



**Full Length Article**

## Association Mapping of QTLs with Yield and Lint Traits in Elite Upland Cotton Germplasm

Imtiaz Ali<sup>1</sup>, Naqib Ullah Khan<sup>1\*</sup>, Samrin Gul<sup>1</sup>, Shahid Ullah Khan<sup>2</sup>, Iqra Tahir<sup>1</sup>, Zarina Bibi<sup>3</sup>, Iftikhar Hussain Khalil<sup>1</sup>, Mehboob-Ur-Rahman<sup>4</sup>, Sheraz Ahmed<sup>1</sup>, Kashif Aslam<sup>5</sup>, Ghulam Shabir<sup>5</sup>, Hafiz Abdul Haq<sup>6</sup>, Ajmalud Din<sup>1</sup> and Shah Fahad<sup>7</sup>

<sup>1</sup>Department of Plant Breeding and Genetics, The University of Agriculture, Peshawar, Pakistan

<sup>2</sup>Institute of Biotechnology and Genetic Engineering, The University of Agriculture, Peshawar, Pakistan

<sup>3</sup>Department of Soil Science, Faculty of Agriculture, Gomal University, Dera Ismail Khan, Pakistan

<sup>4</sup>National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

<sup>5</sup>Institute of Molecular Biology and Biotechnology, Bahauddin Zakarya University, Multan, Pakistan

<sup>6</sup>PCCC Central Cotton Research Institute, Multan, Pakistan

<sup>7</sup>Department of Agricultural Chemistry, The University of Agriculture, Peshawar, Pakistan

\*For correspondence: nukmarwat@yahoo.com; nukmarwat@aup.edu.pk

Received 21 December 2019; Accepted 06 February 2020; Published 20 April 2020

### Abstract

Field testing across different environments can help to identify the cotton genotypes with relatively stable performance. To ascertain quantitative trait locus (QTL) associations with yield and lint traits, 28 upland cotton (*Gossypium hirsutum* L.) genotypes were field tested during two consecutive years at three locations of Pakistan i.e., Peshawar, Dera Ismail Khan and Faisalabad. Combined analysis of variance for genotypes across years and locations ( $G \times Y \times L$ ) interactions showed substantial variances among the genotypes for all the variables. Molecular characterization of the cotton germplasm was conducted at the National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad. Hundred simple sequence repeat (SSR) markers were used to characterize these cotton genotypes wherein 22 markers were polymorphic, which justified the further analysis. Population structure analysis divided all the cotton genotypes into two main groups at delta ( $K = 2$ ). In the first group, three cotton genotypes i.e., NIBGE-115, NN-3 and NIBGE-2472 exhibited some relationship with the genotypes of the second group. Based on association mapping, a total of 23 QTLs were identified in which 22 QTLs associated with yield and lint traits were identified through general linear model (GLM) approach while one QTL was identified through mixed linear model (MLM) approach for bolls per plant. The identified QTLs would be effective in considering the genetic basis of the aforementioned traits and maintaining diversity in the upland cotton germplasm. The favorable QTLs might also facilitate the breeders in genomic selection in upland cotton for efficient breeding. © 2020 Friends Science Publishers

**Keywords:** QTL mapping in cotton; Linkage disequilibrium; Population structure; SSR markers; Cotton

### Introduction

Cotton is the most significant cash crop grown around the world providing innate fiber for the textile trade. There are four species of cotton extensively cultivated around the world. It is very important to improve their genetic baseline especially *G. hirsutum* L. which has 95% share in the world cotton production. Improving its yield potential and resistance to different diseases are the prime objectives of cotton breeding. Though, the synchronous improvement in production and disease resistance is quite challenging for the breeders due to their negative association (Zhang *et al.* 2009). However, various methods have been introduced to

increase seed cotton yield and improve fiber quality traits. Recent advances in molecular and biometrical genetics have made it easy to associate the quantitative trait loci (QTLs) for different parameters like disease resistance, yield, and fiber quality, thus simplifying the use of marker-based selection for genetic enhancement. As a result, several QTLs for fiber quality, disease resistance and yield attributing traits have been recognized (Wang *et al.* 2007; Qin *et al.* 2008; Ali and Awan 2009; Shakoor *et al.* 2010). In these experiments, the QTLs mapping were carried out in tetraploid cotton populations resulting from crosses in which two parents were used (He *et al.* 2005; Yu *et al.* 2013).

It is known that due to inadequate events in

recombination, detection of strictly associated markers for marker-based selection is hard in biparental segregating populations (Fu *et al.* 2017). Generally, the occurrence of polymorphic loci in the segregating populations is low; hence some insignificant QTLs cannot be observed. Inherent genetic diversity within crop germplasm can be exploited through linkage disequilibrium (LD) which is another alternative powerful molecular tool based on QTL association mapping (Zhang *et al.* 2013). In conventional QTL analysis based on biparental populations, the association mapping LD can be used for mapping QTLs in genotypes with broader genetic background (Zhu *et al.* 2008). Thus, due to past recombination the higher number of events can be explored in natural population resulting in a higher resolution of QTL mapping as compared to biparental segregating populations (Ersoz *et al.* 2007). Association mapping based on LD can be employed to recognize polymorphism inside a gene which is accountable for the observable dissimilarities in phenotypes (Yan *et al.* 2010).

The starting idea for the studies on the alternative approach in QTL mapping is based on the non-random association of alleles at different loci, specifically between a marker locus and a phenotypic trait locus (Flint-Garcia *et al.* 2003). There are several factors that could result in LD including unknown population structure, mutation, genetic drifts, genetic bottlenecks, selection and level of inbreeding. Therefore, characterization of population with LD patterns is an important prerequisite to efficiently apply QTL association mapping in crop plants. To lessen the chance of false associations, it is vital to differentiate between physical LD and some other diverse forces that cause LD in broad based genetic populations.

*Gossypium hirsutum* L. a tetraploid species having great economic importance for its high yield and better adaptability and has attracted the interest of many cotton breeders around the world. Annually about 95% of the cotton production is obtained from *G. hirsutum* in the world and more than 150 countries are involved in its trade, generating an economic activity worth ~ \$500 billion per year worldwide (Chen *et al.* 2007; Zhang *et al.* 2012). However, research on LD in *G. hirsutum* and population structure is very restricted due to the complexity of genome structure and lack of best molecular markers. Thus, in this regard cotton is lagging far behind as compare to the several different species. However, some other studies have explored the extent of LD among genetic markers in different cotton populations (Abdurakhmonov *et al.* 2008). In a study of different agronomic traits, the fast LD decay of cotton genotypes showed the substantial potential for LD-based association mapping. Elite germplasm of *G. hirsutum* serves as an important source in cotton breeding, possessing several desirable traits i.e., early maturity, pests resistance, high yield, and good fiber quality. Therefore, classification of the population structure into different groups based on the ancestry information and LD levels in best cotton genotypes could be helpful for

association mapping of the important traits.

For verticillium wilt resistance, 60 QTLs have been described on 10 different chromosomes in cotton genotypes (Bolek *et al.* 2005; Wang *et al.* 2008; Yang *et al.* 2008). In this study four biparental populations of cotton genotypes were used for mapping QTLs, and markers linked to the identified QTLs cannot be directly used in cotton breeding programs. Before extensive application of QTL-linked markers in MAS, the QTL effects needed to be tested in other genetic backgrounds. In contrast to this background, great potential for QTL detection has been identified in association mapping using simple genotypes from the broad based genetic background. For the higher resolution of mapping, association mapping can discover the greater number of historical recombination events in genetically broad-based population than two parent segregating populations. Therefore, cotton breeders have drawn their attention to association mapping for efficient and rapid improvement of yield and fiber traits in different cotton cultivars (Abdurakhmonov *et al.* 2008; 2009).

Association mapping is substitute to traditional QTL mapping because it can efficiently identify QTLs even in natural population of diverse germplasm. Generalized linear model (GLM) can formulate the association between marker-traits using software package TASSEL 2.1 because most of the genotypes in the cotton populations have very weak relationship with each other (Yu *et al.* 2006). In GLM to avoid false association, association is estimated by using the percentages of admixture of each genotype (Q matrix) as covariates to include population structure in interpretation. The significance of marker-trait associations is declared by probability, and magnitude of QTL effects is calculated by phenotype variance of marker ( $R^2$ ). Association mapping based on LD detects markers with significant allelic variation among individuals exhibiting the trait to be mapped along with unrelated individuals in a natural population (Ochieng *et al.* 2007). Rafalski (2002) identified that the confirmation of mapping through association depends on LD level, and under rapidly decaying LD, mapping quality of alleles will be higher and vice versa. Keeping in view the economic position of cotton in Pakistan, the present study was arranged with the objectives to: a) study the genetic grouping in elite upland cotton germplasm, and b) identify the QTLs associated traits of seed cotton yield and lint traits in upland cotton.

## Materials and Methods

### Breeding material, experimental sites and procedure

This study on QTLs association with various yield and lint traits in upland cotton was carried out through molecular assays at NIBGE, Faisalabad, Pakistan. However, the field testing was conducted for two consecutive years during 2012 and 2013 at three diverse locations i.e., i) The University of Agriculture, Peshawar, ii) Cotton Research Station, Dera

Ismail Khan, and iii) NIBGE, Faisalabad, Pakistan. Locations in years were considered as separate environment and the total environments were six ( $3 \times 2$ ). Soil analysis of these three locations revealed that soil was clayey loam at two locations i.e., Faisalabad and D.I. Khan while silty clay loam at Peshawar, Pakistan (Table 1). Germplasm comprising 28 upland cotton genotypes was sown in the mid of May during both years under six different environments (Table 2). The trials at all the sites were carried out in randomized complete block design with three replications. Sub-plots made for each genotype were having four rows, five meters in length, with plant and row spacing of 30 and 75 cm, respectively. All the inputs and cultural practices were applied as per recommended package for cotton production to minimize field environmental variations. Picking was done during the month of November at each environment on single plant basis.

### Crop husbandry

Cotton is a deep-rooted crop which needs fine good tilth and well-prepared soil for successful germination and growth of crop. To get this, field was ploughed with deep plough then harrowed with planking each time to make the soil loose, fine, leveled and pulverized. The stubbles of the previous crop left in the field were also removed. All the fertilizers were applied at the rate of 100:60:60 kg ha<sup>-1</sup> of NPK, respectively. All the P<sub>2</sub>O<sub>5</sub>, K<sub>2</sub>O and 1/3 of N were used at the time of sowing while remaining N was applied in two split doses i.e., with first irrigation and at the pre-flowering stage. However, the doses of N and P were increased or decreased keeping in view the fertility of soil at different locations. Overall, 5–6 irrigations (from June to September) were given to the crops at all the locations. The weeds at all the locations were removed and controlled manually. For the control of sucking pests i.e., whitefly (*Bemisia tabaci*), jassids (*Amrasca biguttula* devastans) and thrips (*Thrips tabaci*), the insecticides viz., Confidor 200 SL (625 mL ha<sup>-1</sup>) and Baythroid TM 525 EC (1250 mL ha<sup>-1</sup>) were used in the experiments at all the locations during both years. In chewing insects, the American bollworm (*Helicoverpa armigera*) was more prominent at all the locations and was controlled by the insecticides i.e., Larvin 80 DF (1125 g ha<sup>-1</sup>) and Deltaphos 36 EC (1500 mL ha<sup>-1</sup>). Picking was done during the month of November on the single plant basis at all the locations during both years.

### Data collection

For data recording on various variables, the ten plants from central two rows were randomly selected in each sub-plot/replication. Effective, mature and open bolls from all the picks were counted and recorded as bolls per plant for each genotype. For boll weight, the seed cotton picked from each plant was divided by the number of effective and open bolls per plant. In each genotype, the seeds were counted in

10 bolls, and then averaged. For seed index, the hundred clean cotton seeds were weighed after ginning. Lint index for each genotype was calculated by applying the following formula.

$$\text{Lint index} = \frac{\text{Seed index} \times \text{lint \%}}{100 - \text{lint \%}}$$

From ten randomly selected plants, the dry and clean seed cotton was picked and weighed. The ginning was made separately with 8-saw gin. The lint obtained from each genotype was weighed and lint percentage was calculated by the following formula.

$$\text{Lint \% (G.O.T.)} = \frac{\text{Weight of lint in a sample}}{\text{Weight of seed cotton sample}} \times 100$$

### DNA extraction

The delinted seeds of all the 28 upland cotton genotypes were sown in disposable glasses filled with sand in glasshouse at NIBGE, Faisalabad, Pakistan. After germination and when the plants reached at 4–5 leaves stage, young leaves from each genotype were excised and stored in freezer for DNA extraction.

Using CTAB method, the DNA was extracted from 2–3 days old seedlings leaves (Iqbal *et al.* 1997). Water bath was set at 65°C to heat 2 x CTAB with 1% 2-mercapthanol. Pestle and mortar were autoclaved first and then pre-cooled with liquid nitrogen. Four to five stored young leaves were grinded to a very fine powder in CTAB solution or in liquid nitrogen. This grinded material was then shifted to a 15 mL falcon tube. Fifteen mL of hot (65°C) 2 x CTAB was added to the grinded material in tube, mixed carefully and incubate at 65°C for half an hour. After half an h, 15 mL of chloroform / isoamylalcohol (24:1) was also added to form an emulsion. Mixture was centrifuged for 10 min at 9000 rpm. Supernatant solution was shifted to a new 15 mL falcon tube, whereas, the remaining chloroform phase was discarded. This step was repeated twice as to ensure the complete digestion of various cell components and phenolic compounds. To precipitate the DNA, 0.6 volumes of chilled 2-propanol was added to the supernatant and then centrifuged at 9000 rpm for five min. The supernatant was discarded. The pellet was washed thrice with 70% ethanol and air-dried. The pellets were re-suspended in 0.5 mL 0.1 x TE buffer. The suspension was transferred into an eppendorf tube (1.5 mL) and then 5 µL of RNAs was added to digest all the RNAs incubating for one h at 37°C. After this, equal volume of chloroform / isoamylalcohol (24:1) was added and mixed gently and, centrifuged for 10 min at 13000 rpm in a microcentrifuge. The supernatant was transferred to a new eppendorf tube and 1/10<sup>th</sup> volume of 3 M NaCl solution was added to supernatant and mixed gently. DNA was precipitated with chilled absolute ethanol (2 volumes), spinned at 13000 rpm for 10 min, pellets were washed with 70% ethanol after supernatant was discarded. Pellets were air dried, re-suspended in 0.1 x TE buffer and quantified.

**Table 1:** Soil analysis of three locations used in the studies

Locations	Soil texture	pH	Organic matter (%)	N (%)	P <sub>2</sub> O <sub>5</sub> (ppm)	K <sub>2</sub> O (ppm)
The Univ. Agric. Peshawar	Silty Clay Loam	8.2	0.81	0.063	7.18	112
ARI, D.I. Khan	Clay Loam	7.9	0.87	0.047	7.8	147
NIBGE, Faisalabad	Clay Loam	7.4	0.93	0.038	9.05	179

**Table 2:** Pedigree of 28 upland cotton genotypes used in the studies

Genotypes	Parentage	Breeding centre	Released / under Approval
IR-NIBGE-901	PGMB-33/FH-90	NIBGE, Faisalabad	2011
IR-NIBGE-1524-4	PGMB-33/NIBGE-2	-do-	2010
IR-NIBGE-3	PGMB-33/FH-100	-do-	2012
IR-NIBGE-4	PGMB-33/CIM-448	-do-	2011
IR-NIBGE-5	PGMB-33/CIM496	-do-	Under approval
IR-3300-24	PGMB-33/BH-160	-do-	Under approval
IR-3300-13	PGMB-33/BH-160	-do-	Under approval
NIBGE-115	S-12/LRA-5166	-do-	2012
NN-3	S-12/LRA-5166	-do-	Under approval
NIBGE-2472	S-12/LRA-5166	-do-	Germplasm
NIBGE-2	LRA/S-12	-do-	2006
IR-2379	PGMB-33/FH-100	-do-	Germplasm
IR-NIBGE-3701-38	PGMB-33/CIM-448	-do-	2010
IR-1526	PGMB-33/NIBGE-2	-do-	Germplasm
NIBGE-314	S-12/LRA	-do-	Under approval
NIBGE-5	S-12/LRA	-do-	Germplasm
NIBGE-4	S-12/ CIM-448	-do-	Germplasm
IR-NIBGE-2620	IR-901/Rajhans	-do-	Germplasm
NIBGE-758-8	S-12/ CIM-448	-do-	Germplasm
IR-NIBGE-3701-33-6	PGMB-33/CIM-448	-do-	2010
SLH-284	-	CRS, Sahiwal	Under approval
CIM-446	CP 15/2 × S 12	CCRI, Multan	1998
CIM-473	CIM-402 × LRA 5166	-do-	2002
CIM-496	CIM-425 × 755-6/93	-do-	2005
CIM-499	CIM-433 × 755-6/93	-do-	2003
CIM-506	CIM-360 × CP 15/2	-do-	2004
CIM-554	2579-04/97 × W-1103	-do-	2009
CIM-707	CIM-243 × 738-6/93	-do-	2004

A total of 20 µL volume was used for polymerase chain reactions (PCR) using 15 ng of cotton DNA, 10 X buffer, 25 mM MgCl<sub>2</sub>, Primer-F 30 ng/µL, Primer-R 30 ng/µL, Taq polymerase 5 U/µL and deoxy-nucleotide triphosphates 2.5 mM. The amplification profile consisted of initial period of denaturation at 94°C for 5 min, followed by cycle (step-1) of 94°C for 30 s, 50°C for 30 s annealing, 72°C extension for 1 min. The PCR amplifications were followed by incubation at 72°C for 10 min. DNA quantification was carried using the NanoDrop®ND-1000. To check the quality and quantity of DNA 50 ng DNA was checked on 0.8% agarose gel. The DNA samples were rejected giving smear in the gel. Moreover, the dilution of 15 ng/µL was made from stock solution. The dilutions were also checked by comparing them with DNA quantification standards in agarose gel. The PCR was carried out using eppendorf master cycler gradient. The bands amplification was verified by omitting genomic DNA from control reaction. No amplification product was detected without genomic DNA in any PCR.

### Genetic markers

For the present study, 100 SSR markers were provided by Plant Genomics and Molecular Breeding (PGMB)

Laboratory, NIBGE, Faisalabad, Pakistan. These markers were selected based on their reproducible nature, PCR based, highly polymorphic, small quantity of genomic DNA requirement, easy interpretation in genotyping and easy automation.

### Agarose gel electrophoresis

The concentration of amplicons after PCR amplification was determined on 1% agarose gel stained with 3–5 µL ethidium bromide. For brightness of bands on 3% agarose gel, the agarose-based gel electrophoresis (PAGE) was made. By using pipette, all the PCR products were loaded into the wells carefully. The gel was loaded at room temperature while immersed in 1 x Tris/Boric acid / EDTA (TBA) buffer. Gels were run at 80 volts. Under these conditions, the PCR products usually separated after 80 min. Voltage gradient can be raised as high as 16 volts/cm to shorten time and improve band resolution. After the run completion, the gel was moved into a large UV illuminator and photographed.

### Scoring of data

In different cotton genotypes, the amplification shapes were

associated with each other and the DNA fragments bands were scored as there (1) or lacking (0). The said data was used to approximate the relationship based on common intensification products (Nei and Li 1979).

### Statistical tools

The following statistical tools were used to analyze the molecular data.

**STRUCTURE V. 2.3.1:** The basic purpose of association mapping was to search out the markers which have association with QTLs controlling the compound traits. In association mapping analysis, population structure is essential part because it reduces the type-1 error between traits of interest in self-pollinated species and molecular markers (Yu and Buckler 2006). The major problem is false positive in association mapping analysis. Population structure is considered as an effective approach to minimize the detection of false positive. Therefore, software 'STRUCTURE V. 2.3.1' was used to determine the population structure of all the cotton genotypes studied in this experiment before marker trait association analysis (Pritchard *et al.* 2000). In software options, a burn in length of 30,000 iterations and run length of 30,000 durations were used to test the K value in the range of 1–28. The populations denoted by K, while Delta-K values determine the sub-populations for K-ranging (Ali *et al.* 2019).

### Association mapping

Association mapping is an alternative to traditional QTL mapping because it is more accommodative in terms of using diverse germplasm. Since, most of the lines have very weak kinship in the cotton populations and conventional QTL mapping becomes ineffective in such case. Therefore, to compute the marker-trait association, the GLM was applied using the software 'TASSEL V. 2.1'. Major steps of association mapping includes i.e., a) collection of diverse germplasm lines, b) phenotypic characterization of selected population across multi-environments for the trait of interest, c) genotyping of the selected breeding material with suitable markers such as SSR, SNP and AFLP, d) assessment of the population structure and kinships based on genotypic data generated through unlinked molecular markers to avoid false positives and spurious associations, and e) correlation between genotypic and phenotypic data to tag the position of QTLs for a specific trait in upland cotton (Abdurakhmonov *et al.* 2008; Mei *et al.* 2013).

**TASSEL V. 2.1:** TASSEL (Trait Analysis by Association, Evolution and Linkage) is highly developed statistical tool used in association genetic study to find out the population structure and kinship information among varied individuals (Yu and Buckler 2006). The analysis of SSR can be carried out through TASSEL V. 2.1. In TASSEL, two approaches GLM and mixed linear model (MLM) are applied to achieve association analysis (Khan 2012). The identification

of QTLs for a trait is confirmed by SSR markers through GLM and MLM approaches.

### GLM

In GLM approach, association between mean phenotypic traits and markers is determined. In this method kinship data is not required which is latent cause of correlation between genotype and phenotype. It includes only population structure for analysis (Yu and Buckler 2006).

### MLM

In MLM approach, both population structure and kinship data are used in association mapping analysis (Ehrenreich *et al.* 2007). In the said method, the Q matrix and kinship data was used in TASSEL software. Therefore, K matrix (pair wise kinship among the studied genotypes) of a population was predictable using selected markers in our experiment. The MLM has an advantage over the GLM method because it collects evidence from both Q and K, while GLM accounts only Q matrix. Yu and Buckler (2006) findings revealed that MLM is an important approach to clear the confusion and identify the strong QTLs in association mapping analysis.

## Results

### Population structure

In present study, for molecular characterization of 28 upland cotton genotypes, 100 SSRs markers were used. Results revealed that 87 out of 100 SSRs markers were amplified in which 22 markers were polymorphic, and 65 were monomorphic in the existing cotton germplasm. However, 13 SSRs were not amplified for any genotype and were excluded. The 22 polymorphic SSR markers justified further analysis. In mixed ancestry the individuals might have inherited some portion of genome from their ancestors to diverse subpopulations. Structure analysis distributed 28 cotton genotypes into two main groups i.e., group-1 (genotypes 1 to 10) and group-2 (genotypes 11 to 28). The three genotypes i.e., NIBGE-115, NN-3, and NIBGE-2472 showed a little admixture (Fig. 1). The ideal numbers of groups (K) were obtained using online program 'Structure Harvester' (Evanno *et al.* 2005; Yu *et al.* 2006). This value reaches to the stability level when less number of groups that best illustrates the population structure by using structure harvester (Pritchard *et al.* 2000; Evanno *et al.* 2005). Based on K values, all the studied cotton genotypes again formed two major groups, where X-axis shows  $\Delta K$  value, and Y-axis shows the number of subpopulations (Fig. 2).

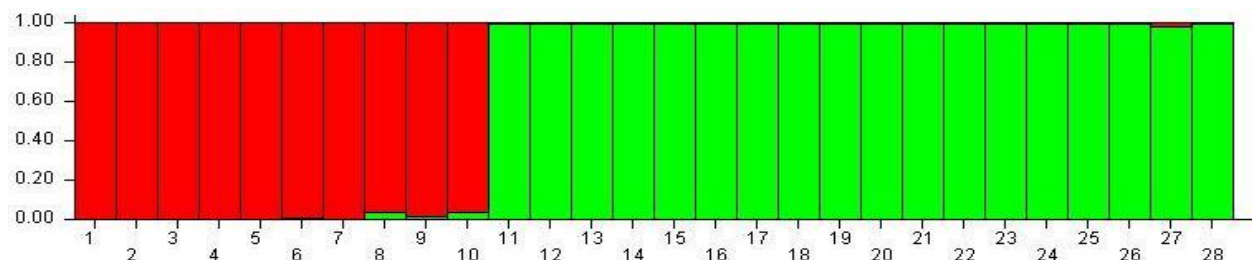
### Association mapping

Analysis of association mapping revealed significant

**Table 3:** Significant SSR markers for each QTL associated with different traits in upland cotton genotypes under general linear model (GLM) and mixed linear model (MLM) approaches

Traits	S.No.	Markers	Chromosome No	Position (cM)	P Value	R <sup>2</sup>	LOD
General linear model (GLM) approach							
Bolls per plant	1	MGHES-20	14	32	0.00	0.49	2.63
	2	BNL-1066	6	131	0.03	0.30	1.48
	3	MGHES67	16	62	0.01	0.44	1.85
	4	BNL-3280	20	111	0.01	0.87	1.86
Boll weight	5	MGHES-63	17	76.8	0.03	0.36	1.58
Seeds per boll	6	BNL-3254	18	145	0.00	0.51	2.64
	7	BNL-4108	16	183	0.00	0.61	2.54
	8	BNL-1667	5	24	0.04	0.40	1.43
	9	BNL-1417	6	71	0.05	0.18	1.33
	10	BNL-1066	6	131	0.02	0.33	1.63
	11	MGHES-18	16	169	0.01	0.33	2.02
Seed index	12	BNL-3280	20	111	0.03	0.82	1.54
	13	MGHES-55	23	162	0.01	0.42	2.19
Lint %	14	BNL-1667	5	24	1.08× 10 <sup>-4</sup>	0.72	3.97
	15	BNL-3254	18	145	1.14× 10 <sup>-4</sup>	0.65	3.94
	16	BNL-4108	16	183	0.00	0.64	2.76
	17	MGHES-53	9	184	0.02	0.30	1.78
	18	MGHES-20	14	32	0.03	0.34	1.55
	19	MGHES-18	16	169	0.01	0.32	1.91
	20	MGHES-3	18	108	0.03	0.28	1.54
	21	MGHES-60	19	48	0.01	0.76	1.87
	22	BNL-3627	23	39	0.00	0.27	2.30
	Mixed linear model (MLM) approach						
Bolls per plant	23	MGHES-3	18	108	0.032	0.32	1.49

\*\* , \* = Significant at p<0.01 and p<0.05, respectively; NS = Non-significant



**Fig. 1:** Q-plot showing clustering of 28 upland cotton genotypes based on analysis of genotypic data using ‘STRUCTURE’. Each genotype is represented by a vertical bar. The colored subsections within each vertical bar indicate membership coefficient (Q) of the genotype to different clusters. Identified subgroups are group 1 (red color) and group 2 (green color)

Legends for 28 upland cotton genotypes: 1: IR-NIBGE-901, 2: IR-NIBGE-1524-4, 3: IR-NIBGE-3, 4: IR-NIBGE-4, 5: IR-NIBGE-5, 6: IR-3300-24, 7: IR-3300-13, 8: NIBGE-115, 9: NN-3, 10: NIBGE-2472, 11: NIBGE-2, 12: IR-2379, 13: IR-NIBGE-3701-38, 14: IR-1526, 15: NIBGE-314, 16: NIBGE-5, 17: NIBGE-4, 18: IR NIBGE-2620, 19: NIBGE-758-8, 20: IR-NIBGE-3701-33-6, 21: SLH-284, 22: CIM-446, 23: CIM-473, 24: CIM-496, 25: CIM-499, 26: CIM-506, 27: CIM-554, 28: CIM-707

(p<0.05) association of 23 QTLs with different traits in the 28 upland cotton genotypes wherein 22 QTLs were identified by using GLM, while one with MLM approach.

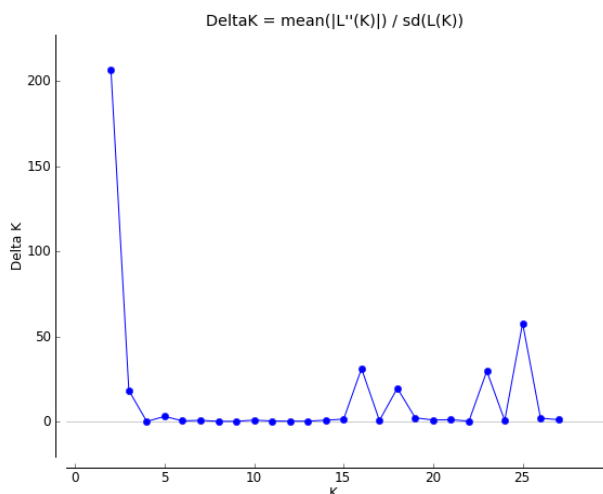
### QTLs associated with bolls per plant

Out of the 22 markers, four markers i.e., MGHES-20, BNL-1066, MGHES-67 and BNL-3280 showed significant (p<0.01) association of markers and traits, and were observed on chromosomes 6, 14, 16, and 20 by GLM approach (Table 3; Fig. 3a). In these four associated markers, the R<sup>2</sup> and P values were ranging from 0.30 to 0.87 and 0.002 to 0.01, respectively. The highest phenotypic variance (0.87) was identified for marker BNL-3280 with P value of 0.01 on chromosome 20 while the lowest R<sup>2</sup> (0.30) was determined for marker BNL-1066 on chromosome 6.

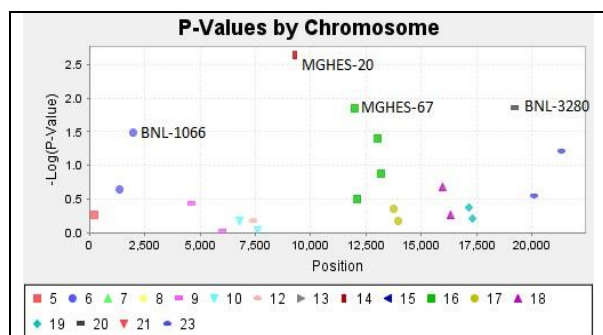
The marker MGHES-20 was found on chromosome 14 and showed strong association with bolls per plant with highest P value (0.002). In MLM approach, one marker MGHES-3 was significantly (p<0.05) associated with bolls per plant having R<sup>2</sup> and P values of 0.32 and 0.03, respectively (Table 3, Fig. 3b).

### QTLs association with boll weight

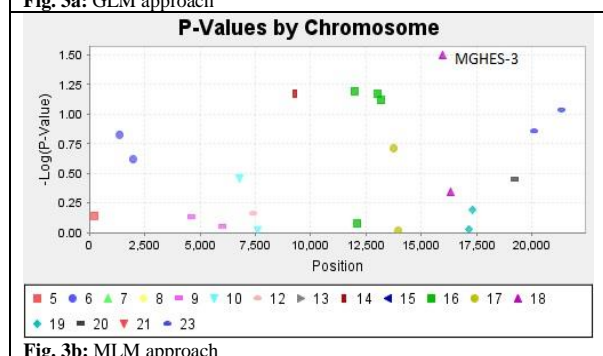
Both GLM and MLM approaches were applied by using SSR markers, and 22 markers showed marker trait association for boll weight. However, under GLM approach, one marker MGHES-63 showed significant (p<0.05) association and found on chromosome 17 with R<sup>2</sup> and P values of 0.36 and 0.02, respectively (Table 3; Fig. 4a). However, in MLM approach, no QTL was identified for the



**Fig. 2:** Estimating number of sub-populations using delta K values for K ranging from 2 to 28 using the method proposed by Evanno *et al.* (2005). K = 2 decided by delta K in 28 upland cotton genotypes



**Fig. 3a:** GLM approach



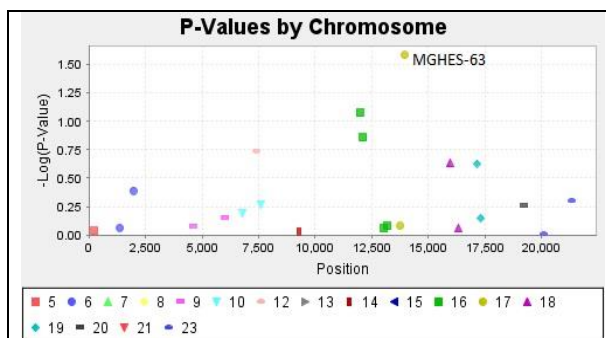
**Fig. 3b:** MLM approach

**Fig. 3a, b:** QTLs detection through SSR markers by applying GLM and MLM approaches for bolls per plant. Position of chromosomes and -Log (P-value) shown along X-axis and Y-axis, respectively

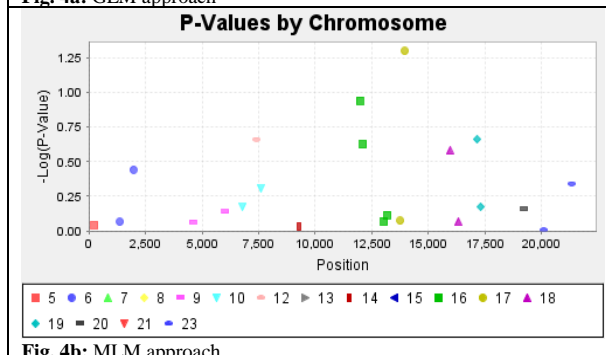
boll weight in these cotton genotypes (Fig. 4b).

### QTLs association with seeds per boll

All the 22 markers exhibited the marker-trait association for the said trait. However, three markers i.e., BNL-3254,



**Fig. 4a:** GLM approach



**Fig. 4b:** MLM approach

**Fig. 4a, b:** QTLs detection through SSR markers by applying GLM and MLM approaches for boll weight. Position of chromosomes and -Log (P-value) shown along X-axis and Y-axis,

BNL-4108, and MGHE-18 revealed highly significant ( $p < 0.01$ ) association, while four other markers (BNL-1667, BNL-1417, BNL-1066 and BNL-3280) showed significant ( $p < 0.05$ ) marker-trait association for seeds per boll through GLM approach (Table 3; Fig. 5a). In highly significant markers, one marker was found on chromosome 18 and other three were found on chromosome 16. In four significant markers, one was found on chromosome 5, two on chromosome 6, while fourth one was observed on chromosome 18. For the above eight significant markers, the  $R^2$  and P values ranged from 0.18 to 0.82 and 0.0022 to 0.047, respectively. Highest phenotypic variance (0.82) was identified for marker BNL-3280 with P value of 0.01 on chromosome 20, while lowest  $R^2$  (0.18) was determined for marker BNL-1417 on chromosome 6. Two markers BNL-3254 and BNL-4108 were found on chromosomes 18 and 16, which showed strong association with seeds per boll with highest P values of 0.0022 and 0.0029, respectively (Table 3). In MLM approach, none of the studied markers revealed significant marker trait association with seeds per boll (Fig. 5b).

### QTLs association with seed index

Both GLM and MLM approaches were applied by using SSR markers to study the marker-trait association for seed index. Twenty-two markers showed marker trait association. However, one marker MGHE-65 showed

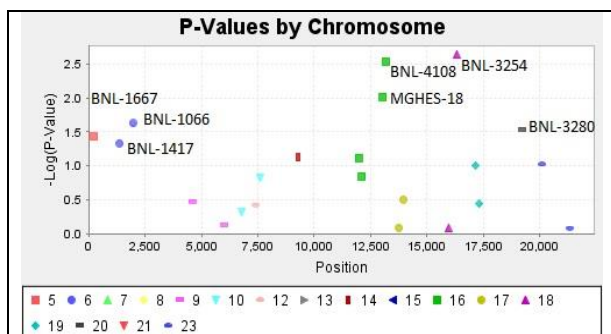


Fig. 5a: GLM approach

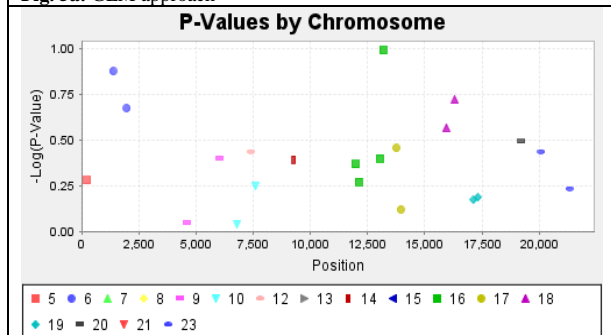


Fig. 5b: MLM approach

**Fig. 5a, b:** QTLs detection through SSR markers by applying GLM and MLM approaches for seeds per boll. Position of chromosomes and  $-\log(P\text{-value})$  shown along X-axis and Y-axis, respectively

highly significant ( $p < 0.01$ ) association with seed index and was found on chromosome 23 through GLM (Fig. 6a). The phenotypic variance and P values for the associated marker were 0.42 and 0.006, respectively (Table 3). In MLM approach, none of the studied markers showed significant association with seed index (Fig. 6b).

### QTLs association with lint index

Twenty-two markers showed marker trait association. However, none of the markers showed significant association with lint index under GLM and MLM approaches (Fig. 7a, b).

### QTLs association with lint percentage

For lint percentage, both GLM and MLM approaches were applied by using SSR markers for marker trait association. In GLM, four markers i.e., BNL-1667, BNL-3254, BNL-4108, and BNL-3627 showed highly significant ( $p < 0.01$ ) association, while five markers (MGHES-53, MGHES-20, MGHES-18, MGHES-3 and MGHES-60) showed significant ( $p \leq 0.05$ ) marker trait association for lint percentage (Table 3, Fig. 8a). The former four markers were found on chromosomes 5, 16, 18, and 23, while later five markers were observed on chromosomes 9, 14, 16, 18 and 19. For above nine significant markers, the  $R^2$  and P values

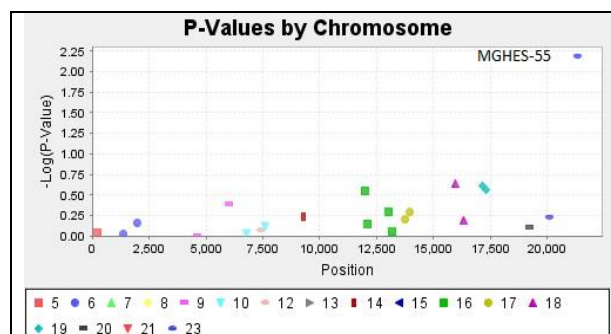


Fig. 6a: GLM approach

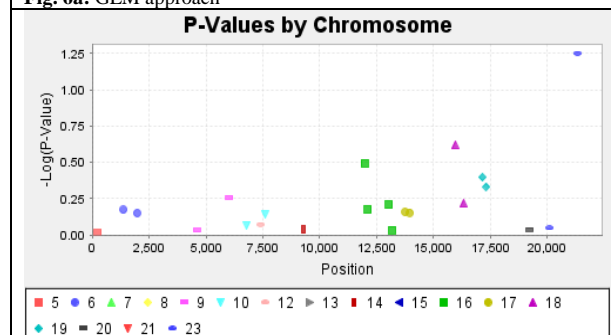


Fig. 6b: MLM approach

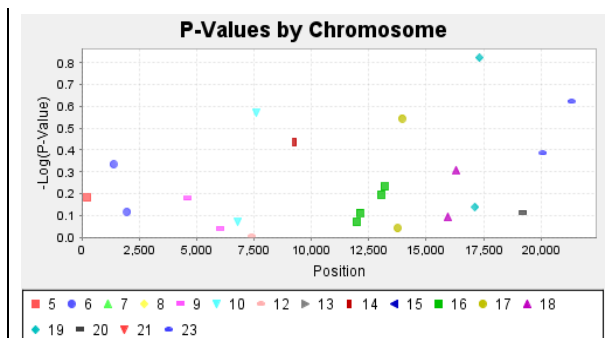
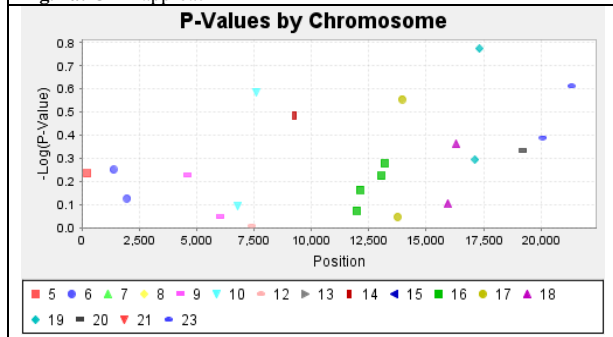
**Fig. 6a, b:** QTLs detection through SSR markers by applying GLM and MLM approaches for seed index. Position of chromosomes and  $-\log(P\text{-value})$  shown along X-axis and Y-axis, respectively

were ranging from 0.27 to 0.76 and  $1.08 \times 10^{-4}$  to 0.029, respectively. The highest phenotypic variance (0.76) was identified for marker MGHES-60 with P value of 0.013 on chromosome 19, while lowest  $R^2$  value (0.27) was determined for marker BNL-3627 on chromosome 23. Three markers BNL-1667, BNL-3254, and BNL-4108 were found on chromosomes 5, 18 and 16 and showed strong association with lint percentage with highest P values of  $1.08 \times 10^{-4}$ ,  $1.14 \times 10^{-4}$  and 0.0017, respectively (Table 3). In MLM approach, none of the studied markers showed marker trait association for lint percentage (Fig. 8b).

### Discussion

In this study, two different sub-populations were identified using model-based population structure analysis in the existing elite upland cotton germplasm. Population structure analysis provides information about origin of accessions used for association analysis (Hussain et al. 2019). Twenty-eight upland cotton genotypes were allocated to mixed group demonstrating a little admixture. Sharing of germplasm among different breeding programs is the possible reason of admixture in the studied cotton genotypes. Another reason could be the recurrent use of few lines in multiple breeding programs with best agronomic traits (Van-Esbroeck et al. 1999). Genotypic data consisting of unlinked markers in upland cotton is the major cause of



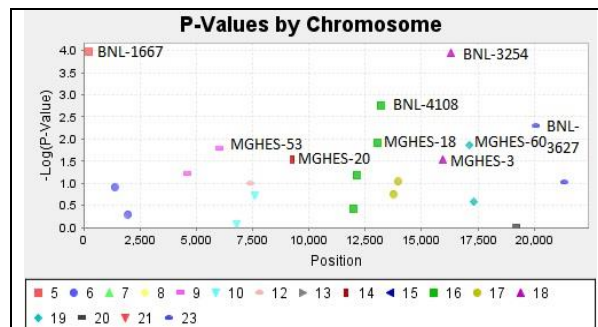
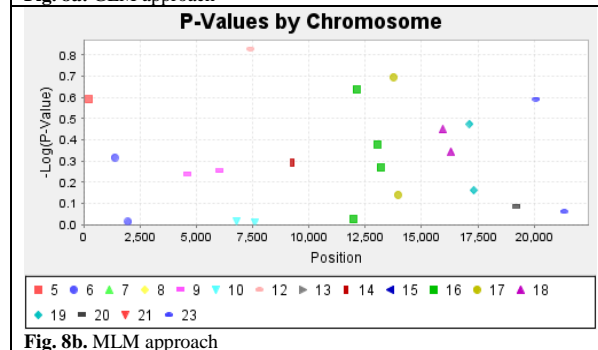

**Fig. 7a:** GLM approach

**Fig. 7b:** MLM approach

**Fig. 7a, b:** QTLs detection through SSR markers by applying GLM and MLM approaches for lint index. Position of chromosomes and  $-\text{Log}(P\text{-value})$  shown along X-axis and Y-axis, respectively

clustering of individuals into subpopulations (Guo *et al.* 2007; Khan *et al.* 2010; Paterson *et al.* 2010). Zhang *et al.* (2019) reported that easily available SSRs could efficiently reduce labor-cost and inefficient processes by providing a best alternative for the identification of molecular markers for MAS breeding in the future.

In this study, four highly significant QTLs were observed for bolls per plant through GLM approach, while in MLM, one marker showed significant association (Table 3; Fig. 3a, b). Past findings revealed that 32 new QTLs were identified in upland cotton genotypes and 10 marker loci were found to be consistent with formerly identified QTLs (Zhao *et al.* 2014). However, in other studies only four QTLs were reported on four different chromosomes i.e., 2, 11, 14 and 21 for bolls per plant in upland cotton (Said *et al.* 2013). Qin *et al.* (2015) reported 25 QTLs for yield traits i.e., bolls per plant, boll weight and lint percentage in upland cotton through MLM approach. However, population structure and kinship can affect the results of association mapping (Wen *et al.* 2019). Similarly, Wang *et al.* (2015) reported 14 QTLs on eight chromosomes i.e., 5, 6, 9, 13, 15, 17, 24, and 25 for bolls per plant in upland cotton germplasm.

With GLM, genotypic, phenotypic and population structure Q, matrix (admixture of population) was used, while in MLM the kinship (relatedness of individuals means brother and sisters) was also used and identified markers


**Fig. 8a:** GLM approach

**Fig. 8b:** MLM approach

**Fig. 8a, b:** QTLs detection through SSR markers by applying GLM and MLM approaches for lint percentage. Position of chromosomes and  $-\text{Log}(P\text{-value})$  shown along X-axis and Y-axis, respectively

related with different agronomic traits (Yu *et al.* 2006). In this study, under GLM approach one marker was observed for boll weight (Table 3; Fig. 4a, b). Previously, 39 QTLs were identified for yield and yield components in upland cotton genotypes (Yu *et al.* 2013). A total of 26 QTLs were identified in which chromosome 14 contained four QTLs, while chromosome 18 and 22 contained three each QTLs, chromosomes 5, 25, and 26 contained two each QTLs, while chromosomes 1, 2, 4, 11, 12, 15, 16, 21, and 24 contained one QTL for boll weight in upland cotton genotypes (Said *et al.* 2013). Zhang *et al.* (2016) identified 16 stable QTLs for boll weight on different chromosomes. Twenty QTLs were reported on seven chromosomes i.e., 4, 5, 12, 15, 16, 21, and 26 for boll weight in upland cotton accessions (Wang *et al.* 2015).

In present study, each of the markers provided highly significant and significant marker trait associations for seeds per boll through GLM approach (Table 3; Fig. 5a, b). However, in past studies, only one QTL was reported for this trait and it was found on chromosome 12 in upland cotton genotypes (Said *et al.* 2013). Hussain *et al.* (2019) also identified one QTL for seeds per boll on chromosome 21 in upland cotton through GLM approach. In support of present findings, for seeds per boll, six QTLs were also identified in different upland accessions (He *et al.* 2005). In successful use of association mapping in plants, the major restrictions are genetic relatedness and population structure,

resulting in false relationships between markers and make it hard to identify and separate the loci that actually affect the targeted variables in upland cotton germplasm (Gupta *et al.* 2005).

One marker was identified with highly significant association with seed index through GLM approach in this research (Table 3; Fig. 6a, b). Past studies revealed QTLs association with the agronomic and fiber traits of *Gossypium hirsutum* L. populations (Shappley *et al.* 1998). He *et al.* (2005) identified five QTLs for seed index, while Wang *et al.* (2015) findings enunciated two QTLs for seed index in upland cotton germplasm. However, in the past, 10 QTLs were identified and reported for seed index on different chromosomes i.e., chromosome 14 contained three QTLs, while chromosomes 3, 7, 17, 22, 23, 24, and 26 contained one each QTL in upland cotton genotypes (Said *et al.* 2013).

In present study, for lint index no QTL was observed for lint index which might be due to less diverse origin of the genotypes and the weak kinship between studied genotypes (Table 3; Fig. 7a, b). However, past studies reported two QTLs (Wang *et al.* 2015) and one QTL (He *et al.* 2005) for lint index in upland cotton populations. Like current investigations, Zheng *et al.* (2008) as well stated that the true potential of MLM was amplified by addition of population structure and kinship data in upland cotton. Li *et al.* (2007) revealed that marker E6M3-266 had sound connection with lint index in cotton. In other past studies, 15 QTLs were reported for lint index in which two each QTLs were present on chromosomes 11, 12, 14 and 26, while chromosomes 4, 5, 7, 9, 10, 22 and 25 contained one each QTL for lint index in cotton germplasm (Said *et al.* 2013). However, due to differences in genetic make-up of the genotypes and environmental conditions, contradictions are expected by comparing the identified QTLs in the present and past studies.

In present study, nine markers showed marker-trait association for the lint percentage under GLM approach. In MLM approach, other than population structure, genotypic and phenotypic variances and kinship components were used, therefore, the contradiction with the previous studies might be due to the high relatedness among studied genotypes (Table 3; Fig. 8a, b). Hussain *et al.* (2019) also reported two markers on chromosome three and five in upland cotton through GLM approach for lint percentage. Past studies revealed that 55 marker-trait relationships were perceived in 26 SSRs for fiber percentage based on MLM approach in upland cotton genotypes (Mei *et al.* 2013). However, other studies revealed that seven (He *et al.* 2005) and 13 QTLs were observed for lint percentage in upland cotton germplasm (Wang *et al.* 2015). Liu *et al.* (2018) structured genetic map holding SSR markers and SNPs and identified 36 QTLs on chromosome 21 across nine environments. Due to many advantages, mapping of QTL through association analysis is extra precise and proficient for assessing major QTLs protecting marked genes

accountable for major variables in *G. hirsutum* germplasm (Zhang *et al.* 2019).

Present investigations confirmed that association analysis is an effective tool to ascertain associations with the important agronomic traits in upland cotton. Previous studies have identified associations with various traits by using many SSR markers in different crop plants (El-Hosary and El-Akkad 2015). These studies can be further exploited for making linkage maps and marker assistant selection. The QTLs identified through association mapping can thus be used to improve cotton cultivars by using techniques like marker assistant selection.

## Conclusion

Association mapping identified 23 QTLs associated with different traits in 28 upland cotton genotypes wherein 22 QTLs were identified through GLM approach while one QTL through MLM approach. The detected QTLs will be effective in identifying and grasping the genetic source of different traits and diversity in upland cotton genotypes. The identified and favorable QTLs might also facilitate the breeders in maintaining the genetic variability in the gene pool of upland cotton genotypes for future breeding program.

## Acknowledgements

The present investigations were financed by the Higher Education Commission (HEC), Islamabad - Pakistan. We also pay thanks to the University of Agriculture, Peshawar - Pakistan for organizational support, and the Department of Plant Breeding and Genetics for different assistances during the studies. We are also thankful to the National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad - Pakistan for their technical support. The observations uttered in the manuscript are solely those of the authors and do not represent the funding agency.

## References

- Abdurakhmonov IY, RJ Kohel, JZ Yu, AE Pepper, AA Abdullaev, FN Kushanov, IB Salakhutdinov, ZT Buriev, S Saha, BE Scheffler, JN Jenkins, A Abdurkarimov (2008). Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics* 92:478–487
- Abdurakhmonov IY, S Saha, JN Jenkins, ZT Buriev, SE Shermatov, BE Scheffler, AE Pepper, JZ Yu, RJ Kohel, A Abdurkarimov (2009). Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica* 136:401–417
- Ali MA, SI Awan (2009). Inheritance pattern of seed and lint traits in cotton (*Gossypium hirsutum*). *Intl J Agric Biol* 11:44–48
- Ali I, NU Khan, S Gul, SU Khan, Z Bibi, K Aslam, G Shabir, HA Haq, SA Khan, I Hussain, S Ahmed, A Din (2019). Genetic diversity and population structure analysis in upland cotton germplasm. *Intl J Agric Biol* 22:669–676
- Bolek Y, KM El-Zik, AE Pepper, AA Bell, CW Magill, PM Thaxton, OUK Reddy (2005). Mapping of verticillium wilt resistance genes in cotton. *Plant Sci* 168:1581–1590

- Chen ZJ, BE Scheffler, E Dennis, BA Triplett, T Zhang, W Guo, *et al.* (2007). Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol* 145:1303–1310
- Ehrenreich IM, PA Stafford, MD Purugganan (2007). The genetic architecture of shoot branching in *Arabidopsis thaliana*: A comparative assessment of candidate gene associations vs. quantitative trait locus mapping. *Genetics* 176:1223–1236
- El-Hosary AAA, TA El-Akkad (2015). Genetic diversity of maize inbred lines using ISSR markers and its implication on quantitative traits inheritance. *Arab J Biotechnol* 18:81–96
- Ersoz ES, J Yu, ES Buckler (2007). Applications of linkage disequilibrium and association mapping in crop plants. In: *Genomics-Assisted Crop Improvement*, Vol 1, pp:97–119. Varshney RV, R Tuberosa (eds). Springer, Dordrecht, The Netherlands
- Evanno G, S Regnaut, J Goudet (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14:2611–2620
- Flint-Garcia SA, JM Thornsberry, ES Buckler (2003). Structure of Linkage Disequilibrium in Plants. *Annu Rev Plant Biol* 54:357–374
- Fu YB, MH Yang, F Zeng, B Biligetu (2017). Searching for an accurate marker-based prediction of an individual quantitative trait in molecular plant breeding. *Front Plant Sci* 8; Article No. 1182
- Guo W, P Cai, C Wang, Z Han, X Song, K Wang, X Niu, C Wang, K Lu, B Shi, T Zhang (2007). A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. *Genetics* 176:527–541
- Gupta PK, S Rustgi, PL Kulwal (2005). Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Mol Biol* 57:461–485
- He DH, ZX Lin, XL Zhang, YC Nie, XP Guo, C Da Feng, JMD Stewart (2005). Mapping QTLs of traits contributing to yield and analysis of genetic effects in tetraploid cotton. *Euphytica* 144:141–149
- Hussain S, M Hussain, M Javed, S Sarwar, M Zubair (2019). Mapping of QTLs responsible for yield related traits in advance lines of cotton (*Gossypium hirsutum* L.). *J Genet Mol Biol* 03:11–18
- Iqbal MJ, N Aziz, NA Saeed, Y Zafar, KA Malik (1997). Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. *Theor Appl Genet* 94:139–144
- Khan MA (2012). *Association Mapping and TASSEL Software Tutorial*. University of Illinois, Urbana-Champaign, Illinois, USA
- Khan AI, FS Awan, B Sadia, RM Rana, IA Khan (2010). Genetic diversity studies among coloured cotton genotypes by using rapd markers. *Pak J Bot* 42:71–77
- Li Y, Y Li, S Wu, K Han, Z Wang, W Hou, Y Zeng, R Wu (2007). Estimation of multilocus linkage disequilibria in diploid populations with dominant markers. *Genetics* 176:1811–1821
- Liu R, J Gong, X Xiao, Z Zhang, J Li, A Liu, *et al.* (2018). GWAS analysis and qtl identification of fiber quality traits and yield components in upland cotton using enriched high-density snp markers. *Front Plant Sci* 9:Article 1067
- Mei H, X Zhu, T Zhang (2013). Favorable QTL alleles for yield and its components identified by association mapping in Chinese upland cotton cultivars. *PLoS One* 8; Article 0082193
- Nei M, WH Li (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76:5269–5273
- Ochieng JW, AWT Muigai, GN Ude (2007). Localizing genes using linkage disequilibrium in plants: Integrating lessons from the medical genetics. *Afr J Biotechnol* 6:650–657
- Paterson AH, J kang Rong, AR Gingle, PW Chee, ES Dennis, D Llewellyn, *et al.* (2010). Sequencing and utilization of the *Gossypium* genomes. *Trop Plant Biol* 3:71–74
- Pritchard JK, M Stephens, NA Rosenberg, P Donnelly (2000). Association mapping in structured populations. *Amer J Hum Genet* 67:170–181
- Qin H, M Chen, X Yi, S Bie, C Zhang, Y Zhang, J Lan, Y Meng, Y Yuan, C Jiao (2015). Identification of associated SSR markers for yield component and fiber quality traits based on frame map and upland cotton collections. *PLoS One* 10; Article e0118073
- Qin H, W Guo, YM Zhang, T Zhang (2008). QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L. *Theor Appl Genet* 117:883–894
- Rafalski A (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* 5:94–100
- Said JI, Z Lin, X Zhang, M Song, J Zhang (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; Article 776
- Shakoor MS, TA Malik, FM Azhar, MF Saleem (2010). Genetics of agronomic and fiber traits in upland cotton under drought stress. *Intl J Agric Biol* 12:495–500
- Shappley ZW, JN Jenkins, J Zhu, JC McCarty (1998). Quantitative trait loci associated with agronomic and fiber traits of upland cotton. *J Cotton Sci* 2:153–163
- Van-Esbroeck GA, DT Bowman, OL May, DS Calhoun (1999). Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. *Crop Sci* 39:323–328
- Wang P, YZ Ding, QX Lu, WZ Guo, TZ Zhang (2008). Development of *Gossypium barbadense* chromosome segment substitution lines in the genetic standard line TM-1 of *Gossypium hirsutum*. *Chin Sci Bull* 53:1512–1517
- Wang B, W Guo, X Zhu, Y Wu, N Huang, T Zhang (2007). QTL mapping of yield and yield components for elite hybrid derived-RILs in upland cotton. *J Genet Genomics* 34:35–45
- Wang H, C Huang, H Guo, X Li, W Zhao, B Dai, Z Yan, Z Lin (2015). QTL mapping for fiber and yield traits in upland cotton under multiple environments. *PLoS One* 10; Article e0130742
- Wen T, B Dai, T Wang, X Liu, C You, Z Lin (2019). Genetic variations in plant architecture traits in cotton (*Gossypium hirsutum*) revealed by a genome-wide association study. *Crop J* 7:209–216
- Yan J, CB Kandianis, CE Harjes, L Bai, EH Kim, X Yang, DJ Skinner, Z Fu, S Mitchell, Q Li, MGS Fernandez, M Zaharieva, R Babu, Y Fu, N Palacios, J Li, D DellaPenna, T Brutnell, ES Buckler, ML Warburton, T Rocheford (2010). Rare genetic variation at *Zea mays crtRB1* increases B-carotene in maize grain. *Nat Genet* 42:322–327
- Yang C, W Guo, G Li, F Gao, S Lin, T Zhang (2008). QTLs mapping for Verticillium wilt resistance at seedling and maturity stages in *Gossypium barbadense* L. *Plant Sci* 174:290–298
- Yu J, ES Buckler (2006). Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17:155–160
- Yu J, G Pressoir, WH Briggs, IV Bi, M Yamasaki, JF Doebley, MD McMullen, BS Gaut, DM Nielsen, JB Holland, S Kresovich, ES Buckler (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208
- Yu J, K Zhang, S Li, S Yu, H Zhai, M Wu, X Li, S Fan, M Song, D Yang, Y Li, J Zhang (2013). Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. *Theor Appl Genet* 126:275–287
- Zhang ZS, MC Hu, J Zhang, DJ Liu, J Zheng, K Zhang, W Wang, Q Wan (2009). Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Mol Breed* 24:49–61
- Zhang C, L Li, Q Liu, L Gu, J Huang, H Wei, H Wang, S Yu (2019). Identification of loci and candidate genes responsible for fiber length in upland cotton (*Gossypium hirsutum* L.) via association mapping and linkage analyses. *Front Plant Sci* 10; Article 00053
- Zhang T, N Qian, X Zhu, H Chen, S Wang, H Mei, Y Zhang (2013). Variations and transmission of QTL alleles for yield and fiber qualities in upland cotton cultivars developed in China. *PLoS One* 8; Article 0057220
- Zhang Z, H Shang, Y Shi, L Huang, J Li, Q Ge, *et al.* (2016). Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to quantitative trait loci (QTL) analysis for boll weight in upland cotton (*Gossypium hirsutum*). *BMC Plant Biol* 16; Article 79

- Zhang K, J Zhang, J Ma, S Tang, D Liu, Z Teng, D Liu, Z Zhang (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
- Zhao Y, H Wang, W Chen, Y Li (2014). Genetic structure, linkage disequilibrium and association mapping of verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. *PLoS One* 9; Article 0086308
- Zheng P, WB Allen, K Roesler, ME Williams, S Zhang, J Li, K Glassman, J Ranch, D Nubel, W Solawetz, D Bhatramakki, V Llaca, S Deschamps, GY Zhong, MC Tarczynski, B Shen (2008). A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Natl Genet* 40:367–372
- Zhu GL, CH Wang, XH Guo, WK Gao YY Gan (2008). The preliminary research on the growth characteristics of Baimian2. *J Henan Agric Sci* 15:47–50