1771 ANALYSIS OF GENETIC DIVERSITY IN COTTON (*GOSSYPIUM HIRSUTUM* L.) VARIETIES USING AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP) MARKERS

Dr. Everina Lukonge , Agriculture Research Institute, Mwanza, Tanzania
Dr. Liezel Herselman , University, Bloemfontein, South Africa
Prof. Maryke Labuschagne , University, Bloemfontein, South Africa

ABSTRACT

Morphological markers in cotton are few and provide the general representation of the cultivars according to their growing environments. Molecular markers are numerous and do not effect phenotypes. Amplified Fragment Length Polymorphism (AFLP) analysis has proved powerful for identification of large numbers of potential polymorphic loci in diverse cotton germplasm. In Tanzania morphological characters had been used to study genetic diversity among cotton material. AFLP was used to assess genetic diversity among 26-selected cotton (G*ossypium hirsutum* L.) cultivars from Tanzania (local and exotics) where eight informative AFLP primer combinations were used. NTSYS-pc software was used to determine genetic similarities and construct a dendrogram. Polymorphic information content values (PIC) calculated to study the capacity of the primer combinations in amplifying polymorphisms in cotton cultivars. Results indicated that primer combinations used had high PIC values (0.37 to 0.57) and could discriminate between cultivars. Genetic similarity between all cultivars based on Dice similarities ranged from 0.894 to 0.979 with total average of 0.939. Dendrogram indicated narrow genetic diversity in Tanzanian cotton cultivars. Molecular data confirmed pedigree and origin, meaning that many cultivars were closely related. Heterotic groups were not of much diverse except for Delcot 344, which was very different from others. The study on genetic diversity for more material has to be done. Diverse cultivated cotton cultivars have to be introduced to increase the cotton genetic base in Tanzania.

Key words*: Heterotic group*, *molecular markers, morphology, similarity*

INTRODUCTION

Cultivated cotton (*Gossypium hirsutum* L.) is a second cash crop in Tanzania after coffee grown by small-scale farmers. The narrow genetic diversity in the available germplasm has led into cotton production hindrance because of insect pests attack, diseases, low yield and poor fibre quality (Lukonge and Ramadhani*,* 1999). Characterisation and maintenance of crop germplasm are important for continuous supply of genetic variability for crop improvement and identification of genetic relatedness of available genetic resources (Kumar, 1999; Ali *et al*., 2003). Genetic diversity within lines and populations is fundamental for breeding and germplasm conservation (Rana and Bhat, 2004; Murtaza *et al*., 2005). Furthermore, classification and assignment of breeding lines to established heterotic groups are important in any breeding programme (Smith and Smith, 1992). Cultivated cotton displays low levels of genetic diversity (narrow genetic base) Abdukarimov *et al*., 2003. Morphological and agronomical characteristics for cotton have traditionally been used to distinguish cultivars and provide useful information to users. However, the expression of the majority of these characteristics is significantly influenced by the environment causing problems for consistent identification (Kumar, 1999). In cotton, improved different molecular markers have been used to characterise the genome for

genetic diversity (Murtaza *et al.,* 2005). AFLP analysis detects a large number of polymorphisms that are distributed across the genome and have a high multiplex ratio (Rafalski *et al*., 1996).

In Tanzania, cotton parental selection for hybridisation is based on morphological characteristics. There exists a need for molecular marker analysis of available germplasm material for cotton improvement. The objectives were to determine the genetic diversity of 26 cotton cultivars using AFLP analysis and identify heterotic groups to be used in hybridisation.

METHODS AND MATERIALS

Plant materials and DNA extraction. Twenty-six cotton cultivars (five developed in Tanzania and 21 exotics from other countries) were chosen to explore the genetic diversity of cotton germplasm. Cultivars used had been previously characterised for morphological variations (Lukonge *et al*., 2006). These cultivars have been extensively used in the cotton-breeding programme in Tanzania. Two plants of each cultivar were grown in two pots in a glasshouse at the University of the Free State (UFS) in Bloemfontein, South Africa and at the Mikocheni Agriculture Research Institute, Dar es salaam, Tanzania for DNA extraction. DNA was extracted using a modified monocot extraction procedure (Edwards *et al*., 1991) as described by Adugna (2002). DNA concentration and purity was determined by measuring absorbancies at 260 nm and 280 nm. The quality, integrity and concentration of the DNA were confirmed by electrophoresis in 0.8 % (w/v) agarose gel.

AFLP analysis. AFLP analysis was performed according to the protocol described by Vos *et al*. (1995) with minor modifications as described by Herselman (2003). Primer combinations of *Eco*RI and *Mse*I are represented as E- and M- respectively followed by the selective nucleotides used. A total of eight primer combinations were used with all 26 cotton cultivars studied. E-ACA and E-AAC were used in combination with M-CAT while E-ACT and E-ACC were used in combination with M-CTG, M-CTA and M-CAC. Primers were selected based on literature (Abdalla *et al.,* 2001; Rana and Bhat, 2004). AFLP fragments were resolved using a Perkin Elmer Prism ABI 310 automatic capillary sequencer (PE Biosystems, 2002) using a GENESCAN-1000 ROX™ standard.

Data analysis. AFLP data for selected primer combinations were coded using a binary unit character (1 as present and 0 as absent) of each polymorphic band. Data was summarised in a data matrix for all cultivars based on both unique and shared fragments. Genetic similarities were calculated on the bases of Dice coefficient method (Dice, 1945) using the similarity of qualitative data (SIM-QUAL) programme of Numerical Taxonomy Multivariate Analysis System (NTSYS-pc) version 2.02i software package (Rohlf, 1993). Cluster analyses were performed using unweighted pair group method of arithmetic averages (UPGMA) clustering (Sokal and Michener, 1958) and utilised to construct a dendrogram using the SAHN programme of NTSYS-pc. Calculations for polymorphic information content (PIC) were done using the formula of the expected heterozygosity (Smith *et al.,* 2000) as: $PIC=1-\sum f^2 i$, where f is the percentage of genotypes in which the fragment is present. The PIC value is an indication of a high probability of obtaining polymorphism using that primer combination.

RESULTS

AFLP analysis. Eight selected AFLP primer combinations generated a total of 835 reproducible amplification fragments across all cotton cultivars among which 309 fragments were polymorphic with an average of 37% polymorphisms per primer combination (Table

1). Primer combinations E-AAC/M-CAT, E-ACA/M-CAT and E-ACT/M-CTA produced the highest number of amplified fragments (132, 126 and 119, respectively) while E-ACC/M-CTG amplified the lowest number (76) of fragments (Table 1). Even though some of the primer combinations amplified low numbers of fragments, they were able to uniquely distinguish some of the cultivars. For example, E-ACT/M-CAC uniquely identified Delcot 344 and E-ACC/M-CTG uniquely discriminated High gossypol and Delcot 344. Primer combination E-AAC/M-CAT uniquely identified eight cultivars followed by E-ACT/M-CTA (six cultivars). Delcot 344 was uniquely discriminated from other cultivars by all primer combinations. High levels of polymorphism were observed for primer combinations E-ACC/M-CAC (51.6%) and E-ACT/M-CTG (45.5%) (Table 1). PIC values ranged from 0.37 (E-ACT/M-CAC) to 0.57 (E-ACC/M-CAC) with an average of 0.47 (Table 1).

Estimates of genetic similarities. Genetic similarities of AFLP analysis data are summarised in Table 2. Genetic similarities between all pairs of 26 cotton cultivars varied from 0.894 to 0.979 with mean of 0.939. Genetic similarities were high between some of the cultivars including McNair 235 and MZ561 (0.979), Frego bract and Reba W296 **(**0.978) and between SG 125 and DP 4049 (0.977). The lowest genetic similarity value was observed between High gossypol and Cyto 12/74 (0.894). Generally, High gossypol, Cyto 12/74, Delcot 344, Super okra leaf and Reba B50 had low similarity with the other cultivars (Table 2).

Cluster analysis. The dendrogram based on AFLP marker analysis revealed two major groups *A* and *B* (Figure 1). Major group *A* contained Delcot 344. Delcot 344 has distinctive characteristics including reddish green coloured leaves with no leaf hairs. The second major group (*B)* contained four clusters. Cluster *I* contained 12 cultivars, and divided into two subclusters. The upper most subcluster divided into two groups. The first group contained cultivars from the USA except for NTA 88-6, which is from Mali, but has traits from Deltapine cultivars in its pedigree. These cultivars had high ginning outturn (GOT) values ranging from 40.5% to 43.9% (data not shown). The second group contained four cultivars, Guazuncho (from Argentina, drought tolerant), Stoneville 506 (bacterial blight resistance from the USA), IL74 and IL85 (bacterial blight resistance from Tanzania). The second subcluster contained four cultivars, McNair 235, Des 119, Auburn 56 (all from the USA and resistant to fusarium wilt) and MZ561 (from Tanzania) (Figure 1).

Cluster *II* contained seven cultivars; NTA 93-15, BJA 592 and Irma 1243 originated from West/Central Africa (might have shared some genes). NTA 93-15 and Irma 1243 are susceptible to bacterial blight and fusarium wilt and have high GOT values. UK82 and UK91 are Tanzanian cultivars for the Western Cotton Growing Area's (WCGA's) and clustered with BJA 592, their ancestor for bacterial blight resistance. Cluster *III* contained High gossypol from Chad and has resistance to insects due to high gossypol content. Cluster *IV* was composed of five cultivars; Frego bract (insect resistant) and Reba W296 (Coker 100 x Allen 51-296) clustered together. Dixie King (resistant to fusarium wilt) and Reba B50 (Stoneville B 1439 x A50T) clustered together. Cyto 12/74 (from Pakistan) joined them as a separate group with a genetic similarity of 0.944. Reba W296 and Reba B50 are bacterial blight and fusarium wilt resistant, have weak fibres and both originated from Central Africa (Figure 1).

DISCUSSION

The consideration of estimated genetic distance is important for comparative analysis of diversity levels (Roldan-Ruiz *et al.,* 2001). AFLP analysis is a powerful tool to discriminate and cluster closely related cultivars as well as to trace origin and pedigree through genepool sharing. AFLP analysis covers the entire genome, compared to morphological analyses that

focus on a few traits. AFLP markers are highly efficient compared to morphology and some other DNA markers since AFLP makers are reproducible and display intraspecific homology (Rana and Bhat, 2004). According to Kumar (1999), morphological traits controlled by a single locus can be used as genetic markers, provided expression does not change over a range of environments.

Lu and Myers (2002) reported the high genetic distance of Delcot 344 with other cultivars. The observed high genetic similarity average (0.939) in this study confirmed results reported by Abdukarimov *et al*. (2003) and Van Becelaere *et al.* (2005) that cotton has low genetic diversity. Roldan-Ruiz *et al.* (2001) observed that when cultivars with shared genepools were examined using AFLP markers, high similarity measures produced were linked to morphological similarities. Therefore, AFLP analysis can be used to confirm cultivar pairs that shared genepools.

In the present study, AFLP analysis exposed useful genetic relationships where cultivars were dispersed more evenly; it provided more accurate and reliable relationships because it dealt with basic DNA sequences thus confirmed the usefulness of AFLP markers in studying genetic relationships. However, Lübberstedt *et al.* (1998) and Swanepoel (1999) suggested that the combination of morphological and molecular markers could serve as a major source of information in separating closely related cultivars. In the current study, closely related cotton cultivars, for example the cultivars developed in Tanzania (IL74 and IL85 for the Eastern Cotton Growing Areas (ECGA's)) and (UK91 and UK82 for the Western Cotton Growing Areas (WCGA's)) were separated because of pedigree relationships. Guazuncho, Stoneville 506 and Des 119 clustered together as reported by Poisson *et al.* (2003) due to similar pedigree relationships.

The overall findings from this study indicated that AFLP analysis sufficiently detected genetic diversity to differentiate Tanzanian cotton cultivars. Apart from the narrow genetic diversity present in cotton, AFLP analysis managed to distinguish all cultivars. However, one cannot undermine the role of morphological characterisation because it has been used extensively for germplasm identification, selection of the parents for cotton cultivar improvement and in germplasm collection, conservation and maintenance and is still useful in Tanzania. Reduced genetic diversity of the studied cultivars as observed emphasises the need to focus on introduction of more diverse cultivated cotton cultivars from other countries to Tanzania. The introduction of germplasm should include other tetraploid (e.g. *Gossypium barbadense* L.) species to enable improvement of the available material through hybridisation. The heterotic groups identified could be used for improving cotton breeding programmes through hybridisation. Application of DNA markers could accelerate the process of finding markers related to specific agronomical and morphological traits of interest (Spielmeyer *et al.,* 1998). Although molecular markers like AFLPs analysis are more efficient, they are limited due to initial costs, inadequate infrastructure and expensive chemicals. Molecular analysis using more primer combinations and different molecular markers, along with costs, should be included.

ACKNOWLEDGEMENT

I am grateful to principle secretary for the Ministry of Agriculture Tanzania for the permission, TWOWS, UFS and TCB for sponsoring this study and the Director ARD Mikocheni for the biotechnology laboratory used in this study.

REFERENCES

Abdalla A.M., O U.K. Reddy, K.M. EL-Zik and A.E. Pepper, 2001. Genetic diversity and relationship of diploid and tetraploid cottons revealed using AFLP. *Theoretical Applied Genetics* **102**: 222-229.

Abdukarimov, A.S., S. Djataev and I. Abdukarimov, 2003. Cotton research in Uzbekistan: Elite cultivars and future cotton breeding. Proceedings World Cotton Research Conference pp 5-15. Cape Town, South Africa.

Adugna, W.G., 2002. Genetic diversity of linseed (*Linum usitatissimum* L.) in different environments. PhD Thesis pp 183-201. Department of Plant Sciences, UOVS, Bloemfontein, South Africa.

Ali, A.M., O.M. Ahmed, B.C. Misaka, A.H. Latif, K. Elsiddig and E.A. Babiker, 2003. Characterisation of cotton germplasm and its utilization in breeding for major production constraints in Sudan. Proceedings World Cotton Research Conference pp 16-21. Cape Town, South Africa.

Dice, L.R., 1945. Measures of amount of ecologic association between species. *Ecology* **26**: 297-302.

Edwards K., C. Johnstone and C. Thompson, 1991. A simple and rapid method for the preparation of plan genomic DNA for PCR analysis. *Nucleic Acids Res***. 19**: 1349.

Herselman L., 2003. Genetic variation among Southern African cultivated peanut (*Arachis hypogaea* L.) genotypes as revealed by AFLP analysis. *Euphytica* **133**: 319-327.

Kumar L.S. 1999. DNA markers in plant improvement. An overview. *Biotechnology Advances* **17:** 143-182.

Lu H. and G.O. Myers, 2002. Genetic relationships and disrimination of the influential upland cotton cultivars using RAPD markers. *Theoretical Applied Genetics* **105**: 325-331.

Lübberstedt T, A.E. Melchinger, C. Dußle, M. Vuylsteke and M. Kuiper, 2000. Relationship among early European maize inbreds: IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD and pedigree data. *Crop Science* **40**: 783-791.

Lukonge E., M.T. Labuschagne and L. Herselman, (2006). A comparison of agronomical, morphological and AFLP markers in a genetic diversity study of Tanzania cotton. *Fiftieth congress of the South African Genetics Society*, Bloemfontein, 3-5 March 2006.

Lukonge, E. and O.S. Ramadhani, 1999. Review of Plant Breeding and agrobiodiversity in the Lake Zone. Breeding and Agrobiodiversity workshop proceedings pp 6-11. Agricultural Research and Development Ukiriguru, Mwanza, Tanzania.

Murtaza N., M. Kitaoka and G.M. Ali, 2005. Genetic differentiation of cotton cultivars by polyacrylamide gel electrophoresis**.** *Journal Central Europe Agriculture* **6**: 69-76.

PE, Perkin-Elmer Applied Biosystems, 2002. http://www.perkin.elmmer.com

Poisson, J., O.M. Royo, L. Erazzu, S. Ibalo and A. Montenegro, 2003. First Super okra leaf cotton cultivar bred and released in Argentina**.** Proceedings World Cotton Research Conference pp 259-262. Cape Town, South Africa.

Rafalski J.A., J.M. Vogel, M. Morgante, W. Powell, C. Andre and S.V. Tingey, 1996. Generating and using DNA markers in plants. In: Birren B. and E. Rai (Eds.), *Non-Mammalian Genomic Analysis: A practical Guide* pp 75-134*.* Academic Press, London.

Rana M.K. and K.V. Bhat, 2004. A comparison of AFLP and RAPD markers for genetic diversity and cultivar identification in cotton. *J. Plant Biochemistry and Biotechnology* **13**: 19-24.

Rohlf F.J., 1993. NTSYS-pc. Numerical taxonomical and multivariate analysis system. Exerter Software, Setauket, New York, USA.

Roldan-Ruiz, I, T.J. Gilliland, E. Calsyn, T.J. Gillilad, R. Coll, M.J.T. Van Eijk and M. De Loose, 2001. Estimating genetic conformity between related ryegrass (*Lolium*) varities. II. AFLP characterization. *Molecular Breeding* **6**: 593-602.

Smith J.S.C. and O.S. Smith, 1992. Fingerprinting crop cultivars. *Advances in Agronomy* **47**: 85-129.

Smith, J.S.C., S. Kresovich, M.S. Hopkins, S.E. Mitchell, R.E. Dean, W.L. Woodman, M. Lee and K. Porter, 2000. Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. *Crop Science* **40**: 226-232.

Sokal R.R. and C.D. Michener, 1958. A statistical method for evaluating systematic relationships. Univ Kansas. *Science Bulletin* **28**: 1409-1438.

Spielmeyer, W., A.G. Green, D. Bittisnich, N. Mendham and E.S. Lagudah, 1998. Identification of quantitative trait loci contributing to Fusarium wilt resistance on AFLP linkage map of flax. *Theoretical and Applied Genetics* **97:** 633-641.

Swanepoel, E., 1999. Identification of genetic distances and heterotic groups of inbred maize (*Zea mays*) lines using DNA fingerprinting. MSc Thesis, Department of Plant Breeding, University of Orange Free State, Bloemfontein, South Africa.

Van Becelaere, G., E.L. Lubbers, A.H. Paterson and P.W. Chee, 2005. Pedigree- vs. DNA Marker-Based Genetic Similarity Estimates in Cotton. *Crop science* **45**:2281-2287.

Vos P., R. Hogers, M. Bleeker, M Reijans, T. Van De Lee, M. Hornes, A. Frijters, J. Pot, J. Pelema, M. Kuiper and M. Zabeau, 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acid Research* **23**: 4407-4414. TABLES Table 1: AFLP primer pairs used for assessment of genetic diversity in Tanzanian cotton

Primer combinations	Total fragments	Polymorphic fragments	% Polymophism	PIC	Unique Cultivars
E-ACA/M-CAT	126	40	31.7	0.46	4
E-AAC/M-CAT	132	33	25.0	0.43	8
E-ACC/M-CTG	76	28	36.8	0.47	2
E-ACT/M-CTG	100	45	45.5	0.54	2
E-ACT/M-CTA	119	42	35.3	0.43	6
E-ACCM-CTA	109	43	39.4	0.49	4
E-ACT/M-CAC	96	29	30.2	0.37	1
E-ACCM-CAC	95	49	51.6	0.57	4
Total	835	309			
Average	104	39	37.0	0.47	

PIC = polymorphic information content

Table 2. AFLP genetic similarities of the 26 cotton cultivars

	Hcb	Rb	Frb	Dk	Dp4	Sg	Cyt	Rb5	N88	Hgp	Des	Mz	Mcn	IL8	IL7	Gua	Del	Stn	Aub	N93	Bja	UK9	Irm	UK8	Okl
Rb	0.961																								
Frb	0.946	0.978																							
Dk	0.954	0.960	0.953																						
Dp4	0.970	0.953	0.946	0.953																					
Sg	0.962	0.951	0.944	0.958	0.977																				
Cyt	0.921	0.943	0.954	0.934	0.920	0.912																			
Rb5	0.943	0.951	0.946	0.964	0.939	0.943	0.946																		
N88	0.957	0.944	0.937	0.949	0.964	0.967	0.910	0.943																	
Hgp	0.936	0.924	0.918	0.931	0.937	0.935	0.894	0.925	0.945																
Des	0.944	0.941	0.947	0.937	0.948	0.939	0.923	0.927	0.950	0.936															
Mz	0.950	0.954	0.953	0.936	0.950	0.937	0.933	0.929	0.941	0.932	0.970														
Mcn	0.944	0.943	0.952	0.934	0.943	0.937	0.925	0.922	0.942	0.935	0.975	0.979													
IL8	0.953	0.939	0.927	0.946	0.957	0.952	0.910	0.939	0.964	0.942	0.956	0.953	0.953												
IL7	0.949	0.936	0.928	0.945	0.958	0.947	0.907	0.932	0.949	0.936	0.943	0.946	0.940	0.966											
Gua	0.945	0.927	0.919	0.944	0.947	0.947	0.900	0.924	0.945	0.950	0.939	0.937	0.940	0.957	0.953										
Del	0.920	0.921	0.927	0.918	0.922	0.915	0.923	0.911	0.911	0.909	0.937	0.936	0.936	0.917	0.919	0.921									
Stn	0.955	0.938	0.931	0.951	0.960	0.953	0.911	0.934	0.952	0.945	0.949	0.949	0.946	0.964	0.967	0.970	0.927								
Aub	0.939	0.938	0.940	0.934	0.946	0.938	0.918	0.920	0.935	0.923	0.947	0.955	0.949	0.943	0.952	0.940	0.940	0.958							
N93	0.947	0.929	0.917	0.936	0.949	0.940	0.901	0.920	0.940	0.930	0.926	0.934	0.930	0.946	0.949	0.944	0.920	0.951	0.938						
Bja	0.948	0.935	0.925	0.940	0.947	0.943	0.903	0.927	0.938	0.934	0.929	0.935	0.931	0.945	0.947	0.945	0.922	0.953	0.951	0.964					
UK9	0.947	0.927	0.926	0.934	0.952	0.940	0.900	0.920	0.941	0.938	0.932	0.935	0.934	0.941	0.950	0.943	0.918	0.946	0.937	0.954	0.954				
Irm	0.937	0.931	0.939	0.931	0.940	0.930	0.918	0.918	0.932	0.927	0.940	0.952	0.947	0.938	0.942	0.939	0.935	0.941	0.946	0.937	0.947	0.955			
UK8	0.950	0.932	0.925	0.939	0.953	0.947	0.903	0.929	0.947	0.940	0.930	0.938	0.931	0.949	0.953	0.950	0.915	0.958	0.948	0.952	0.965	0.957	0.955		
Okl	0.941	0.936	0.929	0.943	0.938	0.936	0.912	0.935	0.935	0.925	0.920	0.929	0.921	0.938	0.939	0.933	0.909	0.940	0.931	0.942	0.952	0.947	0.941	0.961	
Acl	0.943	0.934	0.923	0.938	0.944	0.944	0.907	0.925	0.937	0.928	0.927	0.936	0.927	0.945	0.952	0.937	0.916	0.945	0.940	0.947	0.953	0.952	0.946	0.958	0.945

Hcb= Hc-B4-75, Rb= Reba W296, Frb= Frego Bract, Dk= Dixie King, Dp4= Dp 4049, Sg= Sg 125, Cyt= Cyto 12/74, Rb5= Reba B50, N88= NTA 88-6, Hgp= High Gossypol, Des= Des 119, MZ= MZ561, Mcn= Mcnair 235, IL8= IL85, IL7= IL74, Gua= Guazuncho, Del= Delcot 344, Stn= Stoneville, Aub= Auburn 56, N93= NTA 93-15, Bja= BJA 592, UK9=UK91, Irm= Irma 1243, UK8= UK82, Okl= Super Okra Leaf, Acl= Acala SJ-2

FIGURES

Figure 1: Dendrogram generated based on UPGMA clustering method and Dice coefficient using AFLP analysis among 26 cotton cultivars