© Indian Academy of Sciences

ONLINE RESOURCES

CrossMark

Association between SSR markers and fibre traits in sea island cotton (*Gossypium barbadense*) germplasm resources

QI MA[1,2], JING ZHAO[3], HAI LIN[1], XINZHU NING[1], PING LIU[1], FUJUN DENG[1], AIJUN SI[1] and JILIAN LI[1]*

[1] *Cotton Research Institute, Xinjiang Academy of Agricultural and Reclamation Science/Northwest Inland Region Key Laboratory of Cotton Biology and Genetic Breeding, Shihezi 832000, People's Republic of China*
[2] *State Key Laboratory of Cotton Biology, Anyang 455000, People's Republic of China*
[3] *Crop Research Institute, Xinjiang Academy of Agricultural and Reclamation Science, Shihezi 832000, People's Republic of China*

Abstract

Identification of molecular markers associated with fibre traits can accelerate cotton marker-assisted selection (MAS) programmes. In this study, *Gossypium barbadense* germplasm accessions with diverse origins ($n = 123$) were used to perform association analysis of fibre traits with 120 polymorphic simple sequence repeat (SSR) markers. In total, 120 polymorphic primer pairs amplified 258 loci with a mean of 2.15 loci per primer. Population structure analysis identified three main clusters for the accessions, which indicated agreement of genetic and predefined populations. Marker–trait associations ($n = 58$) were detected for 10 fibre traits with 26 SSR markers located on 15 chromosomes. The R^2 (phenotypic variation explained) ranged from 3.19 to 15.21%. Two markers (NAU5465 and NAU3013) were found to be stably associated with boll number per plant (BNP) and fibre uniformity (UI), respectively. Four markers (BNL252, NAU3424, NAU3324 and CGR5202) associated with fibre quality traits preferentially clustered on the D8 chromosome, which was thus identified as an important candidate region for study molecular mechanisms underlying fibre quality and for use in breeding cotton cultivars for improving fibre quality. This study generated molecular data with a potential for better understanding of the genetic basis of the fibre traits and provided new markers for MAS in *G. barbadense* breeding programmes.

[Ma Q. I., Zhao J., Lin H., Ning X., Liu P., Deng F., Si A. and Li J. 2017 Association between SSR markers and fibre traits in sea island cotton (*Gossypium barbadense*) germplasm resources *J. Genet*. **96**, e55–e63. Online only: http://www.ias.ac.in/jgenet/OnlineResources/96/e55.pdf]

Introduction

Cotton is one of the most important natural fibre crops worldwide. Sea island cotton (*Gossypium barbadense*) and upland cotton (*G. hirsutum*) are two cultivated tetraploid species, accounting for 2% and 95%, respectively, of the annual worldwide cotton production (Cai *et al.* 2014). Although, *G. barbadense* has some shortcomings such as low-fibre yield, poor adaptability and difficulty in picking, it has superior fibre quality traits. The fibre traits of *G. barbadense* offer great potential for progress and development of the textile industry with respect to fibre breeding; therefore, more focus on research of these traits is required. Because fibre yield and quality traits are complex quantitative traits, tagging these traits will accelerate mining of novel genes and enable quick and efficient pyramiding of nonallelic quantitative trait loci (QTLs) by marker-assisted selection (MAS).

Currently, linkage analysis using segregation populations and association analysis using natural populations based on linkage map and linkage disequilibrium (LD), respectively, are the two main methods used for studying QTLs (Wang *et al.* 2013). Since linkage map construction entails selecting appropriate parents and then growing temporary or permanent populations, it is particularly a time-consuming process (Salvi and Tuberosa 2005). An additional limitation of linkage maps is that some loci

*For correspondence. E-mail: xjnkycottons@163.com.
Qi Ma and Jing Zhao contributed equally to this work.

Keywords. germplasm resources; fibre traits; association analysis; simple sequence repeat marker; *Gossypium barbadense*.

often do not separate and recombine when two particular parents are selected for constructing a linkage map (Li *et al.* 2013).

Association analysis is an effective approach for linking phenotypes and genotypes in plants, when information on population structure and LD is available (Thornsberry *et al.* 2001). With rapid development of DNA-based molecular markers, association analysis was first successfully used for identification of alleles at loci contributing to disease susceptibility in humans (Goldstein *et al.* 2003). Recently, this effective approach was widely used in various plant species, such as wheat (Arief *et al.* 2009; Gouis *et al.* 2011; Wang *et al.* 2014), maize (Thornsberry *et al.* 2001), rice (Agrama *et al.* 2007; Park *et al.* 2009), soybean (Gillman *et al.* 2014) and potato (Gebhardt *et al.* 2004; D'Hoop *et al.* 2014) to identify marker–trait associations. Association analysis was also used in cotton (*Gossypium* spp.), especially, *G. hirsutum* to identify associations between markers and a variety of phenotypic traits such as fibre quality traits (Abdurakhmonov *et al.* 2009; Cai *et al.* 2014; Nie *et al.* 2016), yield traits (Abdurakhmonov *et al.* 2007; Wu *et al.* 2008; Wang *et al.* 2013; Iqbal and Rahman 2017), agronomic traits (Yang *et al.* 2013; Liang 2014), early-maturing traits (Li *et al.* 2016) and salt tolerance (Shao *et al.* 2015).

All the above mentioned association analysis studies focussed on the phenotypic traits of *G. hirsutum*. However, *G. barbadense* has superior fibre quality traits and harbours many elite fibre genes. Therefore, it is necessary to perform an association analysis to identify QTLs associated with fibre traits in *G. barbadense* germplasm with the objective of pyramiding elite alleles and promoting MAS in cotton breeding programmes.

Materials and methods

G. barbadense germplasm used in this study was derived from a natural population which consisted of 123 representative *G. barbadense* accessions (see table 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/), including 113 accessions developed in China, seven accessions (Pima 3-79, Pima 90, Pima cotton, Pima 3, Pima 5, Pima Ø and uoc620) introduced from USA and three accessions (Bahatini, Minufei and Luoxiya 1) introduced from Africa.

A field experiment was carried out at the Korla experimental station of the Xinjiang Academy of Agricultural and Reclamation Science, Korla, Xinjiang province, a region most suitable for *G. barbadense* growth in China. In both years, 2014 and 2015, 123 *G. barbadense* germplasm accessions (two replicates each) were planted in a randomized plot design with a single-row plot and 80 individuals per row. A specific wide–narrow planting pattern was used, such that the row spacing was 66 (±10) cm with 12 cm between plants, and the plot size of each accession was 3.75 m^2.

Yield traits were measured in early October of 2014 and 2015 (table 2 in electronic supplementary material). Ten plants growing close to each other were selected to count the total boll number and the average was scored as BN per plant (BNP). A total of 30 bolls were harvested to determine weight per boll (BW) and lint percentage (LP). Seed cotton weight per plant (SWP) and lint weight per plant (LWP) were calculated based on the BNP, BW and LP. After ginning, fibres were mixed well and 10–15 g fibres were randomly sampled for each plant material. Fibre samples were independently tested for fibre-quality traits using HVI system (HFT 9000, Uster Technologies, Switzerland) at 20 °C and 65% relative humidity, at the Cotton Fibre Quality Inspection and Testing Centre, Ministry of Agriculture of China (Anyang, Henan, China). Variables that were tested included upper half mean length of fibre (UHML), fibre strength (STR), micronaire value (MIC), fibre elongation rate (ELO) and fibre uniformity (UI).

Genomic DNA was isolated from the *G. barbadense* germplasm as described by Paterson *et al.* (1993). SSR markers with broad genomewide coverage ($n = 500$) were used to screen the polymorphisms in 12 randomly selected accessions from the 123 *G. barbadense* germplasm accessions. Detailed information on the microsatellite markers, including primer sequences, can be accessed at the cotton microsatellite database (https://www.cottongen.org/search/markers).

Simple sequence repeat polymerase chain reaction (SSR-PCR) amplifications were performed using PCR Veriti 96-well Thermal cycler (ABI, Carlsbad, USA). The total reaction volume for PCR was 10 μL and reaction mix consisted of 1 μL DNA extract (20 ng/μL), each primer (300 nM), 400 μL dNTPs (l M), 0.5 U of *Taq* polymerase and 1 μL PCR reaction buffer. Amplifications were performed under the following conditions: (i) 95°C for 5 min, (ii) 15 cycles of 94°C for 45 s, 65°C for 45 s with a reduction of 1°C per cycle and 72°C for 1 min, (iii) 25 cycles of 95°C for 5 min, 45 s annealing at (optimum annealing temperature (T_m) for each primer pair = 5°C) and 72°C for 1 min, and (iv) a final step of 72°C for 10 min. Electrophoresis and staining were performed as described by Zhang *et al.* (2000).

The STRUCTURE v2.3.4 (Pritchard and Wen (2004), http://pritch.bsd.uchicago.edu/software.html), a model-based Bayesian method, was used to subdivide the *G. barbadense* germplasm accessions into individual clusters based on codominant genotypic data. For each run, the burn-in time was 50,000 and the number of replications was 100,000 (Pritchard and Wen 2007). Pritchard introduced a model-based clustering method to infer population structure and assign individuals to populations using multilocus genotype data.

Table 1. Summary of fibre traits of *G. barbadense* germplasm in the E1 and E2 environments.

Parameter	Environment	Fiber trait									
		BW	BNP	SWP	LWP	LP	UHML	UI	MIC	STR	ELO
Mean	E1	2.92	10.40	29.64	9.50	30.64	36.78	88.20	3.95	45.32	6.96
	E2	3.09	10.20	31.47	10.03	31.78	35.82	86.74	3.68	41.50	6.77
Minimum	E1	1.86	4.60	13.66	4.03	18.21	30.94	84.00	2.59	31.20	6.40
	E2	2.32	4.60	13.82	4.46	21.19	29.56	81.90	2.23	28.60	6.40
Maximum	E1	3.59	31.40	52.79	53.12	37.79	39.74	91.90	4.92	58.50	6.96
	E2	4.19	17.40	58.52	18.21	36.76	39.43	89.80	4.69	54.90	7.10
SD	E1	0.35	3.03	7.40	4.61	2.92	1.89	1.59	0.44	5.88	0.15
	E2	0.33	2.33	7.81	2.81	2.85	2.16	1.44	0.44	5.64	0.15
CV (%)	E1	11.99	29.13	24.97	48.53	9.53	5.14	1.80	11.14	12.97	2.16
	E2	10.68	22.84	24.82	28.02	8.97	6.03	1.66	11.96	13.59	2.22
h_B^2 (%)	E1	42.36	30.43	48.47	54.86	73.14	52.15	40.14	46.32	52.21	60.34
	E2	32.46	41.26	50.32	46.15	80.23	40.36	45.80	52.24	40.36	52.13

SD, standard deviation; CV, coefficient of variation; h_B^2, the broad sense heritability; E1, Korla in 2014; E2, Korla in 2015; BW, weight per boll; BNP, boll number per plant; SWP, seed cotton weight per plant; LWP, lint weight per plant; LP, lint percentage; UHML, upper half mean length of fibre; UI, fiber uniformity; MIC, micronaire value; STR, fibre strength; ELO, fibre elongation rate.

The TASSEL software (ver. 2.1, http://www.maizegenetics.net) was used to perform association analysis of fibre yield and quality traits. The mixed linear model (MLM) approach was used to conduct marker–trait association tests. The MLM association test was performed by simultaneous accounting of multiple levels of population structure (Q-matrix) and relative kinship among the individuals (K-matrix) according to Yu *et al.* (2006). The population structure matrix (Q) was identified by running STRUCTURE after K value was determined. The P value determined whether a QTL was associated with the marker and the R^2-marker evaluated the magnitude of the QTL effects (Agrama *et al.* 2007). The R^2 value represented the correlation between alleles at two loci, which is informative for evaluating the resolution of association approaches (Kantartzi and Stewart 2008).

The SPSS 21.0 software (http://www.spss.com.cn/) was used to conduct variation, correlation and principal component analysis (PCA). The broad-sense heritability (h_B^2) of each trait was estimated using SAS 8.1 software (SAS Institute 1999).

Results

Fibre yield and quality properties of G. barbadense germplasm

Hundred and twenty three *G. barbadense* germplasm accessions used in this study revealed a wide range of phenotypic variations in fibre yield and quality traits including BW, BNP, SWP, LWP, LP, UHML, UI, MIC, STR and ELO. LWP had the highest coefficient of variation (CV) of 48.53% in the E1 (2014 in Korla) environment, whereas UI had the lowest CV (1.66%) in the E2 (2015 in Korla) environment. The CV of most fibre traits was over 10%, which indicated a great variation in fibre traits in *G. barbadense* germplasm. The broad sense heritability (h_B^2) for 10 traits ranged from 30.43 to 73.14% in the E1 environment and ranged from 32.46 to 80.23% in the E2 environment. The highest h_B^2 value was for LP (73.14% in the E1 environment and 80.23% in the E2 environment), indicating that LP was less affected by environmental factors than the other nine traits (table 1).

Correlation analysis of fibre-yield and fibre-quality traits

Correlations among five fibre-yield traits and five fibre-quality traits of *G. barbadense* germplasm are listed in table 2. We observed significant trait correlations among 10 fibre traits in the E1 and E2 environments. The following variables were positively correlated ($P < 0.01$) in both the environments: BW with LWP, LP, MIC and ELO; BNP with LWP; SWP with LWP; LP with MIC; UHML with UI, STR and ELO; UI with STR and ELO; MIC with ELO; and STR with ELO. The following variables were negatively correlated ($P < 0.01$): BNP with LP, STR and ELO. These results indicated that there were positive correlations among fibre-yield traits and among fibre-quality traits, but negative correlations between fibre-yield and fibre-quality traits.

Principal component analysis of phenotypic traits

Results of the PCA of 10 fibre traits of *G. barbadense* germplasm are shown in table 3. Based on the principle of eigenvalue >1, the former three principal components with cumulative rates of 83.475% were selected, which could relatively and comprehensively reflect all the information. To be specific, the first principal component showed the maximum contribution (43.754%). Among the eigenvectors of the first principal component, the major phenotypic traits with relatively high load and positive sign were STR

Table 2. Correlations of fibre-yield and fibre-quality traits of *G. barbadense* germplasm in E1 and E2 environments.

Trait	BW	BNP	SWP	LWP	LP	UHML	UI	MIC	STR	ELO
BWP		0.814**	0.079	0.360**	0.373**	0.316**	0.287**	0.234**	0.351**	0.332**
BN	0.099		0.015	0.364**	−0.538**	0.119	0.233**	0.552**	−0.337**	−0.383**
SWP	0.355**[a]	0.561**		0.898**	0.842**	0.060	0.122	0.035	0.134	0.085
LWP	0.255**	0.836**	0.345**		0.950**	0.198*	0.241**	0.142	0.287**	0.227*
LP	0.268**	−0.392**	0.043	−0.187*		−0.128	0.238**	0.316**	0.300**	0.280**
UHML	0.105	−0.113	−0.108	−0.080	−0.289**		0.716**	−0.093	0.757**	0.616**
UI	0.156	−0.206*	−0.071	−0.159	−0.078	0.632**		0.191*	0.792**	0.711**
MIC	0.353**	−0.095	0.078	0.128	0.419**	−0.235**	0.155		0.290**	0.402**
STR	0.228*[b]	−0.236**	−0.163	−0.092	−0.015	0.693**	0.745**	0.073		0.853**
ELO	0.355**	−0.232**	−0.113	0.003	0.204*	0.426**	0.614**	0.455**	0.703**	

Correlation coefficients on the bottom left were the coefficients of the traits in E1, and those on the top right were the coefficients of the traits in E2. See table 1 for abbreviations.
[a] Significant at $P < 0.01$ level.
[b] Significant at $P < 0.05$ level

Table 3. Eigenvectors and percentages of accumulated contribution of PCA.

Phenotypic trait	Component 1	2	3
BW	0.612	−0.056	**0.546***
BNP	0.653	0.083	**0.701***
SWP	0.447	**0.757***	−0.459
LWP	0.688	**0.680***	−0.193
LP	0.729	0.673	−0.021
UHML	0.622	−0.499	−0.412
UI	0.725	−0.445	−0.287
MIC	0.438	0.022	0.559
STR	**0.814***	−0.441	−0.203
ELO	**0.776***	−0.431	−0.070
Eigen value	4.375	2.325	1.647
Contribution rate (%)	43.754	23.248	16.473
Cumulative percentage (%)	43.754	67.002	83.475

See table 1 for abbreviations.
The relatively high absolute value of each index in all the factors are in bold.
* Significant at $P < 0.05$ level.

and ELO, which mainly reflected the fibre-quality traits. Among the eigenvectors of the second principal component, major phenotypic traits with relatively high load and positive sign were SWP and LWP, which mainly reflected the yield traits. Among the eigenvectors of the third principal component, major phenotypic traits with relatively high load and positive sign were BNP and BW, which also reflected the yield traits (table 3).

SSR-marker polymorphism

Of the 500 SSR markers with broad genomewide coverage, 120 SSR markers (table 3 in electronic supplementary material includes the list of 120 SSR primers with their repeat motif and chromosomal locations, as reported in the literature) showed polymorphism reproducibility and locus specificity and they covered 258 alleles among the 123 *G. barbadense* germplasm accessions assayed. The number of polymorphic alleles per locus ranged from 1 to 4, with a mean of 2.15. The genetic diversity index ranged from 0.000 to 0.693 with an average of 0.673.

Population structure

The population structure of the *G. barbadense* germplasm was determined using the Structure v2.3.4 software. The ($\ln P(D)$) value increased continuously with K values ranging from 1 to 10 (figure 1a). Therefore, the most likely number of subpopulations (K) was determined according to ΔK value (Evanno *et al.* 2005). The maximum (peak) ΔK value (148.84) was observed for $K = 3$ (figure 1b), which indicated that the entire population could be divided into three subpopulations (figure 2).

The Structure model-based analysis showed that the model of three different subpopulations had the highest

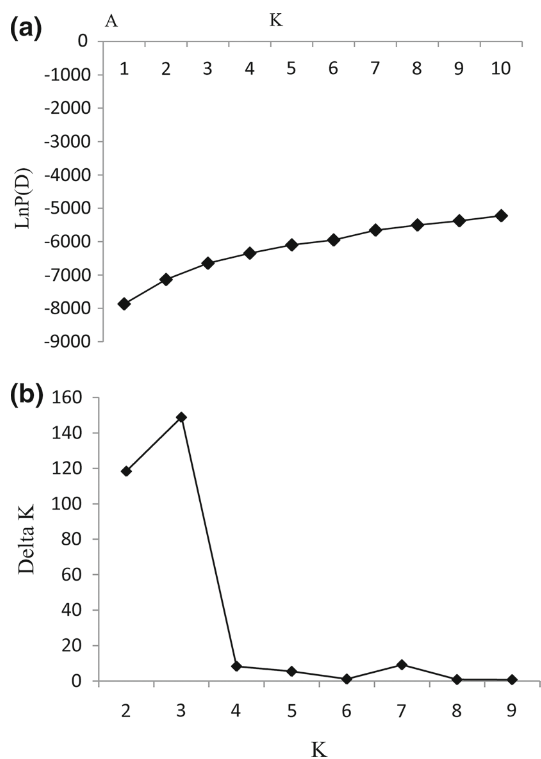

Figure 1. Estimated ln $P(D)$ and ΔK from 10 iterations obtained through Structure 2.3.3 analysis. (a) ln $P(D)$ for K values from 1 to 10 for simulations using the *G. barbadense* germplasm. (b) ΔK for K values from 2 to 9 for the *G. barbadense* germplasm.

posterior probability. The *G. barbadense* germplasm was assigned to three subpopulations with 50% or higher probability. The three subpopulations, designated as group 1, group 2 and group 3, consisted of 43, 41 and 29 germplasm accessions, respectively. A particularly noteworthy finding was that several germplasm lines from USA, such as Pima 90, Pima 3-79, Pima 3 and Pima 5 with long fruit branches and loose plant type were in group 1. The groups 2 and 3 contained almost all germplasm accessions with short fruit branches and compact plant type. In general, the germplasm accessions with compact plant type and short fruit branches had better fibre quality than those with loose plant type and long fruit branches. These results indicated that fibre-quality traits were very important factors for clustering of the *G. barbadense* germplasm, which was substantially consistent with the result that fibre-quality traits were the first principal component analysed by PCA. The remaining 10 germplasm accessions failed to group with a probability >50%. These germplasm accessions with mixed ancestral genetic backgrounds, including a few Xinjaing *G. barbadense* germplasm were artificially assigned to the 'mixed group'.

Marker–trait association

Association analysis identified marker–trait associations ($P < 0.05$) for all the traits examined. In total, 58 marker–trait associations, including 33 marker loci associated with fibre-yield traits and 25 loci associated with fibre-quality traits were identified using MLM of Tassel 2.1 software, with 26 SSR markers located on 15 chromosomes (tables 4 and 5). The R^2 value ranged from 3.19 (BNL226) to 15.21% (HAU2768). There were 38 and 20 marker loci detected in the E1 and E2 environments, respectively. It is worth mentioning that the markers NAU5465 and NAU3013 were found to be associated with BNP and UI, respectively, in both the environments. We also discovered that some SSR markers were simultaneously associated with more than one fibre trait: HAU2146 with BW, SWP and LWP; NAU3013 with BW, BNP and LWP; and NAU3110 with BNP, SWP and LWP. Seven marker loci were significantly associated with two fibre-yield traits each: NAU797 and NAU5465 with BNP and LWP, NAU5120 with BW and LP, HAU2768 with BNP and LP, HAU2146 with BW and SWP, NAU2687 with SWP and LWP, NAU803 with LWP and LP (table 4). In addition, a single marker locus (NAU3110) was significantly associated with four fibre-quality traits (UHML, STR, MIC and UI), whereas another one locus (NAU3791) was significantly associated with two fibre-quality traits (UHML and MIC). Three marker loci were significantly associated with three fibre-quality traits: PGML01548 with UHML, STR and ELO; NAU3013 with UHML, STR and UI; and HAU2828 with STR, MIC and ELO (table 5).

In the present study, we found that some SSR marker loci associated with fibre-yield and fibre-quality traits preferentially clustered on specific chromosomes. Three SSR marker loci (NAU797, NAU3110 and NAU1102) associated with fibre-yield traits (BNP and YWP) clustered on the D5 chromosome, whereas four marker loci (NAU3424, BNL252, CGR5202 and NAU3324) associated with fibre-quality traits (UHML, STR, MIC and ELO) clustered on the D8 chromosome. LD (R^2 values) among markers varied from 3.19 to 15.21% (tables 4 and 5). Of the 58 marker–trait associations, the association of HAU2768 with BNP and LP accounted for 10% or more of the total variation.

Discussion

Association analysis was first used in human populations to identify loci controlling disease susceptibility (Risch and Merikangas 1996; Schafer and Hawkins 1998) and this approach has since been widely used in numerous plants to identify DNA polymorphisms associated with various phenotypes. However, association analysis was almost rarely used in *G. barbadense* because of its scarce resources and poor adaptability. Therefore, identification of associations between markers and phenotypic traits in *G. barbadense* in this study will contribute to promoting the use of the association analysis in cotton research and further, verify the associated markers identified in *G. hirsutum*.

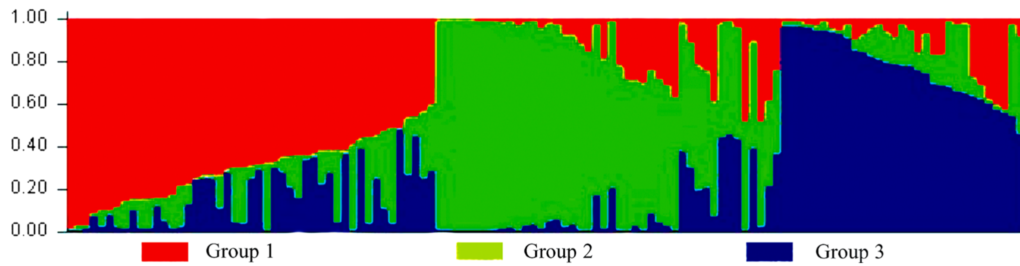

Figure 2. The summary plot of subpopulation structures in the *G. barbadense* germplasm analysed using Structure 2.3.3 by *Q*-matrix estimates ($K = 3$). Groups are represented in different colours as shown in figure legends. Each column represents one *G. barbadense* germplasm accession and partitioned into segments representing admixture of ancestral composition. The length of segments represents the percentage of a single ancestral background in that line. The columns (123 in total) were assigned to three groups.

Table 4. SSR markers associated with the same yield component traits in two different environments using the MLM method.

Trait	Environment	Marker loci	Chromosome	R^2 (%)	P
BW	E1	BNL226	Chr.03(A03)	3.19	0.0499
		HAU2146	Chr.09(A09)	4.01	0.0282
		NAU5120	Chr.07(A07)	3.62	0.0370
		NAU2820	Chr.07(A07)	3.35	0.0446
		NAU5163	Chr.01(A01)	6.35	0.0061
	E2	NAU3189	Chr.26(D12)	4.27	0.0242
		NAU3013	Chr.10(A10)	3.36	0.0452
BNP	E1	JESPR232	Chr.08(A08)	3.62	0.0368
		NAU5465	Chr.14(D02)	3.79	0.0325
		NAU3013	Chr.10(A10)	4.21	0.0245
		NAU797	Chr.19(D05)	4.65	0.0182
		HAU2768	Chr.06(A06)	12.08	0.0002
	E2	**NAU5465**	Chr.14(D02)	6.13	0.0070
		NAU3110	Chr.19(D05)	4.34	0.0225
SWP	E1	HAU2146	Chr.09(A09)	5.26	0.0126
		NAU2687	Chr.25(D06)	7.17	0.0037
	E2	NAU3110	Chr.19(D05)	5.45	0.0109
		NAU5465	Chr.14(D02)	4.96	0.0150
		NAU1102	Chr.19(D05)	3.34	0.0451
LWP	E1	HAU2146	Chr.09(A09)	3.22	0.0491
		NAU797	Chr.19(D05)	3.43	0.0425
		NAU3791	Chr.04(A04)	5.69	0.0094
		NAU3013	Chr.10(A10)	5.89	0.0082
		NAU2687	Chr.25(D06)	6.06	0.0074
	E2	NAU3110	Chr.19(D05)	6.36	0.0060
		NAU5465	Chr.14(D02)	5.18	0.0129
		HAU2828	Unknown	3.94	0.0295
		NAU803	Chr.14(D02)	3.60	0.0374
LP	E1	NAU5120	Chr.16(D07)	3.33	0.0466
		NAU803	Chr.14(D02)	3.87	0.0323
		NAU1322	Chr.24(D08)	6.20	0.0070
		BNL1604	Chr.16(D07)	8.47	0.0017
		HAU2768	Chr.06(A06)	15.21	0.0000

See table 1 for abbreviations; E1, Korla in 2014; E2, Korla in 2015. NAU5465 was found to be associated with BNP in both environment is in bold.

Ideally, an association analysis should include as much phenotypic and genotypic diversities as can be stably measured in a common environment. However, owing to scarce resources worldwide, our sample was restricted to 123 germplasm accessions approximately double of those used in Wang *et al.* (2013). The level of detected diversity was relatively high, with an average value of 0.673 (ranging from 0.000 to 0.693). The mean number of polymorphic alleles per locus was 2.15, which was higher than that reported by Wang *et al.* (2013) (1.66 loci per primer) based on 56 *G. barbadense* accessions. The genetic diversity and the number of SSR loci in the population in our study

Table 5. Simple sequence repeat (SSR) markers associated with the same fibre-quality traits in two different environments using the MLM method.

Trait	Environment	Marker loci	Chromosome	R^2 (%)	P
UHML	E1	PGML01548	Unknown	3.26	0.0482
		NAU3481	Chr.21(D11)	4.28	0.0241
		NAU797	Chr.19(D05)	4.81	0.0168
		NAU3791	Chr.04(A04)	4.90	0.0158
		NAU3013	Chr.10(A10)	6.47	0.0058
	E2	NAU3110	Chr.19(D05)	4.20	0.0253
		NAU3424	Chr.24(D08)	3.56	0.0391
STR	E1	BNL252	Chr.24(D08)	3.70	0.0345
		NAU3110	Chr.19(D05)	3.91	0.0297
		HAU2828	Unknown	4.33	0.0223
		NAU3013	Chr.10(A10)	4.51	0.0198
		PGML01548	Unknown	5.97	0.0075
MIC	E1	HAU2828	Unknown	4.47	0.0222
		NAU2083	Chr.01(A01)	5.50	0.0113
	E2	NAU3791	Chr.04(A04)	6.47	0.0057
		NAU3110	Chr.19(D05)	6.13	0.0071
		CGR5202	Chr.24(D08)	4.03	0.0281
UI	E1	**NAU3013**	Chr.10(A10)	3.39	0.0433
	E2	NAU1102	Chr.19(D05)	3.50	0.0409
		NAU3110	Chr.19(D05)	3.39	0.044
		NAU3013	Chr.10(A10)	3.36	0.045
ELO	E1	HAU2828	Unknown	3.47	0.0404
		NAU3324	Chr.24(D08)	3.48	0.0401
		PGML01548	Unknown	7.68	0.0026
	E2	NAU2820	Chr.07(A07)	4.15	0.0262

See table 1 for abbreviations; E1, Korla in 2014; E2, Korla in 2015.
NAU3013 was found to be associated with UI in both environment is in bold.

Table 6. Comparison of associated SSR markers with other researches.

Our research		Previous researches	
Marker	Trait	Trait	Reference
NAU803	LWP	FF	Shen *et al.* (2005); Cai *et al.* (2014)
NAU5163	BW	UHML	Duan (2015)
NAU2687	SWP, LWP	STR, ELO	Qian (2009)
NAU3110	LWP, UHML	UHML, STR	Duan (2015); Qin *et al.* (2015)
PGML01548	STR, ELO	Senescence-related trait	Li (2015)
BNL226	BW	LP	Qian (2009)
BNL1604	LP	LP	Zeng *et al.* (2009)
JESPR232	BNP	STR	Qian (2009)
NAU797	BNP, UHML	Verticillium wilt resistance	Mei (2012)
NAU2083	MIC	BNP	Said *et al.* (2015)

See table 1 for abbreviations; FF, fibre fineness.

for association analysis were higher and lower, respectively, than those observed in prior studies in *G. hirsutum* (Mei *et al.* 2013; Qin *et al.* 2015), which indicated that the germplasm in our study showed a high level of genetic variation.

Population structure is an important factor that typically leads to spurious associations. Therefore, conducting population structure analysis on natural population is a prerequisite for association analysis (Kline *et al.* 2001; Flintgarcia *et al.* 2005). The results of the Structure analysis ($K = 3$) demonstrated that the *G. barbadense* germplasm was divided into three subgroups and the *G. barbadense* cotton accessions from USA were significantly different from those from Xinjiang, China. Further analysis revealed that the Xinjiang germplasm accessions did not belong to the same subgroup. The results indicated that there was still frequent gene exchange among *G. barbadense* cotton of Xinjiang. Because population structure analysis does not require any prior knowledge of the origin, geo-

graphic distribution, phenotypic characteristics and other factors, it can truly reflect the genetic differences between materials, excluding the interference of human factors on the subpopulation divisions (Kantartzi and Stewart 2008).

In the present study, marker–trait analysis using MLM revealed significant associations between cotton fibre traits and SSRs in the *G. barbadense* germplasm. In total, 58 marker–trait associations were identified, including 33 marker loci associated with fibre-yield traits and 25 marker loci associated with fibre-quality traits. Ideally, the associations between markers and traits should be examined in two ways—significance of marker–trait association (*P* values) using TASSEL software and the marker–trait associations found in other QTL studies (Kantartzi and Stewart 2008). Although, several markers were also found in previous QTL studies in *G. hirsutum*, the traits associated with a particular marker were not the same as in *G. barbadense* (table 6). Therefore, we hypothesize that the marker–trait associations identified in *G. hirsutum* and *G. barbadense* were inconsistent because of their dissimilar genomic structures and phenotypic traits. However, it is noteworthy that the markers on the D8 chromosome (Chr. 24) associated with fibre strength were identified and validated in previous studies (Chen *et al.* 2009; Kumar *et al.* 2012; Cai *et al.* 2014). Simultaneously, our research indicated that there are important markers related to fibre strength (BNL252) and other fibre-quality traits (NAU3424, NAU3324 and CGR5202) on the D8 chromosome. Therefore, we believe that this chromosome is an important candidate region for the study of molecular mechanisms underlying fibre quality and for use in breeding cotton cultivars to improve fibre quality.

In summary, genomewide association studies are advantageous as they enable the entire genome to be assessed for trait-associated variants. These studies were proved to be effective for elucidating the genetic basis of complex traits in plants (Ingvarsson and Street 2011). Application of association mapping to plant breeding appears to be a promising means of overcoming the limitations of conventional linkage mapping. This study demonstrated that SSR markers associated with fibre traits of *G. barbadense* germplasm provided a reference and basis for identification of more elite genes of fibre traits in *G. barbadense* and enhanced the data from QTL studies for the implementation of MAS.

Acknowledgements

This work was financially supported by the State Key Laboratory of Cotton Biology Open Project Fund (CB2015A10), the Youth Fund of Xinjiang Academy of Agricultural and Reclamation Science (YQJ201503), the National Natural Science Fund (31260360) and the Special Fund of Corps Breeding (2016AC027).

Supplementary Information (SI) Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12041-017-0849-9) contains supplementary material, which is available to authorized users.

References

Abdurakhmonov I. Y., Buriev Z. T., Saha S., Pepper A. E. and Musaev J. A. 2007 Microsatellite markers associated with lint percentage trait in cotton, *Gossypium hirsutum*. *Euphytica* **156**, 141–156.

Abdurakhmonov I. Y., Saha S., Jenkins J. N., Buriev Z. T., Shermatov S. E., Scheffler B. E. *et al*. 2009 Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* l. Variety germplasm. *Genetica* **136**, 401–417.

Agrama H. A., Eizenga G. C. and Yan W. 2007 Association mapping of yield and its components in rice cultivars. *Mol. Breed.* **19**, 341–356.

Arief V. N., Delacy I. H., Dieters M. J., Crossa J., Dreisigacker S. and Basford K. E. 2009 Pattern analysis of marker-trait association profiles derived from 25 years of international wheat performance trials. *Sabrao J. Breed. Genet.* **41**, 1–7.

Cai C. P., Ye W. X., Zhang T. Z. and Guo W. Z. 2014 Association analysis of fiber quality traits and exploration of elite alleles in upland cotton cultivars/accessions (*Gossypium hirsutum*). *J. Integr. Plant Biol.* **56**, 51–62.

Chen H., Qian N., Guo W., Song Q., Li B., Deng F. *et al*. 2009 Using three overlapped rils to dissect genetically clustered qtl for fiber strength on chro.d8 in upland cotton. *Theor. Appl. Genet.* **119**, 605–612.

D'Hoop B. B., Keizer P. L., Paulo M. J., Visser R. G., van Eeuwijk F. A. and van Eck H. J. 2014 Identification of agronomically important qtl in tetraploid potato cultivars using a marker-trait association analysis. *Theor. Appl. Genet.* **127**, 731–748.

Duan Z. Q. 2015 *Association mapping of yield and fiber traits in upland cotton*. Zhejiang University, Hangzhou.

Evanno G., Regnaut S. and Goudet J. 2005 Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620.

Flintgarcia S. A., Thuillet A. C., Yu J., Pressoir G., Romero S. M., Mitchell S. E. *et al*. 2005 Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J. Cell Mol. Biol.* **44**, 1054–1064.

Gebhardt C., Ballvora A., Walkemeier B., Oberhagemann P. and Schuler K. 2004 Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Mol. Breed.* **13**, 93–102.

Gillman J. D., Tetlow A., Hagely K. and Boersma J. G. 2014 Identification of the molecular genetic basis of the low palmitic acid seed oil trait in soybean mutant line rg3 and association analysis of molecular markers with elevated seed stearic acid and reduced seed palmitic acid. *Mol. Breed.* **34**, 447–455.

Goldstein D. B., Tate S. K. and Sisodiya S. M. 2003 Pharmacogenetics goes genomic. *Nat. Rev. Genet.* **4**, 937–947.

Gouis J. L., Bordes J., Ravel C., Heumez E., Faure S., Praud S. *et al. 2011* Genome-wide association analysis to identify chromosomal regions determining components of earliness in wheat. *Theor. Appl. Genet*. **124**, 597–611.

Ingvarsson P. K. and Street N. R. 2011 Association genetics of complex traits in plants. *New Phytol.* **189**, 909–922.

Iqbal M. A. and Rahman M. U. 2017 Identification of marker–trait associations for lint traits in cotton. *Front. Plant Sci.* **8**, 86.

Kantartzi S. K. and Stewart J. M. 2008 Association analysis of fibre traits in *Gossypium arboreum*, accessions. *Plant Breed.* **127**, 173–179.

Kline J. B., Moore D. J. and Clevenger C. V. 2001 Activation and association of the tec tyrosine kinase with the human prolactin receptor: mapping of a tec/vav1-receptor binding site. *Mol. Endocrinol.* **15**, 832–841.

Kumar P., Singh R., Lubbers E. L., Shen X., Paterson A. H., Campbell B. T. *et al*. 2012 Mapping and validation of fiber strength quantitative trait loci on chromosome 24 in upland cotton. *Crop Sci.* **52**, 1115–1122.

Liang B. 2014 Association analysis of agronomic traits in upland cotton using SSR markers. *Cotton Sci.* **26**, 377–386.

Li C. Q., Xu X. J., Dong N., Ai N. J. and Wang Q. L. 2016 Association mapping identifies markers related to major early-maturating traits in upland cotton (*Gossypium hirsutum* L.). *Plant Breed.* **135**, 483–491.

Li L. B. 2015 *Association analysis of senescence-related trait, resistance to Verticillium Wilt in upland cotton*. Chinese Academy of Agricultural Sciences. Beijing.

Li W. Y., Zhang B., Zhang J. N., Chang X. P., Li R. Z. and Jing R. L. 2013 Exploring elite alleles for chlorophyll content of flag leaf in natural population of wheat by association analysis. *Acta Agron. Sinica* **38**, 962–970.

Mei H. X. 2012 *Genetic diversity and association analysis of main breeding target traits in upland cotton cultivars of China*. Nanjing Agricultural University, Nanjing.

Mei H. X., Zhu X. F. and Zhang T. Z. 2013 Favorable QTL alleles for yield and its components identified by association mapping in Chinese upland cotton cultivars. *PLoS One* **8**, e82193.

Nie X. H., Huang C., You C. Y., Li W., Zhao W. X., Shen C. *et al*. 2016 Genome-wide ssr-based association mapping for fiber quality in nation-wide upland cotton inbreed cultivars in china. *BMC Genomics* **17**, 352.

Park E. J., Chung J. W., Park Y. J., Chung I. M., Ahn J. K. and Kim G. H. 2009 Association analysis of the amino acid contents in rice. *J. Integr. Plant Biol.* **51**, 1126–1137.

Paterson A. H., Brubaker C. L. and Wendel J. F. 1993 A rapid method for extraction of cotton (*Gossypium*, spp.) genomic dna suitable for rflp or pcr analysis. *Plant Mol. Biol. Rep.* **11**, 122–127.

Pritchard J. K. and Wen W. 2004 *Documentation for Structure software*. The university of Chicago Press, Chicago.

Pritchard J. K. and Wen W. 2007 *Documentation for Structure software, version 2.1*. University of Chicago, Chicago.

Qian N. 2009 *Genetic diversity and association analysis of gene (QTL) of breeding target traits of upland cotton*. Nanjing Agricultural University, Nanjing.

Qin H. D., Chen M., Yi X. D., Bie S., Zhang C., Zhang Y. C. *et al*. 2015 Identification of associated ssr markers for yield component and fiber quality traits based on frame map and upland cotton collections. *PLoS One* **10**, e0118073.

Risch N. and Merikangas K. 1996 The future of genetic studies of complex human diseases. *Science* **9**, 350–354.

Said J. I., Knapka J. A., Song M. Z. and Zhang J. F. 2015 Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between *Gossypium hirsutum* and *G. hirsutum* x *G. barbadense* populations. *Mol. Genet. Genomics* **290**, 1615–1625.

Salvi S. and Tuberosa R. 2005 To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci.* **10**, 297–304.

Schafer A. J. and Hawkins J. R. 1998 DNA variation and the future of human genetics. *Nat. Biotechnol.* **16**, 33–39.

Shao B., Wang H., Zhao Y., Chen W., Guo Z., Gong H. *et al*. 2015 Association analysis of salt tolerance with ssr markers in *Gossypium hirsutum* l. *Cotton Sci.* **27**, 118–125.

Shen X. L., Guo W. Z., Zhu X. F., Yuan Y. L., Yu J. Z., Kohel R. J. *et al*. 2005 Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. *Mol. Breed.* **15**, 169–181.

Thornsberry J. M., Goodman M. M., Doebley J., Kresovich S. and Nielsen D. 2001 Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* **28**, 286–289.

Wang G., Leonard J. M., Von Z. J., James P. C., Ross A. S. and Rieralizarazu O. 2014 Marker–trait association analysis of kernel hardness and related agronomic traits in a core collection of wheat lines. *Mol. Breed.* **34**, 177–184.

Wang X. Q., Yu Y., Li W., Guo H. L., Lin Z. X. and Zhang, X. L. 2013 Association analysis of yield and fiber quality traits in *Gossypium barbadense* with ssrs and sraps. *Genet. Mol. Res.* **12**, 3353–3362.

Wu J., Jenkins J. N., Mccarty J. C., Saha S., and Percy R. 2008 Genetic association of lint yield with its components in cotton chromosome substitution lines. *Euphytica* **164**, 199–207.

Yang X. L., Zhou X. D., Liu H. W., Wang X. F., Wu L. Q., Li Z. K. *et al*. 2013 Aflp markers associated with important agronomic traits in cotton. *Cotton Sci.* **25**, 211–216.

Yu J., Pressoir G., Briggs W. H., Vroh Bi. I., Yamasaki M., Doebley J. F. *et al.* 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208.

Zeng L. H., Jr W. R. M., Gutiérrez O. A. and Boykin D. L. 2009 Identification of associations between ssr markers and fiber traits in an exotic germplasm derived from multiple crosses among *Gossypium*, tetraploid species. *Theor. Appl. Genet.* **119**, 93–103.

Zhang J., Wu Y. T., Guo W. Z. and Zhang T. Z. 2000 Fast screening of microsatellite markers in cotton with PAGE/silver staining. *Cotton Sci.* **12**, 267–269.

Received 3 December 2016, revised form 12 March 2017; accepted 24 march 2017
Unedited version published online: 28 March 2017
Final version published online: 1 December 2017

Corresponding editor: Manoj Prasad