doi: 10.17221/142/2016-CJGPB

Linkage Disequilibrium and Association Mapping of Fibre Quality Traits in Elite Asiatic Cotton (*Gossypium arboreum*) Germplasm Populations

KHUSHBOO SETHI[1], PRIYANKA SIWACH[1]* and SURENDER KUMAR VERMA[2]

[1]Department of Biotechnology, Chaudhary Devi Lal University, Sirsa, Haryana, India;
[2]Central Institute of Cotton Research, Regional Station, Sirsa, Haryana

*Corresponding author: Psiwach29@gmail.com

Abstract

Sethi K., Siwach P., Verma S.K. (2017): Linkage disequilibrium and association mapping of fibre quality traits in elite Asiatic cotton (*Gossypium arboreum*) germplasm populations. Czech J. Genet. Plant Breed., 53: 159–167.

Cotton productivity has been hindered by the narrow genetic base of cultivated cotton. Linkage disequilibrium-based association mapping has become a powerful molecular tool to dissect and exploit genetic diversity. In the present study, population structure and marker-trait associations for fibre quality traits in genotypes belonging to six races of *Gossypium arboreum* were assessed. Out of 300 simple sequence repeat (SSR) markers, 100 were found polymorphic, yielding a total of 240 alleles (all polymorphic). Structure analysis revealed allelic admixtures between genotypes. A Q-matrix exhibited mixed ancestry for the majority of genotypes, the race indicum forming a significant percent ancestry for almost all genotypes. At significant threshold values of $r^2 \geq 0.05$, 7.37% of SSR loci showed significant linkage disequilibrium (LD), while at highly significant threshold of $r^2 \geq 0.1$, the value was reduced to 5.31%. LD clearly decayed within the genetic distance of 9–10 cM, with $r^2 \geq 0.1$. Twenty-eight SSR markers were found associated with six fibre quality traits using general linear model and mixed linear model.

Keywords: general linear model; genetic distance; genetic diversity; mixed linear model; simple sequence repeat

Cotton (the genus *Gossypium*), one of the most important fibre crops in the world, has 45–50 species. Of these only four are cultivated (two allotetraploids – *G. hirsutum* L. and *G. barbadense* L. and two diploids – *G. herbaceum* L. and *G. arboreum* L.). At present, *G. hirsutum* is the most widely cultivated species worldwide (> 90%) because of suitability to industrial production and hence designated as primary cotton. Currently, there is an intense concern over the declines in yields and quality of primary cotton due to the narrow genetic base. A strong need is felt for enrichment of the gene pool with genetic diversity, for molecular understanding of desired traits and efficient transfer of allelic variation to breeding germplasm (ABDURAKHMONOV *et al.* 2008).

The *G. arboreum* (Asiatic cotton) germplasm collection is an important genetic resource for primary cotton improvement. It has inherent resistance to various biotic (AKHTAR *et al.* 2010) and abiotic stresses (TAHIR *et al.* 2011); tetraploids lack these qualities (KULKARNI 2002). Natural *G. arboreum* fibres display various colours (e.g. white, off-white and tan) and some of the accessions also produce fibres with unusually high strength (MEHETRE *et al.* 2003).

With the extensive development and easy availability of simple sequence repeats (SSRs) through cotton marker data base (CMD) (http://www.cottonmarker.org), successful efforts have been made in cotton for genetic diversity analysis and mapping and tagging of desired traits (ABDALLA *et al.* 2001; ABDURAKHMONOV *et al.* 2008). Linkage mapping has some major limitations including high running costs, poor resolution in detecting quantitative trait loci (QTL), consideration of only two alleles per locus and

doi: 10.17221/142/2016-CJGPB

there is an extreme disequilibrium between linked loci (MATHER *et al.* 2004; SALVI & TUBEROSA 2005).

Association mapping, based on linkage disequilibrium (LD), is emerging out as a better alternative. In cotton, association studies are limited; LD has been measured in a collection of tetraploid genotypes (WANG *et al.* 2013; CAI *et al.* 2014; SAEED *et al.* 2014; ZHAO *et al.* 2014; NEI *et al.* 2016), while a single report exists on the association analysis of fibre traits in *G. arboreum* accessions (KANTARTZI & STEWART 2008). In the present study, 95 genotypes representing six races of *G. arboreum* were explored for population structure, kinship, and genome-wide LD measurement using SSR markers. LD-based association mapping for fibre quality traits was applied using mixed linear model (MLM) as well as general linear model (GLM).

MATERIAL AND METHODS

Plant materials and DNA extraction. Ninety-five cotton genotypes (Table S1 in Electronic Supplementary Material (ESM)) were selected. The plants belonging to the races bengalense and cernuum were cultivated in two rows of 6 m length with 30 cm interplant distance in an experimental field. Plants of the remaining 4 races (being non-native to the experimental place) were cultivated in controlled conditions of a polyhouse of the Central Institute of Cotton Research (CICR), Sirsa, Haryana, India, in a completely randomized design (CRD) with 3 replications. Fresh leaves from a single plant, selected randomly from three replicates of each genotype, were used for DNA extraction using the CTAB method (SAGHAI-MAROOF *et al.* 1984) with certain modifications, followed by qualitative and quantitative analysis using gel electrophoresis and UV-spectrophotometer.

Phenotypic analysis. The trials were conducted in the autumn season (starting from early rainy season) from April to October for three consecutive years (2012, 2013 and 2014). The phenotypic traits (2.5% span length, fibre strength, fibre fineness, maturity coefficient, fibre elongation and uniformity ratio) were determined at CIRCOT (Central Institute for Research on Cotton Technology), C.I.C.R., Sirsa as per the calibration-cotton-standards (http://circot.res.in/circot/services/calibration-cotton-standards).

SSR genotyping. Three hundred microsatellite primer pairs based on even distribution across 13 chromosomes (PARK *et al.* 2005; GUO *et al.* 2007; MA *et al.* 2008; YU *et al.* 2012) were taken for initial screening. Of these, 100 primers were selected on the basis of polymorphic and reproducible band pattern (Table S2 in ESM; the sequence information available at http://www.cottonmarker.org). PCR amplification was performed in a volume of 20 µl containing 2 µl of DNA (50 ng/µl), 0.5 µM of primer (Sigma-Aldrich, India), 200 µM of dNTPs (Sigma-Aldrich), 0.5U Taq polymerase (Sigma-Aldrich) and 1X PCR buffer (Sigma-Aldrich). The PCR products were separated by electrophoresis at 100 V for 4 h in 4% metaphor gel and fragment sizes were calculated by interpolation from the migration distance of marker fragments of 100-bp DNA ladder (Thermo-Scientific; India).

Population structure analysis. Population structure analysis was carried out using the STRUCTURE 2.3.4 program (PRITCHARD *et al.* 2000). Evaluation of all possible K's (i.e. K = 2 to 8) was done using the simulation of 10 iterations, with each iteration consisting of 100 000 burn-in followed by 100 000 Markov Chain Monte Carlo (MCMC) replications, with default settings for both the Ancestry Model (Admixture Model) and the Frequency Model (allele frequencies correlated among populations; assumed different standardized variance in allele frequency (F_{ST}) values for subpopulations). The most likely number of clusters (K) was evaluated considering the plateau criterion proposed by PRITCHARD and FALUSH (2007) using the non-parametric Wilcoxon test (ROSENBERG *et al.* 2001) and the ΔK method (EVANNO *et al.* 2005). The CLUMPP 1.1.2 software (JACOBSON & ROSENBERG 2007) was used to find optimal alignments of independent runs and the output was used directly as an input into DISTRUCT 1.1, a program for cluster visualisation (ROSENBERG 2004).

Pairwise linkage disequilibrium and LD decay. TASSEL 4.0 was used to measure the extent of LD as squared allele frequency correlations estimates (r^2) (WEIR 1996) and to measure significance of r^2 for each pair of loci. Only alleles with frequencies equal or greater than 0.05 were considered for LD calculations (THORNSBERRY *et al.* 2001). Significance of LD for SSR pairs was determined by 100 000 permutations for each pair. The LD values between all pairs of SSR loci were plotted as triangle plots. LD decay (at $r^2 < 0.1$) was estimated by plotting r^2 values for pairs of SSR loci plotted as a function of map distances (cM) using GGT 2.0.

Analysis of marker-trait associations (MTAs). MTAs were conducted using both GLM and MLM using TASSEL 4.0 software (BRADBURY *et al.* 2007).

doi: 10.17221/142/2016-CJGPB

The magnitudes of marker-trait associations were evaluated by R^2 values for the markers at a significance level of $P \leq 0.05$, $P \leq 0.01$ and $P \leq 0.001$.

RESULTS AND DISCUSSION

Phenotypic variation. Least significant difference (LSD) (5%) and standard error (SE) were found maximum for fibre length and minimum for maturity coefficient (Table 1). The analysis of variance among genotypes showed significant differences at three *P* values for fibre strength, fibre elongation, 2.5% span length and fineness. Variation among years was found significant for the traits of 2.5% span length, fineness and uniformity, while interaction for genotype × year was obtained as non-significant for all the six traits. Accurate analysis of functional QTLs requires the phenotypic measurement in multiple locations; it could not be carried out during the present investigation due to different approval and administrative reasons.

SSR genotyping and inference of population structure. One hundred SSRs yielded a total of 240 polymorphic alleles. The mean number of alleles per locus was 2.4 while the number of alleles per locus varied from 2 to 6. Each chromosome was covered by 7–8 primer pairs.

Information on the memberships of individuals in specific clusters (Q-matrix) and the relatedness of individuals (K-matrix) are crucial when conducting LD based association mapping (Zhao *et al.* 2007). During the present study, as per Wilcoxon test, the most likelihood was obtained at K = 6 while with the delta K method, maximal ΔK (7.83) occurred at K = 3, with the next largest peak of ΔK occurring at K = 6 (Figure 1). Since we were interested in the structure composition of six races, we selected ΔK at K = 6 for further analysis.

Table 1. Evaluation of six phenotypic traits in the *Gossypium arboreum* genotypes

Trait	Year	Range	Mean	5% LSD	SE
Fiber length (mm)	2012	17.4–28.7	23.04	4.07	1.46
	2013	19.7–29.7	23.86		
	2014	19.7–29.7	24.02		
Fiber strength (g/tex)	2012	16.0–22.7	18.3	1.39	0.50
	2013	16.9–20.0	18.3		
	2014	16.9–22.0	18.4		
Fineness (Mic)	2012	5.0–7.2	6.14	0.73	0.26
	2013	5.2–7.1	6.0		
	2014	5.1–7.2	5.99		
Maturity coefficient	2012	0.72–0.90	0.84	0.052	0.019
	2013	0.79–0.91	0.85		
	2014	0.80–0.91	0.85		
Elongation (%)	2012	4.0–6.4	5.6	0.68	0.24
	2013	4.7–6.4	5.5		
	2014	4.4–6.7	5.5		
Uniformity ratio (%)	2012	46–54	50.8	3.78	1.35
	2013	47–54	49.8		
	2014	46–54	49.8		

LSD – least significant difference; SE – standard error

Individual membership and population membership coefficients at K = 6 were considered by means of CLUMPP software and were used to generate the structure plot using DISTRUCT (Figure 2). The race indicum almost formed a separate group from the other races (Figure 2), and genotypes from all races exhibited a comparatively high membership coefficient in indicum. It supports the most primi-

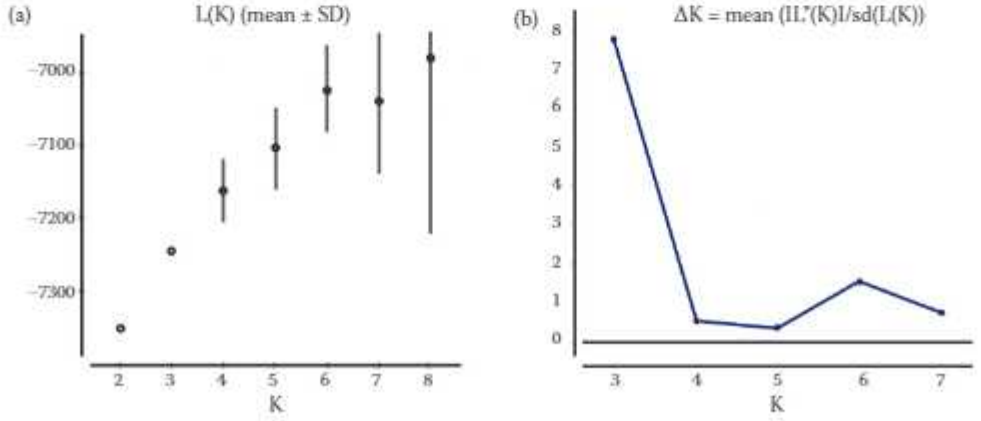

Figure 1. Structure analysis in *Gossypium arboreum* populations by Wilcoxon method (a) and ΔK method (b) (Evanno *et al.* 2005)

doi: 10.17221/142/2016-CJGPB

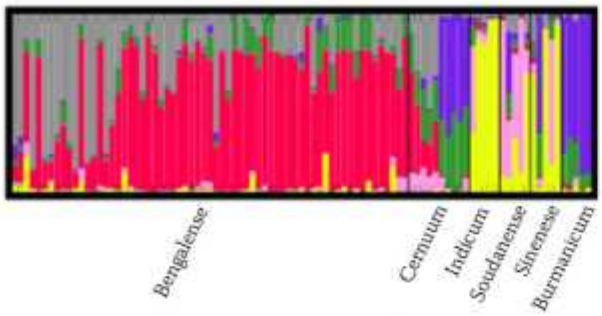

Figure 2. DISTRUCT plot of six *Gossypium arboreum* population groups

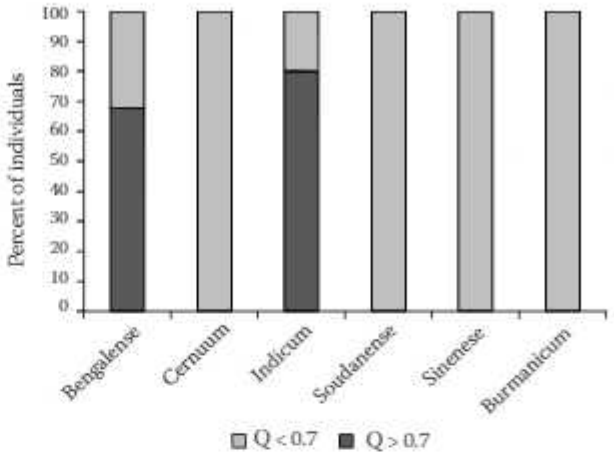

Figure 3. Assignment of *Gossypium arboreum* individuals on the basis of their membership coefficient (Q)

tive status of indicum by the evolutionary studies (Brubaker *et al.* 1999). With the arbitrary cut-off value of 70% ancestry for assignment of individuals to cluster (i.e. membership coefficient Q > 0.7), 44 plants (67.6%) of bengalense and 4 plants (80%) of indicum were attributed to their primary cluster. The remaining genotypes (32.3% of bengalense, 20% of indicum and 100% of the rest of races) exhibited mixed ancestry (Figure 3).

Pairwise linkage disequilibrium and LD decay. At $r^2 \geq 0.05$, 7.37% (365/4950) of SSR loci showed significant LD ($P < 0.05$) (Table 2). In some previous reports on *G. hirsutum*, at the same threshold, different values of SSR loci showing significant LD (ranging from 3% to 12%) have been reported (Abdurakhmonov *et al.* 2008, 2009; Saeed *et al.* 2014). The level of LD depends on the number of SSR loci and factors like mutation, selection, genetic drift. Significant LD at the intra-chromosomal level (Table 2) suggested the existence of LD generating factors other than linkage. Such an observation was also made by Abdurakhmonov *et al.* (2008).

LD blocks were observed as demonstrated by triangle plots for pairwise LD between SSRs (Figure 4). Sizes of intra-chromosomal LD blocks were also calculated; at $r^2 \geq 0.1$, the longest LD block (104.6 cM) was observed on chromosome 3 between markers BNL-3259 and NAU-1167 (Table 3).

Genome-wide LD decay was assessed by plotting r^2 values of LD as a function of genetic distance. LD clearly decayed within the genetic distance of 9–10 cM, with $r^2 \geq 0.1$ (Figure 5). Earlier studies reported genome-wide LD decay in cotton within the genetic distance of nearly 4–7 cM (Saeed *et al.* 2014), < 10 cM (Zhao *et al.* 2014). The extent of LD varies depending on the factors involved in a specific mode of breeding and selection pressure.

Marker-trait associations. Earlier studies on association mapping in cotton suggested the potential of conducting effective LD mapping with a lower number of markers than that needed in the majority of cases because of moderately significant LD blocks

Table 2. Evaluation of linkage disequilibrium (LD) between SSR loci at the whole genome level in 95 *Gossypium arboreum* accessions

Parameters	$r^2 \geq 0.1$	$r^2 \geq 0.05$
Sample size	95	95
Intrachromosomal LD% (LD/non LD)	4.59 (16/348)	7.47 (26/348)
Interchromosomal LD% (LD/non LD)	5.36 (247/4602)	7.36 (339/4602)
Total LD% (LD/non LD)	5.31 (263/4950)	7.37 (365/4950)

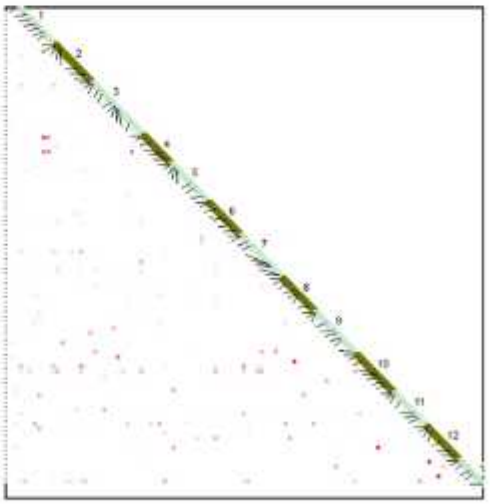

Figure 4. Triangle heat plot showing pairwise locus combinations in 13 chromosomes of *Gossypium arboreum*

doi: 10.17221/142/2016-CJGPB

Table 3. Size of intra-chromosomal linkage disequilibrium blocks in *Gossypium arboreum* loci (at significant threshold of $r^2 \geq 0.1$)

Locus 1	Locus 2	Distance (cM)	r^2	D'	Chromosome
NAU-2083	BNL-3090	32.10	0.11	0.63	1
MUSS-599	NAU-5383	2.40	0.11	0.63	2
BNL-3424	MUSS-73	0.0	0.65	0.73	2
BNL-3259	MUSS-207	0.0	0.15	0.63	3
BNL-3259	NAU-1167	104.6	0.10	0.47	3
NAU-2363	NAU-3093	43.3	0.68	0.73	4
MUSS-99	BNL-542	93.8	0.11	0.30	5
BNL-3359	NAU-2580	81.7	0.15	0.63	6
NAU-933	NAU-4030	13.0	0.24	0.97	7
BNL-1395	NAU-2432	24.6	0.15	0.63	7
BNL-1395	BNL-1531	24.4	0.15	0.63	7
BNL-1694	BNL-1531	20.6	0.11	0.63	7
BNL-1694	NAU-2308	30.1	0.13	0.93	7
BNL-1531	NAU-2432	0.20	0.24	0.65	7
NAU-3590	NAU-3793	15.8	0.15	0.63	8
NAU-3454	BNL-2530	3.0	0.49	0.98	10

D' – linkage disequilibrium coefficient

(Abdurakhmonov *et al.* 2008). Our selected germplasm collection of predefined groups represented by an unbalanced number of accessions suggested the probability of a significant influence of population structure and kinship. Different methods have been proposed to control this structured association (Pritchard *et al.* 2000), stepwise regression (Setakis *et al.* 2006) and mixed linear model (MLM) (Yu *et al.* 2006). We employed the MLM (K + Q) model to generate more accurate correlations with less inflated type-I errors. General linear model (GLM) (K only) was also employed for comparison. Results were compared with reports of Kantartzi and Stewart (2008), Abdurakhmonov *et al.* (2008), Fang *et al.* (2014), Wang *et al.* (2013) only; like in other reports different marker sets were used and for some of them we could not find any common SSRs in our results.

MLM analysis revealed 32 significant marker-trait associations (MTA) (24 markers in different association with six traits); of these only four (12.5%) were validated by earlier studies (Table 4). GLM revealed 48 associations (28 markers in different association with six traits); seven of them have already been published (Table 4).

Comparative analysis of MLM and GLM results revealed a significant reduction of MTAs in MLM, suggesting the effect of population structure in stratified populations on MTA. Similar observations were also made by Qin *et al.* (2015). In the present study, except 3 associations (BNL-1440 with micronaire, BNL 3992 with micronaire and NAU-2363 with uniformity ratio), the remaining 29 associations of MLM were also detected by GLM.

Considering GLM and MLM together, a total of 51 significant MTAs (involving 28 markers – Figure 6) were observed, with 56.8% as common associations; R^2 values for these associations ranged from 3.6% to 25.6% (Table 4). Interestingly, some markers associated with the same traits were located on the same chromosomes, e.g. BNL-3259, MUSS-192 and NAU 1167 present on chromosome 3 (though NAU 1167 is distantly located to the other two) were found associated with elongation; BNL-4047, BNL-530, BNL-4049 present on chromosome 4 and NAU-2580, NAU-3206, NAU-3427 present on chromosome 6 were associated with maturity coefficient; NAU-5383, BNL-3971, BNL-1897 present on chromosome 2 were associated with micronaire; NAU-2363, BNL-4047, BNL-4049 present on chromosome 4 were associated with span length; NAU-2000, BNL-3241, BNL-3992 present on chromosome 5 were associated with fibre strength.

The majority of the SSR markers associated with fibre quality traits in our diverse cotton germplasm were new markers. This could probably be attributed to the mapping of potentially new loci contributing to fibre quality traits because of the inclusion of diverse germplasm for the study. These new markers should be useful for

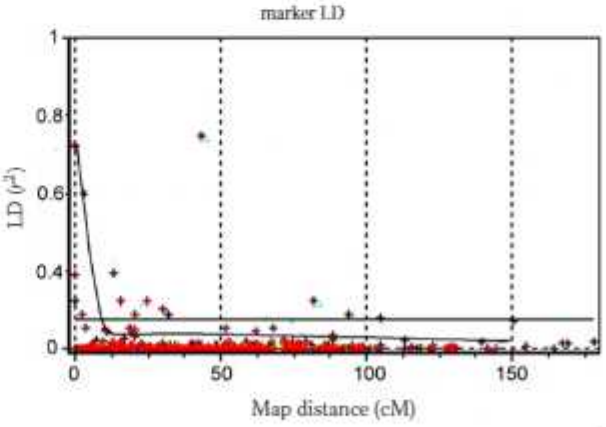

Figure 5. Scatterplot of the linkage disequilibrium (LD) (r^2) against genetic distance (cM) between all locus pairs on the whole genome of *Gossypium arboreum*
Horizontal straight line indicates the 95th percentile of the distribution of unlinked r^2 values

doi: 10.17221/142/2016-CJGPB

Table 4. Association of SSR markers with fibre quality traits in *Gossypium arboreum*

Trait	Marker	GLM results R^2 value (%)	MLM results R^2 value (%)	Chromosome location	Reference
Elongation	NAU-5499	4.49*	–	2	Wang *et al.* (2013)
	NAU-3427	8.92**	8.23**	6	
	NAU-3793	10.5***	9.88**	8	
	BNL-1066	4.85*	–	11	
	NAU-2038	5.15*	4.52**	13	
	NAU-1067	9.66**	7.19**	1	
	BNL-3259	3.79*	–	3	
	BNL1434	6.16**	3.97*	2	Kantartzi and Stewart (2008); Fang *et al.* (2014)
	MUSS-192	10.39***	9.50**	3	
	NAU-1167	25.60***	24.1***	3	
Maturity coefficient	BNL-4047	5.8**	–	4	Kantartzi and Stewart (2008)
	BNL-530	16.7***	14.5***	4	
	BNL-4049	4.03*	–	4	
	NAU-2580	16.7***	15.8***	6	
	NAU-3206	5.72*	5.47*	6	
	NAU-3427	5.54*	–	6	
	NAU-2432	13.1***	6.65**	7	
	BNL-1404	25.6***	24.1***	11	
	NAU-2038	6.1**	–	13	
	BNL-686	5.9**	5.67**	9	
	NAU-5383	4.39*	4.10*	2	
	BNL-1434	3.67*	3.78*	2	
Micronaire	BNL-3259	6.21**	5.95**	3	Kantartzi and Stewart (2008)
	NAU-1167	3.72*	3.82*	3	
	NAU-2000	5.43**	5.13*	5	
	BNL-1440	–	3.82*	6	
	BNL-3992	–	4.78*	5	
	NAU-1067	4.17*	–	1	
	BNL-3580	5.19**	–	1	
	NAU-5383	9.78***	6.73**	2	
	BNL-3971	5.0**	3.6*	2	
	BNL-1897	5.55**	–	2	
	BNL-530	5.99**	4.18*	4	
Span length	BNL-3259	7.92**	7.45**	3	Abdurakhomonov *et al.* (2008)
	NAU-2363	4.14*	–	4	
	BNL-4047	7.03**	5.32*	4	
	NAU-2000	9.58***	7.35**	5	
	BNL-3992	4.39*	–	5	
	BNL-686	3.70*	–	9	
	NAU-2038	3.64*	–	13	
	NAU-5499	3.74*	3.78*	2	
	BNL-4049	4.90*	4.90*	4	
Strength	NAU-2000	7.83**	7.67**	5	Abdurakhomonov *et al.* (2008)
	NAU-3426	3.98*	3.93*	12	
	BNL-686	4.80*	–	9	
	BNL-3241	7.09**	–	5	
	BNL-3992	4.76*	–	5	
Uniformity ratio	BNL-3241	6.14**	–	5	Abdurakhomonov *et al.* (2008)
	BNL-1066	3.68*	–	11	
	BNL-686	5.05*	4.34*	9	
	NAU-2363	–	3.79*	4	

*$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$; GLM – general linear model; MLM – mixed linear model

doi: 10.17221/142/2016-CJGPB

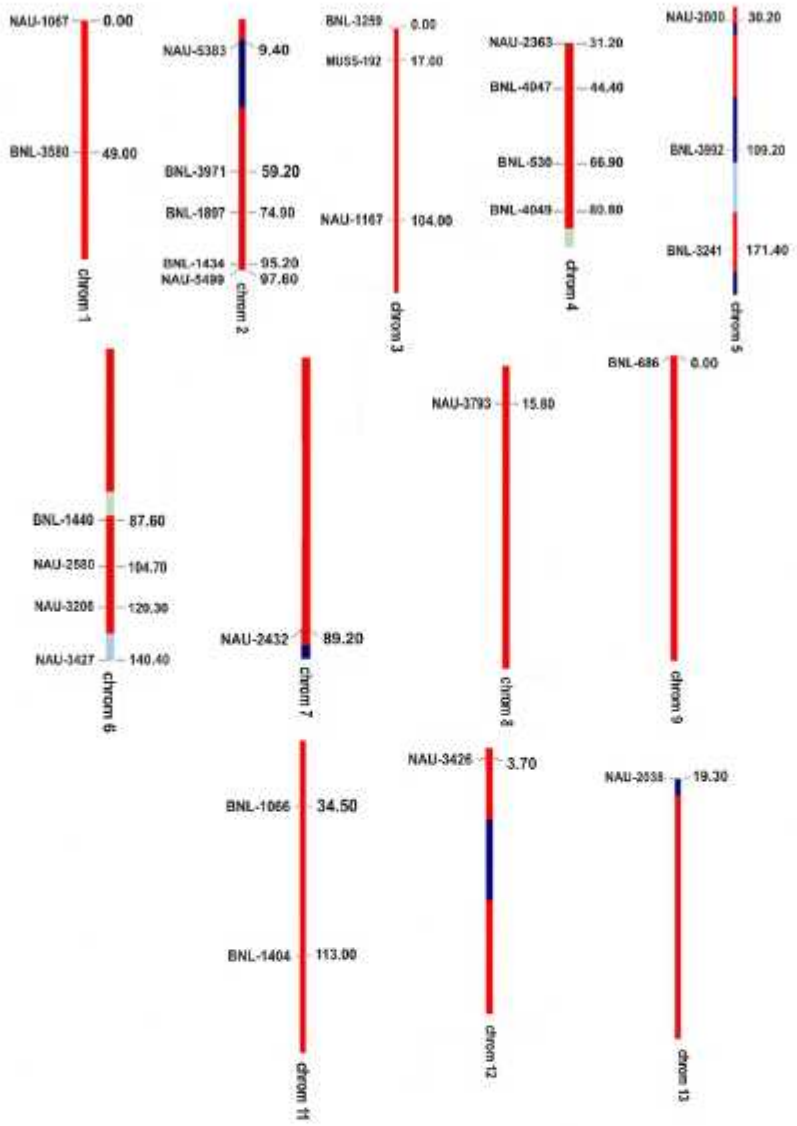

Figure 6. Positions of 28 markers, on chromosomes, involved in 51 marker-trait association as revealed by general linear model (GLM) and mixed linear model (MLM) analysis in *Gossypium arboreum*

future marker-assisted selection (MAS) programs to mobilize potentially new QTLs from underutilized diverse germplasm resources to elite cultivars.

Acknowledgements. Financial support provided by the University Grants Commission (UGC), Government of India, is acknowledged. Acknowledgement is extended to Chaudhary Devi Lal University, Sirsa, for laboratory and infrastructural facilities and to Director, CICR, Sirsa and Director, CICR, Nagpur for providing the germplasm and experimental field facility.

References

Abdalla A.M., Reddy O.U.K., El-Zik K.M., Pepper A.E. (2001): Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theoretical Applied Genetics, 102: 222–229.

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

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

Akhtar K.P., Haidar S., Khan M.K.R., Ahmad M., Sarwar N. (2010): Evaluation of *Gossypium* species for resistance to

doi: 10.17221/142/2016-CJGPB

cotton leaf curl Burewela virus. Annals of Applied Biology, 157: 135–147.

Bradbury P.J., Zhang Z.W., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S. (2007): TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23: 2633–2635.

Brubaker C.L., Bourland F.M., Wendel J.F. (1999): The origin and domestication of cotton. In: Smith C.W., Cothren J.T. (eds): Cotton – Origin, History, Technology and Production. New York, John Wiley: 3–31.

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

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

Fang D.D., Jenkins J.N., Deng D.D., McCarty J.C., Li P., Wu J.X. (2014): Quantitative trait loci analysis of fiber quality traits using a random-mated recombinant inbred population in upland cotton (*Gossypium hirsutum* L.). BMC Genomics, 15: 397.

Guo W., Cai C., Wang W., Han Z., Song X., Wang K., Niu X., Wang C., Lu K., Shi B., Zhang T. (2007): A microsatellite based, gene rich linkage map reveals genome structure, function and evolution in *Gossypium*. Genetics, 176: 527–541.

Jakobsson M., Rosenberg N.A. (2007): CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23: 1801–1806.

Kantartzi S.K., Stewart J. McD. (2008): Association analysis of fibre traits in *Gossypium arboreum* accessions. Plant Breeding, 127: 173–179.

Kulkarni V.N. (2002): Hirsutisation of *G. arboreum* cotton and genetic emendation of *G. hirsutum* for sucking pest resistance. [Ph.D. Thesis.] Dharwad, University of Agricultural Sciences.

Ma X., Ding Y., Zhou B., Guo W., Lu Y., Zhu X., Zhang T. (2008): QTL Mapping in A-genome diploid Asiatic cotton and their congruence analysis with AD genome tetraploid cotton in genus *Gossypium*. Journal of Genetics and Genomics, 35: 751–762.

Mather D.E., Hyes P.M., Chalmers K.J., Eglinton J., Matus I., Richardson K., Zitzewitz J.V., Cedillo L.M., Hearnden P., Pal N. (2004): Use of SSR marker data to study linkage disequilibrium and population structure. In: Proc. Int. Barley Genetics Symposium, *Hordeum vulgare*: Prospects for Association Mapping in Barley, Brno, June 20–26, 2004: 302–307.

Mehetre S.S., Aher A.R., Gawande V.I.., Patil V.R., Mokate A.S. (2003): Induced polyploidy in *Gossypium*: a tool to overcome interspecific compatibility of cultivated tetraploid and diploid cottons. Current Science, 84: 1510–1512.

Nei X., Huang C., You C., Li W., Zhao W., Shen C., Zhang B., Wang H., Yan Z., Dai B., Wang M., Zhang X., Lin Z.. (2016): Genome-wide SSR based association mapping for fiber quality in nation-wide upland cotton inbreed cultivars in China. BMC Genomics, 17: 352.

Park Y.H., Alabady M.S., Ulloa M., Sickler B., Wilkins T.A., Yu J., Stelly D.M., Kohel R.J., Osama M., Shihy E., Cantrell R.G. (2005): Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Molecular Genetics and Genomics, 274: 428–441.

Pritchard J.K., Falush D. (2007): Documentation for Structure Software: Version 2.2. Chicago, University of Chicago.

Pritchard J.K., Stephens M., Rosenberg N.A., Donnelly P. (2000): Association mapping in structured populations. American Journal of Human Genetics, 67: 170–181.

Qin H., Chen M., Yi X., Bie S., Zhang C., Zhang Y., Lan J., 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.

Rosenberg N.A. (2004): DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes, 4: 137–138.

Rosenberg N.A., Burke T., Elo K., Feldman M.W., Freidlin P.J., Groenen M.A., Hillel J., Mäki-Tanila A., Tixier-Boichard M., Vignal A., Wimmers K., Weigend S. (2001): Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics, 159: 699–713.

Saeed M., Wangzhen G., Tianzhen Z. (2014): Association mapping for salinity tolerance in cotton (*Gossypium hirsutum* L.) germplasm from US and diverse region of China. Australian Journal of Crop Science, 8: 338–346.

Saghai-Maroof M.A., Soliman K.M., Jorgensen R.A., Allard R.W. (1984): Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proceedings of National Academy of Sciences, 81: 8014–8018.

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

Setakis E., Stirnadel H., Balding D.J. (2006): Logistic regression protects against population structure in genetic association studies. Genome Research, 16: 290–296.

Tahir M.S., Khan N.U.I., Sajid U. R. (2011): Development of an interspecific hybrid (triploid) by crossing *Gossypium hirsutum* and *G. arboreum*. Cytologia, 76: 193–199.

doi: 10.17221/142/2016-CJGPB

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., Buckler E.S. (2001): Dwarf polymorphisms associate with variation in flowering time. Nature Genetics, 28: 286–289.

Wang X.Q., Yu Y., Li W., Guo H.I.., Lin Z.X., Zhang X.L. (2013): Association analysis of yield and fiber quality traits in *Gossypium barbadense* with SSRs and SRAPs. Genetics and Molecular Research, 12: 3353–3362.

Weir B.S. (1996): Genetic Data Analysis II. 2nd Ed. Sunderland, Sinauer Associates, Inc.

Yu J.M., Pressoir G., Briggs W.H., Bi I.V., Yamasaki M., Doebley J.F., McMullen M.D, Gaut B.S., Nielsen D.M., Holland J.B., Kresovich S., Buckler E.S. (2006): A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38: 203–208.

Yu J.Z., Kohel R.J., Fang D.D., Cho J., Deynze A.V., Ulloa M., Hoffman S.M., Pepper A.E., Stelly D.M., Jenkins J.N., Saha S., Kampatla S.P., Shah M.R., Hugie W.V., Percy R.G. (2012): A high density simple sequence repeat and single nucleotide polymorphism genetic map of the tetraploid cotton genome. Genes Genomes Genetics, 2: 43–58.

Zhao H.H., Fernando R.L., Dekkers J.C. (2007): Power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci. Genetics, 175: 1975–1986.

Zhao Y., Wang H., Chen W., Li Y. (2014): Genetic structure, linkage disequilibrium and association mapping of verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L..) germplasm population. PLoS ONE, 9: e86308.

Received for publication August 27, 2016
Accepted after corrections June 26, 2017
Published online August 9, 2017