AcademicPres

Available online at www.notulaebotanicae.ro

Print ISSN 0255-965X; Electronic 1842-4309

Not Bot Horti Agrobo, 2012, 40(2): 261-265

Notulae Botanicae Horti Agrobotanici Cluj-Napoca

DNA Fingerprinting and Genotyping of Cotton Varieties Using SSR Markers

Mohammad Nazrul ISLAM[1], Md. Rezwan MOLLA[2*], Md. Motiar ROHMAN[3], Mirza HASANUZZAMAN[4], Shah Mohammad Naimul ISLAM[5], Lutfur RAHMAN[6]

[1]*Sher-e-Bangla Agricultural University, Department of Biotechnology, Dhaka, Bangladesh*

[2]*Agricultural Research Institute, Plant Genetic Resources Centre, Molecular Biology Laboratory, Joydebpur, Gazipur-1701, Bangladesh; mhzsauag@yahoo.com (*corresponding author)*

[3]*Bangladesh Agricultural Research Institute, Plant Breeding Division, Molecular Breeding Laboratory, Joydebpur, Gazipur-1701, Bangladesh*

[4]*Sher-e-Bangla Agricultural University, Department of Agronomy, Dhaka-1207, Bangladesh*

[5]*Bangabandhu Sheikh Mujibur Rahman Agricultural University, Department of Biotechnology, Gazipur, Bangladesh*

[6]*Bangladesh Agricultural University, Department of Genetics and Plant Breeding, Mymenisingh, Bangladesh*

Abstract

DNA fingerprinting and genetic diversity analysis helps direct selective breeding and conservation of plant species. Since simple sequence repeats (SSR) markers are co-dominant, they can predict level of genetic diversity and thereby protect plant genetic resources of a region. Keeping the aforesaid rationale in mind, we worked on molecular characterization of eight cotton varieties in Bangladesh using simple sequence repeat (SSR) or microsatellite DNA markers. All the three microsatellite DNA markers were found to be polymorphic, extracting a total of eight alleles with an average of 2.67 alleles per locus in the present study. Allele sizes were as 149-155 bp, 178-198 bp and 140-202 bp for the loci BNL1551, BNL1721 and BNL2960, respectively. Polymorphism Information Content (PIC) values were ranged from 0.469 to 0.531. UPGMA dendrogram separated 8 varieties of cotton into two clusters. One cluster contained six varieties CB-1, CB-2, CB-3, CB-7, CB-9 and CB-10 while other two varieties CB-5 and HC 1 formed another cluster. The findings of this study would provide a useful guide for selecting specific germplasm with distinct genetic background for diversifying cotton breeding program in Bangladesh.

Keywords: cotton, crop biotechnology, DNA fingerprinting, microsatellite markers, molecular breeding

Introduction

Cotton, major fibre in preparing human apparel, has played a key role in the development of civilization. For centuries, cotton has been clothing the human kind. From several decades, it has been competing with variety of man-made fibres, most fiercely polyester. Cotton is particularly grown for fibre, which is the principal raw material for textile industry. Thus, fibre quality parameters play a decisive role in cotton marketing. According to Cotton Development Board in Bangladesh, present annual requirement of raw cotton for local textile industry is estimated to be 4.0 million bales. Country's present (2011-12 FY) upland (*Gossypium hirsutum*) and hill cotton (*G. utrients*) production is around 1.03 lakh bales from about 35675 ha of which 6800 bales of hill cotton has been produced from 15650 ha area. This production can meet only a small portion of the total annual requirement of local spinning mills. In order to increase the production and productivity of the cotton, organized cotton breeding programmes for OPs, hybrids and Bt cottons must be developed. Such a programme will require development of techniques to determine the genetic diversity of the cotton varieties/land races available in Bangladesh, where molecular markers can also play very important roles.

Molecular markers have been widely used in genetic analyses, breeding studies and investigations of genetic diversity and the relationship between cultivated species and their wild parents (Kurt *et al*., 2011; Mokrani *et al*., 2012; Sofalian *et al*., 2009). This is because they have several advantages as compared to morphological markers, including detecting high polymorphism and independence from effects related to environmental conditions and the physiological stage of the plant. Simple sequence repeat (SSR) markers (microsatellites) have been successfully employed in many genetic diversity studies (Gutiérrez *et al*., 2002; Király *et al*., 2012; Liu *et al*., 2000b; Shah *et al*., 2009) and are useful for a variety of applications in plant genetics and breeding because of their reproducibility, multi-allelic nature, codominant inheritance, relative abundance and good genome coverage (Powel *et al.,* 1996). Microsatellite variation is thought to be due to slippage of the DNA polymerase during replication or unequal crossing over re-

sulting in differences in the copy number of the core nucleotide sequence (Yu and Kohel, 1999).

Limited pools of cotton germplasm have been characterized previously with random amplified polymorphic DNA (RAPDs) (Iqbal *et al*., 1997; Tatineni *et al*., 1996) and restriction fragment length polymorphisms (RFLPs) (Brubaker and Wendel, 1994). However, the level of polymorphism detected was relatively low. The availability and abundance of microsatellite markers throughout the cotton genome coupled with the fact that they are polymorphic, co-dominant. In addition, they are based on polymerase chain reaction (PCR) make them particularly useful in genetic diversity studies (Reddy *et al.*, 2001). According to Nguyen *et al.* (2004) more than 1000 microsatellite primers have already been isolated from cotton DNA genome and are available in the genome libraries. 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*., 2005), but no SSR analysis of Bangladeshi cotton genotypes has been made till 2007 when 157 varieties of 20 crop genetic materials including cotton were studied and recorded in a government publication by Rahman *et al.* (2007). Zhang *et al.* (2005) reported genetic distance among 'Acala1517' genotypes ranged from 0.06 to 0.38 with an average of 0.18 on the basis of 189 SSR marker alleles, indicating a substantial genetic diversity among 'Acala1517' cotton germplasm.

In Bangladesh, there are several cotton varieties which are lacking molecular pedigree records or characterization at DNA level. Due to lack of patent application (patenting studies) at national level, the country may lose genetic resource and varietal identity of this important crop plant. It is essential to have proper genetic records and identity in maintaining conservation and cultivation of this plant species. Hence, the proposed plant variety protection lends added urgency to the search for solutions to the conservation of plant genetic diversity. Molecular markers have been successfully applied in registration activities including cultivar identification (Mailer *et al*., 1994). However, with the necessity of varietal identity, present study was an attempt to reveal genetic diversity and varietal identification of eight cotton varieties in Bangladesh by DNA fingerprinting using microsatellite markers.

Materials and methods

DNA isolation/extraction

Seeds of eight cotton varieties were obtained from Cotton Development Board, Bangladesh. Genomic DNA was isolated following protocol described by Saghai-Maroof *et al*. (1984) with some modifications. Juvenile leaves (unfolded) of 30 days old plants were used in genomic DNA isolation. Leaf tissues were cut into small pieces, homogenized and digested with extraction buffer (pH= 8.0): 50 mM Tris-HCl, 25 mM EDTA (Ethylenediaminetetraacetic acid), 300 mM NaCl and TEN buffer + 5% SDS (Sodium Dodecyl Sulfate) +10% PVP (Poly Vinyl Pyrolideone) +20% CTAB (Cetyl Trimethyl Ammonium Bromide). After incubation for 20 minutes at 65°C with intermittent swirling, the mixture was emulsified with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1, v/v/v). DNA was precipitated using two volume of absolute alcohol in presence of 0.3 M sodium acetate and pelleted by centrifugation. The pellets were then washed with 70% ethanol, air dried and resuspended in an appropriate volume of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH=8.0) and finally treated with 2 µl of RNAse. Quality and quantity of DNA were controlled via gel electrophoresis and spectrophotometer, repectively (Spectronic Genesis™, Spectronic Instruments Inc., USA).

Microsatellite Markers and PCR amplification

A set of five microsatellite loci (BNL1551, BNL1721, BNL 2572 and BNL2960) have been selected from the literature cited by Rungis *et al.* (2005) and Liu *et al.* (2006) to determine the potential of these markers for variety identification. Finally three primers, BNL1551, BNL1721 and BNL2960 were selected based on their performance for SSR data analysis. Polymerase Chain Reactions were done in a volume of 10 µl containing 10 x PCR Buffer, 0.25 mM each of the dNTPs, 1 µM of each of primer, 1 unit ampli *Taq* DNA polymerase, 50 ng template DNA. Amplification were carried out in oil free thermal cycler (Thermal cycler gradient, Eppendorf) with the following thermal profile: initial denaturation step at 94°C for 3 min followed by 35 cycles at 95°C for 1 min, 57°C for 45 sec, and 72°C for 1.5 min and a final cycle at 70°C for 7 min. PCR was the confirmed by electrophoresis on 2% agarose gel.

Electrophoretic separation and visualization of PCR products

PCR-products were electrophoresed on a 6% denaturing polyacrylamide gel containing 19:1 acrylamide: bisacrylamide and 8 M urea. Electrophoresis was conducted using the SequiGen GT sequencing gel electrophoresis system (BIO-RAD Laboratories, Hercules, CA.). A prerun of the gel for 30 mins at 120 W was followed by a final run at 60 W and 50°C upon loading of denatured PCR products for a specified period of time depending on the size of amplified DNA fragment (usually 1 hour for 100 bp). After completion of electrophoresis, the DNA fragments were visualized following the Promega (Madison, WI) silver-staining protocol.

Analysis of microsatellite data

The bands representing particular alleles at the microsatellite loci were scored manually and designated the bands as A, B, C etc. from the top to the bottom of the gel. The genotypes of different strains were scored as AA,

BB, CC, etc. for homozygous or as AB, AC, BC etc. for heterozygous. A single genotypic data matrix was constructed for all loci. Polymorphism Information Content (PIC) was computed by adding the square values for all the frequencies of different alleles produced by a single marker locus and then deducted from one (PIC= 1- ΣX_i^2, where, X_i is the frequency of the i-th allele of a particular locus). PIC provides an estimate of the discriminatory power of a marker by taking into account, not only the number of alleles that are expressed, but also the relative frequencies of those alleles. PIC values range from 0 (monomorphic) to 1 (very high discriminative, with many alleles in equal frequencies (Smith and Register, 1998). The software DNA FRAG version 3.03 (Nash, 1991) was used to estimate marker length and allelic length. Estimation of Nei's (1972) genetic distance and construction of a dendrogram was done using the computer programme POPGENE (Version 1.31) (Yeh *et al*., 1999).

Results

All the three microsatellite DNA markers were found to be polymorphic, extracting a total of eight alleles with an average of 2.67 alleles per locus in the present study. Allele sizes were ranged between 149-155 bp, 178-198 bp and 140-202 bp for the loci BNL1551, BNL1721 and BNL2960, respectively. Microsatellite allele profiles are shown in Fig. 1. Allele frequency ranged from 0.1250 to 0.6250 in the present study (Tab. 1). Polymorphism Information Content (PIC) values in the present study ranged from 0.469 to 0.531. The observed level of PIC values indicated remarkable varietal difference in their pedigree levels.

Genetic distance values between cotton varieties were as 0-1.098. Genetic distance values between the cotton varieties CB-1 and CB-5, CB-1 and HC-1, CB-2 and HC-1, CB-3 and CB-5, CB-3 and CB-10, CB-3 and HC-1, CB-5 and CB-9, CB-7 and CB-9, CB-7 and HC-1, CB-9 and CB-10, CB-9 and HC-1 and CB-10 and HC-1 were highest (1.098). However, the least (nil) genetic distance values were found between the cotton varieties CB-1 and CB-2, CB-1 and CB-1 and CB-9, CB-2 and CB-9, CB-3 and CB-9, CB-5 and HC-1 and CB-7 and CB-10 (Tab. 2). UPGMA dendrogram based on Nei's genetic distance (Nei, 1972) separated 8 varieties of cotton initially into two clusters (Fig. 2). One cluster contained CB-1, CB-2, CB-3, CB-7, CB-9 and CB-10 while other CB-5 and HC 1 formed another cluster. Each main cluster subsequently separated into two sub clusters where CB-7 and CB-10 formed one whereas CB-5 and HC-1 were clustered together. Although Hill Cotton (HC-1) showed closeness with C5 cotton variety but forms a distinct group from all others. There was no correlation between genetic distance and geographic origin. Morphologically HC-1 variety is very different from all the others in the study (data not shown).

Tab. 1. Size and frequency of alleles and diversity index at three microsatellite loci in 8 cotton (*Gossypium spp.* L.) varieties

Locus	Allele Size (bp)	Allele frequency	Diversity Index (PIC=1-$\sum Xi^2$)
BNL1551	155	0.625	0.531
	151	0.125	
	149	0.250	
BNL1721	198	0.625	0.469
	178	0.375	
BNL2960	202	0.125	0.531
	169	0.625	
	140	0.250	

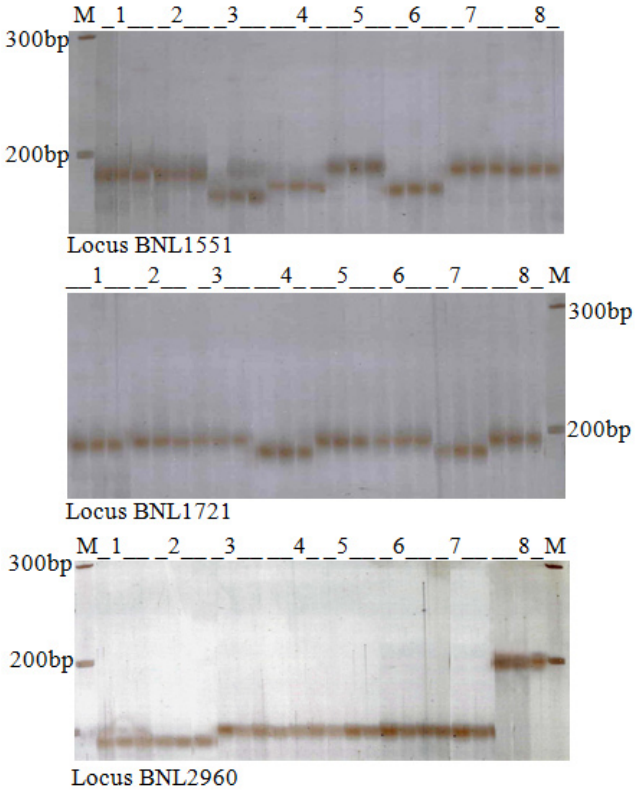

Fig. 1. Microsatellite profiles of eight cotton (*Gossypium spp.* L.) varieties at loci BNL1551, BNL1721 and BNL2960; Lane, 1= CB-1, 2= CB-2, 3= CB-3, 4= CB-5, 5= CB-7, 6= CB-9, 7= CB-10 and 8= HC-1; M: Molecular weight Marker (100 bp DNA ladder)

Discussion

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). These characterizations have provided useful information for understanding the genetic diversity and structure of various cotton gene pools found in different geographic regions. This

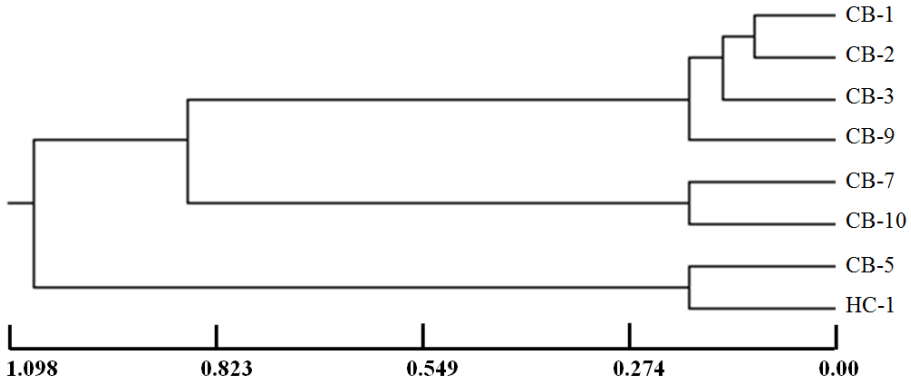

Fig. 2. UPGMA dendrogram based on Nei's (1972) genetic distance, summarizing the data on differentiation between eight cotton cultivars according to Microsatellite marker analysis

Tab. 2. Summary of Nei's (1972) genetic distance values between 8 (*Gossypium spp.* L.) varieties for all loci

Cultivars	CB-1	CB-2	CB-3	CB-5	CB-7	CB-9	CB-10
CB-2	0.0000						
CB-3	0.0000	0.0000					
CB-5	1.0986	1.0986	1.0986				
CB-7	0.4055	0.4055	1.0986	0.4055			
CB-9	0.0000	0.0000	0.0000	1.0986	1.0986		
CB-10	0.4055	0.4055	1.0986	0.4055	0.0000	1.0986	
HC-1	1.0986	1.0986	1.0986	0.0000	1.0986	1.0986	1.0986

information could be incorporated into effective management of cotton germplasm in some cotton breeding programs for control of genetic diversity.

Bertini *et al*. (2006) reported 2.13 alleles per microsatellite locus using 31 primer sets in 53 cotton varieties. The average number of alleles of the present study was comparable with that of the previous studies in cotton. With only three pairs of primers our investigation yielded somewhat larger number of average alleles than the previous study. The possible reason for this result might be due to the less number of diverse cotton varieties used in this study. Observed allelic lengths in our study were almost similar to the previous study. However, some variation did occur, which might be due to mutation of di-nucleotide repeat units. Bertini *et al*. (2006) obtained allele sizes ranged between 170-180 bp and 140-150 bp for the loci BNL1721 and BNL2960, respectively and average PIC value of 0.40 which were lower than the present observation. In an investigation on Pakistani cotton cultivars by Khan *et al.* (2009), values of PIC for each SSR marker ranged from 0.05 to 1.00 with an average of 0.46. This is not unexpected, as mentioned above, from previous diversity analyses (e.g., Bertini *et al*., 2006; Khan *et al*., 2009). UPGMA dendrogram revealed only one unusual clustering of Hill cotton, HC-1 with one of the varieties of cotton registered by the Cotton Board. This indicates that the Hill cotton grouped with non-hill one variety. Possible explanation for this could be that the three primers which were used in the determination might have identified the close alleles of these two materials, whereas, in other cases the primers have identified different alleles of different areas. Such variations between cluster analysis in cultivars and known parentage should not surprise because these are tetraploids of different combinations and has narrow genetic base (Meredith, 2000). The limited sampling of the cotton genome revealed by only 3 SSR primer pairs may contribute to such inconsistencies.

The findings of this study indicate the need for continued effort to widen more the diversity range for both cotton germplasm conservation and future breeding. The characterization of cotton cultivars using SSR markers generated not only essential information for understanding genetic diversity of elite Bangladeshi cotton germplasms, but also provided a useful guide for selecting specific germplasm with distinct genetic background for diversifying cotton breeding program in Bangladesh.

Acknowledgements

The authors thank the Danish International Development Agency (DANIDA) for supporting this fingerprinting of fundamental research work of national interest through the Seed Industry Development Project of the Ministry of Agriculture, Government of the Peoples' Republic of Bangladesh. They also express their thanks to Prof. Dr. Md. Samsul Alam of the Department of Fisheries Biology and Genetics at Bangladesh Agricultural University for all possible technical support during work.

References

Bertini CHC de Magalhaes, Schuster I, Sediyama T, de Barros EG, Moreira1 MA (2006). Characterization and genetic diversity analysis of cotton cultivars using microsatellites. Genetics Mol Biol 29:321-329.

Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape J-M, Yu JZ, Jesudurai C, Jung S, Muthukumar S, Yellambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu S, Zhang T, Fang D, Pepper A, Kumpatla S, Jacobs J, Tomkins J, Cantrell R, Main D (2006). CMD: a cotton microsatellite database resource for *Gossypium* genomics. BMC Genomics 7:132.

Brubaker CL, Wendel JF (1994). Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*) using nuclear restriction fragment length polymorphisms (RFLPs). American J Bot 81:1309-1326.

Fu YB (2006). Genetics redundancy and distinctness of flax germplasm as revealed by RAPD dissimilarity. Plant Genet Resour 4:177-124.

Gutierrez OA, Basu S, Saha S, Jenkins JN, Shemaker DB, Cheatham CL, McCarty JC Jr (2002). Genetic distance among selected cotton genotypes and its relationship with F_2 performance. Crop Sci 42:1841-1847.

Iqbal MJ, Aziz N, Saeed NA, Zafar Y, Malik KA (1997). Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. Theor Appl Genet 94:139-144.

Karp A (2002). The new genetic era: will it help us in managing genetic diversity?, 43-56 p. In: Engels JMM, Rao VR, Brown AHD, Jackson MT (Eds.). Managing Plant Genetic diversity. International Plant Genetic Resources Institute, Rome, Italy.

Khan AI, Fu YB, Khan IA (2009). Genetic diversity of Pakistani cotton cultivars as revealed by simple sequence repeat markers. Commun Biomet Crop Sci 4:21-30.

Király, I, Redeczki, R, Erdélyi, E , Tóth, M (2012). Morphological and molecular (SSR) analysis of old apple cultivars. Not Bot Horti Agrobo 40(1):269-275.

Kurt, Y, Bilgen, BB, Kaya, N, Isik, K (2011). Genetic comparison of *Pinus brutia* Ten. populations from different elevations by RAPD markers. Not Bot Horti Agrobo 39(2):299-304.

Liu S, Cantrell RG, McCarty-Jr JC, Stewart J McD (2000b). Simple sequence repeat based assessment of genetic diversity in cotton race stock accessions. Crop Sci 40:1459-1469.

Liu DQ, Guo XP, Lin ZX, Nie YC, Zhang XL (2006). Genetic diversity of Asian cotton (*Gossypium arboreum* L.) in China evaluated by microsatellite analysis. Genet Resour Crop Evol 53:1145-1152.

Mailer RJ, Scarth R, Fristensky B (1994). Discrimination among cultivars of rapeseed (*Brassica napus* L.) using DNA polymorphism amplified from arbitrary primers. Theor Appl Genet 87:697-704.

Meredith WR Jr (2000). Cotton yield progress-why has it reached a plateau? Better Crops 84:6-9.

Mokrani L, Jawdat D, Esselti MN, Fawaz I, Al-Faoury H (2012). Molecular characterization of Syria commercial and introduced cotton germplasm using AFLP and SSR for breeding applications. J Plant Biol Res 1:65-75

Nash JHE (1991). DNAfrag, Version 3.03. Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada.

Nei M (1972). Genetic distance between populations. Am Nat 106:283-292.

Nguyen TB, Giband M, Brottier P, Risterucci AM, Lacape JM (2004). Wide coverage of the tetraploide cotton genome using newly developed microsatellite markers. Theor Appl Genet 109:167-175.

Powell W, Machray GC, Provan J (1996). Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215-222.

Rahman L, Molla MR, Sultana S, Islam MN, Ahmed NU, Rahman MS, Nazim-ud-Dowla M, Shah-E-Alam M, Hassan L, Alam MS (2007). Plant varieties of Bangladesh: Morphological and molecular characterization. Seed Wing, Ministry of Agriculture, Government of the People's Republic of Bangladesh, Vol. 1, 486 p.

Reddy OU, Pepper AE, Abdurakhmonov I, Saha S, Jenkins JN, Brooks T, Bolek Y, El-zik KM (2001). New dinucleotide and trinucleotide microsatellite marker resources for cotton genome research. Crop Sci 5:103-113.

Rungis D, Llewellyn D, Dennis ES, Lyon BR (2005). Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars. Aust J Agric Res 58:301-307.

Saghai-Maroof MA, Soliman KM, Jonsensan RA, Allard RW (1984). Ribosomal spacer length polymorphism in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 81:8014-8019.

Shah Z, Munir I, Ali S, Iqbal A, Mumtaz S, Nwaz R, Swati ZA (2009). Genetic diversity of Pakistani maize genotypes using chromosome specific simple sequence repeat (SSR) primer sets. Afr J Biotechnol 8:375-379.

Smith JSC, Register JC (1998) Genetic purity and testing technologies for seed quality: a company perspective. Seed Sci Res 8:285-293.

Sofalian O, Chaparzadeh N, Dolati M (2009). Genetic diversity in spring wheat landraces from Northwest of Iran assessed by ISSR Markers. Not Bot Horti Agrobo 37(2):252-256.

Tatineni V, Cantrell RG, Davis DD (1996). Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. Crop Sci 36:186192.

Yeh FC, Yang RC, Boyle T (1999). POPGENE VERSION 1.31: Microsoft Window-based free software for population genetic analysis. ftp://ftp.microsoft.com/Softlib/HPGL.EXE.

Yu J, Kohel RJ (1999). Update of the cotton genome mapping, 439 p. In: Proc. Beltwide Cotton Res. Conf., Orlando, FL. 3-7 Jan. 1999. Natl Cotton Council Am, Memphis, TN.

Zhang JF, Lu Y, Adragna H, Hughs E (2005). Genetic improvement of New Mexico Acala cotton germplasm and their genetic diversity. Crop Sci 45:2363-2373.