Mol Genet Genomics
DOI 10.1007/s00438-017-1346-9

ORIGINAL ARTICLE

Association mapping analysis of fiber yield and quality traits in Upland cotton (*Gossypium hirsutum* L.)

Mulugeta Seyoum Ademe[1] · Shoupu He[1] · Zhaoe Pan[1] · Junling Sun[1] · Qinglian Wang[2] · Hongde Qin[3] · Jinhai Liu[4] · Hui Liu[5] · Jun Yang[6] · Dongyong Xu[8] · Jinlong Yang[4] · Zhiying Ma[7] · Jinbiao Zhang[9] · Zhikun Li[7] · Zhongmin Cai[4] · Xuelin Zhang[10] · Xin Zhang[2] · Aifen Huang[11] · Xianda Yi[3] · Guanyin Zhou[4] · Lin Li[9] · Haiyong Zhu[1] · Baoyin Pang[1] · Liru Wang[1] · Yinhua Jia[1] · Xiongming Du[1]

Received: 12 September 2016 / Accepted: 29 June 2017

Abstract Fiber yield and quality are the most important traits for Upland cotton (*Gossypium hirsutum* L.). Identifying high yield and good fiber quality genes are the prime concern of researchers in cotton breeding. Association mapping offers an alternative and powerful method for detecting those complex agronomic traits. In this study, 198 simple sequence repeats (SSRs) were used to screen markers associated with fiber yield and quality traits with 302 elite Upland cotton accessions that were evaluated in 12 locations representing the Yellow River and Yangtze River cotton growing regions of China. Three subpopulations were found after the estimation of population structure. The pair-wise kinship values varied from 0 to 0.867. Only 1.59% of the total marker locus pairs showed significant linkage disequilibrium (LD, $p < 0.001$). The genome-wide LD decayed within the genetic distance of ~30 to 32 cM at $r^2 = 0.1$, and decreased to ~1 to 2 cM at $r^2 = 0.2$, indicating the potential for association mapping. Analysis based on a mixed linear model detected 57 significant ($p < 0.01$) marker–trait associations, including seven associations for fiber length, ten for fiber micronaire, nine for fiber strength, eight for fiber elongation, five for fiber uniformity index, five for fiber uniformity ratio, six for boll weight and seven for lint percent, for a total of 35 SSR markers, of which 11 markers were associated with more than one trait. Among marker–trait associations, 24 associations coincided with the previously reported quantitative trait loci (QTLs), the remainder were newly identified QTLs/genes. The QTLs identified in this study will potentially facilitate improvement of fiber yield and quality in the future cotton molecular breeding programs.

Keywords Upland cotton · SSRs · Linkage disequilibrium · Association mapping · Fiber quality · QTLs

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-017-1346-9) contains supplementary material, which is available to authorized users.

✉ Yinhua Jia
jiayinhua_0@sina.com

✉ Xiongming Du
dujeffrey8848@hotmail.com

[1] State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences (ICR, CAAS), P. O. Box 455000, Anyang, Henan, China

[2] Henan Institute of Science and Technology, Xinxiang, China

[3] Cash Crop Institute, Hubei Academy of Agricultural Sciences, Wuhan, China

[4] Zhongmian Cotton Seed Industry Technology Co., Ltd, Zhengzhou, China

[5] Jing Hua Seed Industry Technologies Inc, Jingzhou, China

[6] Cotton Research Institute of Jiangxi Province, Jiujiang, China

[7] Key Laboratory of Crop Germplasm Resources of Hebei, Agricultural University of Hebei, Baoding, China

[8] Guoxin Rural Technical Service Association, Hebei, China

[9] Zhongli Company of Shandong, Shandong, China

[10] Hunan Cotton Research Institute, Changde, China

[11] Sanyi Seed Industry of Changde in Hunan Inc, Changde, China

Published online: 26 July 2017

Introduction

Cotton is one of the most important fiber crops in the world and provides the main products of lint for the industry. The genus *Gossypium* includes approximately 45 diploid ($2n = 2x = 26$) and 5 allotetraploid ($2n = 4x = 52$) species distributed mostly in tropical and subtropical regions worldwide (Fryxell 1992). Diploid species ($2n = 26$) are classified into eight genomic groups (A–G and K), occurring naturally in Africa, Asia, Americas, and Australia (Wendel and Cronn 2003). Tetraploid species, including *G. hirsutum* and *G. barbadense*, arose in the New World from inter-specific hybridization between A and D genome species, which are believed to have originated from *G. herbaceum* race *africanum* and *G. raimondii*, respectively (Brubaker et al. 1999). As complexity of tetraploid crop, which is considered as "doubled interspecific hybrid" that lead to the permanent fixation of heterozygosity, *G. hirsutum* and *G. barbadens*e have inherited excellent genes from the A and D genomes species. Identification of these genes may facilitate improvements of fiber traits in cotton (Saha et al. 2006). With the improving and upgrading of textile industry, the demand for good fiber quality is increasing rapidly. Therefore, breeding new cultivars with high yield and elite fiber qualities is becoming one of the goals of modern cotton breeding.

Evaluating the valuable germplasm will benefit improvements of fiber yield and quality traits in cotton the breeding programs. The primary objective of worldwide cotton breeding programs is the genetic improvement of fiber yields and quality. However, improving cotton fiber quality is challenging due to the narrow genetic base of modern cotton cultivars (Iqbal et al. 2001; Rungis et al. 2005; Abdurakhmonov et al. 2012) and negative genetic correlation between fiber yield and quality traits (Culp and Lewis 1973). Both drawbacks to cotton improvement illustrate that it is necessary to effectively explore existing germplasm resources for important quantitative traits of agronomic importance, including valuable genes from wild species. It showed that the breeding of every new and typical cultivar greatly benefitted from the identification of valuable germplasm from different areas with different backgrounds (Van-Esbroeck and Bownam 1998). Tagging economically important quantitative trait loci (QTLs) from *G. hirsutum* accessions in different genetic backgrounds will accelerate the mining of novel genes, enable the quick and efficient pyramiding of nonallelic QTLs by marker-assisted selection (MAS), and enhance the breeding of elite commercial cotton cultivars.

Many QTLs related to fiber yield and quality traits have been identified based on QTL linkage mapping methods by assembling intra-specific segregating populations of *G. hirsutum* (Shappley et al. 1998; Ulloa and Meredith 2000, 2005; Zhang et al. 2003, 2012; Wang et al. 2006; Shen et al. 2005, 2006, 2007; Gore et al. 2014) in inter-specific populations from crosses between *G. hirsutum* and *G. barbadense* (Jiang et al. 1998; Kohel et al. 2001; Paterson et al. 2003; Mei et al. 2004; Lacape et al. 2003, 2005, 2010; Park et al. 2005), and intra-specific *G. arboreum* populations (Ma et al. 2008) in the past years. The precise detection of detecting QTL locations depends on the recombination fraction between the QTLs for the traits of interest and adjacent markers (Mackay and Powell 2007). Those QTLs identified in different linkage groups have been summarized in one genetic map by Said et al. (2013, 2015). In linkage mapping, the major limitation hampering fine mapping is the availability of few meiotic events (Jannink et al. 2001; Flint-Garcia et al. 2005; Hall et al. 2010).

Association mapping, based on linkage disequilibrium (LD), offers an alternative method for mapping QTLs that evaluates whether certain alleles within a population are found with specific phenotypes more frequently than expected (Flint-Garcia et al. 2003). Since the first association analysis in *G. arboreum* for the evaluation of fiber quality (Kantartzi and Stewart 2008), association mapping study has been widely employed in Upland cotton germplasm resources to identify QTLs associated with important fiber yield and quality traits (Abdurakhmonov et al. 2008, 2009; Zeng et al. 2009; Mei et al. 2013; Cai et al. 2014; Fang et al. 2014; Qin et al. 2015), with epistasis and environmental interaction of fiber yield traits (Jia et al. 2014a), with Verticillium wilt resistance (Jia et al. 2014b; Zhao et al. 2014), and with drought and salinity tolerance (Saeed et al. 2014; Jia et al. 2015). Some stable QTLs about fiber quality identified in different genetic backgrounds were also detected in association mapping (Fang et al. 2014; Hugie et al. 2016). The number of QTLs detected in a given study depends on the type and size of the mapping population, genetic markers used, traits investigated, number of environments used for phenotyping, and genome coverage (Semagn et al. 2010). Once QTLs affecting a trait of interest are accurately identified, marker tags are effective tools that allow the mobilization of the genes of interest from donor lines to the breeding material to develop superior cotton cultivars through marker-assisted selection (MAS) programs. In addition, exploring the genetic basis of fiber quality traits using association mapping will increase the scope of effective utilization of germplasm resources. The main objective of the present study was to explore QTLs underlying fiber quality traits in diverse Upland cotton genotypes using SSR markers. The QTLs identified in this study will facilitate fiber yield and quality improvement in future breeding programs.

Materials and methods

Plant material

A total of 302 elite Upland cotton (*G. hirsutum* L.) accessions obtained from the Institute of Cotton Research, Chinese Academy of Agricultural Sciences (ICR, CAAS) germplasm banks, were used for this study. These materials were selected on the basis of phenotypic expression of important agronomic and fiber quality traits, such as fiber yield, boll number and size, maturity, fiber quality, and biotic and abiotic stress tolerance (Du et al. 2007). Among these accessions, 253 (83.8%) were collected from diverse cotton growing areas of China. The remaining 49 (16.2%) were introduced from 13 different countries (USA, Australia, Burundi, Chad, Ivory Coast, Kenya, Pakistan, Russia, Sudan, Turkmenistan, Uganda, Uzbekistan, and Vietnam). A detailed description of the 302 Upland cotton accessions is provided in Supplemental File 1.

Field experiment locations

The field experiments were conducted during the main rain seasons of 2012 and 2013 at the 12 different locations of the Yellow River (YeR) and Yangtze River (YaR) cotton growing regions of China. Among the 12 experimental sites, six of them, i.e., Anyang (Baibi), Anyang (ICR), Baoding, Dongying, Hejian, and Xinxiang were selected in the YeR region. The remaining six locations, i.e., Changde, Changsha, Hefei, Jingzhou, Jiujiang, and Wuhan were located in the YaR region. The agro-ecological characteristics of the two major cotton growing regions vary greatly in terms of soil type and fertility, rainfall amount and distribution, temperature, length of growing season, and cotton production and management practices (Hsu and Gale 2001; Wu and Guo 2005). In general, early and intermediate maturing cotton varieties are usually grown in the YeR region, whereas medium and late maturing cotton varieties are favored in the YaR region due to the availability of abundant rainfall and the long growing season.

Experimental treatments, design and phenotypic data collection

In this study, the 302 Upland cotton accessions were considered treatments in each experimental location. The field experiments were arranged in a randomized complete block design (RCBD) with three replications. Each accession was grown in a single row plot 8.0 m in length, with an inter-row spacing of 0.8 m. Intra-row spacings were 0.3 and 0.5 m in the YeR and YaR region, respectively. Each replication was set 1.0 m apart. The cotton crop was established depending on locally recommended practices by either direct sowing of linted or delinted seeds or transplanting using seedlings. Planting was conducted in the months of April and May in each production season. Other recommended agronomic practices, including seedbed preparation, seeding rate, seed treatment, thinning, fertilizer rate and application, irrigation, weeds and pest control, were followed to maintain a good crop stand in each location.

Fiber yield and quality traits were measured on ten representative plants selected randomly from each plot (genotype) and tagged for identification. At the stage of plant maturity (approximately 70% ball open), 30 well developed open boll samples (3 bolls per plant) from the middle branches of tagged plants were harvested in each plot and weighed for seed cotton yield. After ginning of the sample bolls using a roller gin, approximately 150 g lint samples were used from each plot to evaluate various fiber traits. The data for fiber quality traits were measured using high volume instruments (HVI) in the Laboratory of Quality & Safety Risk Assessment for Cotton Products (Anyang), Ministry of Agriculture, People's Republic of China.

Genomic DNA extraction, SSR genotyping, and allele recording

A total of 198 genomic SSR markers selected previously (Pan et al. 2008) including BNL (13), CIR (2), CM (1), DPL (14), GH (9), HAU (25), JESPR (7), MGHES (3), MUCS (2), MUSS (4), NAU (100), STV (3), and TMB (15), were used for genotyping. The primer sequences for the SSR markers were obtained from the publicly available CottonGene (http://www.cottongene.org) and Cotton Marker (http://www.cottonmarker.org) databases. The SSR markers are distributed across the 26 chromosomes (AD genome) of Upland cotton with a mean of 7.6 markers per each chromosome (Supplemental File 2).

Genomic DNA was extracted from the sample leaves of each cotton genotype using the cetyltrimethyl ammonium bromide (CTAB) method described by Zhang and Stewart (2000). In the field, 2–3 young and fully expanded sample leaves were randomly collected from ten seedlings of each cotton genotype. All sample leaves were stored at −70 °C. During genomic DNA extractions, sample leaves were freeze-dried in liquid nitrogen. The quality of the genomic DNA extraction was assessed using 1% agarose gel electrophoresis.

PCR solution preparation, PCR amplification, and gel electrophoresis were performed as described by Zhang and Stewart (2000). The PCR amplification was run in a total volume of 10 μL [1.0 μL 10× PCR buffer, 0.2 μL dNTP mix, 0.2 μL Taq DNA polymerase, 0.65 μL forward primer, 0.65 μL reverse primer, 6.1 μL ddH_2O, and 1.2 μL DNA (50 ng μL^{-1})]. The PCR amplification protocol was performed in a thermal cycler (MJ Research,

Waltham, USA) with initial denaturation at 95 °C for 3 min, 30 cycles of denaturation at 94 °C for 30 s, annealing at 57 °C for 50 s, and an extension at 72 °C for 50 s, followed by a final extension at 72 °C for 7 min, after which the samples were maintained at 4 °C. PCR amplification products were separated via gel electrophoresis using 8% polyacrylamide gel (PAGE) with 1× TBE electrolyte solution. Electrophoresis was conducted using a double-sided vertical gel casting apparatus (HTJY, China) with a capacity of 96 comb lanes. A standard 50-bp DNA ladder was used as a reference to determine the sizes of amplified DNA bands. The gels were stained with silver solution to easily identify polymorphisms during DNA band reading. The visualization and recording of DNA bands were performed manually using a UV light board.

For each SSR marker locus, amplified alleles with clear band separations were scored manually as either '1' (present) or '0' (absent) binary characters. For each SSR marker, the patterns of amplified alleles were denoted by letter codes as A, B, C, and so on according to molecular weight.

Population structure and kinship

The genetic structure of the cotton population was evaluated using model-based Bayesian (MBB) analysis provided in Structure version 2.3 (Pritchard et al. 2010). To infer the number of subpopulations (Q-matrix), an admixture model and correlated allele frequencies between populations were applied according to Falush et al. (2003). Structure was run under 10,000 burn-ins and 100,000 Markov Chain Monte Carlo iterations after the burn-ins. Five replicates of Structure were run by fixing the number of groups (K) from 1 to 10. An average likelihood value, $L(K)$, was calculated for each K across all runs. The number of group was determined according to Evanno et al. (2005) by estimating ΔK, an ad hoc quantity based on the second order change of the log probability of data with respect to the number of subpopulations inferred by the Structure software. The uppermost likelihood of ΔK was taken to represent the optimal value of K and the cotton accessions were clustered into different subpopulations. Cotton accessions were assigned to different subpopulations based on the maximum probability of membership and a threshold probability of membership ($\geq$0.50) as inferred from the MBB structure analysis.

To estimate the genetic relationships between individuals, the relative kinship coefficients (K-matrix) among all pairs of Upland cotton population were estimated using SPAGeDi version 1.3, with the negative kinship value set to zero (Hardy and Vekemans 2002).

Linkage disequilibrium

The genome-wide LD was analyzed for each pair of SSR marker loci using TASSEL version 2.1 (Bradbury et al. 2007; http://www.maizegenetics.net), based on the genetic map of Shen et al. (2005, 2007). Alleles present at less than 5% frequency were removed using the TASSEL site filtration function before LD analysis to avoid spurious associations on low-frequency markers because minor alleles are usually problematic and biased for LD estimates between pairs of loci (McRae et al. 2002; Abdurakhmonov et al. 2008; Myles et al. 2009). LD was estimated by calculating the square value of the correlation coefficient (r^2) between all pairs of SSR loci (Hill and Robertson 1968). The significance of r^2 among all pairs of SSR loci was evaluated using TASSEL version 2.1 with the rapid permutation test (Churchill and Doerge 1994) in 1000 shuffles. Each pair of loci was grouped as linked (marker loci located on the same chromosome) or unlinked (marker loci located on different chromosomes). The LD was estimated for linked and unlinked markers in the cotton population and in the three subpopulations (P-I, P-II, and P-III) inferred from the structure analysis. The r^2 values for pairs of SSR loci were plotted as a function of map distances (centimorgan, cM) between loci and LD decay (at $r^2 < 0.001$). A nonlinear logarithmic regression curve was drawn to describe an overall relationship between r^2 and the cM of SSR markers on the same chromosome using TASSEL (http://www.maizegenetics.net/tassel). The linkage map performing LD was obtained from a linkage map constructed by Shen et al. (2007).

Phenotypic data analysis

The phenotypic data for fiber yield and quality traits collected on the 302 Upland cotton accessions from each environment for two consecutive years (2012 and 2013) were subjected to analysis of variance (ANOVA) for RCBD, according to Gomez and Gomez (1984). The homogeneity of variance tests was evaluated to determine whether data from individual environments (E) could be pooled to evaluate $G \times E$ using a combined ANOVA. The linear model that describes the factors or sources contributing to the variance of each trait in a multi-environment experiment was used according to Freeman (1973). In this model, each trial was considered a separate environment: $Y_{ijk} = \mu + G_i + B(E)_{jk} + E_j + GE_{ij} + e_{ijk}$ where, Y_{ijk} is the measurement of the ith genotype, jth environment, and kth replication; μ is the overall mean; G_i is the effect of the ith genotype; $B(E)_{jk}$ is the effect of the jth block in the kth environment; E_j is the effect of the jth environment; GE_{ij} is the effect of interaction of the ith genotype with the jth environment; and e_{ijk} denotes the experimental error (plot residual).

The general linear model (GLM) procedure of Statistical Analysis System version 9.2 (SAS Institute 2007) was used for the analysis of variance in the field experiment data to test the genotype (G), environment (E) and genotype by environment ($G \times E$) effects. The mean values of the genotypes for each phenotypic trait were compared using the least significance difference (LSD) test at 5% level of probability. The Pearson linear correlation coefficients (r) were estimated among all phenotypic traits according to Kwon and Torrie (1964). For each trait, the phenotypic variance, genotypic variance, and broad-sense heritability (H^2) were calculated from the ANOVA mean squares according to Burton (1951) as $H^2 = \frac{V_G}{V_P} \times 100$, where V_G is genetic variance that is estimated as $\frac{M_g - M_e}{r}$; V_P denotes phenotypic variance and is calculated as $V_G + V_e$; r is the number of replicates; M_g is the mean squares of genotypes (treatment); and M_e represent the mean squares of error.

Marker–trait association

The MLM, which takes into account both population structure (Q-matrix) and relative kinship (K-matrix) and better controls spurious associations (Yu et al. 2006), was used to determine the marker–trait associations using TASSEL version 2.1 (Bradbury et al. 2007). The MLM ($Q + K$) is a statistical model containing both fixed effects (observed to be quantitative in terms of explanatory variables that are treated as if the quantities were nonrandom) and random effects (variance components model) that can be described in Henderson's notation (Henderson 1975) as $y = X\beta + Zu + \varepsilon$ where, y is the known vector of observations; β is an unknown vector of fixed effects; u is an unknown vector of random effects; X and Z are known design matrices; and ε is an unknown vector of random errors. A data file of the Q-matrix was created using Structure 2.3 (Pritchard et al. 2010), while the K-matrix was constructed using SPAGeDi (Hardy and Vekemans 2002). The p value ($p < 0.01$) of each SSR marker was used to decide whether a marker was significantly associated with any fiber quality traits studied. Despite the use of MLM Type I error rates would be increased with multiple tests.

Results

Population structure and kinship

The population structure was analyzed with Structure version 2.3, and population structure was clearly observed. The estimated likelihood, $L(K)$, values consistently increased with increasing K values from 1 to 10 (Fig. 1). However, based on an estimation of the ΔK peak, the highest number of populations occurred at $K = 3$ (Fig. 2). This result suggested that this Upland cotton population could be grouped into three subpopulations (Fig. 3). The maximum probability of membership in the three MBB subpopulations (P-I, P-II, and P-III) is indicated in Supplementary Material 1. A total of 127 accessions were assigned to P-I with probabilities ranging from 0.353 to 0.978. P-II contained 83 accessions with probabilities ranging from 0.356 to 0.910. The P-III group consisted of 92 accessions with probabilities ranging from 0.358 to 0.912. Based on the threshold probability (≥0.5) of membership, 223 (73.8%) accessions were grouped into the three inferred subpopulations, while the remaining 79 (26.2%) were considered to have

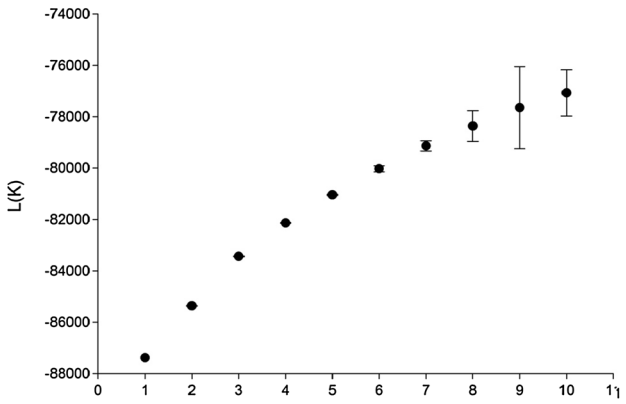

Fig. 1 $L(K)$, likelihood of data, for subpopulations (K) from 1 to 10 in Upland cotton (*Gossypium hirsutum*) population. For each K value, five independent runs were considered. The estimated $L(K)$ values consistently increased with increasing K values from 1 to 10

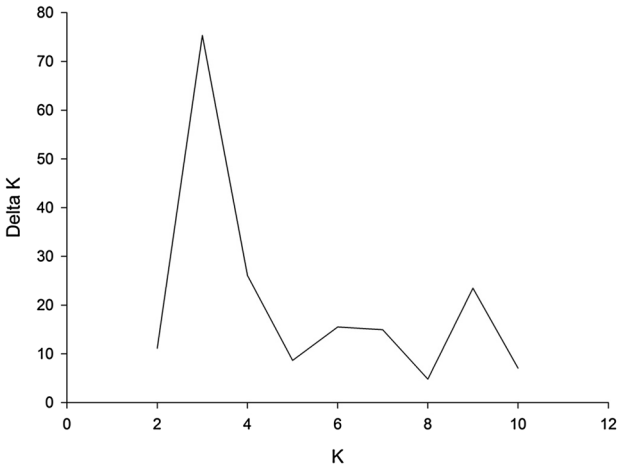

Fig. 2 Detection of the number of subpopulations (K) using delta K (ΔK) in Upland cotton (*Gossypium hirsutum*) population. K (ΔK) is an ad hoc quantity based on the second order change of the log probability of data with respect to the number of subpopulations, for K ranging from 1 to 10. The ΔK peak occurred at $K = 3$

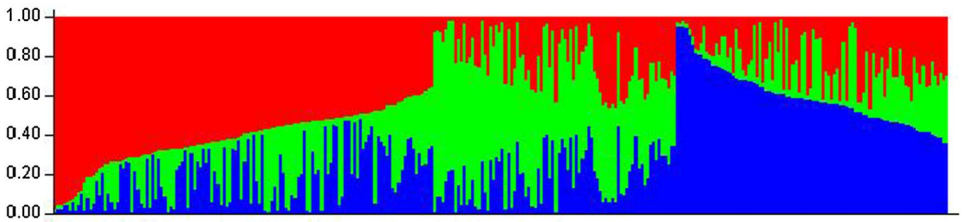

Fig. 3 *Q*-plot showing model-based Bayesian (MBB) structure clustering for the Upland cotton (*G. hirsutum* L.) population. The *colored subsection within each vertical bar* (from *top to bottom*) indicates membership coefficient (*Q*) of the germplasm and the three inferred sub-populations: P-I (*red*), P-II (*green*), and P-III (*blue*)

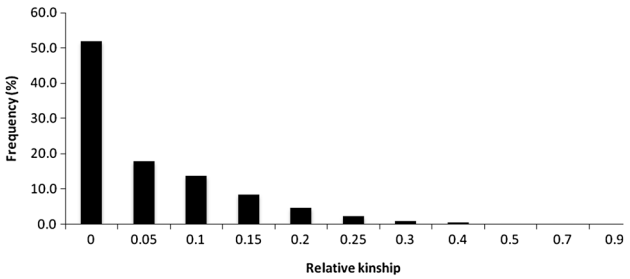

Fig. 4 Frequency of pair-wise relative kinship estimates between the Upland cotton (*Gossypium hirsutum*) population. Relative kinship coefficient ranged from 0 to 0.867. The majority of the pairs of cotton accessions (51.91%) showed zero estimated kinship values

mixed parentage from more than one population. Therefore, three subpopulations were considered correct structure of the population.

Pair-wise kinship was estimated based on the 198 informative SSR markers. The values of pair-wise kinship varied from 0 to 0.867. The majority of the pairs of cotton genotypes (51.91%) had zero estimated kinship values, while 31.41% of the kinship values ranged from 0 to 0.1, 15.15% from 0.1 to 0.25, and the remaining pairs of genotypes (1.53%) had kinship values greater than 0.25 (Fig. 4). The results indicated involvement of some common parental genotypes in the breeding history of the Upland cotton genotypes studied.

Linkage disequilibrium

The linkage disequilibrium (LD) was calculated for total, linked, and unlinked markers loci in the cotton population and MBB subpopulations (Table 1). In this Upland cotton population, the average r^2 of marker locus pairs was 0.0083, and only 1.59% (2174/137,026) of the total marker locus pairs had significant LD ($p < 0.001$). In the linked marker pairs, a mean r^2 of 0.5582 ($p < 0.001$) was obtained, which was higher than the LD estimates of unlinked marker pairs and the whole population. Among the linked marker pairs, 82.8% showed LD values greater than 0.2, while for the unlinked marker pairs, the percentage was 13.6%. The mean r^2 of marker locus pairs in the MBB subpopulations ranged from 0.4210 to 0.4669 ($p < 0.001$), which were higher than the mean LD estimated in the population. Further analysis revealed that the mean r^2 values of marker locus pairs for linked markers in the MBB subpopulations were in the range of 0.5886 to 0.6096 ($p < 0.001$) and high than those of unlinked markers. Among the three subpopulations, P-I had the lowest mean LD.

The LD decay rate of the population was investigated using the r^2 values ($p < 0.05$) of the linked SSR marker

Table 1 Linkage disequilibrium (r^2) analysis in the whole Upland cotton (*Gossypium hirsutum* L.) population and model-based Bayesian (MBB) sub-populations

Population	Overall r^2 (mean)[b]	Significant r^2 (%)[a]			Significant r^2 (mean)[a]		
		Linked[c]	Unlinked[d]	Total[e]	Linked	Unlinked	Total[f]
Panel	0.0083	0.46	1.13	1.59	0.5582	0.1105	0.2401
P-I	0.0132	0.43	0.46	0.89	0.5886	0.2668	0.4210
P-II	0.0177	0.43	0.38	0.80	0.6096	0.3049	0.4669
P-III	0.0162	0.37	0.34	0.72	0.6045	0.3147	0.4659

[a] Level of significance: $p < 0.001$

[b] Overall mean r^2 was calculated for the total SSR pairs of loci generated for each population group

[c] Linked: pairs of loci on the same chromosome

[d] Unlinked: pairs of loci from different chromosomes

[e] Total: sum for linked and unlinked

[f] Total: mean for total SSR pairs in each population

pairs. To identify the average genome-wide LD decay, r^2 values of the SSR marker pairs on the same chromosomes were plotted as a function of genetic distance in cM (Fig. 5). The non-linear regression curve exhibited a clear decay of LD as genetic distance increased. In the scatter diagram, genome-wide LD extended within the genetic distance of ~30 to 32 cM at $r^2 = 0.1$, whereas it was reduced to ~1 to 2 cM at $r^2 = 0.2$, demonstrating the potential for association mapping.

Variations of fiber yield and quality traits

Phenotypic performance data of the 302 accessions evaluated from 2012 to 2013 across 18 locations (year as environment) were used in this study. Homogeneity of variance tests allowed for a combined analysis across the environments. The ANOVA for fiber yield and quality traits (Table 2) revealed significant ($p \leq 0.01$) variations in the main effects of genotypes (G), environments (E), and genotype by environment ($G \times E$) interactions. ANOVA also indicated that except for fiber elongation (FE) and fiber uniformity index (UI), the genotype effects accounted for the largest portion of the sums of squares for boll weight (BW), lint percent (LP), fiber length (FL), micronaire (Mic), fiber strength (FS), and fiber uniformity ratio (UR), followed by environment and $G \times E$ interactions.

The broad-sense heritability (H^2) values and correlation coefficients of all fiber yield and quality traits were estimated (Table 2). The H^2 values were ranged from 75.18 to 97.08%. The lowest H^2 was recorded for UI, while the highest was for LP. Majority of fiber yield and quality traits showed higher H^2 values. Simple linear correlation coefficients (r) calculated among all pairs of fiber quality traits (Table 3) revealed that among all traits, FL had positive significant ($p < 0.01$) correlations with several yield and quality traits, including BW, FS, FE, UI, and UR. Alternatively, Mic exhibited negative significant ($p < 0.01$) relationships with BW, FS, FE, and UR. While, LP was negative relationships with the important quality traits, FL, and FS Among all pairs of correlations, 60% were positive suggesting that simultaneous improvement of quality traits might be used to achieve high lint yield.

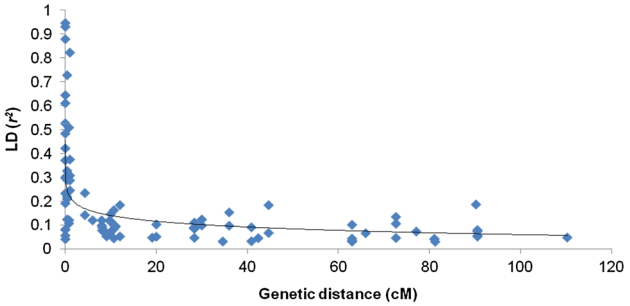

Fig. 5 LD (r^2) decay against genetic distance among linked SSR marker loci in the Upland cotton (*G. hirsutum* L.) population. Level of significance was determined at $p < 0.001$. The inner fitted trend line is a non-linear logarithmic regression curve of r^2 on genetic distance. The genome-wide LD decayed within the genetic distance of ~30 to 32 cM at $r^2 = 0.1$, and reduced to ~1 to 2 cM at $r^2 = 0.2$

SSR loci associated with fiber quality traits

LD-based association mapping was performed with the MLM of TASSEL version 2.1. The analysis revealed 57 significant ($p < 0.01$) marker–trait associations, involving 35 SSR markers (Table 4). The QTLs detected include seven for FL, ten for Mic, nine for FS, eight for FE, five for UI, five for UR, six for BW and seven for LP. The

Table 2 Analysis of variance for fiber yield and quality traits in Upland cotton (*Gossypium hirsutum* L.) population evaluated from 2012 to 2013 in China

Variations	Fiber yield and quality traits							
	BW	LP	FL	Mic	FS	FE	UI	UR
Rep (E)	0.51^{ns}	29.32**	26.49**	1.93**	166.17**	0.53**	41.97**	4138.99**
Genotype (G)	7.25**	377.31**	57.83**	5.04**	126.51**	0.79**	16.99**	4903.51**
Environ (E)	251.48**	3207.49**	779.58**	83.57**	1909.19**	32.06**	342.27**	78,766.67**
$G \times E$	0.52**	8.47**	2.03**	0.25**	5.09**	0.11**	2.37**	214.78**
Pooled error	0.24	3.75	1.24	0.13	3.05	0.05	1.69	120.59
Mean	5.11	36.05	29.46	4.80	29.77	6.80	85.18	140.76
CV	9.60	5.37	3.78	7.44	5.86	3.36	1.52	7.80
LSD 0.05	0.19	0.73	0.42	0.13	0.66	0.09	0.49	4.14
H^2	90.64	97.08	93.82	92.75	93.10	82.46	75.18	92.97

BW (boll weight, g), LP (lint percent, %), FL (fiber length, mm), Mic (micronaire, μg in.$^{-1}$), FS (fiber strength, cN tex^{-1}), FE (fiber elongation, %), UI (fiber length uniformity index, %), UR (fiber uniformity ratio, %), CV (coefficients of variation, %), H^2 (broad-sense heritability, %)

ns non-significant

*, ** Significant at the 0.05 and 0.01 level of probability, respectively

Table 3 Pearson's linear correlation coefficients (*r*) among fiber yield and quality traits of Upland cotton (*Gossypium hirsutum* L.) population

Traits	BW	LP	FL	Mic	FS	FE	UI
BW (g)							
LP (%)	0.154**						
FL (mm)	0.185**	−0.049**					
Mic (μg in.$^{-1}$)	−0.018*	0.262**	−0.276**				
FS (g tex^{-1})	0.105**	−0.068**	0.535**	−0.258**			
FE (%)	0.167**	0.109**	0.314**	−0.130**	0.355**		
UI (%)	0.123**	0.043**	0.303**	0.020*	0.305**	0.199**	
UR (%)	0.154**	−0.121**	0.619**	−0.404**	0.795**	0.343**	0.608**

largest numbers of SSR markers were associated with Mic (10) and FS (9). The proportion of phenotype variation explained by the markers ranged from 1.7 to 8.9%, with a mean of 3.5%. In the present study, SSR markers were associated with different numbers of fiber yield and quality traits. Out of the 35 SSR markers, 24 (68.5%) were associated with only one type of fiber yield and quality traits, while the remaining 11 (31.5%) were simultaneously associated with two, three, four, or seven different traits. NAU749 was simultaneously associated with all fiber quality traits and lint percent. Among the 57 marker–trait associations, 16 were highly significantly associated with these eight fiber yield and quality traits ($p < 0.001$), including five SSR markers (HAU1185, HAU423, NAU749, NAU874 and TMB436) associated with FL, two SSR markers (NAU1102 and NAU749) associated with Mic, one SSR marker (TMB436) associated with Fs, two SSR markers (HAU1185 and NAU749) associated with FE, one SSR marker (NAU749) associated with UI, two SSR markers (NAU749 and TMB436) associated with UR, one SSR marker (JESPR274) associated with BW, and two SSR markers (DPL212 and DPL715) associated with LP.

Discussion

LD is the basis of association analysis (Flint-Garcia et al. 2003). In this Upland cotton population, the average r^2 of marker locus pairs was 0.0083 while only 1.59% (2174/137,026) of the total marker locus pairs were determined to have significant LD ($p < 0.001$). The current result was lower than previous reports (Abdurakhmonov et al. 2009; Mei et al. 2013; Cai et al. 2014), but it was similar to the result of Zhao et al. (2014). The discrepancy between different results may be due to differences in the type and density of markers, types of cotton genotypes, and probability of statistical significance used. Compared to other cultivated crops, the percentage of SSR loci pairs in LD observed in cotton is much lower than the 100% reported for cultivated barley germplasm (Malysheva-Otto et al. 2006), the 52–86% reported for durum wheat (Maccaferri et al. 2005), and the 27–93% reported for an elite maize germplasm (Stich et al. 2005). Since out-crossing is common and a high rate of recombination exists in allopolyploid cottons (Brubaker et al. 1999), this might be one of the various factors associated with the observed low level of LD along with mutation, selection and genetic drift that occurred during the domestication of Upland cotton germplasm (Abdurakhmonov et al. 2009).

LD between the molecular marker and causative polymorphisms was strongly influenced by the physical linkage and the population structure. Further examination of the LD analysis results showed that the extents of LD for linked markers in the whole population and sub-populations were higher than those of unlinked markers in the present study, which may suggest the suitability of the cotton population used in this study for association mapping of important QTLs. Cai et al. (2014), Zhao et al. (2014) and Qin et al. (2015) also reported similar results. According to Abdurakhmonov et al. (2009), in addition to physical linkage that mainly causes LD between linked markers, selection, relatedness, genetic drift, bottleneck, and population stratification might generate LD between unlinked marker loci pairs. In agreement with the results reported by Mei et al. (2013) and Zhao et al. (2014), the present study showed varying extents of LD among the population and MBB subpopulations. The LD level was elevated when the cotton population was grouped into subpopulations. Flint-Garcia et al. (2003) stated that population structure is one of the important factors generating LD within specific population groups. However, it can also cause spurious marker–trait associations (Pritchard et al. 2000). This approach entails a serious consideration of population structure, along with relatedness, when undertaking population-based association mapping studies in cotton genetic resources (Abdurakhmonov et al. 2009).

LD decay rates affected the resolution of QTL mapping from deciding the number of markers in the association analysis (Gaut and Long 2003). In general, the genetic distance at which r^2 decays to 0.1 and 0.2 is considered to be the extent of LD in a species (Zhu et al. 2008). In the present study, the decay rate was very fast (~1–2 cM)

Table 4 SSR marker loci significantly ($p < 0.01$) associated with fiber yield and quality traits of Upland cotton (*Gossypium hirsutum* L.) population

Traits	Marker loci	Chr no.	*F* value	*p* value	R^2	
FL	BNL2449	10, 13	9.36	0.00240	0.0357	Said et al. (2013, 2015 (10c, 42.14)
	HAU1185**	19	11.15	0.00095	0.0317	
	HAU2759	25	10.84	0.00110	0.0381	
	HAU423**	11	12.96	0.00037	0.0426	
	NAU749**	10	18.38	0.00002	0.0647	
	NAU874**	6	14.20	0.00020	0.0376	
	TMB436**	25	14.83	0.00014	0.0571	Fang et al. (2014)
Mic	DPL209	11	9.07	0.00280	0.0319	
	HAU2786	17	9.15	0.00270	0.0376	
	MUSS138	19	10.97	0.00100	0.0443	
	NAU1102**	19	11.80	0.00068	0.0462	Shen et al. (2006)
	NAU1167	3, 4	9.97	0.00180	0.0413	Hugie et al. (2016) and Cai et al. (2014)
	NAU3995	3	7.25	0.00750	0.0320	
	NAU749**	10	11.97	0.00062	0.0393	
	NAU874	6	9.40	0.00240	0.0290	
	TMB10	24	9.42	0.00230	0.0397	
	TMB436	25	10.39	0.00140	0.0304	Fang et al. (2014)
FS	BNL2449	10, 13	8.39	0.00410	0.0368	Said et al. (2013)
	HAU2056	1	11.04	0.00100	0.0311	
	HAU423	11	7.34	0.00710	0.0242	
	NAU1102	19	10.28	0.00150	0.0428	Hugie et al. (2016), Cai et al. (2014) and Shen et al. (2006)
	NAU1302	2, 24	10.52	0.00130	0.0449	Shen et al. (2005, 2006)
	NAU2508	10	8.16	0.00460	0.0368	Cai et al. (2014)
	NAU749	10	8.08	0.00480	0.0311	
	STV31	10	8.17	0.00460	0.0369	
	TMB436**	25	11.46	0.00081	0.0481	Fang et al. (2014)
FE	DPL513	1	7.83	0.00550	0.0255	
	HAU1185**	19	11.44	0.00082	0.0334	
	HAU2056	1	9.17	0.00270	0.0323	
	JESPR50	22	6.85	0.00930	0.0201	Shen et al. (2007)
	NAU1215	13, 18	8.02	0.00500	0.0246	Said et al. (2013, 2015)
	NAU1366	21	10.53	0.00130	0.0314	
	NAU2265	2	7.73	0.00580	0.0373	Said et al. (2013) and Fang et al. (2014)
	NAU749**	10	11.78	0.00068	0.0448	
UI	DPL715	11	8.39	0.00410	0.0268	Fang et al. (2014)
	HAU1185	19	10.72	0.00120	0.0216	
	HAU878	5	7.27	0.00740	0.0292	
	NAU1070	14	6.78	0.00970	0.0212	
	NAU749**	10	27.65	0.00000	0.0888	
UR	HAU1185	19	7.60	0.00620	0.0176	
	NAU1102	19	7.65	0.00600	0.0309	Shen et al. (2006)
	NAU749**	10	20.05	0.00001	0.0715	
	NAU874	6	7.89	0.00530	0.0178	
	TMB436**	25	15.39	0.00011	0.0556	Fang et al. (2014)
BW	BNL4030	22	6.80	0.00960	0.0183	Shen et al. (2007) and Said et al. (2013, 2015)
	CIR328	5	7.05	0.00840	0.0274	Said et al. (2013, 2015)
	HAU880	17	9.34	0.00240	0.0294	Hugie et al. (2016), Tan et al. (2015) and Wang et al. (2013)
	JESPR274**	9, 23	14.60	0.00016	0.0517	Fang et al. (2014)

Table 4 (continued)

Traits	Marker loci	Chr no.	*F* value	*p* value	R^2	
	NAU1190	3	7.01	0.00850	0.0231	Hugie et al. (2016), Kim et al. (2013), Wang et al. (2013) and Shen et al. (2007)
	NAU1302	2, 24	7.15	0.00790	0.0218	Shen et al. (2005, 2006, 2007)
LP	BNL1694	7	8.54	0.00370	0.0222	Said et al. (2015)
	DPL212**	19	12.69	0.00043	0.0323	
	DPL715**	11	14.86	0.00014	0.0376	Fang et al. (2014)
	HAU250	13	8.26	0.00440	0.0180	
	HAU2759	25	7.97	0.00510	0.0206	
	NAU3377	21	10.93	0.00110	0.0279	
	NAU749	10	8.68	0.00350	0.0205	

** Significant ($p < 0.001$) QTLs identified with Bonferroni correction

when r^2 was set at 0.2, indicating a high mapping resolution, possibly due to the fairly large and diverse cotton accessions and SSR markers used in the study. The result suggests that LD decay over short distances will facilitate association mapping for agronomic and fiber quality traits in cotton with a relatively large number of various types of markers (Abdurakhmonov et al. 2009; Zhang et al. 2013), while LD decay over longer distances will facilitate the initial association of trait data with haplotypes in chromosome regions (Thudi et al. 2014). According to Zhao et al. (2014), genome-wide LD in the elite Upland cotton germplasm decayed to 10–15 and 1–3 cM when the r^2 thresholds were set at 0.1 and 0.2, respectively.

Association mapping is a powerful method to detect QTLs underlying complex traits and is more efficient than linkage analysis because it exploits abundant genetic variations in diverse genetic backgrounds (Flint-Garcia et al. 2003; Ziegler et al. 2008). In this study, association mapping was used to analyze the genetic bases of fiber yield and quality traits. A total of 57 associations were detected using the means of fiber yield and quality traits across 2 years at the 12 different locations of China. Of these, only 24 associations were consistent with QTLs identified in previously studies (Table 4). The additional marker–trait associations that did not coincide with previously reported QTLs were novel QTLs/genes associated with fiber yield and quality traits. For fiber length, seven associated markers were identified, among which two markers, BNL2449 and TMB436, had been reported to be located on the chromosome 10 at the position 36.95 cM and chromosome 25 at 74.3 cM, respectively (Said et al. 2013, 2015; Fang et al. 2014). For fiber strength, NAU1102, NAU1302 and NAU2508 were determined to coincide with the previous QTLs, and they were located on the chromosome 19, chromosome 2 at 0 cM and chromosome 10 at 126.827 cM, respectively (Shen et al. 2005, 2006; Cai et al. 2014; Hugie et al. 2016). For fiber elongation, NAU1215 and NAU2265 were previously reported markers that were indentified using the linkage population, which mapped them on the chromosome 18 at the position of 49.42 cM and the chromosome 2 at 8.5 cM (Said et al. 2013, 2015). For lint percent, BNL1694 and DPL715 were previously indentified markers (Said et al. 2015; Fang et al. 2014). For boll weight, all the six markers have been reported in previous research (Shen et al. 2005, 2006, 2007; Wang et al. 2013; Kim et al. 2013; Said et al. 2013, 2015; Tan et al. 2015; Hugie et al. 2016). These stably inherited QTLs that were detected in different populations with different methods would be very useful for marker-assisted selection of the fiber-related traits in MAS breeding.

Some associated markers that were mentioned in previous reports were determined to be associated with more than one traits. In previous reports, NAU1102 was indentified to be linked or associated with fiber strength by Shen et al. (2006), Cai et al. (2014) and Hugie et al. (2016). However, the present study determined that it was associated with three fiber quality traits, fiber strength, fiber micronaire and fiber uniformity ratio. We found that TMB436 was associated with fiber length, fiber strength, fiber micronaire and fiber uniformity ratio, but this QTL namely qMIC-c25 was associated with fiber micronaire only in the report of Fang et al. (2014). NAU1167 was found to be associated with fiber length in the previous reports of Cai et al. (2014) and Hugie et al. (2016), while we found it to be associated with fiber micronaire. In the reports of Fang et al. (2014), NAU2265 and DPL715 were associated with fiber length and fiber elongation, respectively, the corresponding QTLs named as *qUHM-c2-2* at position 19.8 cM of chromosome 11 and *qELO-c11* at position 8.5 cM of chromosome 2. However, our results showed they were associated with fiber elongation and fiber uniformity index, respectively. JESPR50 linked the yield traits QTL named *qLP-D6-1* at position 146.96 cM of chromosome 22 reported by Shen et al. (2006), while present study

determined that it was associated with fiber elongation. The markers associated with more than one trait may be efficiently utilized to simultaneously improve different fiber yield and qualities through various breeding programs.

The heritability of fiber quality traits affected the efficiency of screening in the association mapping. The phenotypes of complex traits often result from the combined actions of multiple genes and environmental factors (Mackay et al. 2009). Only those traits with high heritability can be stably detected by genetic mapping study. This view was confirmed in the present study that all phenotypic traits had a very high heritability (>80%) across all environments, except for UI. The lowest number of QTLs detected in the UI might be due to the relatively large contribution of the environment to the phenotype ($G \times E$ interaction). The number of traits associated with each of the SSR markers varied in the present study. The genetic mechanisms of the association of markers with more than one trait could be due to pleiotropy of a single causal gene (Scholl and Miller 1976) or tight linkage of multiple causal genes (Meredith 1977), or both.

Population structure and familial relatedness usually affected the accuracy of association mapping, and led to false positive results if not controlled correctly in the statistical analysis (Pritchard et al. 2000; Yu et al. 2006). A number of reports have revealed the spurious effects of population structure and ancestry in association mapping of QTLs, causing the false positive-associated markers (Abdurakhmonov et al. 2009; Mei et al. 2013; Qin et al. 2015). To correct the effects of population structure and significantly reduce false positive associations (Type I error) in the marker–trait association analysis, the MLM (Yu et al. 2006) has been widely used (Abdurakhmonov et al. 2008; Mei et al. 2013; Qin et al. 2015). In the present study, the MLM analysis identified 58 SSR markers that were significantly associated with six of the fiber quality traits, which filtered amount of false positive markers caused by the family stratification. The results suggest that stably inherited QTLs/genes are present near these markers, which may be used to improve fiber quality traits with broad adaptability across different environments. Recent studies have demonstrated the feasibility of applying association mapping to explore complex traits in Upland cotton population and detected a number useful markers/QTL (Abdurakhmonov et al. 2008, 2009; Mei et al. 2013; Cai et al. 2014; Qin et al. 2015) for various breeding programs.

Using multiple testing methods will increase precision in QTLs screening. The major concerns in both linkage and association mapping studies are the statistical power and the control of false positive associations. Although the MLM provides a robust method to correct for relatedness in population mapping studies, attempts to map phenotypes that are strongly correlated with relatedness remain problematic (Myles et al. 2009). A genome-wide Type I error occurs if at least one false QTL is declared, thus multiple tests are required to control the Type I error rate (Bu et al. 2011). The use of stringent probability thresholds will reduce the rate of false positives. However, it seems difficult to ascertain which multiple testing method is acceptable to lessen an inflated Type I error in a given association study. When a stringent threshold with $p < 0.001$ was adopted, only sixteen marker–trait associations were determined to be significant in the present study, suggesting only a few markers could represent highly reliable associations. However, corrections for false positives using conservative multiple adjustment methodologies may also risk the rejection of true positives caused by setting the thresholds too high (Abdurakhmonov et al. 2009; Yan et al. 2011). This limitation suggests cautious application of multiple testing methods in association mapping study.

QTLs loci could be a better way to understand the molecular basis of complex traits related to fiber yield and fiber quality improvement in Upland cotton. Molecular markers that are tightly linked to QTLs/genes can be utilized to develop improved cultivars through various breeding programs. The markers associated with more than one trait may be efficiently utilized for simultaneous improvement of different fiber quality traits. The information generated from the current study is believed to provide the scientific grounds to identify, conserve, and efficiently utilize favorable alleles residing in the natural population to improve the genetic base of Upland cotton cultivars.

Acknowledgements This project was financially supported by the National Natural Science Foundation of China (31301365), the Crop Germplasm Conservation Program of Ministry of Agriculture China (2014NWB034) and the National Science and Technology Support Program of China (2013BAD01B03). We thank the national mid-term cotton gene bank of ICR-CAAS for providing the cotton genotypes.

Compliance with ethical standards

Funding This study was funded by the National Natural Science Foundation of China (31301365), the Crop Germplasm Conservation Program of Ministry of Agriculture China (2014NWB034) and the National Science and Technology Support Program of China (2013BAD01B03).

Conflict of interest All the authors declare that they have no any conflict of interest.

Ethical statement This article does not contain any studies with human participants or animals performed by any of the authors. There is additional documentation related to this study. You may login to the system and click the 'View Attachments' link in the Action column.

References

Abdurakhmonov I, Kohel R, Yu J, Pepper A, Abdullaev A, Kushanov F, Salakhutdinov I, Buriev Z, Saha S, Scheffler B, Jenkins J, Abdukarimov A (2008) Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics 92:478–487

Abdurakhmonov I, Saha S, Jenkins J, Buriev Z, Shermatov S, Scheffler B, Pepper A, Yu J, Kohel R, Abdukarimov A (2009) Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. Genetica 136:401–417

Abdurakhmonov IY, Buriev ZT, Shermatov SE, Abdullaev AA, Urmonov K, Kushanov F et al (2012) Genetic diversity in *Gossypium* genus. In: Caliskan M (ed) Genetic diversity in plants. InTech, pp 313–338. http://www.intechopen.com/books/genetic-diversity-inplants/genetic-diversity-in-gossypium-genus. Accessed 05 Apr 2013

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Brubaker CL, Paterson AH, Wendel JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184–203

Bu M, Stich B, Piepho HP (2011) A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants. Heredity 106:825–831

Burton GW (1951) Quantitative inheritance in pearl millet (*Pennisetum glaucum*). Agron J 43:409–417

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

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Culp TW, Lewis CF (1973) Breeding methods for improving yield and fiber quality of Upland cotton (*Gossypium hirsutum*). Crop Sci 13:686–689

Du XM, Zhou ZL, Jia YH, Liu GQ (2007) Collection and conservation of cotton germplasm in China. Cotton Sci 19:346–353

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

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Fang FD, Jenkins JN, Deng DD, McCarty JC, Li P, Wu JX (2014) Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in Upland cotton (*Gossypium hirsutum* L.). BMC Genom 15:397

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Flint-Garcia SA, Thuillet AC, Yu JM, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a higher solution platform for quantitative trait locus dissection. Plant J 44:1054–1064

Freeman GH (1973) Statistical methods for the analysis of genotype–environment interactions. Heredity 31:339–354

Fryxell PA (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2:108–165

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1503–1506

Gomez KA, Gomez AA (1984) Statistical procedures for agricultural research. Willey, New York

Gore MA, Fang DD, Poland JA, Zhang JF, Percy RG, Cantrell RG, Thyssen G, Lipka AE (2014) Linkage map construction and quantitative trait locus analysis of agronomic and fiber quality traits in cotton. Plant Genome. doi:10.3835/plantgenome2013.07.0023

Hall D, Tegstrom C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. Brief Funct Genom 9:157–165

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Hsu HH, Gale F (2001) Regional shifts in China's cotton production and use. Cotton and wool situation and outlook. Economic Research Service, USDA. http://www.ers.usda.gov/ContentPages/4093190.pdf. Accessed 15 Nov 2014

Hugie KL, Fang DD, Wayne Smith C, Li P, Hinze LL, Hague SS, Jones DC (2016) Utility assessment of published microsatellite markers for fiber length and bundle strength QTL in a cotton breeding program. Crop Sci 56:2983–2995

Iqbal J, Reddy OUK, El-Zik KM, Pepper AE (2001) A genetic bottleneck in the 'evolution under domestication' of Upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. Theor Appl Genet 103:547–554

Jannink JL, Bink MC, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci 6:337–342

Jia Y, Sun X, Sun J, Pan Z, Wang X et al (2014a) Association mapping for epistasis and environmental interaction of yield traits in 323 cotton cultivars under 9 different environments. PLoS One 9:e95882

Jia Y, Wang X, Sun J, Zhou Z, Pan Z, He S, Pan B, Wang L, Du X (2014b) Association mapping of resistance to Verticillium wilt in *Gossypium hirsutum* L. germplasm. Afr J Biotechnol 13:3165–3172

Jia YH, Sun JL, Wang XW et al (2015) Molecular diversity and association analysis of drought and salt tolerance in *G. hirsutum* L. germplasm. J Integr Agric 30:1845–1853

Jiang CX, Wright RJ, El-Zik KM, Paterson A (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proc Natl Acad Sci USA 95:4419–4424

Kantartzi SK, Stewart JM (2008) Association analysis of fiber traits in *Gossypium arboreum* accessions. Plant Breed 127:173–179

Kim HJ, Moon HS, Delhom CD, Zeng LH, Fang DD (2013) Molecular markers associated with the immature fiber (im) gene affecting the degree of cell wall thickening in cotton (*Gossypium hirsutum* L.). Theor Appl Genet 126:23–31. doi:10.1007/s00122-012-1956-x

Kohel RJ, Yu J, Park YH, Lazo GR (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121:163–172. doi:10.1023/A:1012263413418

Kwon SH, Torrie JH (1964) Heritability and interrelationship among traits of two soybean population. Crop Sci 4:194–198

Lacape JM, Nguyen TB, Thibivilliers S, Bojinnov TB, Courtois B, Cantrell RG, Burr B, Hau B (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* × *Gossypium barbadense* backcross population. Genome 46:612–626

Lacape JM, Nguyen TB, Courtois B, Belot JL, Giband M, Gourlot JP, Gawryziak G, Rouques S, Hau B (2005) QTL analysis of cotton-fiber quality using multiple *Gossypium hirsutum* × *Gossypium barbadense* backcross generations. Crop Sci 45:123–140

Lacape JM, Llewellyn D, Jacobs J, Arioli T, Becker D, Calhoun S et al (2010) Meta-analysis of cotton fiber

quality QTLs across diverse environments in a *Gossypium hirsutum* × *G. barbadense* RIL population. BMC Plant Biol 10:132. doi:10.1186/1471-2229-10-132

Ma XX, Ding YZ, Zhou BL, Guo WZ, Lv YH, Zhu XF, Zhang TZ (2008) QTL mapping in A-genome diploid Asiatic cotton and their congruence analysis with AD-genome tetraploid cotton in genus *Gossypium*. J Genet Genom 35:751–762. doi:10.1016/S1673-8527(08)60231-3

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a drum wheat elite collection. Mol Breed 15:271–289

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10:565–577

Malysheva-Otto LV, Ganal MW, Roder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). BMC Genet 7:6

McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J (2002) Linkage disequilibrium in domestic sheep. Genetics 160:1113–1122

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor Appl Genet 108:280–291. doi:10.1007/s00122-003-1433-7

Mei H, Zhu X, Zhang T (2013) Favorable QTL alleles for yield and its components identified by association mapping in Chinese Upland cotton cultivars. PLoS One 8:e82193

Meredith WR Jr (1977) Backcross breeding to increase fiber strength of cotton. Crop Sci 17:172–175

Myles S, Peiffer J, Brown P, Ersoz E, Zhang Z, Costich D, Buckler E (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202

Pan ZE, Sun JL, Wang XW, Jia YH, Zhou ZL, Pang BY, Du XM (2008) Screening of SSR core primers with polymorphism on a cotton panel (English abstract). Biodivers Sci 16:555–561

Park YH, Alabadt MS, Ulloa M, Sicker B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an inter-specific recombinant inbred line cotton population. Mol Genet Genom 274:428–441

Paterson AH, Saranga Y, Menz M, Jiang CX, Wright RJ (2003) QTL analysis of genotype × environment interaction affecting cotton fiber quality. Theor Appl Genet 106:384–396. doi:10.1007/s00122-002-1025-y

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pritchard JK, Wen X, Falush D (2010) Documentation for structure software version 2.3. University of Chicago. http://pritch.bsd.uchicago.edu/structure.html. Accessed 07 Sep 2014

Qin H, Chen M, Yi X, Bie S, Zhang C, Zhang Y, Lan Y, Meng Y, Yuan Y, Jiao C (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

Rungis D, Llewellyn D, Dennis ES, Lyon BR (2005) Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars. Aust J Agric Res 56:301–307

Saeed M, Guo W, Zhang T (2014) Association mapping for salinity tolerance in cotton (*Gossypium hirsutum* L.) germplasm from US and diverse regions of China. Aust J Crop Sci 8:338–346

Saha S, Raska DA, Stelly DM (2006) Upland cotton (*Gossypium hirsutum* L.) × Hawaiian cotton (G. *tomentosum* Nutt. Ex. Seem) F1 hybrid hypoaneuploid chromosome substitution series. J Cotton Sci 10:146–154

Said JI, Lin ZX, Zhang XL, Song MZ, Zhang JF (2013) A comprehensive meta QTL analysis for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease resistance in tetraploid cotton. BMC Genom 14:776

Said JI, Song MZ, Wang HT, Lin ZX, Zhang XL, Fang DD, Zhang JF (2015) A comparative meta-analysis of QTL between intraspecific *Gossypium hirsutum* and interspecific *G. hirsutum* × *G. barbadense* populations. Mol Genet Genom 290:1003–1025. doi:10.1007/s00438-014-0963-9

Scholl RL, Miller PA (1976) Genetic association between yield and fiber strength in Upland cotton. Crop Sci 16:780–783

Semagn K, Bjørnstad A, Xu Y (2010) The genetic dissection of quantitative traits in crops. Electron J Biotechol 13:5

Shappley ZW, Jenkins JN, Zhu J, McCarty JC (1998) Quantitative trait loci associated with agronomic and fiber traits of Upland cotton. J Cotton Sci 2:153–163

Shen XL, Guo WZ, Zhu XF, Yuan YL, Yu JZ, Kohel RJ, Zhang TZ (2005) Molecular mapping of QTLs for qualities in three diverse lines in Upland cotton using SSR markers. Mol Breed 15:169–181

Shen XL, Zhang TZ, Guo WZ, Zhu XF, Zhang XY (2006) Mapping fiber and yield QTLs with main, epistatic, and QTL × environment interaction effects in recombinant inbred lines of Upland cotton. Crop Sci 46:61–66. doi:10.2135/cropsci2005.0056

Shen XL, Guo WZ, Zhu XF, Yuan YL, Zhang TZ (2007) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155:371–380

Statistical Analysis System (2007) SAS version 9.2. SAS Institute Inc., Cary

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111:723–730

Tan ZY, Fang XM, Tang SY, Zhang JA, Liu DJ, Teng ZH, Li L, Ni HJ, Zheng FM, Liu DX, Zhang TF, Paterson AH, Zhang ZS (2015) Genetic map and QTL controlling fiber quality traits in Upland cotton (*Gossypium hirsutum* L.). Euphytica 203:615–628. doi:10.1007/s10681-014-1288-9

Thudi M, Upadhyaya HD, Rathore A, Gaur PM, Krishnamurthy L, Roorkiwal M, Nayak SN et al (2014) Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. PLoS One 9:e96758

Ulloa M, Meredith WR (2000) Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intra-specific population. J Cotton Sci 4:161–170

Ulloa M, Saha S, Jenkin N, Meredith WR, McCarty JC, Stelly DM (2005) Chromosomal assignment of RFLP linkage groups harboring important QTL on an intra-specific cotton (*Gossypium hirsutum* L.) joinmap. J Hered 96:132–144. doi:10.1093/jhered/esi02

Van-Esbroeck G, Bownam DT (1998) Cotton germplasm diversity and its importance to cultivar development. J Cotton Sci 2:121–129

Wang BH, Guo WZ, Zhu XF, Wu YT, Huang NT, Zhang TZ (2006) QTL mapping of fiber quality in an elite hybrid derived-RILs in Upland cotton. Euphytica 152:367–378

Wang FR, Xu ZZ, Sun R, Gong YC, Liu GD, Zhang JX, Wang LM, Zhang CY, Fan SJ, Zhang J (2013) Genetic dissection of the introgressive genomic components from *Gossypium barbadense* L. that contribute to improved fiber quality in *Gossypium hirsutum* L. Mol Breed 32:547–562. doi:10.1007/s11032-013-9888-y

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:140–186. doi:10.1016/S0065-2113(02)78004-8

Wu KM, Guo YY (2005) The evolution of cotton pest management practices in china. Annu Rev Entomol 50:31–52

Yan J, Warburton M, Crouch J (2011) Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. Crop Sci 51:433–449

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genet 38:203–208

Zeng LH, Meredith WR, Gutierrez QA, Boykin DL (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. doi:10.1007/s00122-009-1020-7

Zhang J, Stewart JM (2000) Economical and rapid method for extracting cotton genomic DNA. J Cotton Sci 4:193–201

Zhang TZ, Yuan TL, Yu J, Guo WZ, Kohel RJ (2003) Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. Theor Appl Genet 106:262–268

Zhang K, Zhang J, Ma J, Tang S, Liu D, Teng Z, Liu D, Zhang Z (2012) Genetic mapping and quantitative trait locus analysis of fiber quality traits using a three-parent composite population in Upland cotton (*Gossypium hirsutum* L.). Mol Breed 29:335–348

Zhang T, Qian N, Zhu X, Chen H, Wang S, Mei H, Zhang Y (2013) Variations and transmission of QTL alleles for yield and fiber qualities in Upland cotton cultivars developed in China. PLoS One 8:e57220

Zhao YL, Wang HM, Chen W, Li YH (2014) Genetic structure, linkage disequilibrium and association mapping of Verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. PLoS One 9:e86308

Zhu C, Gore M, Buckler E, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20

Ziegler A, Konig IR, Thompson JR (2008) Biostatistical aspects of genome-wide association studies. Biom J 50:8–28