CBCS
http://agrobiol.sggw.waw.pl/cbcs

Communications in Biometry and Crop Science
Vol. 4, No. 1, 2009, pp. 21–30

International Journal of the Faculty of Agriculture and Biology,
Warsaw University of Life Sciences, Poland

REGULAR ARTICLE

Genetic diversity of Pakistani cotton cultivars as revealed by simple sequence repeat markers

Azeem I. Khan[1*], Yong-Bi Fu[2], Iftikhar A. Khan[1]

[1] Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture Faisalabad, Pakistan.
[2] Plant Gene Resources of Canada, Saskatoon Research Centre, Agriculture and Agri- Food Canada, 107 Science Place, Saskatoon, SK, Canada S7N 0X2.
* Corresponding author: Azeem I. Khan, E-mail: ranaazeemiqbal@yahoo.com

CITATION: Khan, A.I., Fu, Y.B., Khan, I.A. (2009). Genetic diversity of Pakistani cotton cultivars as revealed by simple sequence repeat markers. *Communications in Biometry and Crop Science* 4 (1), 21–30.

Received: 19 November 2008, Accepted: 28 February 2009, Published online: 11 March 2009

ABSTRACT

Cotton (*Gossypium hirsutum* L.) is the most important cash crop of Pakistan. More than 80 cotton cultivars have been developed since 1914 through conventional breeding, but no comprehensive study has been made on the genetic diversity of these cultivars. Simple sequence repeat (SSR) markers were used to assess the genetic diversity of 40 representative cotton cultivars released from 1914 to 2005. Thirty-four of the 57 SSR primer pairs screened displayed polymorphism and 122 of the 204 SSR bands detected by these polymorphic primer pairs were polymorphic across the cultivars. The frequencies of these polymorphic bands ranged from 0.02 to 0.98 and averaged 0.27. Analysis of molecular variance revealed 12.4% of the total SSR variation residing among the cultivars over five specified breeding periods. The cotton cultivars released after 2000 displayed slightly more SSR variation than those released earlier. Clustering 40 cultivars resulted in three major clusters mixed with cultivars released from various breeding periods at different research stations. The average dissimilarity (AD) of a cultivar ranged from 0.191 to 0.365 with the mean AD of 0.248, and genetically distinct cultivars were identified. These results are useful for conserving elite cotton germplasm and developing future cotton breeding programs in Pakistan.

Key Words: *Gossypium hirsutum; genetic diversity; genetic relationship; genetic distinctiveness; Simple Sequence Repeat (SSR) marker.*

INTRODUCTION

Cotton, *Gossypium hirsutum* L, is an important crop in Asia, particularly in the sub-continent (Hussain et al., 2005). In Pakistan, it is the most important cash crop, plays an

important role in the economy through export, and provides livelihood to millions from its sowing through processing and fabrication to its consumption (Federal Bureau of Statistics, 2004). Thus, cotton has always received priority for research and development in Pakistan. Formal cotton breeding was started with the introduction of the American upland cotton. The first cotton cultivar (4F) was released in 1914 for commercial cultivation by Department of Agriculture, Lyallpur (Faisalabad). Since then, several breeding programs have been established across the country, but the breeding focus remained on the improvement of yield, resistance against disease and insect pests, and fibre quality traits with specific objectives to meet the requirements of national and international textile industry. More than 80 cultivars have been released for different cotton growing areas of Punjab and Sindh provinces (Hussain et al., 2005). Cotton production gradually increased to 2.8 million bales in 1991 (MINFAL, 1992) but decreased significantly in the next couple of years with the infestation of cotton leaf curl virus (CLCuV). The incidence and devastating effect of CLCuV may be partially due to the narrow genetic base of the prevailing cotton cultivars for the disease resistance. Thus, understanding the genetic diversity of the Pakistani cotton cultivars or germplasm is of great importance for the future cotton breeding programmes (Iqbal et al., 1997 and 2001) and a comprehensive diversity assessment of Pakistani cotton cultivars is warranted.

Characterization of plant germplasm using molecular techniques has an important role in the management and utilization of plant genetic resources (Karp, 2002). It can also enhance plant breeding in selection of diverse parents to widen the breeding gene pool (Fu, 2006). Efforts have been made to characterize cotton germplasm using allozymes (Wendel et al., 1992), restriction fragment length polymorphism (RFLP) (Wendel and Brubaker, 1993), random amplified polymorphic DNA (RAPD) (e.g., see Multani and Lyon, 1995; Iqbal et al., 1997), amplified fragment length polymorphism (AFLP) (e.g., see Iqbal et al., 2001; Rana et al., 2005) and simple sequence repeat (SSR) markers (Liu et al., 2000; Reddy et al., 2001; Zhang et al., 2005; Lacape et al., 2007). These characterizations have provided useful information for understanding the genetic diversity and structure of various cotton gene pools found in different geographic regions. This information has been incorporated into effective management of cotton germplasm in some cotton breeding programs for control of genetic diversity. In general, low levels of genetic diversity have been found in modern cotton cultivars, which is consistent with the hypothesized narrow genetic base of upland cotton germplasm used in breeding (Meredith, 2000).

The recent development of abundant cotton SSR markers has stimulated more effort in molecular characterization of cotton germplasm released from specific cotton breeding programs across the world (Blenda et al., 2006; Zhang et al., 2008), but no SSR analysis of Pakistani cotton genotypes has been made (Iqbal et al., 1997 and 2001). The objective of this study was to investigate genetic diversity of 40 representative Pakistani cotton cultivars released from 1914 to 2005 using SSR markers.

MATERIALS AND METHODS

PLANT MATERIALS

Forty commercial cotton cultivars released from different breeding stations in Pakistan from 1914 to 2005 were selected for this study (Table 1). The seed of the selected cultivars were collected from Central Cotton Research Institute Multan, Pakistan. To assess genetic diversity changes over the long-term breeding effort, these Pakistani cotton cultivars were classified into five groups based on year of release (period 1 from 1914 to 1934; period 2 from 1945 to 1959; period 3 from 1970 to 1980; period 4 from 1983 to 2000; and period 5 from 2001 to 2005).

Table 1. Forty Pakistani cotton cultivars with origin of release, parentage, year of release (YOR), derived cluster and average dissimilarity (AD).

Cultivar a	Origin of release b	Parentage	YOR	Cluster	AD
4-F (P1)	Dept of Agric, Lyallpur	Selection from stray plants of American cotton	1914	III	0.263
289-F (P1)	Dept of Agric, Lyallpur	4F selection	1921	III	0.278
289-F/K25 (P1)	Dept of Agric, Lyallpur	289 Bulk selection	1930	I	0.322
LSS (P1)	CRS, AARI Faisalabad	4F selection, a single off-type plant: a natural hybrid found in the field of 4F cotton	1934	II	0.207
124-F (P2)	CRS, AARI Faisalabad	289F/43 selection (1936) and bulked as "Victory"	1945	III	0.241
199-F (P2)	CRS, AARI Faisalabad	4F-98 (material from sakrand) selection, natural hybrid, an off-type plant in 4F-98 cotton field at Multan	1946	III	0.270
238-F (P2)	CRS, AARI Faisalabad	28-F/ 43 selection	1948	I	0.365
268-F (P2)	CRS, AARI Faisalabad	LSS selection	1948	II	0.197
AC-134 (P2)	CRS, AARI Faisalabad	148-F x 199-F	1959	II	0.214
LASANI-11 (P2)	CRS, AARI Faisalabad	181 x F selection	1959	III	0.279
MS-39 (P3)	CRS, Multan	Natural hybrid in L-11 (Lasani 11)	1970	II	0.194
MS-40 (P3)	CRS, Multan	A single variant selection from AC 252	1970	II	0.232
149-F (P3)	CRS, Multan	124-F x Babdal	1971	III	0.225
B-557 (P3)	CRS, AARI, Faisalabad	208-F x L-5	1975	III	0.256
MNH-93 (P3)	CRS, Multan	389/65 x C 158	1980	III	0.224
NIAB-78 (P4)	NIAB, Faisalabad	DPL-16 x AC-134 F1 Irradiate-30 Kr Gamma rays (60 Co)	1983	III	0.262
S-12 (P4)	CRS, Multan	MNH-93 x 7203-14-4-Arizona	1988	II	0.224
RH-1 (P4)	CRS, Rahim yar khan	LH-62 x W-1104	1990	I	0.278
CIM-109 (P4)	CCRI, Multan	NIAB-78 x A 89/ FM	1990	II	0.216
M-4 (P4)	CRS, Tandojam	289F selection	1992	II	0.220
CIM-240 (P4)	CCRI, Multan	CIM-70 x W-1106	1992	I	0.241
BH-36 (P4)	CRS, Bahawalpur	BS 1 x TX Bonhem-76 C	1992	II	0.223
S-14 (P4)	CRS, Multan	H 2102-1/ 83 x H 2105-1/83	1995	I	0.287
KARISHMA (P4)	NIAB, Faisalabad	NIAB-86 x W-83-29-Mex	1996	III	0.191
CIM-448 (P4)	CCRI, Multan	492/87 x Cp 15/2	1996	I	0.262
CIM-1100 (P4)	CCRI, Multan	492/87 x Cp 15/2	1996	I	0.265
CIM-443 (P4)	CCRI, Multan	CIM-109 x CRA-5166	1998	II	0.277
CIM-446 (P4)	CCRI, Multan	CP 15/2 x 512	1998	II	0.269
MNH-554 (P4)	CRS, Multan	MNH-439 x LRA-5166	2000	II	0.198
BH-118 (P4)	CRS, Bahawalpur	BS-48 x 829-4/90	2000	II	0.199
CIM-482 (P4)	CCRI, Multan	CIM-229 x CP 15/2	2000	II	0.203
CRIS-134 (P5)	CCRI, Sakrand	NIAB-78 x DPL (1988)	2001	II	0.216
CIM-473 (P5)	CCRI, Multan	CIM-402 x CRA 5/66	2002	II	0.245
CIM-506 (P5)	CCRI, Multan	CIM-360 x CP 15/2 (1993)	2002	II	0.235
CIM-499 (P5)	CCRI, Multan	CIM-443 x 755-6/93 (1992)	2002	II	0.240
NIAB-111 (P5)	NIAB, Faisalabad	(NIAB-313/12 x CIM 100) F1 300 Gy	2004	II	0.236
CIM-707 (P5)	CCRI, Multan	CIM-243 x 738-6/93 (1993)	2004	II	0.299
BH-160 (P5)	CRS, Bahawalpur	Cedi x FDW 496 x 673/93 (1994)	2004	I	0.319
FH-1000 (P5)	CRS, AARI, Faisalabad	SIZ x CIM 448 (1994)	2004	II	0.300
CIM-496 (P5)	CCRI, Multan	CIM-425 x 755-6/93 (1993)	2005	II	0.245

[a] Five breeding periods (P1-P5) given in parenthesis following cultivars were defined in the text. [b] CRS=Cotton Research Station; AARI=Ayub Agricultural Research Institute; CCRI=Central Cotton Research Institute; NIAB=Nuclear Institute for Agriculture and Biology.

DNA EXTRACTION AND SSR ANALYSIS

Seeds of each cultivar were sown in a greenhouse at the Saskatoon Research Centre, Agriculture and Agri-Food, Canada. Fresh leaf tissues 100-120 g were collected from the growing seedlings of each cultivar, bulked, freeze-dried (in a Labconco Freeze Dry System for 3-5 days), and stored at –80°C. The procedures for DNA extraction and quantification for each bulked sample are described in Fu et al. (2007).

Based on reported polymorphism and genomic coverage, a set of 57 cotton microsatellite primers (Table 2) were selected: 40 CIR genomic and 17 NAU EST-derived primer sets of Nguyen et al. (2004) and Han et al. (2006), respectively. These primers were custom synthesized by Integrated DNA Technologies, Coralville, IA, USA. The polymerase chain reactions (PCRs) for CIR primer sets contained 25 ng DNA as template, 0.5 U *taq* polymerase (New England Biolabs), 2.5 mM $MgCl_2$, 0.2 mM dNTPs and 0.6 mM of each primer, 1X PCR buffer and final volume made up to 25µl with double distilled deionized water. PCR amplification profile consisted of denaturation at 94˚C for 3 min and 33 cycles of 94˚C for 10 s, 51˚C for 20 s, 72˚C for 40 s and then final extension at 72˚C for 10 min. For NAU Primer sets, the PCR reaction differed only in concentrations of $MgCl_2$ (2.0 mM) and each primer (0.2mM), leading to final reaction volume of 20 µl, made up with double distilled deionized water. PCR amplification profile consisted of denaturation at 94 ˚C for 3 min and 33 cycles of 94˚C for 10 s, 50˚C for 20 s (step down to 45˚C in 4 cycles) then 45˚C for 20 s and 72˚C for 1 min and then final extension at 72˚C for 10 min. The PCR procedures, including separation and visualization of PCR products, are described in Fu et al. (2007).

DATA ANALYSIS

To generate a set of SSR data for each cultivar, DNA fragments amplified by SSR primer pairs were identified based on their sizes in base pairs measured with a 50 and 10-bp DNA ladder (Invitrogen, Carlsbad, CA), compared with the fragment sizes reported in the literature (Nguyen et al., 2004; Han et al., 2006), and scored as presence (1) or absence (0). Frequencies of the scored alleles were calculated with respect to primer and breeding period. The likely number of loci (or likely loci) detected for each primer pair were determined by the observed allelic pattern and the possible number of alleles per single tetraploid plant. The polymorphic information content was calculated for each primer pair, as described in Roussel et al. (2004) with adjustment to the number of likely loci, to assess the informativeness of each marker.

To assess the genetic relationships of the cotton cultivars, the similarity matrix of pairwise cultivars was calculated using simple matching coefficient (Sokal and Michener, 1958) and clustered using NTSYS-PC 2.01 (Rohlf, 1997) with the algorithm of unweighted pair-group methods using arithmetic averages (UPGMA). A principal component analysis of 40 cotton cultivars was also conducted using NTSYS-PC 2.01 (Rohlf, 1997), and plots of the first three resulting principal components were made to assess the genetic associations of the cotton cultivars. To assess the genetic distinctiveness of the cotton cultivars, the dissimilarities of each cultivar with the remaining cultivars assayed were calculated using the simple matching coefficient, as described in Fu (2006).

An analysis of molecular variance was also performed using Arlequin version 3.0 (Excoffier et al., 2005) to assess the genetic structure of the cotton cultivars. Two models of genetic structuring were examined for five breeding periods and three derived clusters of cotton cultivars. Significance of resulting variance components and inter-group genetic distances was tested with 10,100 random permutations.

Table 2. SSR polymorphism revealed by 34 SSR primer pairs.

Primers[a]	Linkage group or Chromosome label[a]	Number of bands[b]		Size range (bp)	Likely loci[c]	PIC[d]
		Mono	Poly			
CIR027	A01,c4	5	2	190-400	1	0.51
CIR080	c20	3	3	290-310	1	0.19
CIR277	c18,D08	5	2	210-300	3	0.88
CIR291	A01,A02,c4	3	1	380-460	2	0.18
CIR061	D02,D03	4	4	111-400	3	0.82
CIR165	D08	8	1	210-460	1	1.00
CIR207	A03	8	2	240-425	1	0.52
CIR216	c18	2	2	146-150	1	0.40
CIR224	c5,c22,D08	10	6	150-330	1	0.40
CIR234	c15,unl-D03	10	7	210-350	3	0.50
CIR280	c5,c25	9	6	212-305	3	0.61
CIR305	c10,c20	5	2	156-210	1	0.52
CIR329	unknown	9	5	210-500	2	0.26
CIR340	c20	11	3	190-385	1	0.06
CIR372	c10	6	3	157-194	1	0.32
CIR376	A02,c2,c5	11	8	145-206	3	0.45
CIR410	D02	4	2	240-340	1	0.14
CIR009	c1,c15	9	8	216-320	4	0.62
CIR032	c26	8	7	142-198	3	0.62
CIR156	D02	5	4	120-138	3	0.69
CIR166	c20	4	4	118-134	2	0.74
CIR167	c26	4	2	210-260	1	0.10
CIR222	c4,D08	6	2	280-340	1	0.10
CIR237	A02	7	4	94-160	3	0.57
CIR294	c5,c22	6	1	197-420	1	0.05
NAU2017	c20	5	4	165-225	2	0.58
NAU2082	c10	2	1	190-200	1	0.91
NAU2083	c1	6	4	158-250	1	0.38
NAU2096	c12	4	1	200-440	1	0.19
NAU2108	c7	5	2	218-290	1	0.10
NAU2110	D02	8	8	210-370	4	0.05
NAU2138	c18	3	3	600-800	2	0.36
NAU2139	unknown	3	2	168-196	1	0.98
NAU2152	A03,c16,c17	6	6	201-320	3	0.72
Total or mean		204	122		63	0.46

[a] Primer information and linkage group or chromosome label can be obtained from Nguyen et al. (2004) for CIR primer sets and from Han et al. (2006) for NAU primer sets.
[b] The number of bands observed; mono=monomorphic band; and poly=polymorphic band.
[c] The number of likely loci was determined by the observed band pattern and the possible number of alleles per single tetraploid plant.
[d] PIC=polymorphic information content.

RESULTS AND DISCUSSION

SSR VARIATION

Twenty-three (40%) of the 57 SSR primer pairs screened revealed monomorphic bands across the 40 cultivars, indicating substantial homogeneity of the cotton genome. The other 34 SSR primer pairs detected a total of 63 likely loci on possible 6 homologous linkage groups and/or 15 chromosomes of cotton (Table 2; Nguyen et al., 2004; Han et al., 2006). The 16 likely loci detected by EST-derived SSR markers should represent transcribed chromosomal regions, while the other 47 likely loci detected by genomic SSR markers may sample non-coding chromosomal segments (Table 2). A total of 204 SSR alleles were detected, but only 122 (60%) alleles were polymorphic across the 40 cultivars (Table 2). These polymorphic alleles could include some null alleles, because it was difficult to separate non-amplification due to experimental errors from null alleles. The number of polymorphic alleles detected per primer pair ranged from one to eight with an average of 3.6 alleles per primer pair. Values of polymorphic information content (PIC) for each marker ranged from 0.05 to 1.00 with an average of 0.46 (Table 2). Note that these PIC values could be biased by the determination of likely loci, as this variation was not significantly (P = 0.69) associated with the number of polymorphic alleles detected at each primer pair. Also, the use of bulk sampling might have introduced some bias of sampling SSR variation (Fu, 2000). Clearly, the SSR markers revealed considerable amount of variation in the sampled genome, although the overall polymorphism detected for these cotton cultivars was relatively low. This is expected, as mentioned above, from previous diversity analyses (e.g., Iqbal et al., 2001; Lacape et al., 2007).

The observed frequencies of the 122 alleles ranged from 0.02 to 0.98 with an average of 0.47. There were 26 alleles with frequencies of 0.9 or larger in the assayed cultivars and 59 with frequencies greater than 0.47, while 48 alleles were detected with frequencies less than 0.2, and 15 alleles with frequencies equal to or less than 0.03. Some of the rare alleles may be useful as diagnostic markers for some of the assayed cotton cultivars.

GENETIC RELATIONSHIPS OF COTTON CULTIVARS

Clustering of the 40 cultivars based on SSR similarity revealed several variation patterns (Figure 1). First, there were three major clusters of cultivars at the similarity level of 0.75 or higher. Cluster I consisted of eight cultivars, Cluster II 22 cultivars, and Cluster III 10 cultivars (Table 1). It appears that Cluster 1 was less similar to (or more distinct from) the other two clusters. Second, clustering was not associated with the periods of cultivar release, as each cluster consists of cultivars released from different breeding periods. Third, cultivars were not clustered according to the breeding station, since cultivars released from different breeding stations resided in various clusters. Fourth, cultivars were not fully clustered according to their parentages. For example, two cultivars CIM-448 and CIM-1100 from Central Cotton Research Institute at Multan shared the same parents and were grouped in Cluster I. Two sister cultivars CIM-109 and CRIS-134 sharing parent NIAB-78 were in Cluster II, not in Cluster III with NIAB-78. The cultivar M-4 released from Cotton Research Station at Tandojam was derived from 289F, but grouped in Clusters II and III, respectively.

Inconsistencies between cultivar clustering and known parentage should not be surprised for these cotton cultivars with such a narrow genetic base (Meredith, 2000; Iqbal et al., 2001; Zhang et al., 2005). Also, the limited sampling of the cotton genome revealed by only 34 SSR primer pairs may contribute to such inconsistencies. Applications of more mapped markers across the genome would improve the resolution to the genetic relationships of these cotton cultivars. However, the estimated genetic relationships still offers a useful guide for cotton breeding, as they are more informative than parental selection and traditional pedigree analysis (Bowman et al., 1996).

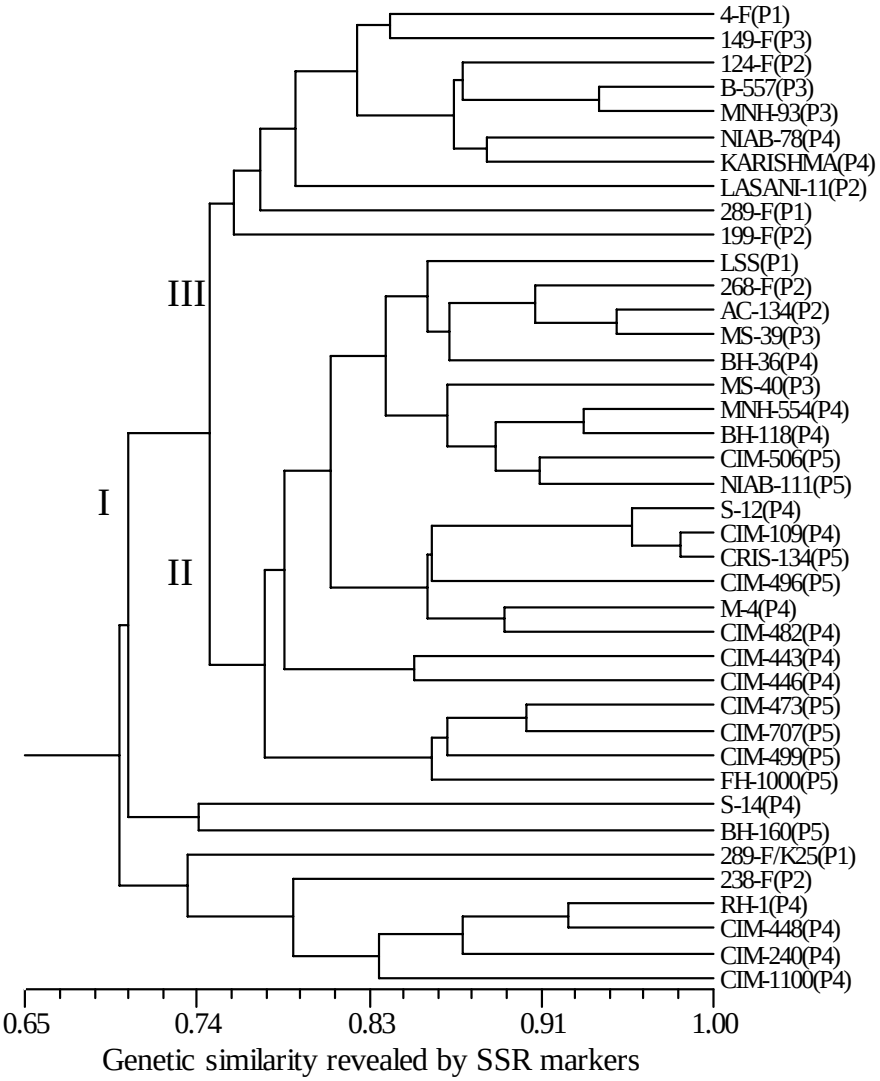

Figure 1. Genetic relationships of 40 Pakistani cotton cultivars as revealed by 122 SSR markers. The breeding period of the cultivar is indicated in parenthesis along with cluster label.

GENETIC DISTINCTIVENESS OF COTTON CULTIVARS

The genetic distinctiveness of the cotton cultivars was measured by the average dissimilarity (AD) of the cultivars against the remaining cultivars assayed. Higher AD value suggests greater distinctiveness of the genetic background for the cultivar. The average distinctiveness of a cultivar ranged from 0.191 for KARISHMA (developed from Nuclear Institute for Agriculture and Biology at Faisalabad) to 0.365 for 238-F (developed at Cotton Research Institute, Faisalabad). The mean value of the AD was 0.248 (Table 1). The genetically most distinctive 10 cotton cultivars are 238-F, 289-F/K25, BH-160, FH-1000, CIM-707, S-14, LASANI-11, 289-F, RH-1, and CIM-443. The least distinctive 10 cultivars are KARISHMA, MS-39, 268-F, MNH-554, BH-118, CIM-482, LSS, AC-134, CRIS-134, and CIM-109. Comparing these two groups of cultivars revealed that the genetic distinctiveness of a cultivar is not associated with the period and station of the cultivar release. Regression of the AD values over the year of release found no statistically significant relationship (results not shown). This may reflect the consequence of the long-term breeding effort in the introgressions of diverse germplasm into the breeding programs.

The values of AD as shown in Table 1 can change if more cotton cultivars are evaluated. Also, the measure indicates the distinctiveness, but not necessarily the relatedness, of

cultivars (Fu, 2006). For example, two closely related cultivars that were quite distinct from the remaining cultivars could have similar, higher levels of AD than the others but both cultivars would have been identified as genetically distinct. In spite of these limitations, the relative measure of genetic distinctiveness reported here provides useful information for selecting specific germplasms with distinct genetic background for a cotton-breeding program (Fu et al., 2007).

GENETIC STRUCTURES OF COTTON CULTIVARS

Analysis of molecular variation with respect to breeding period revealed that 12.4% variation was detected among cultivars of the five specified breeding periods while 87.6% variation was detected for cultivars within a specific breeding period (Table 3). Such low among-cultivar variation was not expected with a predominantly self-pollinating crop such as cotton, but might reflect the narrow genetic base of these released cultivars. Also, cultivars released after 2000 had slightly more SSR variation than those released earlier. Analysis of molecular variation with respect to derived cluster revealed only 3.4% of the total SSR variation residing among three major clusters described above (Table 3). Cultivars in Cluster III appear to have more variation than those in the other two clusters. These observations were supported by the assessment of the genetic associations of these cotton cultivars using principal coordinate analysis (Figure 2). The two components accounted for 34.7% and 17.1% of the total SSR variation, respectively. The cultivars released after 1981 were widely spread over the plot. The cultivars released after 2000 appear to have genetic shift from the cultivars released earlier. Also, the pattern of cultivar spreading over the plot is consistent with the finding of three major clusters for these 40 cultivars (Figure 1).

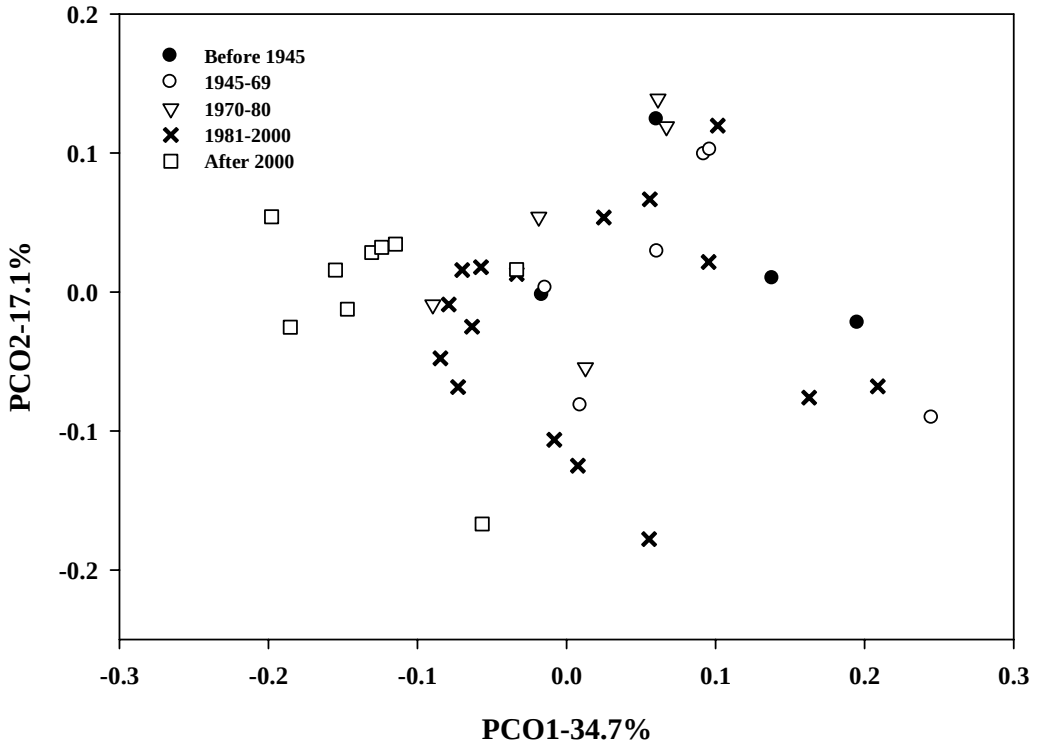

Figure 2. Genetic associations of 40 Pakistani cotton cultivars as revealed by plot of the first two principal component scores based on the Euclidean distances converted from the simple matching coefficient matrix of 122 SSR alleles for 40 cotton cultivars. These two components accounted for 34.7% and 17.1% of the total SSR variation, respectively.

IMPLICATIONS FOR COTTON GERMPLASM CONSERVATION AND BREEDING

The findings of this study have several practical implications. First, the overall genetic diversity residing in these cotton cultivars was low, implying the need for contined effort to widen the diversity range both for cotton germplasm conservation and future breeding. Second, the cultivars released after 2000 displayed slightly more SSR variation than those released earlier, suggesting that genetic diversity has been maintained in Parkistan's long-term cotton breeding programs. Third, this study revealed three major clusters of cotton cultivars and identified the genetically most distinctive cotton cultivars. The findings on genetic relationship and distinctiveness are useful for parental selection of diverse plants for cotton breeding. Fourth, the characterization of cotton cultivars using SSR markers generated not only essential information for understanding genetic diversity of elite Pakistani cotton germplasm, but also provided a useful guide for selecting specific germplasm with distinct genetic background for diversifying cotton breeding program.

CONCLUSIONS

This SSR analysis revealed that the genetic diversity of the Pakistani cotton cultivars released since 1914 was relatively low, but has been maintained over the long-term breeding. The analysis also generated information on genetic relationships and identified genetically unique cotton cultivars. These findings are useful for conserving elite cotton germplasm and developing future cotton breeding programs in Pakistan.

ACKNOWLEDGMENTS

The authors thank Mr. Muhammad Arshad, Director of Central Cotton Research Institute, Multan, for providing seeds of cotton cultivars; Mr. Gregory W. Peterson for his technical assistance and data proofing; and two anonymous reviewers for their helpful comments on the early version of the manuscript. The research was financially supported in part by the Higher Education Commission, Pakistan.

REFERENCES

Blenda, A., Scheffler, J., Scheffler, B., Palmer, M., et al. (2006). CMD: a cotton microsatellite database resource for *Gossypium* genomics. *BMC Genomics* 7, 132.

Bowman, D.T., May, O.L., Calhoun, D.S. (1996). Genetic base of upland cotton cultivars released between 1970 and 1990. *Crop Science* 36, 577–581.

Excoffier, L., Laval, G., Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1, 47–50.

Federal Bureau of Statistics. (2004). *Economic Survey of Pakistan (2003-04)*. Government of Pakistan.

Fu, Y.B. (2000). Effectiveness of bulking procedures in measuring population-pairwise similarity with dominant and co-dominant genetic markers. *Theoretical and Applied Genetics* 100, 1284–1289.

Fu, Y.B. (2006). Genetics redundancy and distinctness of flax germplasm as revealed by RAPD dissimilarity. *Plant Genetic Resource*s 4, 177–124.

Fu, Y.B., Peterson, G.W., Morrison, M.J. (2007). Genetic diversity of Canadian soybean cultivars and exotic germplasm revealed by simple sequence repeat markers. *Crop Science* 47, 1947–1954.

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, 430–439.

Hussain, A., Ibrahim, M., Tajammal, M.A., Naz, M.A. (2005). *Cotton Varieties of Pakistan*. Federal Seed Certification and Registration Department. Ministry of Food, Agriculture & Livestock (MINFAL), Government of Pakistan.

Iqbal, M.J., Aziz, N., Saeed, N.A., Zafar, Y., Mailk, K.A. (1997). Genetic diversity of some elite cotton varieties by RAPD analysis. *Theoretical and Applied Genetics* 94, 139–144.

Iqbal, M.J., Reddy, O.U.K., El-Zak, K.M., Pepper, A.E. (2001). A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theoretical and Applied Genetics* 103, 547–554.

Karp, A. (2002). The new genetic era: will it help us in managing genetic diversity? In: Engels, J.M.M., Rao, V.R., Brown, A.H.D., Jackson, M.T. (Eds): *Managing Plant Genetic diversity*. International Plant Genetic Resources Institute, Rome, Italy, 43–56.

Lacape, J.M., Dessauw, D., Rajab, M., Noyer, J.L., Hau, B. (2007). Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Molecular Breeding* 19, 45–58.

Liu, S., Cantrell, R.G., McCarty, J.C. Jr., Stewart, J.M. (2000). Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. *Crop Science* 40, 1459–1469.

Meredith, W.R. Jr. (2000). Cotton yield progress-why has it reached a plateau? *Better Crops* 84, 6–9.

Ministry of Food, Agriculture & Livestock (MINFAL). (1992). *Annual Report*. Government of Pakistan.

Multani, D.S., Lyon, B.R. (1995). Genetic fingerprinting of Australian cotton cultivars with RAPD markers. *Genome* 38, 1005–1008.

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

Rana, M.K., Singh, V.P., Bha, K.V. (2005). Assessment of genetic diversity in upland cotton (*Gossypium hirsutum* L.) breeding lines by using amplified fragment length polymorphism (AFLP) markers and morphological characteristics. *Genetic Resources and Crop Evolution* 52, 989–997.

Reddy, O.U., Pepper, A.E., Abdurakhmonov, I., Saha, S., Jenkins, J.N., Brooks, T., Bolek, Y., El-Zik, K.M. (2001). New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. *Crop Science* 5, 103–113.

Rohlf, F.J. (1997). *NTSYS-pc 2.1. Numerical taxonomy and multivariate analysis system*. Exeter Software, Setauket, NY.

Roussel, V., Koenig, J., Bechert, M., Balfourier, F. (2004). Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. *Theoretical and Applied Genetics* 108, 920–930.

Sokal, R.R., Michener, C.D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38, 1409–1438.

Wendel, J.F., Brubaker, C.L. (1993). RFLP diversity in *Gossypium hirsutum* L. and new insights into the domestication of cotton. *American Journal of Botany* 80 (SUPPL.), 71.

Wendel, J.F., Brubaker, C.L., Percival, A.E. (1992). Genetic diversity in Gossypium hirsutum and the origin of upland cotton. *American Journal of Botany* 97, 1291–1310.

Zhang, H.B., Li, Y., Wang B., Chee, P.W. (2008). Recent advances in cotton genomics. *International Journal of Plant Genomics* Volume 2008, Article ID 742304 (doi:10.1155/2008/742304).

Zhang, J.F., Lu, Y., Cantrell, R.G., Hughs, E. (2005). Molecular marker diversity and field performance in commercial cotton cultivars evaluated in the Southwestern USA. *Crop Science* 45, 1483–1490.