volume instrument (HVI) 900 (Cotton Fiber Quality

Inspection and Testing Center of Ministry of Agriculture and Rural Affairs, Anyang, Henan, China). Statistical analyses, including descriptive statistics and correlation analysis of FM in 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 various environments was estimated by QTL IciMapping 4.2.0 (Meng et al. 2015).

QTL mapping for FM

SNP genotyping

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

Identification of SNPs associated with FM

A previous study analyzed the population structure and linkage disequilibrium (LD) (Ma et al. 2018a, b). The best linear unbiased predictions (BLUPs) for FM across five environments were estimated using R 3.6.3. 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 (Ma et al. 2019). The Bonferroni threshold of SNP significance was $P < 6.51 \times 10^{-5}$ ($P = 1/n$, n = the number of SNPs, $-\log_{10}(1/15369) \approx 4.19$) (Li et al. 2013; Yang et al. 2014; Liu et al. 2016). Manhattan plots were generated using the "CMplot" package for R (Turner 2014). Stable QTLs were identified in two or more different environments in the present study. The QTLs were named as follows: q + trait abbreviation - chromosome - QTL number (Zhang et al. 2015a).

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 acquired from a previous study (Hu et al. 2019). The screening criteria were defined based on the expression levels in the fiber at 20 or 25 DPA, which should be significantly higher than those at 0~15 DPA. The genes related to FM were subsequently analyzed according to functional gene annotation. R software was used to generate the haplotype blocks based on the stable *qFM-A12–1* SNPs, corresponding to the transcriptome database and functional gene annotations, and differentially expressed genes were screened using intervals (Su et al. 2016).

The expression trends of the candidate genes were validated by quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis. Total RNA was extracted from the fibers collected from five developmental time points (5, 10, 15, 20, and 25 DPA) of the high-micronaire (Msco-12) and low-micronaire lines (Chuangyou-9) using a FastPure Plant Total RNA Isolation Kit (polysaccharides & polyphenolics–rich) (Vazyme Biotech Co., Ltd., Nanjing, China). And cDNA was synthesized using a reverse transcription kit (HiScript II Q RT SuperMix for qPCR). ChamQ universal SYBR quantitative fluorescent kit and a quantitative fluorescent PCR instrument (ABI7500, Eppendorf, Germany) were used for qRT-PCR validation. The qRT-PCR conditions were as follows: DNA polymerase activation as the first step at 94 °C (30 s); the second step included 40 cycles at 94 °C (5 s), 58 °C (15 s), and 72 °C (12 s); the third step included melting curve analysis; the final step included incubation at 12 °C (1 min). 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 $2^{-\Delta\Delta C_T}$ method was used to calculate the relative gene expression (Livak and Schmittgen 2001).

Results

Phenotypic variation of FM

Significant variation in FM was observed in 83 upland cotton accessions; FM varied 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_Ale, and 16_Sy environments, the natural population had average FM values of 5.00, 4.76, 4.92, 4.32, and 4.23, respectively. The coefficients of variation

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_Ale	3.08	5.50	4.32	0.60	13.80	−0.14	−0.52	
16_Sy	2.90	6.50	4.23	0.62	14.77	0.29	1.13	

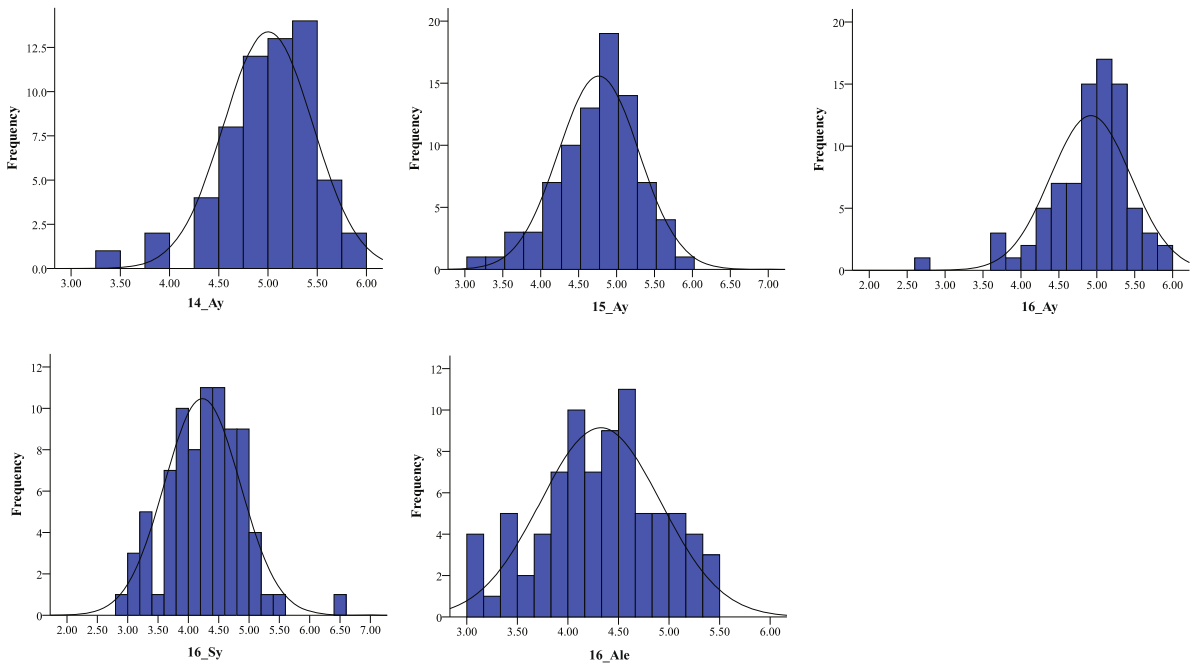

Fig. 1 Frequency distribution of fiber micronaire in upland cotton in five environments

(CVs) for FM ranged from 9.09% to 14.77% in five environments, indicating that the variation of the cotton FM index was adequate (Table 1). As shown in Fig. 1, FM was normally distributed in five environments. The analysis results indicated that cotton fiber in the Yellow River Valley was relatively thicker, and that in the northwest inland area was relatively finer. Correlation analysis demonstrated a very significantly positive correlation of the FM trait in five environments, and the correlation coefficient varied from 0.4 to 0.7 (Fig. 2). Analysis of variance indicated that the genotypes and environments had significant effects on FM ($P < 0.01$), and the estimated broad-sense heritability was 86.27% (Table 1, Table S1). These results suggested that FM in upland cotton is highly heritable; however, the role of environmental factors in fiber development cannot be ignored. The results indicated that this population is suitable for the GWAS of the cotton FM trait.

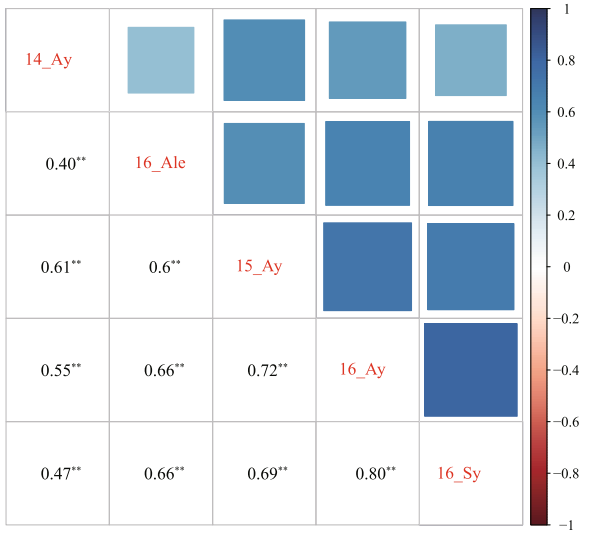

Fig. 2 Correlation analysis of fiber micronaire in upland cotton in five environments

GWAS

GWAS for FM was performed using BLUPs across five environments, and individual environments were evaluated based on a GLM considering Q-matrix (GLM (Q)). A threshold of $-\log_{10}(P) > 4.19$ was considered to indicate significant SNPs. A total of 18, 37, 67, 7, 21, and 25 SNPs were significantly associated with FM in 14_Ay, 15_Ay, 16_Ay, 16_Ale, 16_Sy, and BLUP, respectively (Fig. 3; Fig. S2). Analysis of significantly related SNPs identified a total of 55 common SNPs, including 16 SNPs 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, chromosome A12 has the most SNPs of 12. The QTL interval was calculated as significant SNP position ± LD based on LD specific for each chromosome (Ma et al. 2018a, b). Thus, a total of 11 stable QTLs related to FM were identified. These QTLs explained 18.10%~31.64% of the phenotypic variation in FM. The phenotypic contributions of these 11 markers

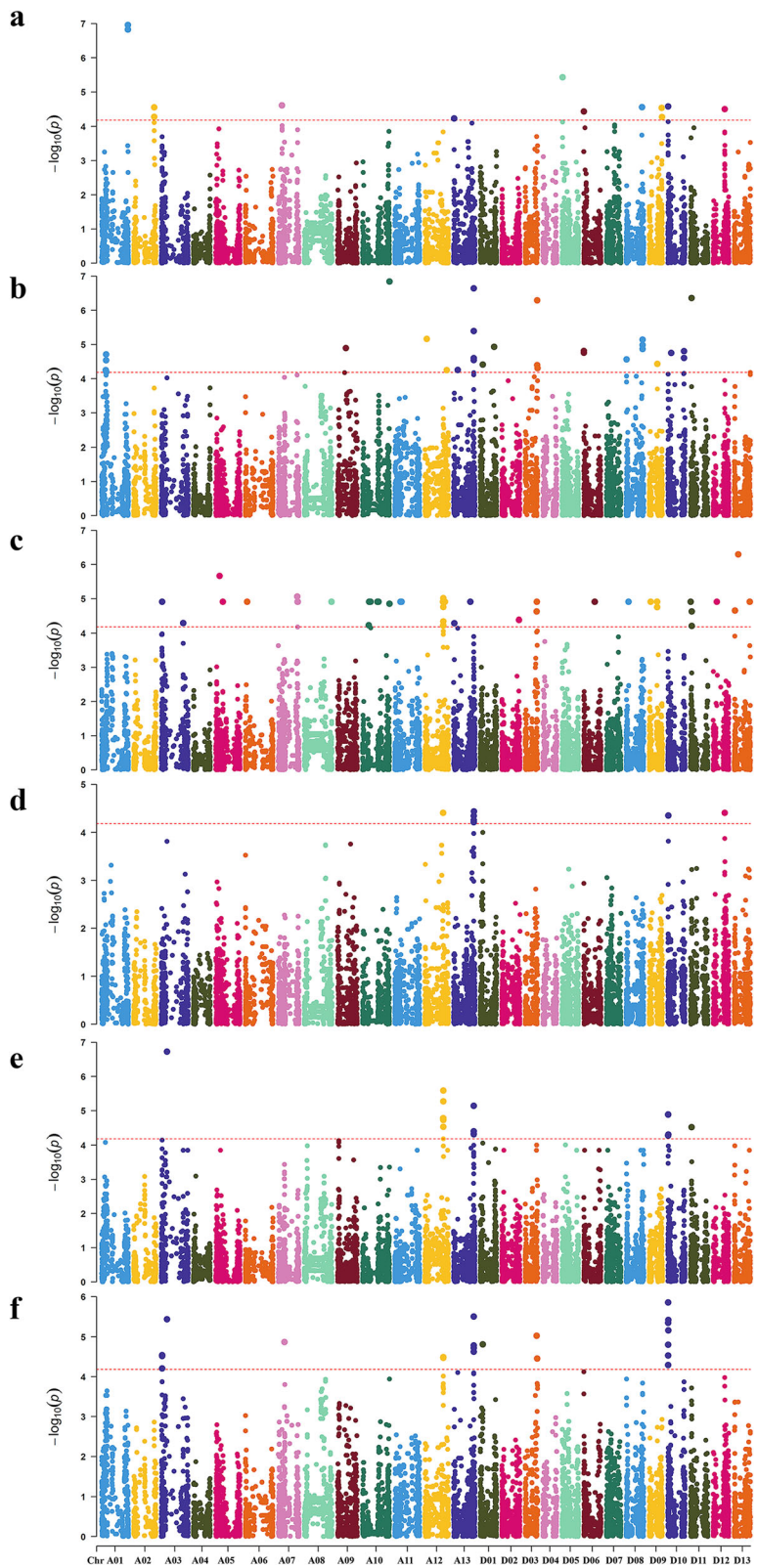

Fig. 3 Manhattan plots of fiber micronaire in upland cotton in five environments and BLUP. The lowercase letter, **a**, **b**, and **c,** represents the Manhattan plots of the GLM in Anyang in 2014, 2015, and 2016, respectively; the letter, d and e represent the Manhattan plots of the GLM in Aral and Sanya in 2016, respectively; and f represents the Manhattan plots of the GLM for BLUP

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. 2017a, b; Jamshed et al. 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	
	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. 2017a, b; 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	
	i45347Gh	D09	30 133 934	4.91	20.9	16_Ay	
	i27999Gh	D09	30 418 262	4.75	23.66	16_Ay	

Table 2 Summary of SNPs significantly associated with fiber micronaire in five environments *(Continued)*

QTL	Marker	Chr.	Position /bp	-log$_{10}$(*P*)	R^2/%	Environment	Reported previously
	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	

in *qFM-A12–1* were more than 20%. The results of GWAS were compared with the data of previous association and linkage studies for FM to verify the reliability of 11 identified stable QTLs for FM. The simple sequence repeats (SSR) and SNP markers of QTLs previously identified as related to FM were aligned to known marker sequences of the *G. hirsutum* TM-1 genome using local basic local alignment search tool (Zhang et al. 2015b). Comparison indicated that 8 stable QTLs for FM were reported in previous studies and 3 new QTLs were detected in the present study (Table 2).

Identification of the candidate genes in 11 stable QTLs

Combined functional annotation of the genes in stable QTLs and the analysis of the *Arabidopsis* genes homologous to the cotton fiber transcriptome database were used to identify the candidate genes related to FM (Du et al. 2018). A total of 1 594 genes were present in 11 QTLs according to local BLAST of the upland cotton genome from Zhejiang University (http://ibi.zju.edu.cn/cotton). The fiber quality of *G. barbadense* was significantly better than that of *G. hirsutum*, and significant differences in FM between *G. barbadense* Hai7124 and *G. hirsutum* TM-1 were detected. The Hai7124 and TM-1 transcriptome databases were used to determine the differences in the 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 identified (Table 3, Fig. 4). Annotation analysis of 27 differentially expressed genes versus previously reported genes were found to be related to fiber development, such as *GhTUB6* and *GhFLA*.

Mining of the candidate genes in *qFM-A12–1*

qFM-A12–1 contained the most significant SNPs, and the phenotypic contribution of this locus was stable and high. However, previous studies have not located any genes related to FM at this locus. Therefore, the present study focused on the analysis of *qFM-A12–1* to identify the candidate genes influencing FM. A haplotype block of *qFM-A12–1* was constructed to narrow the QTL interval, which was subsequently narrowed to location from i27811Gh to i41399Gh based on direct strong LD between significantly related SNPs and the corresponding genes (Fig. 5a). The i41990Gh marker in the interval was used to genotype upland cotton. The accessions carrying the AA (44) allele had significantly higher micronaire than that in the accessions with the GG (35) allele (Fig. 5b), indicating that this interval may significantly influence FM; Chuangyou-9 and Msco-12 harbor the AA and GG alleles, respectively. *GhFLA9* is located between the i27811Gh and i41990Gh markers, and the expression levels of *GhFLA9* in the fiber at 20 and 25 DPA were significantly higher than those in Hai7124 and TM-1 at 5 and 10 DPA. The highest expression level was observed at 20 DPA, and the expression levels of *GhFLA9* in the fiber in Hai7124 at 20 and 25 DPA were significantly higher than those in TM-1 (Fig. 4). In this study, qRT-PCR was used to determine the expression levels of *GhFLA9* at various fiber development stages (5, 10, 15, 20, and 25 DPA) in the high-micronaire and low-micronaire genotypes. The expression level of *GhFLA9* increased during fiber development. Moreover, the expression levels in the fiber in Msco-12 at 20 and 25 DPA were significantly higher than those in Chuangyou-9 (Fig. 6). These data indicated that *GhFLA9* may be involved in the regulation of FM and may act as a positive regulator.

Discussion

The quality of the cotton fiber mainly depends 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 derived products. A

Table 3 Candidate genes for micronaire

Gene	Physical position	Homologs in *A. thaliana*	Functional annotation
Gh_A10G2157	A10:100434822–100 435 289	*AT3G11110*	RING/U-box superfamily protein
Gh_A10G2169	A10:100571847–100 572 047	–	–
Gh_A10G2174	A10:100618393–100 647 910	*AT4G27190*	NB-ARC domain-containing disease resistance protein
Gh_A10G2176	A10:100676093–100 688 392	–	–
Gh_A12G1287	A12:68269264–68 269 998	*AT1G03870*	fasciclin-like arabinoogalactan 9 (FLA9)
Gh_A13G0120	A13:1383789–1 385 806	*AT1G52340*	NAD(P)-binding Rossmann-fold superfamily protein
Gh_A13G0121	A13:1394586–1 395 093	*AT2G41430*	dehydration-induced protein (ERD15)
Gh_A13G0133	A13:1519272–1 521 026	*AT5G17540*	HXXXD-type acyltransferase family protein
Gh_A13G0188	A13:2123316–2 124 229	*AT2G37590*	DNA binding with one finger 2.4 (DOF2.4)
Gh_A13G0220	A13:2631476–2 632 431	*AT3G01390*	vacuolar membrane ATPase 10 (VMA10)
Gh_A13G1555	A13:74370130–74 370 630	*AT1G09310*	protein of unknown function, DUF538
Gh_A13G1563	A13:74415625–74 417 872	*AT1G09380*	nodulin MtN21/EamA-like transporter family protein
Gh_A13G1570	A13:74527856–74 528 134	–	–
Gh_A13G1573	A13:74594101–74 600 768	*AT2G24520*	H(+)-ATPase 5 (HA5)
Gh_A13G1640	A13:75324608–75 326 119	*AT5G25820*	exostosin family protein
Gh_D03G1332	D03:41738284–41 744 222	*AT5G22920*	CHY-type/CTCHY-type/RING-type zinc finger protein
Gh_D03G1452	D03:43497008–43 499 238	*AT5G12250*	beta-6 tubulin (TUB6)
Gh_D03G1533	D03:44563981–44 565 739	*AT1G62790*	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
Gh_D06G0032	D06:322107–322 808	*AT1G75390*	basic leucine-zipper 44 (bZIP44)
Gh_D06G0041	D06:479879–483 912	*AT5G42800*	dihydroflavonol 4-reductase
Gh_D09G0646	D09:29655772–29 660 487	*AT3G17180*	serine carboxypeptidase-like 33 (scpl33)
Gh_D10G0078	D10:622188–623 235	*AT4G16380*	Heavy metal transport/detoxification superfamily protein
Gh_D11G0597	D11:5211532–5 212 612	*AT3G51520*	diacylglycerol acyltransferase family
Gh_D11G0639	D11:5582090–5 583 296	*AT5G13870*	xyloglucan endotransglucosylase/hydrolase 5
Gh_D11G0718	D11:6198554–6 200 040	*AT3G28150*	trichome birefringence-like 22 (TBL22)
Gh_D11G0750	D11:6489218–6 489 475	*AT5G54100*	SPFH/Band 7/PHB domain-containing membrane-associated protein family
Gh_D12G1409	D12:43357297–43 358 031	*AT1G03870*	fasciclin-like arabinoogalactan 9 (FLA9)

moderate FM provides for high yarn strength and good spinning quality of cotton fiber. Therefore, the genetic improvement of cotton FM has become an important area of cotton research. In the early stage of the project, a 63 K Illumina Infinium SNP array was used to obtain 15 369 high-quality SNPs in 83 upland cotton accessions. The population was used to perform an association analysis of dynamic fiber length and oil traits, and the candidate genes were identified (Ma et al. 2018a; Ma et al. 2019), so it is also suitable for GWAS of related traits. FM of 83 upland cotton accessions used in the present study was characterized by high phenotypic variation and a significant variation range in 5 environments (Table 1). The H^2 value of FM obtained in the present study was relatively high (86.27%) and very similar to the H^2 values (76.37%~94.00%) obtained in previous studies (Ma et al. 2020; Sun et al. 2010). These results indicated that the FM trait is mainly influenced by the genotype and is suitable for GWAS. GWAS of FM in 5 environments identified 11 stable QTLs containing 55 significant SNPs, including 4 QTLs located in the At subgenome and 7 QTLs located in the Dt subgenome (Table 2). The mapped interval was compared with the results of QTL mapping obtained in previous studies of micronaire traits using genetic population linkage analysis or natural population association analysis. Comparison indicated that 8 QTLs identified in the present study were similar to those identified in previous studies (Chen et al. 2018; Diouf et al. 2018; Huang et al. 2017; Jamshed et al. 2016; Jia et al. 2018; Wang et al. 2017a; Wang et al. 2017b; Zhang et al. 2016).Huang et al. (2017) used an Illumina Infinium SNP array for GWAS

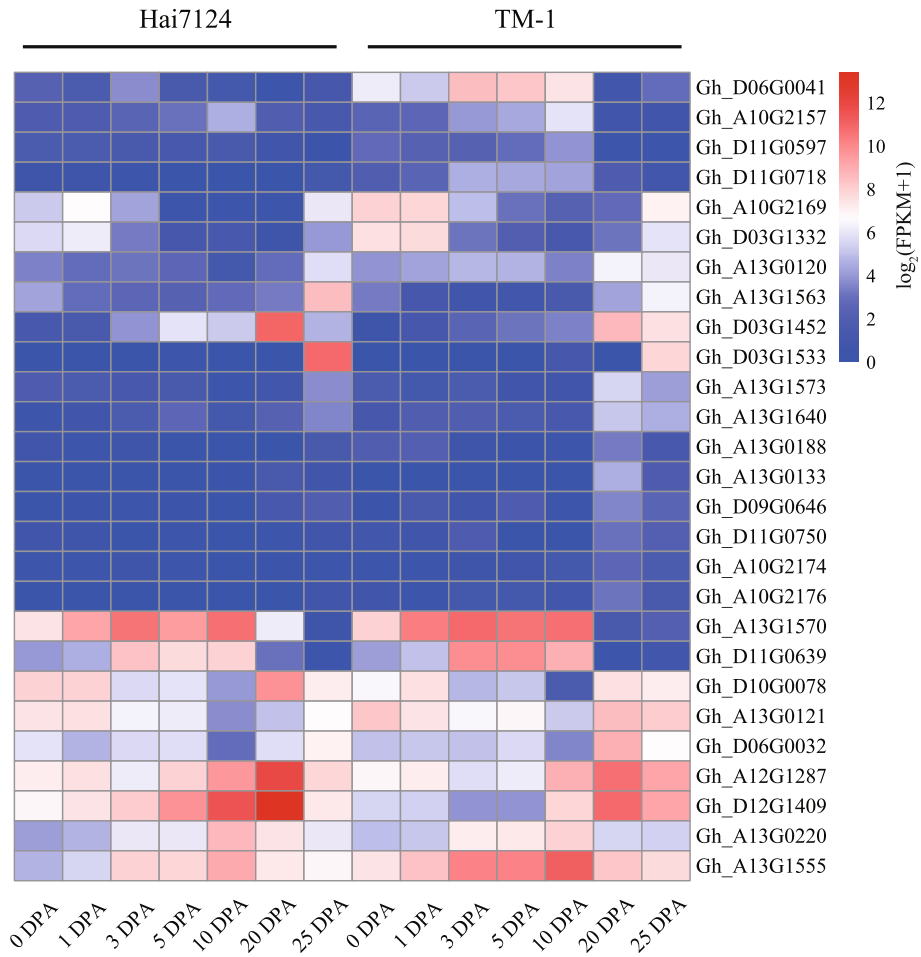

Fig. 4 The expression patterns of 27 genes in the fiber at seven developmental time points in Hai7124 and TM-1

of 503 upland cotton accessions and identified the stable QTL *qGhMV-c26–1*, which coincides with the relatively small *qFM-D12–1* described in the present study. This result indicated that GWAS of FM in 83 upland cotton accessions is credible. The results of GWAS combined with transcriptome and qRT-PCR data analyses identified the genes related to fiber development, including *GhTUB6* and *GhFLA*. *GhTUB6* is predominantly expressed during the rapid elongation stage of the cotton fiber cells and synthesis of the primary cell wall (He et al. 2008). The *GhFLA* gene is involved in cotton fiber development and can promote the expression levels of the genes responsible for the synthesis of the primary cell wall (Huang et al. 2008). New QTLs identified in the present study were stable in multiple environments and provided a reference for cotton molecular marker-assisted selection with targeting micronaire.

The stable QTL *qFM-A12–1* identified in the present study has not been reported previously. This QTL contains 12 adjacent significant SNPs. Therefore, we focused on the differential expression of the genes located within this interval. The interval was narrowed according to a haplotype block and transcriptome data. The *GhFLA9* gene encoding fasciclin-like arabinogalactan protein was identified. This gene is mainly expressed in the secondary wall thickening stage (20~25 DPA) during fiber development, and the expression was positively correlated with micronaire (Fig. 4). The expression level of this gene is the highest in the 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 between different materials. Subsequently, we performed a genotyping analysis based on the significant SNP i41990Gh. The results indicated that the genotypes of the high-micronaire line (Chuangyou-9) and the low-micronaire line (Msco-12) had significant differences. Subsequent qRT-PCR analysis of *GhFLA9* in the fiber tissues of the two materials at five stages (5, 10, 15, 20, and 25 DPA) indicated that the expression of this

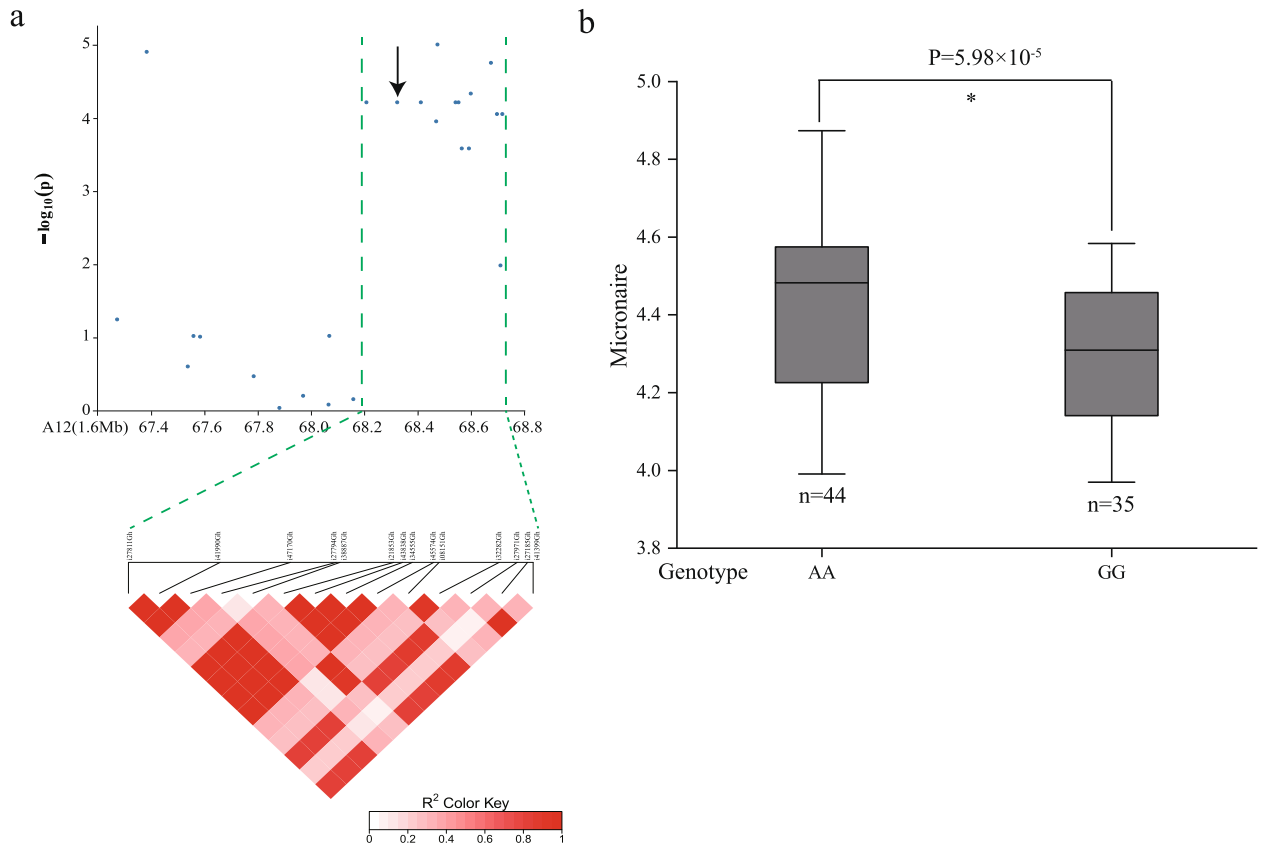

Fig. 5 Identification of *GhFLA9* on chromosome A12. **a** Local Manhattan plot (top) and heat map (bottom) surrounding the peak on chromosome A12. Black arrows indicate the positions of the significant SNP i41990Gh. **b** Box plot for micronaire based on two haplotypes of i41990Gh. n indicates the number of accessions with the same genotype. * indicates significant differences at $P = 0.05$

gene was consistent with the expression in the whole transcriptome assay. 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. Thus, a trend of *GhFLA9* expression identified in the present study is relatively reliable.

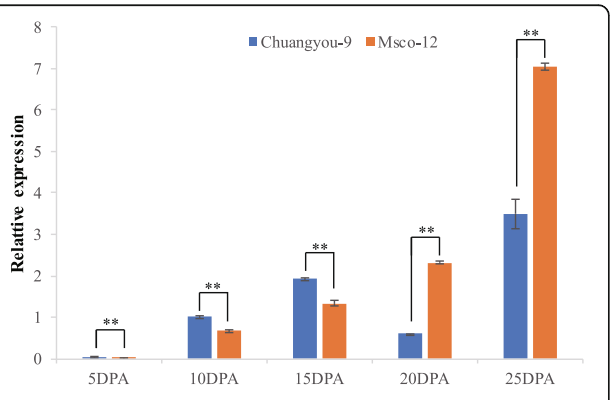

Fig. 6 qRT-PCR analysis of *GhFLA9* during fiber development. Expression levels of *GhFLA9* are shown at five developmental stages of the fiber. ** indicates significant differences at $P = 0.01$

Conclusions

GWAS of FM was performed in 5 environments, and 11 QTLs related to FM were identified, including 8 QTLs colocalized with QTLs identified in previous studies. Three QTLs were newly identified in the present study. Twenty-seven candidate genes related to FM were identified within 11 stable QTLs, and haplotype block analysis and qRT-PCR assay of the stable QTL *qFM-A12–1* indicated that *GhFLA9* may positively 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

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42397-021-00089-1.

Additional file 1: Table S1. Analysis of variance for FM.

Additional file 2: Fig. S1. Box plots of fiber micronaire in upland cotton accessions in different environments. **Fig. S2.** QQ plots against

genotypic and phenotypic data represent the normal distribution for genotypic and phenotypic data. The lowercase letters a, b, and c represent the QQ plots in Anyang in 2014, 2015 and 2016, respectively; the letters d and e represent the QQ plots in Aral and Sanya in 2016, respectively; and f represents the QQ plots for BLUP.

Acknowledgments
The authors would like to thank New Mexico Agricultural Experiment Station, New Mexico, USA.

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.

Funding
The present study was funded by National Key Research and Development Program of China (grants nos. 2018YFD0101402, 2018YFD0100300 and 2016YFD0101400), the Natural Science Foundation of Xinjiang Uygur Autonomous Region of China (grant no. 2020D01A135), and Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences.

Declarations

Consent to publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]State Key Laboratory of Cotton Biology, Ministry of Agriculture and Rural Affairs/Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China. [2]Engineering Research Centre of Cotton, Ministry of Education / College of Agriculture, Xinjiang Agricultural University, Urumqi 830052, Xinjiang, China. [3]Department of Plant and Environmental Sciences, New Mexico State University, Las Cruces 880033, NM, USA.

Received: 25 November 2020 Accepted: 2 April 2021
Published online: 03 May 2021

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. https://doi.org/10.1186/s12864-018-5294-5.

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. https://doi.org/10.3389/fpls.2018.01023.

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. https://doi.org/10.3390/ijms19020441.

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. https://doi.org/10.1186/s12864-018-4837-0.

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. https://doi.org/10.1186/s12864-018-4890-8.

Han LB, Li YB, Wang HY, et al. The dual functions of WLIM1a in cell elongation and secondary wall formation in developing cotton fibers. Plant Cell. 2013; 25(11):4421–38. https://doi.org/10.1105/tpc.113.116970.

He JX. Micronaire value is an important parameter for cotton quality. China Cotton Processing. 2005(3):25–6.

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):2687–95. https://doi.org/10.1093/jxb/ern127.

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):739–48. https://doi.org/10.1038/s41588-019-0371-5.

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): 1374–86. https://doi.org/10.1111/pbi.12722.

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):348–59. https://doi.org/10.1111/j.1399-3054.2008.01139.x.

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. https://doi.org/10.1534/g3.115.018416.

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(1):197. https://doi.org/10.1186/s12864-016-2560-2.

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 Gen Genomics. 2018;293(4):831–43. https://doi.org/10.1007/s00438-018-1424-7.

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. https://doi.org/10.1038/ng.2484.

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):1203–15. https://doi.org/10.1007/s00122-016-2697-z.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. Methods. 2001;25(4):402–8. https://doi.org/10.1006/meth.2001.1262.

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. 2018a;19(1):882. https://doi.org/10.1186/s12864-018-5309-2.

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. https://doi.org/10.1016/j.plantsci.2019.05.019.

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(1):22. https://doi.org/10.1186/s42397-020-00060-6.

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. 2018b;50(6):803–13. https://doi.org/10.1038/s41588-018-0119-7.

Meng L, Li HH, Zhang LY, Wang JK. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. The Crop Journal. 2015;3(3):269-283. https://doi.org/10.1016/j.cj.2015.01.001

Meng L, Li HH, Zhang LY, Wang JK. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. The Crop Journal. 2015;3(3):269-283. https://doi.org/10.1016/j.cj.2015.01.001Duplicate, please delete one.

Michailidis G, Argiriou A, Darzentas N, Tsaftaris A. 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):403–16. https://doi.org/10.1016/j.jplph.2008.06.013.

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):2019–33. https://doi.org/10.1074/mcp.M110.000349.

Said JI, Knapka JA, Song MZ, Zhang JF. Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between *Gossypium hirsutum* and *G. hirsutum* × *G. barbadense* populations. Mol Gen Genomics. 2015a;290(4):1615–25. https://doi.org/10.1007/s00438-015-1021-y.

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(1):776. https://doi.org/10.1186/1471-2164-14-776.

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 Gen Genomics. 2015b;290(3):1003–25. https://doi.org/10.1007/s00438-014-0963-9.

Showalter AM. Arabinogalactan-proteins: structure, expression and function. Cell Mol Life Sci. 2001;58(10):1399–417. https://doi.org/10.1007/PL00000784.

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. https://doi.org/10.3389/fpls.2016.01576.

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 Sci. 2010;22(04):319–25. https://doi.org/10.3969/j.issn.1002-7807.2010.04.005.

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):3110–7. https://doi.org/10.1007/s11434-007-0461-0.

Turner SD. Qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. Biorxiv. 2014:005165. https://doi.org/10.1101/005165.

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. https://doi.org/10.3389/fpls.2017.01848.

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. https://doi.org/10.1111/j.1467-7652.2008.00367.x.

Wang MJ, Tu LL, Lin M, et al. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. Nat Genet. 2017b;49(4):579–87. https://doi.org/10.1038/ng.3807.

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 Sci. 2020;32(1):52–62. https://doi.org/10.11963/1002-7807.wmyjw.20200112.

Yan HC. Analysis of the digital gene expression profiling of cotton fiber development during cell elongation and secondary cell wall biogenesics. Wuhan, China: Huangzhong Agricultural University; 2010. p. 13–4.

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 Genet. 2014;10(9):e1004573. https://doi.org/10.1371/journal.pgen.1004573.

Yu SX. The development of cotton production in the recent hundred years of China. J Agric. 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(1):31954. https://doi.org/10.1038/srep31954.

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. 2015a;33(5):531–7. https://doi.org/10.1038/nbt.3207.

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. 2015b;10(8):e0135430. https://doi.org/10.1371/journal.pone.0135430.