

Association mapping of economic traits based on QTL clusters/hotspots linked SSR markers in Upland cotton.

Muhammad Atif Wahid¹, Muhammad Mahmood Ahmed¹, Huang Cong², Waqas Malik², Zhongxu Lin^{1*}

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

²Department of Plant Breeding and Genetics, Faculty of Science and Agriculture, Bahaiddin Zakariya University, Multan, Pakistan

Received: 21-Dec-2023, Manuscript No. RNAI-23-123246; **Editor assigned:** 23-Dec-2023, RNAI-23-123246 (PQ);

Reviewed: 06-Jan-2024, QC No. RNAI-23-123246; **Revised:** 08-Jan-2025, Manuscript No. RNAI-23-123246 (R);

Published: 15-Jan-2025, DOI: 10.35841/2591-7781.21.1.219

Abstract

Cotton productivity has decreased over last decade owing to various agronomic factors including narrow genetic base of cultivated Upland cotton. The yield of cotton in China has been decreased from 26000 tons to 5.34 million tons. A 4.6 percentage decrease according to natural bureau of statistics alternatively, natural diversity conserved in the germplasm could be dissected by Linkage Disequilibrium (LD) based association mapping. Therefore cotton breeders have focused on detecting QTLs that could enhance cotton yield by improving the genetics of yield and fiber quality. Alternatively, the natural diversity conserved in the germplasm collection could be dissected by linkage disequilibrium based association mapping. To date, numerous QTL clusters/hotspots have been detected based on meta-analyses. In this study, QTL clusters/hotspots-linked Simple Sequence Repeat (SSR) markers were employed to validate QTL effects in 503 upland cotton accessions using association mapping, so that these markers could be efficiently employed in marker-assisted breeding programs. Among 298 clusters/hotspots-linked SSRs, 81 polymorphic ones were utilized to explore genetic variation and Linkage Disequilibrium (LD) decay in the diverse set of 503 cultivars. Association mapping using Generalized Linear Model (GLM) Q approach detected 56 significant loci associated with 14 economic traits. These markers were relegated to forty-eight QTL clusters which were widely distributed on 17 chromosomes. Amongst them, 6 QTL hotspots and 22 meta-QTL were associated to different traits. The significantly associated markers identified through association mapping will facilitate substantial gains in marker-assisted breeding and further research on QTL fine mapping and candidate gene prediction.

Keywords: *Gossypium hirsutum*, Meta-QTL, Polymerase chain reaction, Epistatic effect, Upland cotton.

Accepted on 25 October, 2024

Introduction

Cotton (*Gossypium hirsutum* L.) a tetraploid specie (AD₁), belongs to *Gossypium* tribe under the Malvaceae family, and accounts for more than 95% of world cotton productiv. China remained one of the largest cotton producers, but the most cotton cultivars planted were derived from germplasm featuring narrower genetic base. Although introduction of exotic strains contributed to broaden the genetic base of the national cotton germplasm, leading to the elite cotton breeding programs in China. Since upland cotton is preferred for cultivation featuring valuable economic characteristics and improvement of yield and fiber quality traits, a more diverse foundation could be laid for marker-assisted selection and breeding.

Meta-QTL analysis involves sorting putative QTL into clusters, hotspots and meta-QTL (mQTL). A QTL cluster is a densely populated chromosome region containing multiple QTL associated to different traits. In contrast, a QTL hotspot is also a densely populated region of the chromosome, but possesses multiple QTL with significant association to a single trait termed as mQTL. Various studies have reported meta-QTL analysis which aided in marker-assisted breeding in crops. Cotton breeders have successfully employed meta-analysis to locate QTL clusters and hotspots, thus identifying important regions on the linkage maps [1]. Rapid advances in crop genomics and availability of cultivated cotton genome assembly made it possible to precisely locate QTL cluster hotspots on physical maps. Association analysis of quantitative genes could facilitate understanding the variation observed in

complex traits, and could assist in identification of novel alleles in diverse populations.

Association mapping studies are based on the non-random association of alleles at different loci between a marker locus and a phenotypic trait and are a successful technique to identify DNA markers that are in LD with a locus controlling a trait of interest. Quantification of Linkage Disequilibrium (LD) and association mapping has been performed in numerous plant species. In cotton, several studies have been conducted on traits related to agronomy, fiber quality, yield, and growth stages, using association analysis and multiple marker loci, but no such study has been reported especially for QTL clusters/hotspots revealed through meta-analysis in cotton.

This study employed 298 SSR markers derived from vicinity of QTL clusters and hotspots to exploit genetic variation in a set of natural populations comprising 503 accessions, which was developed during 1920-2011. Eighty-one QTL cluster/hotspots polymorphic SSRs were utilized for association mapping of 16 economic traits. The General Linear Model (GLM) Q detected statistically significant SSRs in two years across eight diverse environments. Thus, the purpose of this study aimed:

- To evaluate genetic diversity of SSR markers in QTL clusters/hotspots.
- To identify SSR marker loci associated with economic traits in cotton by association mapping.
- To integrate linkage and association mapping procedures and to evaluate their usefulness for complex traits in cotton.

Materials and Methods

Plant materials and phenotypic data

The natural population contained 503 upland cotton accessions originated from diverse geographic regions of the world [2]. The phenotypic data of 16 agronomic traits were from Huang, et al. including Plant Height (PH), Flowering Period (FP), Whole Growth Period (WGP), First Fruit Spur Height (FFSH), First Fruit Branch Number (FFBN), First Fruiting Spur Branch Number (FFSBN), Effective Boll Number (EBN), Seed-Cotton Weight (SW), Lint Weight (LW), Lint Percentage (LP), Fiber Upper Half Mean Length (FUHML), Fiber Uniformity (FU), Fiber Strength (FS), Fiber Elongation (FE), Micronaire Value (MV), and Short Fiber (SF) [3].

DNA extraction and SSR markers selection

Total genomic DNA of the 503 accessions was extracted from young leaves from each accession according to procedures described by. A total of 298 SSRs flanking and covering QTL clusters and hotspots were selected. The selected markers were spaced at an average distance of 5 cM, facilitating physical separation as in the study conducted by. The primer pair sequences of SSR markers were retrieved from <http://www.cottongen.org/>.

Screening for polymorphism

The Polymerase Chain Reaction (PCR) for SSRs was performed for 30 repetitive cycles after initial denaturation at 94°C for 5 min. Each cycle covered denaturation (at 94°C for 45 sec), annealing (56°C for 45 sec) and extension (at 72°C for 60 sec) and, a final extension step at 72°C for 5 min. Amplified products were separated using FRAGMENT analyzer (TM-automated ce system advanced analytical technology). Prosize 2.0 software was used to visualize the amplified products and data points were scored for each SSR. The presence of polymorphic DNA fragments was scored as '1'; while, the absence of fragments was scored as '0.' Multiple polymorphic DNA fragments in the panel were identified as multiple marker loci.

Statistical analysis

Powermarker (v3.25) was used to reveal genetic diversity, number of alleles and Polymorphic Information Content (PIC). The genotypic data of polymorphic markers were used to determine genetic diversity in the natural population. Principal Component Analysis (PCA) and cluster analysis were used to confirm a reasonable population structure. The LD (r^2 value and p-value) was calculated with TASSEL 2.1 [4]. The parameter r^2 was used to graphically represent the LD curves with the R software (version R-3.5.0). LD was calculated by squared allele frequency correlations (r^2) between SSR loci. Loci were considered to be significant at $P \leq 0.005$ among all possible SSR loci. Values of LD between all pairs of SSR loci were plotted as LD plots using TASSEL 2.1 for estimating a general view of LD and pairwise r^2 .

TASSEL 2.1 software package was used to analyze the association between molecular marker data and the phenotypic data of economic traits using a General Linear Model (GLM)(P +G+Q) which ignores the population structure ($p \leq 0.01$). The Q-matrix was derived from clump software [5]. In association analysis, P-values describe the association between the marker and QTL while r^2 describes the effects of QTL.

Results

Molecular genetic diversity

Eighty-one SSR markers were polymorphic (27.18% of the total) among the 503 inbred lines with about three markers on each chromosome. Eighty-seven marker loci were detected, with an average of 1.08 loci per marker (ranging from 1-3). The average genetic diversity ranged from 0.051 to 0.695 with an average value of 0.350, while the average PIC values of inbred lines were 0.049 to 0.063 with an average value of 0.292.

Association mapping QTL for economic traits

The LD decay distance for the 503 accessions between all SSR markers was ~10 cM when the value of the cut-off for r^2 was set at 0.1 revealing the potential of association mapping for agronomical traits in upland cotton. A total of 56 SSR markers

were found to be associated with economic traits using GLM (Q). Several significant markers showing significant associations in both crop years were identified.

These are stable and consistent across multiple environments and populations plays an essential role in Marker-Assisted Selection (MAS). These markers were relegated to forty-eight QTL clusters which were widely distributed on 17 chromosomes. Amongst them, 6 QTL hotspots and 22 mQTL were associated to different traits. The sequence of significantly associated SSR markers were searched from the cottongen database and NAU-genome database of AD1 (*G. hirsutum* L.) and AD2 (*G. barbadense* L.) using a BLAST search with $E \leq 1e-10$ [6]. An average of 2.5 markers were detected on each chromosome (ranging from 1 to 6), with maximum of (5) markers on A01 (Chr01) and (06) D06 (Chr06).

In previous studies, a number of loci were identified as associated with yield components and fibre quality using SSRs under the standard threshold at $P \leq 0.05$ or $P \leq 0.01$. One marker was generally associated with several traits, presenting a pleiotropic effect. Moreover, some traits were found to be associated with multiple SSR loci, presenting epistasis. Effective Boll Number (EBN) is associated with BNL3649, CIR086 and BNL0836. Lint Percentage (LP) is associated with NAU0884. Whereas Fibre Percentage (FP) is associated with BNL1227, CIR311 and CIR082. Fibre Strength (FS) is associated with CIR086, BNL1897, NAU5411, BNL1053, BNL1047, BNL3649, BNL1227, NAU0895, BNL3649,

BNL1227, NAU0895, BNL3649 and BNL3281. Fiber Uniformity (FU) is associated with BNL1047, BNL3649, BNL1227, NAU0895, BNL3649 and BNL3281. There were 56 loci associated with economic traits at the 47 $P \leq 0.05$ significance level, among which were significant at the $P \leq 0.01$ level. The range of Phenotypic Variation Explained (PVE) observed was from (0.01%) NAU2083 to BNL3281 (2.74%), with an average of 0.61%.

Among the 16 traits, MV is associated with most loci upto 12 ($P \leq 0.05$) and 1 ($P \leq 0.01$). The remarkable contribution loci were BNL3482 (2.83 %), BNL0836 (1.52%) and MUSB0984 (0.3) at $P \leq 0.01$ significant level. Fiber elongation (mm) was associated with the second number of loci, up to a maximum of 7 ($P \leq 0.05$) and 4 ($P \leq 0.01$); NAU2083 (1.35%) at $P \leq 0.01$, NAU0895 (1.38) $P \leq 0.01$, NAU0884 (2.75 %) had a significant contribution at $P \leq 0.001$. The results have revealed that both the genotype and environment have genotypic effect.

The phenotypes are affected by multigenes; in addition, one gene can decide multiphenotypes [7].

Comparison of linkage and association mapping

At the $\alpha=0.01(-\log_{10}p)$ level, a total of 56 marker loci ($P \leq 0.01$) were found to be associated with 14 economic traits based on a linkage map with 298 SSR markers in cotton. The proportion of phenotypic variation explained by SSR markers ranged from 0.14-0.19 with an average of 0.14% (Table 1).

Table 1. Summary of significant SSRs by GLM (Q+K).

	#SSR	#QTL	$-\log(P)$ range	Average R^2	$R^2\%$ Range
PH	01	01	0.007	0.0181	0.001-0.019
FP	03	03	0.008	0.0337	0.013-0.015
WGP	03	03	0.012	-	0.122-0.145
FFSH	0	-	0.012	-	0.013-0.015
FFBN	07	07		-	0.014-0.015
FFSBN	0	02	0.014	-	0.015-0.014
EBN	02	01	0.014	0.02	0.001-0.019
SW	01	01	0.014	0.014	0.014
LW	02	02	0.013	0.013-0.014	0.013-0.014
LP	02	02	0.049	0.014	0.014
FUHML	01	02	0.564	0.006-0.002	0.006-0.224
FU	04	-	0.585	0.014-0.015	0.014-0.021
FE	06	05	0.014	0.013-0.021	0.014-0.012
FS	02	02	0.013	0.014-0.0149	0.143-0.149
MV	04	02	0.014	0.014-0.0154	0.014-0.015
SF	02	02	0.014	0.014	0.014-0.014
**Significant at the $P < 0.01$ level					

Four marker loci related to SCW, including BNL2440b with QTL named qSCW-A01-1, JESPR232 with QTL qSCW-A01-2, BNL3649b qSCW-A11-1, qSCW-A12-1, qSCW-D01-1, qSCW-A01-2 previously reported by. The GLM (Q) model is a reliable method for identification of phenotypic data after being BLUPed in order to reduce the environmental impact.

QTL clusters

A QTL cluster is a densely populated QTL region of the chromosome which contains multiple QTL associated with various traits [8]. In this study, we found 48 clusters, 6 hotspots, and 22 mQTL on 17 chromosomes. These results will allow cotton breeders to focus on regions containing the greatest number of QTL and having the highest phenotypic variance. Clusters were named on the basis of their position on the chromosomes. There were four QTL on A01-clusters, two QTL on A04-clusters, two QTL on A05-cluster, two QTL on A07-cluster and three QTL on A08-cluster, four QTL on D5-cluster, four QTL on D06-cluster, seven QTL on D07-cluster, four QTL on D08-cluster, one QTL on D04-cluster, 10 QTL on D09-cluster and one QTL on D06-cluster. These QTL clusters

and hotspots will allow breeders to detect multiple regions of the chromosomes.

Discussion

In this research, the SSR markers in QTL clusters and hotspots were analyzed by association mapping of a *G. hirsutum* L. collection introduced from abroad to explore the genetic basis of economic traits. In plants, association mapping is considered a powerful tool to identify the genetic basis of complex traits. The accuracy of association analysis is determined by the powerful genetic basis of upland cotton (*G. hirsutum* L.) and by reasonable statistical methods.

Our study successfully identified 56 QTLs for yield and fiber quality traits. The ratio of QTLs between the *G. hirsutum* L. (4 QTL) population and interspecific Gh × Gb (27 QTL) population and therefore illustrate clear similarities and differences between QTL cluster and Hotspot placement. The significantly associated SSR markers are shown in (Tables 2-4).

Table 2. SSR marker showing significant marker trait association GLM (Q); $P \leq 0.01$ for yield traits.

SSR	Trait	Trait type
NAU2083	WGP	Phenology
NAU5411	LP	Agronomical
NAU5411	WGP	Phenology
CIR263	LW	Agronomical
NAU0884	FSBN	Agronomical
NAU2235	FFSBN	Agronomical
BNL3994	FFSBN	Agronomical
BNL0836	WGP	Agronomical
BNL3281	EBN	Agronomical
BNL0786	WGP	Agronomical
BNL1604	FSBN	Agronomical
At a significance level of $P \leq 0.01$		

Table 3. Markers showing significant associations at GLM (Q); $P \leq 0.01$ for fiber traits.

SSR marker	Trait	Trait type
NAU5411	LP	Fibre
CIR263	LW	Fibre
NAU2265	SF	Fibre
CIR105	SF	Fiber
BNL2440	FUHML	Fiber
CIR272	FE	Fiber

NAU5411	FP	Fiber
CIR263	LW	Fiber
BNL3261	LP	Agronomical
BNL1897	LP	Quality
BNL1897	LP	Quality
CIR347	LP	Quality

Table 4. Common QTLs detected with previously studied traits.

S. no.	SSR marker [*]	F (GLM)	P (GLM)	Match	Chromosome	Association with other traits
1	NAU2083	2.214	0.012	WGP	A01	Morphological
2	NAU5411	0.473	0.014	LP	A01	Fibre
3	NAU5411	0.688	0.018	WGP	A01	Fibre
4	CIR263	0.209	0.018	LW	A03	Fibre
5	NAU0884	1.516	0.036	FSBN	A03	Gland/Gossypol
6	NAU2235	0.178	0.017	FFSBN	A04	
7	BNL3994	1.808	0.018	FSBN	A04	Yield trait
8	BNL0836	1.449	0.016	WGP	A11	Resistance locus
9	BNL3281	3.225	0.050	EBN	A13	
10	BNL3261	0.028	0.031	LP	A13	VW
11	BNL1897	1.703	0.011	LP	D01	EBN
12	BNL1897	7.224	0.011	LP	D01	VW
13	CIR347	0.022	0.021	LP	D01	Agronomic trait
14	BNL0786	4.883	0.011	WGP	D01	Fibre
15	BNL1122	0.027	0.014	EBN	D07	
16	BNL1604	1.004	0.008	FSBN	D07	
17	NAU2265	0.421	2.151	SF	D07	
18	BNL1604	4.785	5.431	FSBN	D07	

This detailed investigation of a *G. hirsutum* L. population revealed the presence of significant ($p \leq 0.001$) levels of variation for all morphological and fibre quality traits. Most of the traits were highly heritable, showing the presence of broad variation among the accessions.

In two different crop years (2012 and 2013), the means of some morphological and fibre quality traits were significantly higher as compared to the others, intimating that genetic variation owed to such traits might be more colligated to genotype $\times$ year interaction than genotypes only. The *Gossypium hirsutum* L. varieties belonging to Yangtze River ecotype is characterized by high lint percentage and larger bolls [9].

Molecular diversity revealed by SSRs in QTL clusters/hotspots

SSR markers are considered more diverse compared to other co-dominant markers and have been successfully used to find genetic diversity in plant species. Eighty one SSR primer pairs showed polymorphism among the 298 genome-wide SSR markers, accounting for about 27% of total primers. This corresponds to an average of about three markers per chromosome. Eighty-seven allelic loci have been detected, with an average of 2.379 loci per marker. The genetic diversity ranged from 0.049-0.302 with the average being 0.695. The range of polymorphic information content was 0.049-0.584 with the average being (0.639) which was higher than that of. These values were similar to those found in one study involving 53 *G. hirsutum* L. cultivars [10]. The germplasm is more diverse, and the number of alleles is higher. In fact, the

number of alleles observed per marker depends on the selection of the marker, collection of germplasm to be genotyped and the platform needed for resolution of amplified products.

Upland cotton from seven different ecotypes of China were identified in this study for maintaining the high level of genetic diversity needed for association mapping. The results revealed that genetic diversity, locus number, and PIC were consistent with previous reports. In this report, higher molecular diversity was revealed as compared to the previous studies, thus indicating the presence of sufficient genetic relationship among these accessions. Moreover, low genetic diversity was not only found in the Chinese upland cotton collection, but also collections of upland cotton from America and other countries.

Linkage disequilibrium

The precise LD status of a population predicts the success of an association analysis. Genome wide analysis of complex traits requires a large number of markers; moreover the sizes of SSR markers are sufficient for marker assisted selection. LD block determine that identified SSR marker is sufficient for marker assisted selection in upland cotton breeding programme in order to conduct the number of markers per chromosome provided several possible explanations moreover the size of the LD block between non-collinear markers, which include selection and co-selection of loci for population stratification, relatedness, and genetic drift [11]. Sometimes these elements lead to LD values resulting in false marker trait associations. Therefore, conducting population structure based association mapping in cotton germplasm resources can prevent spurious associations.

In this study, association mapping of economic traits and meta-analysis of QTL were used to validate the QTL clusters and hotspots for morphological traits, plant type, yield, and fiber and seed traits.

The names and ranges of the clusters and hotspots with linked markers detected in current association mapping has been mentioned. The present study also identified 48 QTL clusters and 27 QTL hotspot regions affecting two or more different fiber quality or yield component traits. This phenomenon of QTL clustering represents the linkage of genes and QTL or result from pleiotropic effects of a single QTL in the same genomic region.

QTL identified by SSR markers in QTL clusters and hotspots through association mapping

In the present research, association mapping using a naive model (GLM), and the Q model (GLM (Q)), model laid the foundation for a potential genomic assisted breeding in upland cotton. We detected SSR markers significantly associated with

genotypes and phenotypes of cultivars in an average environment and detected a number of elite alleles associated with agronomic and fiber quality traits.

The association analysis was based on BLUPed traits and 298 SSR markers. Significantly associated SSRs were detected for all traits using different models at $-\log_{10}P > 2.08$.

The positions of all QTLs were specified on the chromosome to identify significant chromosomal regions associated with same yield and fiber quality parameters.

A QTL cluster can be assigned to a 20 cM region with presence of more than two stable QTLs [12]. Estimated that on average 1 cM equated to a ~0.5 Mb physical region on cotton genome. Here we considered a cluster to be a 10 Mb (~20 cM) physical region enclosing two or more QTLs. Twelve QTL clusters on twelve chromosomes were detected with at least three QTLs in each cluster (Table 4). Eight clusters were in the A, and four were in D genome. All of these clusters contained QTLs for more than one different fiber quality or yield trait. The highest number of QTLs was five in D03-cluster, each for FSNB, WGP. These QTLs with increasing effects could simultaneously improve the yield with acceptable for growth period in upland cotton. Evaluated that on average 1 cM equated to a ~0.5 Mb physical region on the cotton genome [13]. Here we considered a cluster to be 10 Mb (~20 cM) physical region enclosing two or more QTLs. Twelve QTL clusters on twelve chromosomes were detected with at least three QTLs in each cluster (Table 3). Eight clusters were in the A, and four were D genome. All of these clusters contained QTLs for more than one different fiber quality or yield trait. The highest number of QTLs was five in the D03-cluster, each for FSNB or WGP. These QTLs with increasing effects could simultaneously improve the yield while maintaining an acceptable growth period in upland cotton.

Of the 56 QTLs reported in the current association map of the Gh × Gb population, chromosomes D02, D01, and D07 carried more QTL (5-7) followed by, A01, A04, A13 and D03 with 3-4 QTL together accounting for 30.95% of QTL detected. Chromosomes D09 and A08 carried the highest number of QTLs (seven and five QTLs). A04, D02, D01, D10, and A01 each carried two QTLs; whereas A05, D07, D09 and D06 each carried two QTLs. GLM ($p \leq 0.05$) detected 56 associated markers. As with previous studies in QTL mapping in cotton using SSR markers published in an online database, we also submitted our raw and associated SSR markers published in an online cotton EST database to facilitate access to genetic information [14]. Four markers coincided with previous results; BNL3281 (qBN-D01-1), JESPR300 (qWGP-A12-2), BNL3281 (qLP-D01-1), and BNL2440 (qFE15.1) (Table 5).

Table 5. Distribution of Quantitative Trait Loci (QTL) clusters on the chromosomes for fiber quality and yield traits.

S. no.	QTL cluster/hotspot	Approximate position on chromosome (Mb)	Number of QTLs	Name of QTL
1	A04-cluster	2.03-31.24	03	qFSNB-A04-1

				qFP-A04-1
				qWGP-A04-1
2	A12-cluster	2.33-2.48	02	qFP-A12-1
				qWGP-A12-1
3	A15-cluster	3.46-59.49	04	qFSBN-D01-1
				qFSBN-D01-2
				qFUHML-D01-1
				qFSBN-D03-2
4	D05 cluster	12.52-36.82	02	qFSBN-D05-1
				qFP-D05-1

Noteably, we also compared the map position of closely associated markers derived from the GLM (Q) analysis. These will be used for marker assisted selection to develop cultivars with high yield and superior fiber quality. For example, NAU2083 qNB-A01-1 was linked with an increase in the number of bolls per plant, BNL2440 qSCW-A11-1 was found to be associated with qSCW-A01-2, (NAU2474) was found to be associated with fibre strength, and qFL-D07-1 (CIR086) was found to be linked with fibre length. In view of the link between phenotype and genotype, these SSRs could be used in marker-assisted selection of economically important traits. Moreover, we can also use them to improve fiber quality; one should hybridize the material with alleles possessing qFS-A01-1 (NAU2474) and qFS-D07-1 (CIR086) controlling fiber strength. This will allow selection of suitable cultivars through marker assisted selection. Noteably, we also compared the map position of closely associated markers derived from the GLM

(Q) analysis. These will be used for marker assisted selection to develop cultivars with high yield and superior fiber quality. For example, qBN-A01-1 (NAU2083) was linked with an increase in the number of bolls per plant, (BNL2440) was found to be associated with qSCW-A01-1, NAU2474 was found to be associated with fibre strength, and (qFS-A01-1) CIR086 was found to be linked with FL [15]. In view of the link between phenotype and genotype, these SSRs could be used in marker-assisted selection of economically important traits. Moreover, we can also use them to improve fiber quality; one should hybridize the material with alleles possessing qFS-A01-1 (NAU2474) and qFS-D07-1 (CIR086) controlling fiber strength. This will allow selection of suitable cultivars through marker assisted selection (Table 6).

Table 6. Traits and associated SSRs in QTL clusters and hotspots of Upland cotton.

Yield	SSR	$-\log_{10}P$	R^2
EBN	BNL3281	1.53	1.15
	BNL1122	1.81	1.11
	NAU2859	1.42	1.12
Morphological			
FFSBN	NAU2083	2.05	1.6
	NAU2235	2.52	1.4
FB Num	CIR105	1.8	1.15
Growth			
WGP	NAU0884	0	1.3
	BNL3994	1.81	1.15
	BNL3031	2.98	1.15
	BNL0786	1.55	1.15
	BNL1604	5.42	1.19
	NAU2859	1.42	1.12
Fiber	NAU2265	2.16	1.15

SF	CIR105	1.8	1.15
FUHML	BNL2440	2.02	1.12
FE	CIR272	1.39	1.12
FP	NAU5411	0.01	1.4
LP	NAU5411	0.01	1.8
LW	CIR263	1.53	1.81

Detection of pleiotropy and correlation among yield and fiber quality traits

In this study, co-localization of QTL on chromosomes, or QTL clusters, were detected for fiber quality and yield traits in this study, indicating that the pleiotropic loci may control these traits. QTL co-localization on chromosomes, referred to as “QTL cluster/hotspots”, have been previously reported in cotton.

Relationships between different economic traits could be due to pleiotropy or gene linkage. SSRs anchored to different chromosomes were found to be associated with the same trait presenting the epistatic effect. One of the interesting factors in the current study is that the majority of the QTL were aggregated in 1-20 cM on every chromosome with exception of a few. These tight QTL clusters suggest a pleiotropic effect of QTL and present that they have the same pleiotropic effect. Of the 40 and 2 QTLs reported in the Gh and Gh × Gb populations, 6 QTL hotspots and 22 mQTLs were identified, respectively. The clusters in the former were a yield and fiber quality clusters including a (SI) and a (FL) QTL hotspot, while clusters in the latter were a yield and a fiber quality clusters which include a FL QTL hotspot. Fifteen markers demonstrated a pleiotropic effect controlling more than two economic traits in our current study [16].

High localization QTL was observed in the Gh population (D01 and D03). These yield and fiber quality traits were also found to be located in the same region with overlapping of hotspots between these two economic traits. Overlapping of FL, FS and LP QTL hotspots has been observed between two clusters on A07, with the same phenomena observed for BW and FL hotspots on D11, and D08 co-localization for LP, FL, FS, and FE. This close correlation between these yield and fibre quality parameters can be explained.

Candidate genes in QTL hotspots and functional annotation

The reference sequence of 56 elite allele loci associated with economic traits in QTL clusters, hotspots and mQTL were explored to find the corresponding position on NAU assembly for *G. hirsutum* L. Homologs within *Gossypium* and across other representative plant species were defined by a BLAST

search with an e-value of $E \leq 10$ and $1e-10$ respectively. In addition, we collected functional annotation data from the original sequencing projects and the CottonGen database. Among the identified significant SSRs, some have been anchored on physical positions which provided candidate regions in a confidence interval based on LD decay distance. Thirty six loci were found on the AD1 genome *G. hirsutum* L. whereas 22 loci were found on the AD2 genome (*G. barbadense* L.) as both are polyploids. Screening of candidate genes can be done by genome annotation. Five allelic loci were related to gene functional annotation of fiber quality traits in (*G. hirsutum* L.). For example, CIR311 was associated with FUHML traits on Chr15 (D01); its homologous genes in *G. hirsutum* L. and *Arabidopsis thaliana* L. were Gh_A01G0441 and AT5G53940, respectively, which is considered as cotton fiber expressed protein Yippee like which is considered as Cotton fiber expressed protein, while NAU5411 was found to be associated with SCW on Chr01 (A01) in *G. hirsutum* L.; its homologous genes in *G. hirsutum* L. and *A. thaliana* L. were Gh_A01G1576 and AT4G08570, respectively, which were annotated as a heavy metal detoxification super family protein and is considered as candidate gene for *Verticillium* wilt resistance. JESPR179b was found to be significantly associated with FS on Chr01 (A01) Gh_A02G0534 and AT3G06610, respectively, and annotated as a DNA binding enhancer protein related to promote fiber epidermal cell initiation which is a key factor in early fiber cotton fiber development. CIR082b was found FSN (N) to be significantly associated with FSN (N) Gh_A08G1186 (RVE5) and AT4G01280 Homeodomain like Super family protein and is considered with cotton fiber elongation and initiation. JESPR179 was found to be significantly associated with Gh_D08G1470 (RVE5) and with AT4G01280, and is a homeodomain like superfamily protein involved in transcriptional regulation of cellular protein in upland cotton verified that this protein could control fiber elongation [17]. These allelic variations and candidate genes in our study will be a valuable source for marker assisted selection in order to develop accessions with high yield and superior fiber quality (Table 7).

Table 7. Genomic location of polymorphic SSRs in QTL clusters and hotspots with functional association in cotton fibre development.

Marker	Genomic location	(Gene ID from <i>Gossypium hirsutum</i> L.)	Protein type function
CIR311	D01	Gh_A01G0441	Cotton fibre expressed protein
NAU5411	D02	Gh_A01G1576	Heavy metal detoxification protein
JESPR179b	D02	Gh_A02G0534	Fiber epidermal cell initiation
CIR082b	A10	Gh_D08G1470	Cotton fiber elongation and initiation
JESPR179b	D02	Gh_A08G1186 (RVE5)	Fiber elongation

The majority of available marker genetic information was deduced from populations which were created from biparental crosses having limited genetic background, thus resulting in marker assisted selection [18]. Numerous studies on the association mapping of upland cotton explored the feasibility for applying association analysis in order to understand the inheritance of complex traits [19]. Moreover, the QTL hotspots and clusters identified by current association mapping will be utilized by cotton breeders for marker assisted selection for increasing commercially valuable traits and can be used for further studies of high resolution mapping of QTL candidate gene and cloning.

Potential usages of QTL alleles identified in genomics assisted breeding in cotton

The results of this study could be helpful for exploiting diverse genetic diversity in local cotton stocks. These QTL could be used as selection tags for genetic fragments of introgression lines, and could be used to characterize different selection parents for crossing with different selection lines for developing accessions having superior values for the mentioned economic traits [20]. Based on these association results, the elite performing germplasm can be selected as parents in cotton breeding programmes. The data from this study can be compiled into a functional database that can be updated to keep the data from this study current and therefore useful in the cotton community.

Conclusion

In conclusion, the natural population panel of 503 accessions of upland cotton inbred lines from China was genotyped by 81 genome wide SSRs in QTL cluster/hotspots identified and phenotyped in two crop years and therefore revealed abundant phenotypic and genotypic diversity. The population represented different genotypic and phenotypic variation as well as a very low LD level. Twelve QTL clusters were identified that could be used in further breeding programs to improve the fiber quality having high yield in the upland cotton. A related issue to identify QTLs or molecular markers associated with yield, fiber and seed traits was.

This natural population can be used for the mapping and cloning of the novel QTL/genes that control the corresponding desired traits and can serve as a rich source of plant materials for the cotton research community.

Acknowledgement

We thank Dr. Lin of Group of Cotton Genetic Improvement, Huazhong Agricultural University Wuhan for providing some cotton breeding seeds and Huang Cong for his field work. This work was financially supported by the Fundamental Research Funds for the Central Universities (Grant No. 2014PY015).

Author Contributions

Conceived and designed the experiments: ZXL. Performed the experiment: MAW, WTW, CH. Analyzed the Data MAW, MMA, WTW. Wrote the paper MAW. All authors read the manuscript.

References

1. Said JI, Song M, Wang H, 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.
2. Nie X, Huang C, You C, et al. Genome-wide SSR-based association mapping for fiber quality in nation-wide upland cotton inbred cultivars in China. *BMC Genomics*. 2016;17:352.
3. Huang C, Nie X, 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.
4. Bradbury PJ, Zhang Z, Kroon DE, et al. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007;23(19):2633-5.
5. Liu F, Wang YH, Gao HY, et al. Construction and characterization of a bacterial artificial chromosome library for the allotetraploid *Gossypium tomentosum*. *Genet Mol Res*. 2015;14(4):16975-80.
6. Consortium U. UniProt: A hub for protein information. *Nucleic Acids Res*. 2015;43(D1):D204-12.
7. Crowell S, Korniliev P, Falcao A, et al. Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nat Commun*. 2016;7(1):10527.

8. Lacape JM, Llewellyn D, Jacobs J, et al. Meta-analysis of cotton fiber quality QTLs across diverse environments in a *Gossypium hirsutum* × *G. barbadense* RIL population. *BMC Plant Biol*. 2010;10:132.
9. Qin H, Chen M, Yi X, et al. Identification of associated SSR markers for yield component and fiber quality traits based on frame map and upland cotton collections. *PLoS One*. 2015;10(1):e0118073.
10. Atwell S, Huang YS, Vilhjálmsson BJ, et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*. 2010;465(7298):627-31.
11. Abdurakhmonov IY, Saha S, Jenkins JN, et al. Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica*. 2009;136(3):401-17.
12. Said JI, Lin Z, Zhang X, 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.
13. Li F, Fan G, Wang K, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet*. 2014;46(6):567-72.
14. Fang L, Wang Q, Hu Y, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat Genet*. 2017;49(7):1089-98.
15. Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23(14):1801-6.
16. Cai C, Ye W, Zhang T, et al. Association analysis of fiber quality traits and exploration of elite alleles in Upland cotton cultivars/accessions (*Gossypium hirsutum* L.). *J Integr Plant Biol*. 2014;56(1):51-62.
17. Shang L, Liang Q, Wang Y, et al. Identification of stable QTLs controlling fiber traits properties in multi-environment using recombinant inbred lines in Upland cotton (*Gossypium hirsutum* L.). *Euphytica*. 2015;205:877-88.
18. Wu J, Gutierrez OA, Jenkins JN, et al. Quantitative analysis and QTL mapping for agronomic and fiber traits in an RI population of upland cotton. *Euphytica*. 2009;165:231-45.
19. Khan MK, Chen H, Zhou Z, et al. Genome wide SSR high density genetic map construction from an interspecific cross of *Gossypium hirsutum* × *Gossypium tomentosum*. *Front Plant Sci*. 2016;7:436.
20. Ahmed MM, Guo H, Huang C, et al. Selection of core SSR markers for fingerprinting upland cotton cultivars and hybrids. *Aust J Crop Sci*. 2013;7(12):1912-20.

*Correspondence to

Zhongxu Lin

National Key Laboratory of Crop Genetic Improvement,

College of Plant Science and Technology,

Huazhong Agricultural University,

Wuhan 430070, Hubei,

China

E-mail: linzhongxu@mail.hzau.edu.cn