Indian J. Genet., **64**(2): 91-93 (2004)

Amplified fragment length polymorphism (AFLP) based diversity in advanced breeding lines of cotton (*Gossypium hirsutum* L.)

M. K. Rana, Sonika Singh and K. V. Bhat

National Research Center on DNA Fingerprinting NBPGR, Pusa Campus, New Delhi 110 012

(Received: March 2004; Revised: August 2004; Accepted: August 2004)

Abstract

Amplified fragment length polymorphism (AFLP) analysis was carried out in 29 advanced breeding lines of cotton (*G. hirsutum* L.) for diversity analysis. Total genomic DNA isolated following CTAB method was purified and subsequently used for AFLP analysis employing fluorescent dye labeling and detection technology. Using 308 amplicons generated by three AFLP primer combinations, Jaccard's similarity estimates between all possible pairs of 29 lines were calculated. High level of diversity was observed in the studied material. Cultivars Pusa 8-6 and RS 810 had the maximum dissimilarity between them, whereas lines RAH 30 and NDH 1001 were having the highest similarity. Cluster analysis revealed lines AKH 081, NDH 1010, RS 810, RST 13 and Pusa 8-6 to be the most distinct ones. Advanced breeding lines from different sources were found to be interspersed and no source-wise clustering was evident.

Key words: Cotton, amplified fragment length polymorphism (AFLP), genetic diversity

Introduction

Cotton (*Gossypium hirsutum* L.) is an important cash crop of the country. Since the inception of All-India Co-ordinated Cotton Improvement Project (AICCIP) in 1967, a large number of cultivars have been released for growing under different agro-climatic conditions through systematic cotton improvement work. Newer genotypes are being constantly bred to substitute and overcome the genetic defects of the older ones and this process has led to the development of cotton lines, many of which are in the pipeline to be released as cultivars. Estimating the extent of genetic diversity in them is of paramount importance for germplasm enhancement and for obtaining genetic gains.

Morphological markers are routinely used for genetic diversity analysis and for establishing relationships among cultivars. But recently, various PCR based molecular marker techniques like Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Sequence Tagged Microsatellites (STMS) etc. have been developed into powerful tools for such studies. Among these, AFLP has the advantage over others since many markers can be generated in short span of time and throughput can be greatly enhanced using multiplexing of primers. There are only a few reports on the use of this technique for genetic diversity analysis in cotton [1-3]. The present study aimed at establishing genetic relationships among 29 advanced breeding lines of cotton using AFLP markers.

Materials and methods

Plant material and DNA extraction: Twenty-nine advanced breeding lines of cotton including popular commercial cultivars were included in the present study (Table 1). Most of these genotypes are specifically adapted to the Northern cotton-growing zone of the country. Equal number of fresh, young leaves (3-4 g) from at least twenty plants from each genotype was bulked for DNA extraction. Total genomic DNA was isolated following CTAB method [6]. Isolated DNA was purified and the concentration of DNA in the samples was determined with a Hoefer DNA Fluorometer Model DQ 200 (Hoefer Pharmacia Biotech Inc., San Francisco, CA) using Hoechst 33258 as the dye and calf thymus DNA as the standard. DNA samples for PCR were further diluted in 10:1 TE and stored at 4°C until AFLP analysis.

AFLP fingerprinting: AFLP electropherograms were produced for each genotype using the ABI PRISM™ fluorescent dye labeling and detection technology (Perkin-Elmer) [4]. A kit supplied by Perkin Elmer Applied Biosystems (USA), optimized for a genome size of 500 to 6000 Mb, was used according to the manufacturer's instructions. A 0.5µg quantity of genomic DNA was digested with 1U of *Mse* I and 5U of *EcoR* I restriction endonucleases. *EcoR* I and *Mse* I adaptors were ligated with 1U of T4 DNA ligase (all enzymes New England Biolabs, Beverly, Mass., USA). Restriction and ligation were done simultaneously in a single step by incubating at 37°C for 3 hours. Pre-selective and selective amplification cocktails and thermal cyclic conditions were

Table1. Cotton genetic material used for AFLP analysis

S. No.	Line	Source
01	CISV 15	CICR, Sirsa
02	CISV 50	CICR, Sirsa
03	NDH 1001	CRS, Nanded
04	NDH 1010	CRS, Nanded
05	AKH 081	HAU, Hisar
06	H 1226	HAU, Hisar
07	H 1228	HAU, Hisar
08	H 1250	HAU, Hisar
09	H 1261	HAU, Hisar
10	Pusa 802	IARI, New Delhi
11	Pusa 8-6	IARI, New Delhi
12	Pusa 4513	IARI, New Delhi
13	LH 1948	PAU, Ludhiana
14	LH 1953	PAU, Ludhiana
15	LH 1960	PAU, Ludhiana
16	LH 1961	PAU, Ludhiana
17	LH 1995	PAU, Ludhiana
18	F 1861	PAU, Ludhiana
19	F 1945	PAU, Ludhiana
20	F 1946	PAU, Ludhiana
21	RS 810	RAU, Bikaner
22	RS 2013	RAU, Bikaner
23	RST 11	RAU, Bikaner
24	RST 13	RAU, Bikaner
25	RSY 9	RAU, Bikaner
26	RS 2257-1	RAU, Bikaner
27	RACH 108	UAS, Dharwad
28	RAH 30	UAS, Dharwad
29	CPD 745	UAS, Dharwad

chosen as per the manufacturer's instructions. *Mse* I primers were unlabeled whereas *EcoR* I primers were labeled with either FAM, HEX or NED fluorescent dyes. All PCRs were performed on a Perkin Elmer 9600 thermocycler. Selectively amplified products (3.0µl) were mixed with 0.5µl of a Gene Scan 500 ROX internal size standard and 20µl of de-ionized formamide and were denatured at 96°C for 5 min. The denatured samples were electrophoresed and analyzed on an automated DNA sequencer (ABI model 310, Perkin-Elmer Applied Biosystems).

Data scoring and statistical analyses: Amplicons ranging in size from 50 to 450 base pairs were scored for the presence (1) or absence (0) across all the lines using Genotyper (Version 2.5, Perkin Elmer/ABI) software. The binary AFLP data was analyzed to produce a matrix of similarity values based on Jaccard's coefficient of similarity (J_{ij}) [5]. Clustering of the lines was done using unweighted pair-group method based on arithmetic averages (UPGMA) analysis using the programmme Numerical Taxonomy and Multivariate Analysis System (NTSYS-pc) software version 2.02 [6]. PIC (polymor- phism information content) for each AFLP band was calculated as per Anderson *et al.* [7] and averaged for each primer-combination.

Results and discussion

AFLP analysis was done using nine different combinations of *EcoR* I and *Mse* I primers but final statistical analysis was based on three combinations only (Table 2). *Mse* I primers having 3′ nucleotide extensions CTT or CTG were also tried with *EcoR* I primers having 3′ nucleotide extensions ACT, ACC or AAG but due to large amount of missing data with these combinations were not included in the analysis. The primer pair M-CAT/E-AGC gave the highest per cent polymorphism, but the number of amplicons having PIC value equal to or greater than 0.20 was the lowest for this combination. Primer combination M-CAT/E-ACT produced the highest number of fragments having PIC value equal to or greater than 0.20 and PIC value as such was also highest for this combination. Primer combination M-CAT/E-AAG exhibited the lowest similarity or in other words the maximum diversity among 29 advanced breeding lines of cotton. The extent of polymorphism observed in our study is quite high compared with earlier AFLP investigation of Abdalla *et al.* [1] who found 31% AFLP bands to be polymorphic. Since the resolution power using automated DNA sequencer with capillary electrophoresis, as against conventional vertical or horizontal electrophoretic systems, is very high, single base pair differences could be resolved easily which subsequently led to generation of large number of amplification products and consequently polymorphic ones in our study. Moreover, our choice of primers was based on initial screening of a large number of primer combinations on a set of cotton cultivars for polymorphism detection.

Table 2. AFLP primer-combinations studied and band characteristics

Primer-combination	NB	Per cent polymorphism	PIC	AJS
M-CAT/E-AAG	114	97.84	0.343	0.362
M-CAT/E-AGC	47	99.21	0.361	0.396
M-CAT/E-ACT	147	93.86	0.392	0.501

M = *Mse I*; E = *EcoR I*; NB = number of bands with polymorphism information content ≥ 0.20; PIC = polymorphism information content; AJS = average Jaccard's similarity

Pair-wise genetic similarity estimates among all advanced breeding lines were calculated based on 308 polymorphic amplification products (29 × 29 matrix not presented here). Similarity was found to be maximum for the pair of lines RAH 30 - NDH1001 (66%), followed by Pusa 802 - F 1945 (65%). Lines Pusa 8-6 - RS 810 showed least similarity (20%), followed by lines RST 13 - RS810 (24%). Pair-wise average genetic similarity among all the lines was found to be 43%. This value is much lower than that recorded in another

set of 24 advanced breeding lines where average genetic similarity was found to be 66% [8]. This difference in the magnitudes of estimates might be either due to different lines in both the studies (only four were common) and/ or due to altogether different primer combinations used in these two studies. Two released cultivars Pusa 8-6 and RS 810, commonly used as checks showed maximum genetic diversity with lines RSY 9 (42%) and F 1946 (35%), respectively. These two cultivars between them had the highest diversity compared to any other pair-wise combination of lines. Jaccard's similarity estimates calculated using three primer combinations were found to have very high correlation with Jaccard's similarity estimates calculated using the primer combination M-CAT/ E-ACT (r = 0.959) alone. Correlation using other two primer combinations with the pooled estimates was low. This underlines the importance of primer combination M-CAT/ E-ACT in such kind of analyses.

Clustering pattern of the lines has been shown in dendrogram in Fig. 1. At average genetic similarity value (43%) this dendrogram shows lines AKH 081, NDH 1010, RS 810, RST 13 and Pusa 8-6 each to be a cluster of unique entries. Line AKH 081 was found to be the most diverse which could be due to alien genetic material from *Gossypium anomalum* in the course of genesis of this line. These five lines are genetically diverse at AFLP loci analysed and could be useful genetic resource. Furthermore, combining ability of these identified diverse lines can be tested in order to identify the ones having better nicking ability to produce heterotic cross-combinations. Lines CISV 20 and H 1228, both from the state of Haryana, grouped together in a single cluster. Two lines LH 1960 and LH 1961 from Punjab and line RSY 9 from Rajasthan were quite distinct from rest of the lines. Nineteen lines, irrespective of the source, were grouped in a single largest cluster. These lines might have arisen from some common population, and since plant breeders mostly share the elite material of other breeding stations in cotton improvement programmes such sort of clustering becomes obvious.

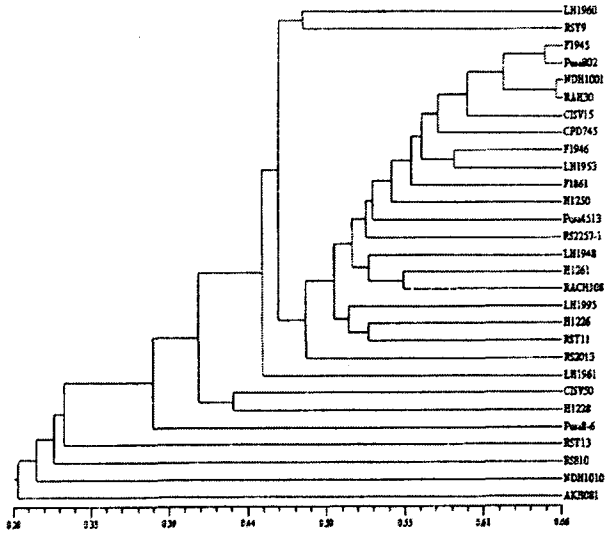

Fig.1. Dendrogram constructed using UPGMA cluster analysis depicting genetic relationships among 29 cotton advanced breeding lines. Scale at bottom denotes Jaccard's genetic similarity.

Other workers have reported narrow range of diversity using AFLP markers in cotton. Abdalla *et al.* (2001) [1] using 16 AFLP primer-combinations found 86% genetic similarity in seven *G. hirsutum* accessions. Iqbal *et al.* (2001) [2] found many pair-wise similarity indices to be greater than 96% using 3178 markers generated by 20 AFLP primer-combinations. The presence of substantial genetic diversity in our study, in contrast, might be due to the use of very powerful resolving system, wherein single nucleotide differences could be detected. Secondly, it might be due to the difference in the primer sequences and consequent genome section that has been scanned for investigation, besides the difference in the genetic material used.

Acknowledgments

This work was supported by the financial grant under the project Technology Mission on Cotton-Mini Mission I, Indian Council of Agricultural Research, New Delhi.

References

1. **Abdalla A. M., Reddy O. U. K., El-Zik K. M. and Pepper A. E.** 2001. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor. Appl. Genet., **102**: 222-229.
2. **Iqbal M. J., Reddy O. U. K., El-Zik K. M. and Pepper A. E.** 2001. A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. Theor. Appl. Genet., **103**: 547-554.
3. **Rana M. K. and Bhat K. V.** 2001. Genetic diversity analysis and fingerprinting of diploid cotton cultivars using automated amplification fragment length polymorphism (AFLP). Indian J. Plant Genet. Resour., **14**: 150-152.
4. **Saghai-Maroof M. A., Soliman K. M., Jorgensen R. A. and Allard R. W.** 1984. Ribosomal DNA spacer-length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics, Proc. Natl. Acad. Sci. USA, **81**: 8014-8018.
5. **Jaccard P.** 1908. Nouvelles recherchis sur la distribution florale. Bull. Soc. Vaud. Sci. Nat., **44**: 223-270.
6. **Rohlf F. J.** 1997. NTSYS-pc Numerical taxonomy and multivariate analysis system, version 2.02. Exeter Softwae, Setauket, New York.
7. **Anderson J. A., Churchuill G. A., Autrique J. E., Tanksley S. D. and Sorrells M. E.** 1992. Optimizing parental selection for genetic linkage maps. Genome, **36**: 181-186.