ESTIMATION OF MORPHOLOGICAL AND MOLECULAR DIVERSITY OF SEVENTY-TWO ADVANCED PAKISTANI COTTON GENOTYPES USING SIMPLE SEQUENCE REPEATS

SHOUKAT, S.[1*] – ANWAR, M.[1*] – ALI, S.[1] – BEGUM, S.[1] – AQEEL, M.[1] – ZIA, M. A.[1] – SHAH, S. H.[2] – ALI, G. M.[1]

[1]National Institute for Genomics and Advanced Biotechnology (NIGAB), National Agricultural Research Centre (NARC), Islamabad, Pakistan

[2]Department of Agricultural Sciences, Allama Iqbal Open University, Islamabad, Pakistan

[*]Corresponding authors
e-mail: f_shehla@yahoo.com; muhammadanwar1964@gmail.com

Abstract. Genetic diversity is the base of biological polymorphism and also the basic part of biological diversity. Significant role is played by genetic diversity and paternity of germplasm in cotton breeding. Genetic diversity in representative sets of advanced cotton genotypes was studied at molecular level employing polymorphic simple sequence repeats. Of seventy-two advanced cotton genotypes, thirty microsatellites primer pairs were studied, 29 showed to be polymorphic. DNA markers amplified fragments of 100 to 300 bp in size. The polymorphism information contents ranged between 0.76 and 0.99. Cluster analysis clearly grouped the 72 genotypes into seven groups A, B, C, D, E, F, G based on Nei's coefficient similarity matrix. Group B was found to be the most genetically distinct group when compared to other groups. However, group B had most of the candidates of under observed genotypes. This study exposed two major clusters of cotton cultivars and recognized the cotton genotypes which were genetically different. These results can be used as reference and baseline information to other researchers/breeders. The resolution to the genetic associations of these cotton accessions could be improved by increasing the number of more mapped markers.
Keywords: *cotton, genetic diversity, SSR markers, Pakistan, Gossypium, genome wide characterization*

Introduction

Man has been refining the domestication of crop plants by giving selection emphasis to the traits that suit to his agro-ecological and socioeconomic needs. Like many other crops, selection preference has been made for the improvement of the different traits in cotton. The process is being continued for centuries that has resulted in cultivars far different from the wild types in particular habitat and potential. During the last 30 years, Excessive breeding emphasis has led to some sort of genetic uniformity among the currently cultivated high yielding varieties and advanced lines (Van Becelaere et al., 2005). There is no precise knowledge of extent of exploitable variability beyond simple inherited traits during breeding of improved crop traits. Also it is not clear yet that what extent the different breeding strategies have been facilitating to broaden or narrow the genetic diversity in the breeding nurseries of cotton. Cotton belongs to Gossypieae tribe of the Malvaceae family. The main area for its cultivation is tropical as well as subtropical regions. Cotton genus Gossypium consists of almost 50 species that is contributing in making it the largest species in the Gossypieae tribe (Ashokkumar et al., 2011).

Cotton crop has great importance in agriculture, industry and trade in continents like Africa, Asia and South America. The cotton genus Gossypium has always been attracting the scientists' attention (Puspito et al., 2015). As far as the commercial

species of cotton are concerned, the *G. hirsutum* accounts for greater than 90% of world production, *G. barbadense* accounts for 3-4%, *G. arboreum* and *G. herbaceum* together accounts for 2%. By using selective breeding and hybridization of the above mentioned species many varieties of cotton have been developed. Humans use cotton as a source of natural fiber. However, cotton is also known as an oilseed crop and it is also the protein source for animals feed. Most of the cotton fiber is found in white color but it may also be found in brown, green and some mixing of these colors naturally (Chapagain et al., 2006). Above and beyond cotton fiber, edible oil is also extracted from cotton seeds. Cotton raw material is also used in poultry and livestock feed (Moiana et al., 2012). Cotton is being produced in many countries of the world, but China, US, India and Pakistan are producing two third of the global production (Ehsan et al., 2013).

In developed as well as developing countries, cotton is the cash crop (Puspito et al., 2015). The credit of being the oldest seed known so far goes to cotton seed. As far as Pakistan is concerned, the cotton is known as cash crop and its fiber is known as "White Gold" (Ahmad, 2018). The cropped area of cotton in Pakistan is about 2961 thousand hectares. In Pakistan, increase in cotton bales' production has been observed to the last years i.e. 14.98 million bales against 12.8 million bales (Economic Survey, 2014-15). In Indus Valley it has been cultivated since 3500 BC. In Balochistan it is cultivated near Bolan Pass (Li et al., 2008). Old-style variety identification technology using the old test and method involves rising plants to maturity and evaluating various phenotypic features which differentiate individual plants. It is difficult to collect morphological data because of environmental influence on morphological characters and also the time factor; other limitation includes explicit differentiation of genotypes (Rana et al., 2007).

Cultivated species of cotton are showing low level of genetic diversity as indicated by previous research. Due to lack of genetic diversity, cotton cultivars with stress resistance, improved yield and quality potential are in great danger (Li et al., 2008). It has been observed that genetic diversity is decreasing in breeding programs. This decrease in genetic diversity is leading to decline in yield improvement through crop breeding, since diversity is essential for selection (Rana et al., 2007). Orthodox breeding techniques are insufficient to explain the multifaceted traits. The choice of markers used for the estimation of diversity is of great alternative now a days as it gives alternatives to variety selection in the initial phase of breeding approach. The marker assisted selection is used to identify the superior parents in a cross. The advantage of this technique is to reduce the time and costs required for development of novel accessions (Bukhari, 2014). Genomics have come into being by doing advances in DNA marker technology. DNA repeats that comprise two to five nucleotide core units are known as simple sequence repeats (SSR) that are short and uncertain repetitive DNA sequences (Dongre et al, 2005). SSR markers have the advantage of being a PCR based marker, co-dominant, multi allelic, genetic marker system (Ullah et al., 2012). DNA markers (SSRs) have been used widely for the genetic diversity analysis as well as identification researches. Just because of high polymorphism rates, SSR markers are highly informative. This marker system is cheap to analyze when compared with other marker system. SSR regions are dispersed all over the genomes (Varshney et al., 2005). By using molecular markers, genetic diversity of cotton can excellently be estimated and this type of study offers beneficial information on the selection of parents in the expansion of cotton varieties and crossbreeds as well (Wu et al., 2006; Ullah et al., 2012). Fewer studies have been reported so far in Pakistani cotton germplasm. So this study was designed by keeping in view the importance of DNA markers to study

Shoukat et al.: Estimation of morphological and molecular diversity of seventy-two advanced Pakistani cotton genotypes using
simple sequence repeats
- 2045 -

diversity among local cotton advanced lines, the genetic variability and estimation of relatedness and to illustrate genotypes with diverse fingerprints of selected advance cotton genotypes using SSR markers.

Materials and methods

The present research was conducted in National Institute for Genomics and Advanced Biotechnology (NIGAB), National Agricultural Research Center (NARC), Islamabad to estimate the genetic diversity among 72 Pakistani advanced cotton lines using thirty simple sequence repeats (SSR) markers. The seeds of 72 advanced cotton genotypes were acquired from National Coordinated Varietal Trial (NCVT), Pakistan Central Cotton Committee (PCCC) at NIGAB, NARC (*Table 1*). The 72 advanced cotton lines were categorized into two populations 1 and 2 based on unweighted pair group method with arithmetic mean (UPGMA). Thirty SSR primer pairs were selected for genotyping (*Table 2*). The Cotton Marker Database (CMD; http:// www.cottongen.org) was explored to select primers. All the markers were selected by reviewing the literature. Only those markers were selected which were polymorphic in others cotton genome.

Table 1. *List of National Coordinated Varietal Trial (NCVT) accessions*

S. No.	Genotypes	S. No.	Genotypes	S. No.	Genotypes	S. No	Genotypes
1	NCVT 1601 (Bahar-07)	19	NCVT 1619 (MNH-1016)	37	NCVT 1637 (GS-Ali-1)	55	NCVT 1655 (Weal-AG-Gold)
2	NCVT 1602 (Deebal)	20	NCVT 1620 (PB-896)	38	NCVT 1638 (GH-Hammad)	56	NCVT 1656 (Weal-AG-Shahkar)
3	NCVT 1603 (NIAB-878-B)	21	NCVT 1621 (NIAB-545)	39	NCVT 1639 (MNS-992)	57	NCVT 1657 (Weal-AG-1606)
4	NCVT 1604 (FH-152)	22	NCVT 1622 (BS-15)	40	NCVT 1640 (CIM-602 (S-1)	58	NCVT 1658 (Tahafuz-5)
5	NCVT 1605 (FH-142 (S-2)	23	NCVT 1623 (RH-668)	41	NCVT 1641 (QM-IUB-65)	59	NCVT 1659 (BH-201)
6	NCVT 1606 (Cyto-179)	24	NCVT 1624 (Eagle-2)	42	NCVT 1642 (NIA-86)	60	NCVT 1660 (Thakkar-808)
7	NCVT 1607 (Cyto-313)	25	NCVT 1625 (TH-20)	43	NCVT 1643 (RH-667)	61	NCVT 1661 (GS-Ali-5)
8	NCVT 1608 (CEMB-88*)	26	NCVT 1626 (IR-NIBGE_9)	44	NCVT 1644 (NS-181)	62	NCVT 1662 (Tarzan-5)
9	NCVT 1609 (FH-326)	27	NCVT 1627 (NIAB-1048)	45	NCVT 1645 (CIM-610)	63	NCVT 1663 (CIM-602 (S-1)
10	NCVT 1610 (CIM-632)	28	NCVT 1628 (Sahara-Buraq)	46	NCVT 1646 (FH-142 (S-2))	64	NCVT 1664 (Sitara-15)
11	NCVT 1611 (CRIS-543)	29	NCVT 1629 (RH-662)	47	NCVT 1647 (MPS-29)	65	NCVT 1665 (SAU-1)
12	NCVT 1612 (BPC-11)	30	NCVT 1630 (Bakhtawar-1)	48	NCVT 1648 (MPS-61)	66	NCVT 1666 (SLH-12)
13	NCVT 1613 (Crystal-12)	31	NCVT 1631 (Zakariya-1)	49	NCVT 1649 (CRIS-600)	67	NCVT 1667 (NIAB-Bt-2)
14	NCVT 1614 (BPC-10)	32	NCVT 1632 (CIM-573 (S)	50	NCVT 1650 (FH-Kehkashan)	68	NCVT 1668 (Shaheen-1)
15	NCVT 1615 (Thakkar-214)	33	NCVT 1633 (CEMB-55-S)	51	NCVT 1651 (VH-Gulzar)	69	NCVT 1669 (Tipu-1)
16	NCVT 1616 (IR_NIBGE-8)	34	NCVT 1634 (FH-142 (S-2)	52	NCVT 1652 (CIM-625)	70	NCVT 1670 (Sitara-14)
17	NCVT 1617 (CIM-602 (S-1)	35	NCVT 1635 (Saim-32)	53	NCVT 1653 (Tahafuz-7)	71	NCVT 1671 (VH-363)
18	NCVT 1618 (NIAB-444)	36	NCVT 1636 (Tipu-2)	54	NCVT 1654 (Suncrop-4)	72	NCVT 1672 (GH-Mubarak)

Table 2. *List of markers with anealing temperature, band size, number of allele per locus and expected/ observed heterozygosity*

	Marker List	Sequence 5'….3'	Annealing temp. C°	Band size (bp)	Repeat motif	Expected heterozygosity	Observed heterozygosity	No. of alleles	PIC value
1	BNL 1551	F: CGCAAGCCACCTGTAAAAC R: TCGAATTTTCTCTCTCTCTCTCT	62	195	(AG)22	0.28	0	2	0.99
2	BNL 1053	F: AGGGTCTGTCATGGTTGGAG R: CATGCATGCGTACGTGTGTA	60	190	(AC)16	0.48	0	2	0.97
3	BNL1414	F: AAAAACCCCTTTCCATCCAT R: GGGTGTCCTTCCCAAAAATT	62	180	(AG)16	0.46	0	2	0.98
4	BNL 1672	F: TGGATTTGTCCCTCTGTGTG R: AACCAACTTTTCCAACACCG	62	200	(AG)14	0.49	0	2	0.97
5	BNL 2634	F: AACAACATTGAAAGTCGGGG R: CCCAGCTGCTTATTGGTTTC	62	200	(AG)11	0.43	0	2	0.97
6	BNL 3031	F: AGGCTGACCCTTTAAGGAGC R: AACCAACTTTTCCAACACCG	61	160	(AG)27	0.61	0	3	0.98
7	BNL 3089	F: TCGAACTTAACAAAAGAGAGAGAGA R: AAATCCGGATTCAGCAATACTT	62	150	(AG)10	0.62	0	3	0.97
8	BNL 3255	F: GACAGTCAAACAGAACAGATATGC R: TTACACGACTTGTTCCCACG	62	215	(GC)6AT(AC)14	0.6	0	4	0.99
9	BNL 3649	F: GCAAAAACGAGTTGACCCAT R: CCTGGTTTTCAAGCCTGTTC	55	200	(TC)20	0.59	0	3	0.99
10	BNL 4108	F: TCCACCATTCCCGTAAATGT R: TGGCCAAGTCATTAGGCTTT	55	150	(GA)31	0.61	0	3	0.99
11	CIR 122	F: AATGTGGGCTGATACG R: CAGACACAATCCACAAAG	43	-	(TG)16	0.59	0	1	1
12	CIR 133	F: TAGCCATTCTCACCCA R: AGGCAGTCAGAGTCAAAG	55	112	(TG)11	0.58	0	3	0.99
13	CIR 148	F: CTAATCTTTGGATTCTACCC R: TCCAAGCCCAGATAAGT	55	190	(TG)8	0.6	0	3	0.99
14	CIR 203	F: AGTTCAAGGGCACAAA R: ATCTCCAAGTCCCACC	41	200	(TG)11(N)1(TG)5	0.6	0	3	0.99
15	CIR 246	F: TTAGGGTTTAGTTGAATGG R: ATGAACACACGCACG	55	200	(TG)6	0.62	0	3	0.93
16	CIR 253	F: CCAACCAAGAAACCAG R: GTAAGCATGGGCATTT	43	205	(TC)15(N)8(AC)5(N)7(CA)8	0.48	0	2	0.99

	Marker List	Sequence 5'….3'	Annealing temp. C^o	Band size (bp)	Repeat motif	Expected heterozygosity	Observed heterozygosity	No. of alleles	PIC value
17	CIR 320	F: CCTCCATAAACCCTCTT R: TCACATACGAAGACAACC	45	200	(T)9(N)2(TG)8	0.56	0	3	0.99
18	CIR 364	F: ACTTTCATTTCGTGTGGT R: TGTTCATGCTTGTATCGT	45	100	(AC)10	0.59	0	3	0.70
19	CIR 376	F: ATACACAAGTCATTCACACA R: TGACTATGACACGAGTGG	49	200	(CA)15	0.62	0	3	0.97
20	CIR 393	F: GACCACACAGACAGACAA R: TCCACAACCAAACTAACA	58	200	(TG)7(T)9	0.62	0	3	0.84
21	JESPER 0065	F: CCACCCAATTTAAGAAGAAATTG R: GGTTAGTTGTATTAGGGTCGTTG	58	190	(GAA)25	0.62	0	3	0.92
22	JESPER 0092	F: GGGACCTCTATTGAATAGCTGGAG R: CTCTTGGCATCATTAGTTCCTGG	67	300	(GAA)23	0.1	0	2	0.99
23	JESPER 0101	F: CCAAGTCAAGGTGAGTTATATG R: GCTCTTTGTTACTGAAATGGG	57	190	(TA)3(GT)15	0.57	0	3	0.99
24	JESPER 0114	F: GATTTAAGGTCTTTGATCCG R: CAAGGGTTAGTAGGTGTGTATAC	51	105	(GT)12	0.62	0	3	0.98
25	JESPER -0135	F: CAAAACCATCATCACTCTCAAG R: CGAGAGCCCACTAACAGAAAAG	57	100	(CT)11	0.52	0	3	0.99
	JESPER 0152	F: GATGCACCAGATCCTTTTATTAG R: GGTACATCGGAATCACAGTG	59	250	(GAA)50	0.53	0	3	0.99
27	JESPER 0153	F: GATTACCTTCATAGGCCACTG R: GAAAACATGAGCATCCTGTG	57	150	(CTA)18	0.22	0	2	0.76
28	JESPER 0220	F: CGAGGAAGAAATGAGGTTGG R: CTAAGAACCAACATGTGAGACC	55	160	(GA)20	0.61	0	3	0.99
29	JESPER 0251	F: CAACTAGAATGATAAGACAC R: CTTTAAGTACGTATGCATC	49	100	(CA)15	0.61	0	3	0.80
30	JESPER 0300	F: CGCATCACAAACCAAACAC R: CGGAAAATGATGATGATGAAGAAG	51	210	(CTT)5(CAT)6	0.62	0	3	0.96

PIC denotes polymorphism information content

Experiment 1: Morphological characterization of cotton

Cotton seeds of each genotype were sown in the field using Randomized Complete Block Design (RCBD) with four replications under natural environmental conditions. Recommended cultural practices were adopted to grow the crop. Experimental layout was designed in four replications i.e. four rows for each line and in each row 20 plants were maintained. Plant to plant and row to row distances were maintained as 30 cm and 75 cm, respectively. Following traits were evaluated on 180 days after sowing. The data on plant height were recorded on five randomly selected plants using meter rod. Each plant was measured from soil to top of the main stem at crop maturity. Internodal distance was measured from one node to the other node on the same stem. Data on number of bolls per plant were recorded on five randomly selected plants by counting the number of bolls on each selected plant. Data on number of branches per plant were collected on five randomly selected plants by counting the number of branches of each selected plant. Data on boll weight were recorded on five randomly selected plants using weighing balance. Data on plant type were recorded whether they were conical or bushy type. Data on yield per plant were also recorded. The data recorded on each trait were subjected to analysis of variance (ANOVA) by using Minitab 16 software. Tukey's Test was applied to access the significance among the mean differences.

Experiment 2: Molecular characterization of cotton on the basis of SSR markers

Cotton seeds of each genotype were sown in the field of NIGAB at NARC, Islamabad using RCBD with three replications under natural environmental conditions. Recommended cultural practices were adopted to grow the crop. Young leaf tissue was collected from all the genotypes for DNA extraction. DNA extraction was carried out as per the procedure described by (Doyle et al., 1987). DNA was checked by measuring absorbance on nano-spectrophotometer. Purity was investigated by the absorbance ratios; 230/260 while concentration at 1 optical density. Polymerase chain reaction (PCR) amplification was performed in an automatic Applied Bio Systems Thermal Cycler (Veriti 96 well) at 94ºC for 5 min, afterward 35 cycles each containing of three phases: one phase of denaturation at 94ºC for 40 sec, one phase of annealing at specified temperature for each primer pair (*Table 2*) for 40 sec and an extension phase for 1 min at 72ºC followed by final extension of 10 min at 72ºC. On 1.5% agarose gel stained with ethidium bromide, amplified products was electrophoresed at 80 V for 45 min and then visualized by means of gel documentation system (Alpha Innotech).

DNA amplifications with each SSR primer were repeated three times to guarantee reproducibility. In three separate amplifications for each primer, bands were considered reproducible and scorable only after careful observation and comparison. Scoring was made on visualizing the bands. Each DNA fragment was transformed into binary character matrices i.e. 1 for presence, 0 for absence. The band size was determined using the software available with gel documentation system. The Minitab 16 was used to build dendrogram for cluster analysis. Sub population statistics, AMOVA, number of allele per locus and expected/observed heterozygosity was calculated using R language citation (locus_table, mlg, nei.dist and plot_poppr_msn) via poppr-package.

Shoukat et al.: Estimation of morphological and molecular diversity of seventy-two advanced Pakistani cotton genotypes using
simple sequence repeats
- 2049 -

Results

Experiment 1: Morphological characterization of cotton

Grand mean of each trait have been shown in *Table 3*. Highly significant results were observed at $P<0.05$ among 72 NCVT lines of cotton germplasm for five traits i.e. plant height (cm), internodal distance (cm), number of bolls, total number of branches and boll weight (*Table 4*). The NCVT 1624 was the tallest (138.43 cm), while genotype NCVT 1668 produced the shortest (44.09 cm) plants. Highly significant differences were also observed within 72 NCVT lines of cotton germplasm. The NCVT 1630 showed minimum internodal distance (1.4 cm), while genotype NCVT 1624 produced the maximum internodal distance (4.76 cm). Highly significant differences were also observed within 72 NCVT lines of cotton germplasm for this trait as depicted in *Table 4*. The NCVT 1624 showed maximum number of bolls (59) per plant, while genotype NCVT 1651 produced the minimum (2) number of bolls per plant. Non-significant differences were observed among seventy-two NCVT lines of cotton germplasm regarding their number of branches per plant (*Table 4*). However, NVCT 1634 showed maximum number of branches per plant (26), while genotype NCVT 1609 produced the minimum (4) number of branches per plant.

Table 3. *Grand means of six traits in 72 cotton varieties with standard error at 0.05 confidence level*

	Plant height (cm)	Internodal distance (cm)	No. of bolls per plant	Total no. of branches	Boll weight/5	Yield per plant (g)
Grand total of mean of all varieties	76.61[a]	2.88[d]	18.08[c]	13.62[c]	3.67[d]	64.95[b]
Alpha	0.05	0.05	0.05	0/05	0.05	0.05
SE	7.69	0.30	4.85	4.86	0.06	4.49

SE: Standard error; a, b, c and d: Grouping by lettering to each accession according to Tukey's test of significance

***Table 4.*** *ANOVA for five trait of Pakistani cotton germplasm at p = 0.05*

Source of variance	Degree of freedom	Plant height (cm)	Internodal distance	Number of branches per plant	No of boll per plant	Boll weight (g)
Replication	3	535.7	0.07	97.89	109.36	0.021
Genotypes	71**	1432.6**	1.68**	130.3[n.s]	286.9**	1.82**
Error	141	177.6	0.28	71.10	70.85	0.01

**= highly significant; [n.s] = non-significant

Experiment 2: Molecular characterization of cotton

A total of 30 primers consisting of di, tri, tetra and penta nucleotide repeat motifs were used for initial screening with 6 genotypes, namely NCVT 1601, NCVT 1602, NCVT 1603, NCVT 1604, NCVT 1605 and NCVT 1606 were used for PCR amplification using all the 30 primers. Out of these, one primer gave no amplification, while rest of twenty-nine primers gave clear band patterns, and subsequently used to analyze the entire set of 72 genotypes. Thirty simple sequence repeats (SSRs) belonging to BNL, CIR and JESPER series were used to analyze genetic diversity among 72 cotton varieties. Out of these, six markers i.e. BNL 1551, BNL 3089, BNL 1672, BNL

2634, BNL 3031 and BNL1414 contained (AG)n repeat motif, CIR 246, CIR 133, CIR 148 and CIR 122 primer contained (TG)n repeat motif, JESPER 0065, JESPER 0152 and JESPER 0092 primers contained (GAA)n repeat motif, CIR 376 and JESPER 0251 primer contained (CA)n repeat motif, BNL 1053 and CIR 364 primers contained (AC)n motif, BNL 4108 and JESPER 0220 primers contained (GA)n repeat motif and BNL 3649 primer contained (TC)n repeat motif,. Apart from these, JESPER 0153 primer contained (CTA)18, JESPER 0135 primer contained (CT)11, JESPER 0114 primer contained (GT)12, JESPER 0101 primer contained (TA)3(GT)15, CIR 393 primer contained (TG)7(T)9, CIR 320 primer contained (T)9(N)2(TG)8, CIR 253 primer contained (TC)15(N)8(AC)5(N) 7(CA)8, CIR 203 primer contained (TG)11(N)1(TG)5, BNL 3255 primer contained (GC)6AT(AC)14 repeat motifs (*Table 2*).
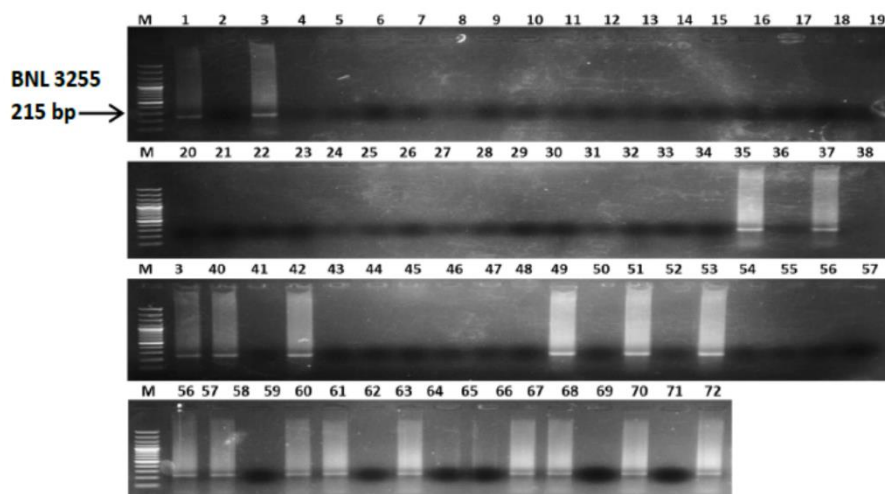

Figure 1. *BNL 3255 representative gel picture, 1-72 genotypes, 100bp ladder DNA, band size of 215 bp*

Out of thirty markers, twenty-nine markers were reproducible. These twenty-nine DNA markers amplified fragments of 100 to 300 bp in size through SSR profiling. On the basis of amplification percentage, markers were categorized into three groups. Sixteen markers i.e. BNL 1551, BNL 1414, BNL 3031, BNL 3255 (*Figure 1*), BNL 3649, BNL 4108, CIR 133, CIR 148, CIR 203, CIR 253, CIR320, JESPER 0092, JESPER 0101, JESPER 0135, JESPER 0152, JESPER 0220 gave low amplification (0-35%). Nine markers that are BNL 1053, BNL 1672, BNL 2634, BNL3089, CIR 246, CIR276, JESPER 0065, JESPER 0114, JESPER 0300 gave average amplification (36-60%). Four markers i.e. CIR 393, JESPER 0251, JESPER 0153, CIR 364 gave high amplification rate that was above 60% while one marker CIR 122 gave no result in any of the accession. The 72 advanced cotton lines were categorized into two populations 1 and 2 based on UPGMA. The population 1 contains 30 genotypes and population 2 was comprised of 42 genotypes.

The Nei's coefficient of similarity was used to construct the cluster diagram (*Figure 2*). All the genotypes were categorized into seven groups A, B, C, D, E, F, G based on Nei's coefficient. Group B was found to be the most genetically distinct/ dissimilar group when compared to other groups. However, group B have most of the candidates of under observed genotypes. Conversely group D, E and F, G have

similarity index of 1.3. The *Figure 3* shows significant population differentiation at all levels given that the observed data does not fall within the distribution expected from the permutation. The expected and observed heterozygosity is listed in *Table 2* along with the number of alleles per locus for each primer along with their polymorphism information content. The greatest number of allele observed was 4 and minimum number of allele per locus is 1 as depicted by population genetics. *Table 5* shows the variation within samples, between samples and population. Here, F value is -35.49, and p-value is very high in case of variation within samples. In other words, the variation within samples means and among different groups variation is much larger than the variation of allele presence within each group, at p-value is less than 0.01.
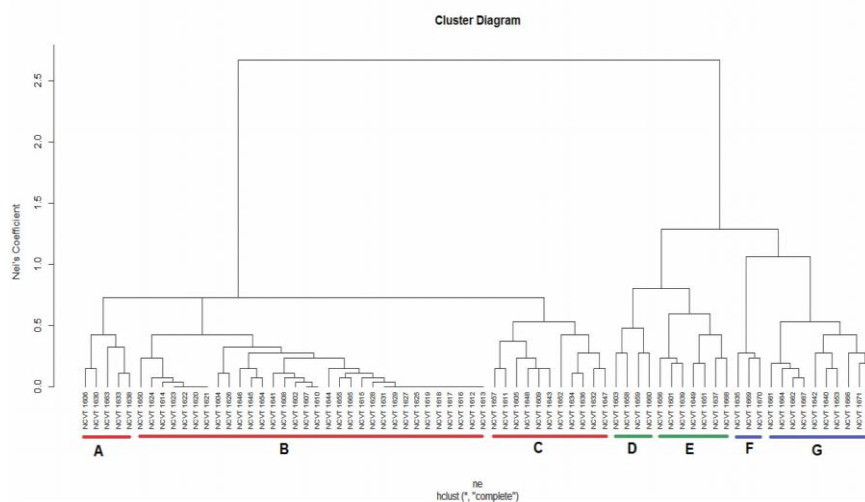
Figure 2. *Cluster dendrogram based on Nei's coefficient for two populations (Pop 1 indicates population 1 and Pop 2 indicates population 2)*
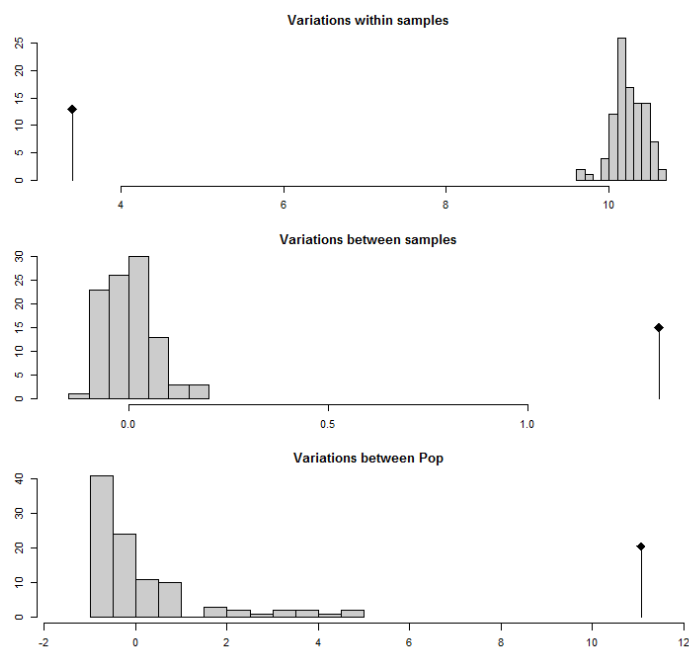

Figure 3. *AMOVA to check variations within samples, between samples and population*

Table 5. *AMOVA test to check variation within samples, between samples and population*

|	Test	Obs	Std. Obs	Alter	P-value
1	Variations within samples	3.411277	-35.494080	Less	0.01
2	Variations between samples	1.328667	21.379898	Greater	0.01
3	Variations between Pop	11.063446	8.363728	Greater	0.01

Population Genetics parameters i.e. MLG multilocus genotypes showed that the genotypic richness is indeed higher in population 2 than that of population 1 when considering equal sample sizes with the standard error of 1.76 in both poulation. Diversity measures incorporate both genotypic richness and abundance. There are three measures of genotypic diversity employed by poppr used for statistical analysis, the Shannon-Wiener index (H), Stoddart and Taylor's index (G), and Simpson's index (lambda). Graphical representation of two population with respect to H, G, Lambda and E.5 constants are depicted in *Figure 4*. In our study, comparing the diversity of pop1 to pop 2 shows that H is greater for pop 1 (3.40 vs. 2.80), but G is also greater (30.0 vs. 9.0). Thus, our expectation that diversity is lower for population 1 than that of population 2 is accepted in the case of H, which is likely due to the sensitivity of the Shannon-Wiener index to genotypic richness in the even sample sizes, and rejected in the case of G *Figure 4*.
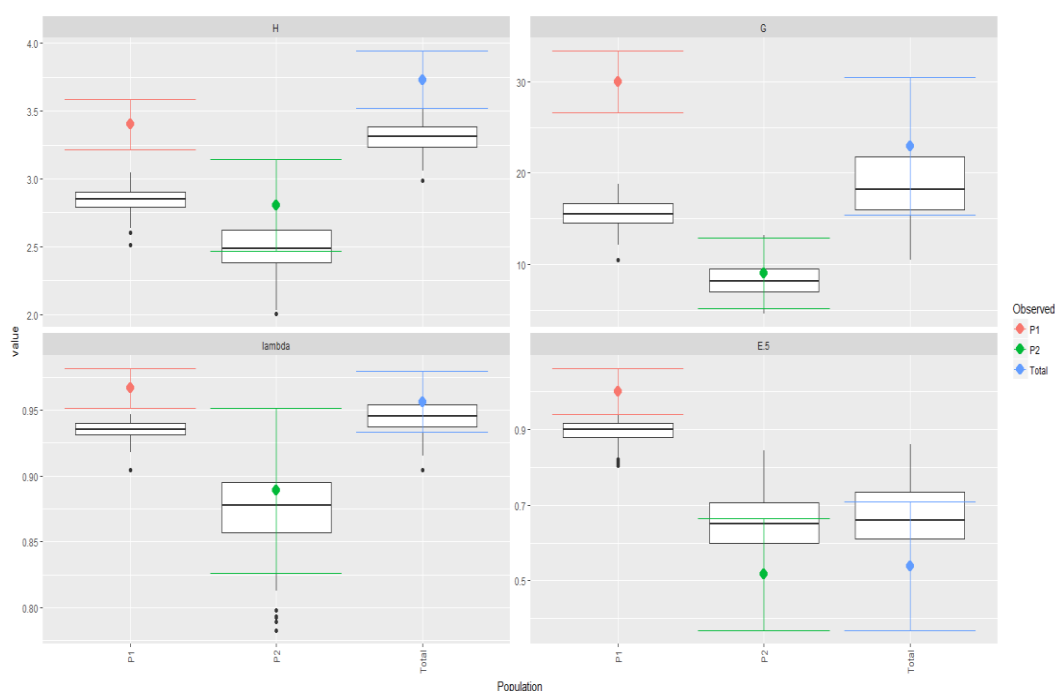
Figure 4. *Graphical representation of two populations with respect to H, G, Lambda and E.5 constants*

Evenness is a measure of the distribution of genotype abundances, wherein a population with equally abundant genotypes yields a value equal to 1 and a population dominated by a single genotype is closer to zero. In our study, the pop 1 has E.5 = 1.00 and the population 2 has E.5 = 0.516 (*Figure 4*). This indicates that the MLGs observed

in the population 1 are closer to equal abundance than those of population 2 (*Figure 4*). Indeed, when we look at a distribution of the MLGs for each population type it shows there are many more unique locus in population 1 as compared to the population 2 *Figure 5*.
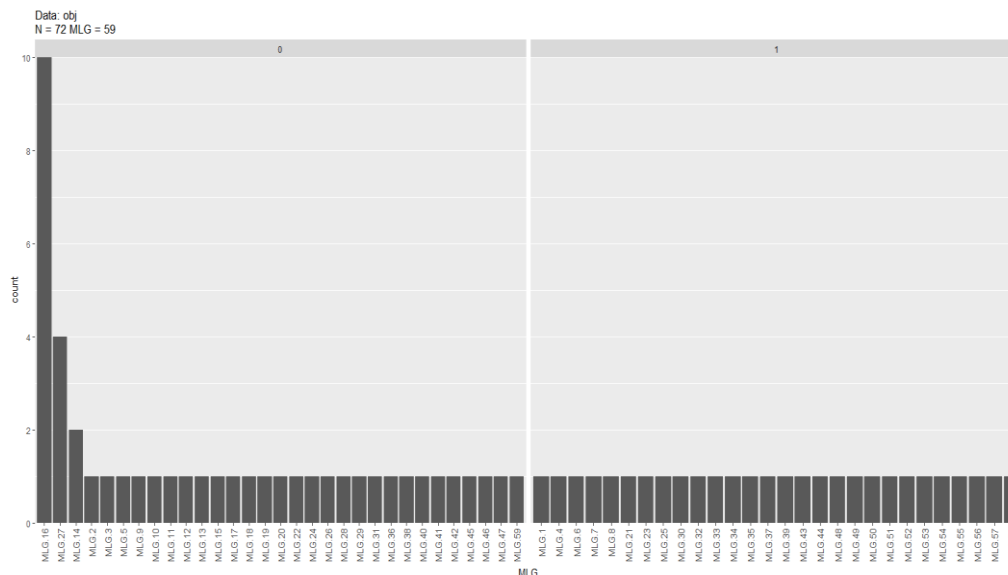
Figure 5. *Bootstrap analysis for MLG calculation against two populations*

Discussion

For characterization of any germplasm first step is to examine the material morphologically (Bidinger et al., 2009; Shen et al., 2017). All plant breeders still understand the value of enhancing plant genetic level by breeding available genetic variability. The results from different crops suggested that genetic diversity is very significant in breeding (Smartt, 1990; Reddy et al., 2005; Rasul et al., 2007). In this research the dendrogram clearly indicates two main clusters. The two groups A and B with the least value of 0% similarity were observed. The band sizes that obtained through SSR primer pairs were matched with the sizes of bands described in literature (Nguyen et al., 2004; Han et al., 2006). The scoring was based on bivariate as 1 is for presence and 0 is for absence. With respect to primer, the frequencies of scored bands or alleles were calculated (Upadhyaya et al., 2001).

Similar loci identified for every marker were assessed by scoring pattern. Twenty-nine (96.6%) out of the 30 SSR markers were evaluated and the results showed monomorphic fragments through the 72 varieties. Thus, the results are representing homogeneity of the cotton genome (Lacape et al., 2006). In sampled genome, SSR markers exhibited significant amount of variation on the basis of monomorphic bands as one genotype is giving the band, while the same band is absent in the other genotype (Fu, 2006). As far as polymorphism is concerned, no polymorphism was noticed among the experimental genotypes. This was expected from previous diversity analyses (Iqbal et al., 1997; Lacape et al., 2006). Irregularities among cultivar clustering on the basis of morphological and marker based dendrogram is just because of the markers that were not trait specific used for this study (Zhang et al., 2005). However, the SSR markers are depicting the genetic diversity to great extent. Inconsistencies between genotypes

clustering and known parentage should not be surprising for these cotton genotypes with such a narrow genetic base (Iqbal et al., 1997; Zhang et al., 2005).

The limited sample size of the genome of cotton exposed by only 29 SSR markers may subsidize to this type of variations (Fu, 2006). There is still need of using more mapped markers across the experimental genotypes that will progress the resolution to the genetic associations of these cotton varieties. On the other hand, the genetic relatedness estimated by this study still exhibit a beneficial direction for cotton breeding. This data is more helpful than that of other techniques used for parental selection and old-style pedigree analysis (Bowman et al., 1996). The values of similarity among cotton germplasm can change if more cotton varieties along with the experimental genotypes or varieties are evaluated or more markers applied. Also, the findings specify the distinctiveness, but not necessarily the relatedness of varieties (Fu, 2006). Regardless of some restrictions, the relative measure of genetic distinctiveness stated here offers useful information for selecting specific germplasm with diverse genetic background for a cotton-breeding program (Fu et al., 2007).

It has been observed that the current study is applicable in many ways. Like, considering polymorphism, the diversity among these cotton cultivars was less suggesting that the range of diversity for preservation of germplasm and breeding in future should be broadened. For selecting diverse plants for cotton breeding, the findings on genetic distinctiveness are beneficial. As this study exposed two major clusters of cotton cultivars and recognized the cotton cultivars which were genetically different. The classification of cotton varieties by the use of SSR primer pairs provides significant information to considerate the genetic diversity of top Pakistani cotton germplasm and to expand the cotton breeding program. It also gives a useful guide for selecting particular germplasm with diverse genetic background.

Conclusion

Relatively low genetic diversity has been revealed by this SSR analysis among experimental Pakistani cotton cultivars. Unique cotton cultivar was identified using the morphological analysis eagle 2 i.e. NCVT 1624. Molecular characterization depicted the homogeneity of the under study cotton genome as monomorphic bands were observed, while polymorphism was detected across the experimental genome via 29 markers. To conserve the elite cotton germplasm and developing future cotton breeding programs in Pakistan, these findings are useful. On the other hand, the genetic associations that have been estimated exhibit a beneficial direction for cotton breeding. It is recommended that more mapped markers across the genome should be used in future research that would improve the resolution to the genetic relationships as well as associations of the experimental cotton accessions. Regular assessment of National Cotton Varietal Trial should be carried out in order to update the diversity level of national cotton germplasm for cotton breeders.

REFERENCES

[1]   Ahmad, I. (2018): Future of Bt cotton in Asia. – Pakissan.com. [cited 17 April 2018]. Retrieved from http://pakissan.com/english/ advisory/biotechnology/ future.of.bt.cotton.in.asia.shtml.
[2]   Ashokkumar, K., Ravikesavan, R. (2011): Morphological diversity and per se

Shoukat et al.: Estimation of morphological and molecular diversity of seventy-two advanced Pakistani cotton genotypes using
simple sequence repeats
- 2055 -

performance in upland cotton (*Gossypium hirsutum* L.). – Journal of Agricultural Science 3(2): 107-113.

[3] Bidinger, F. R., Yadav, O. P. (2009): Biomass heterosis as the basis for grain and stover yield heterosis in arid zone pearl millet (*Pennisetum glaucum* (L.) R. Br.) hybrids. – Crop Science 49(1): 107-112.

[4] Bowman, D., May, O., Calhoun, D. (1996): Genetic base of upland cotton cultivars released between 1970 and 1990. – Crop Science 36(3): 577-581.

[5] Bukhari, S. (2014): Studies on genetic diversity of cotton using RAPD markers. – Pure and Applied Biology 3(3): 95-100.

[6] Chapagain, A., Hoekstra, A., Savenije, H., Gautam, R. (2006): The water footprint of cotton consumption: An assessment of the impact of worldwide consumption of cotton products on the water resources in the cotton producing countries. – Ecological Economics 60(1): 186-203.

[7] Dongre, A., Parkhi, V. (2005): Identification of cotton hybrid through the combination of PCR based RAPD, ISSR and microsatellite markers. – Journal of Plant Biochemistry and Biotechnology 14(1): 53-55.

[8] Doyle, J., Doyle, L. (1987): Isolation of plant DNA from fresh tissue. – Focus 12(1): 13-15.

[9] Economic Survey (2014-15): Retrieved from http://finance.gov.pk/survey/chapters_15/Highlights.pdf [Accessed 12 Jun. 2019].

[10] Ehsan, B., Haque, M., Younas, T., Shaheen, T., Huma, S., Sattar, S., Idrees, S., Iqbal, Z. (2013): Assessment of genomic diversity of cotton (*Gossypium hirsutum*) genotypes using simple sequence repeats markers through genetic analysis software. – International Journal of Agriculture & Biology 15(2): 968-972.

[11] Fu, Y. (2006): Redundancy and distinctness in flax germplasm as revealed by RAPD dissimilarity. – Plant Genetic Resources 4(2): 117-124.

[12] Fu, Y., Peterson, G., Morrison, M. (2007): Genetic diversity of Canadian soybean cultivars and exotic germplasm revealed by simple sequence repeat markers. – Crop Science 47(5): 1947-1954.

[13] Han, Z., Wang, C., Song, X., Guo, W., Gou, J., Li, C., Chen, X., Zhang, T. (2006): Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. – Theoretical and Applied Genetics 112(3): 430-439.

[14] Iqbal, M., Aziz, N., Saeed, N., Zafar, Y., Malik, K. (1997): Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. – Theoretical and Applied Genetics 94(1): 139-144.

[15] Lacape, J., Dessauw, D., Rajab, M., Noyer, J., Hau, B. (2006): Microsatellite diversity in tetraploid Gossypium germplasm: Assembling a highly informative genotyping set of cotton SSRs. – Molecular Breeding 19(1): 45-58.

[16] Li, Z., Wang, X., Zhang, Y., Zhang, G., Wu, L., Chi, J., Ma, Z. (2008): Assessment of genetic diversity in glandless cotton germplasm resources by using agronomic traits and molecular markers. – Frontiers of Agriculture in China 2(3): 245-252.

[17] Moiana, L. D., Vidigal Filho, P. S., Gonçalves-Vidigal, M. C., Lacanallo, G. F., Galvan, M. Z., de Carvalho, L. P., Maleia, M. P., Pacheco, C. M., Ribeiro, T., Neto, H. Z., Coimbra, G. K. (2012): Genetic diversity and population structure of cotton (*Gossypium hirsutum* L. race latifolium H.) using microsatellite markers. – African Journal of Biotechnology 11(54): 11640-11647.

[18] Nguyen, T., Giband, M., Brottier, P., Risterucci, A., Lacape, J. (2004): Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. – Theoretical and Applied Genetics 109(1): 167-175.

[19] Puspito, A., Rao, A., Hafeez, M., Iqbal, M., Bajwa, K., Ali, Q., Rashid, B., Abbas, M., Latif, A., Shahid, A., Nasir, I., Husnain, T. (2015): Transformation and evaluation of *Cry1Ac+Cry2A* and *GT* gene in *Gossypium hirsutum* L. – Frontiers in Plant Science, doi: 10.3389/fpls.2015.00943.

[20] Rana, M., Singh, S., Bhat, K. (2007): RAPD, STMS and ISSR markers for genetic diversity and hybrid seed purity testing in cotton. – Seed Science and Technology 35(3): 709-721

[21] Rasul, M., Hiramatsu, M., Okubo, H. (2007): Genetic relatedness (diversity) and cultivar identification by randomly amplified polymorphic DNA (RAPD) markers in teasle gourd (*Momordica dioica* Roxb.). – Scientia Horticulturae 111(3): 271-279.

[22] Reddy, L., Upadhyaya, H., Gowda, C., Singh, S. (2005): Development of core collection in pigeon pea [*Cajanus cajan* (L.) Millspaugh] using geographic and qualitative morphological descriptors. – Genetic Resources and Crop Evolution 52(8): 1049-1056.

[23] Shen, Z., Duan, J., Ma, L. (2017): Genetic diversity of *Xanthoceras sorbifolium* bunge germplasm using morphological traits and microsatellite molecular markers. – PLOS ONE 12(6): e0177577. https://doi.org/10.1371/journal. pone.0177577.

[24] Smartt, J. (1990): Grain legumes: evolution and genetic resources. – Cambridge University Press, p. 333.

[25] Ullah, I., Iram, A., Iqbal, M., Nawaz, M., Hasni, S., Jamil, S. (2012): Genetic diversity analysis of Bt cotton genotypes in Pakistan using simple sequence repeat markers. – Genetics and Molecular Research 11(1): 597-605.

[26] Upadhyaya, H., Ortiz, R. (2001): A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. – Theoretical and Applied Genetics 102(8): 1292-1298.

[27] Van Becelaere, G., Lubbers, E., Paterson, A., Chee, P. (2005): Pedigree- vs. DNA marker-based genetic similarity estimates in cotton. – Crop Science 45(6): 2281-2287.

[28] Varshney, R., Graner, A., Sorrells, M. (2005): Genic microsatellite markers in plants: features and applications. – Trends in Biotechnology 23(1): 48-55.

[29] Wu, Y., Machado, A., White, R., Llewellyn, D., Dennis, E. (2006): Expression profiling identifies genes expressed early during lint fiber initiation in cotton. – Plant and Cell Physiology 47(1): 107-127.

[30] Zhang, J., Lu, Y., Cantrell, R., Hughs, E. (2005): Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the Southwestern USA. – Crop Science 45(4): 1483-1490.