Crop Breeding and Applied Biotechnology 5:1-10, 2005
Brazilian Society of Plant Breeding. Printed in Brazil

Analysis of cotton genetic diversity by microsatellites and pedigree

Cândida Hermínia Campos de Magalhães Bertini[1*], Ivan Schuster[2], Tocio Sediyama[3], Everaldo, Gonçalves de Barros[1], and Maurílio Alves Moreira[1]

Received 14 March 2005

Accepted 2 July 2005

ABSTRACT - *The main objective of this study was to verify the association between the coefficient of parentage estimates and multivariate techniques which were used as measures of genetic diversity in cultivars. Thirty cotton cultivars were used for this purpose among which genetic diversity was estimated by means of the parentage coefficient (CP) and also through multivariate techniques employing microsatellite markers (SSR). The correlation between genetic distances obtained by CPs and SSRs for cultivars was positive and significant, with a value of 0.25. The 18 ancestors evaluated in the current study contributed with 69% to the genetic constitution of the 30 cotton cultivars. The evidence that few ancestors actually contribute to the genetic constitution of the cotton cultivars used in Brazil indicates that new alleles should be introduced into the gene pool of these cultivars in order to broaden the genetic base of cotton.*

Key words: coefficient of parentage, genetic distance, *Gossypium hirsutum* L., molecular markers.

INTRODUCTION

Plant breeders have been assessing the levels of genetic diversity either through pedigree analysis or multivariate analysis which is undertaken based on some plant attributes (morphological features, agronomic performance, isozymes and polymorphisms at DNA level) (Van Esbroeck et al. 1999).

Similarity or genetic diversity have been estimated by means of pedigree analysis for a great number of crops, such as wheat (Kim and Ward 1997), barley (Graner et al. 1994), maize (Messmer et al. 1993), rice (Xu et al. 1999) and cotton (Bowman et al. 1996, Van Esbroeck et al. 1998, Van Esbroeck et al. 1999). In a comparison with other methods, the estimation of genetic distances between genotypes by pedigree analysis is advantageously cheaper. On the other hand, the method features the following disadvantages: 1) lack of detailed information in connection with pedigree genotype, and 2) the fact that the parentage coefficient, which is used to estimate genetic similarity, is calculated based on unrealistic pre-assumptions, such as (i) all lines, including parental and ancestral lines, are homozygous and homogeneous; (ii) lines without known common parentage are unrelated to each other and (iii) a line derived from a cross obtained half of its genes from each parent.

The main traits used for multivariate cotton analysis include morphological traits (Tatineni et al. 1996), isozymes

[1] Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), Universidade Federal de Viçosa, 36.570-000, Viçosa, MG, Brasil. *E-mail: candida_bertini@yahoo.com.br
[2] COODETEC, C. P. 301, 85.813-450, Cascavel, PR, Brasil
[3] Departamento de Fitotecnia, UFV

(Wendel et al. 1992), and polymorphism at DNA level (Tatineni et al. 1996). Polymorphism at DNA level is currently considered the most accurate marker to estimate genetic distance because it is found in a higher number and is unaffected by environmental influences (Van Esbroeck et al. 1999). Low polymorphism at the DNA level, as observed in cotton, points to a lower genetic diversity level than that obtained through pedigree analysis (Wendel et al. 1992). Relationships between methods that use molecular markers and parentage coefficient to estimate genetic diversity varied depending on the species under study and the sampled plant material (Kim and Ward 1997).

Van Esbroeck et al. (1999) found no relation between pedigree and similarity measurements based on morphological and agronomic features in cotton. Tatineni et al. (1996) detected that the genetic similarity presented a correlation of 0.63 between values based on RAPD (Random Amplified Polymorphic DNA) markers and on morphological features between cotton lines. There is very little information concerning the correlation between genetic distances calculated with molecular markers and the coefficient of parentage for cotton.

The objectives of this study were: (1) to verify the association between the estimates of coefficient of parentage and the multivariate technique calculated based on microsatellite markers employed as measures of genetic diversity in cotton cultivars and, (2) to evaluate the genetic contribution of ancestors to some cotton cultivars used in Brazil.

MATERIAL AND METHODS

Plant material

Thirty cotton (*Gossypium hirsutum* L.) cultivars from public and private breeding companies in Brazil, Argentina, and Paraguay were used (Table 1). These cultivars are indicated for cultivation in several regions of Brazil. Some of them are of great commercial importance, while the others are employed in breeding programs.

DNA extraction and SSR analysis

DNA from cultivars was extracted from a bulk of ten seeds. The extraction was carried out based on a protocol described by McDonald et al. (1994). DNA quality was evaluated by agarose gel (0.8%) as well as by spectrophotometry, taking the A260/A280 ratio into account. The concentration was estimated from an absorbance of 260 nm, according to Sambrook et al. (1989). The cotton microsatellite primers we used were from Brookhaven National Laboratory (BNL) and 31 primer pairs were used to evaluate the cultivars.

Microsatellite reactions were carried out in 0.2 mL micro-tubes and the total volume of the reaction contained 15 µL, consisting of 30 ng DNA template, 0.2 µM of each primer, 1 unit of Taq DNA polymerase, 0.2 mM each dNTP, 2.0 – 3.0 mM of $MgCl_2$ (Table 2), and 1X reaction buffer (10 mM Tris-HCl and 50 mM KCl, pH 8.3). The amplification was carried out in a Perkin Elmer thermocycler (GeneAmp PCR System 9600) and consisted of a denaturation step of 4 min at 94 ºC followed by a touchdown profile. The profile started with 10 cycles of 40 s at 94 °C, a pairing step of 40 s at 65 °C, (decreasing 1 °C per cycle until 55 °C) and 1 min at 72 °C. After that, 30 cycles of 40 s at 94 °C, 40 s at 55 °C and 1 min at 72 °C were performed. The program ended with one polymeration cycle at 72 °C for 7 min.

The amplified fragments were separated in polyacrylamide 7% denaturing gel (w/v), with acrylamide/bis-acrylamide in a 19:1 proportion, 32% formamide and 5.6 M urea, according to the protocol proposed by Litt et al. (1993). The running conditions consisted of 30 min of prerunning and 3 h of running at constant power (80 W). The running buffer consisted of 1X TBE (10.8 g Tris base, 5.5 g boric acid and 0.83 g EDTA). After electrophoresis, the denaturing gels were immersed in an ethidium bromide (1 µg ML^{-1}) solution for 30 minutes. The gels were photographed under ultraviolet light (Eagle Eye II) after staining.

Coefficient of parentage (CP)

Malecot's (1948) coefficients of parentage were estimated in 30 cultivars, matched in pairs, amounting to totally 435 cultivar pairs. We used cultivars with known information about parents and/or grandparents. The CP values were calculated according to assumptions described in detail by Van Esbroeck et al. (1999). Each CP was used to estimate the contribution of an ancestor cultivar to the genetic constitution of all 30 cultivars. For each ancestor cultivar, the relative genetic mean contribution to all cultivars resulted in the relative genetic contribution (RGC) of each ancestor to the gene pool of the evaluated cotton cultivars. The ancestor cultivars were classified in an order of decreasing magnitude corresponding to each RGC. The ancestor cultivar frequency in the genealogy (FAG) of the 30 cultivars was also calculated.

Table 1. Cotton cultivars analyzed in this study with their descriptive data

Nr	Origin	Cultivars	Pedigree	Region of planting or state (Country)
1	IAC	IAC 17	Selection in IAC RM_3	São Paulo (Brazil)
2	IAC	IAC 19	Yucatanense/N1-HOA//IACRM_3	São Paulo (Brazil)
3	IAC	IAC 20	Selection in IAC 17	São Paulo (Brazil)
4	IAC	IAC 21	Selection in IAC 19	São Paulo (Brazil)
5	IAC	IAC 22	IAC 20/GH 11–9–75	SP/ Central West (Brazil)
6	EPAMIG	Redenção	Selection in IAC 17	Minas Gerais (Brazil)
7	EPAMIG	Epamig 5	Selection in C-25-1-80	Minas Gerais/ Central West (Brazil)
8	EPAMIG	Alva	Double Haploid (C-25-1-80)	Minas Gerais (Brazil)
9	EMBRAPA	CNPA 7H	TAMCOT SP 37/IAC 17	NE/Central South (Brazil)
10	EMBRAPA	CNPAPrec.1	Selection in GH-11-9-75	North/NE/Central West (Brazil)
11	EMBRAPA	CNPAPrec.2	C-100-7-81/PNH3	North/NE/Central West (Brazil)
12	EMBRAPA	CNPA ITA90	Selection in Deltapine AC-90	Central West (Brazil)
13	FundaçãoMT	BRS 96	Selection in EPAMIG 3	Central West (Brazil)
14	EMBRAPA	BRS Ipê	Selection in CNPA ITA 90	Central West (Brazil)
15	EMBRAPA	BRS Itaúba	Selection in CS 50	Central West (Brazil)
16	EMBRAPA	BRS 96-148	Selection in CS 50	Central West (Brazil)
17	EMBRAPA	BRS 96-227	Selection in CS 50	Central West (Brazil)
18	IAPAR	IPR 94	IAPAR 71/Deltapine Acala 90	Paraná (Brazil)
19	IAPAR	IPR 95	CNPA ITA 90/IAPAR 71	Paraná (Brazil)
20	IAPAR	IPR 96	CNPA ITA 90/IAPAR 71	Paraná (Brazil)
21	COODETEC	CD 401	SP86/ISA205	MS, PR and SP (Brazil)
22	COODETEC	CD 402	DP Ac 90//IAC 20/S295*IAC20	BA, GO, MT, MS, MG, SP (Brazil)
23	COODETEC	CD 403	DP Ac 90//IAC 20/S295*IAC20	BA, GO, MT, MS, MG, SP (Brazil)
24	COODETEC	CD 404	CHACO 520/DP Ac90	MS, MT and PR (Brazil),
25	COODETEC	CD 98-440	DP Ac 90//IAC 20/S295*IAC20	-
26	COODETEC	CD98-383	DP Ac90//IAC 20/S295	BA, GO, MG, SP, MT, PR, MS and Northern Region (Brazil),
27	Paraguay	IAN 338	CHACO 510/ISA 205//Reba P279	(Paraguay)
28	Argentina	Cacique	MATACO/GUAZUNCHO	(Argentina)
29	Argentina	Guazuncho2	Guazuncho/SP 8535	(Argentina/ Paraguay)
30	Argentina	Oro Blanco	SP2473/SIOKRA	(Argentina)

/, // - refer to the order in which the crossings were realized; (*) refer to backcross

Data analysis

Genetic diversity of each SSR locus was obtained from allele frequency using the following formula

$$PIC = 1 - j = 1 - \sum_{j=1}^{n} p_{ij}^2$$

Where PIC is Polymorphism Information Content and p_{ij} is the frequency of the j^{th} allele for primer *i* (Anderson et al. 1993).

The genetic distances between the cultivars obtained with SSR markers were evaluated through a dissimilarity matrix built using the similarity index

complement (SI) for codominant/multiallelic variables. The software Genes (Cruz 2001) was employed to calculate this index which was used to estimate the similarity between genotypes with scores of 0, 1 and 2, for the absent allele, heterozygous and homozygous, respectively. The coefficients of parentage values obtained for cultivars were subtracted from one to obtain the dissimilarity matrix.

The method employed to develop the cluster analysis based on the dissimilarity matrix obtained through SSR markers and coefficient of parentage was UPGMA (Unweighted pair-group method using an arithmetic average) of the agglomerative hierarchic type. The dendrogram obtained by matrix dissimilarity was developed using the Statistica software package (StatSoft Inc. 1999). The association between genetic distances, calculated based on the molecular data and coefficient of parentage, was evaluated by Pearson's coefficient of correlation (r). The correlation significance level was evaluated by the Mantel Z statistic (Mantel 1967). Significance of Z was determined by comparing the observed Z values with a critical Z value obtained by calculating Z for one matrix with 5000 permuted variants of the second matrix.

RESULTS AND DISCUSSION

The 31 primer pairs used to evaluate the 30 cultivars amplified 65 alleles with an average of 2.10 alleles per SSR locus. The PIC value calculated to estimate the informativeness of each primer varied from 0.10 to 0.62 with an average of 0.39 (Table 2). In the study developed by Liu et al. (2000) the PIC value varied from 0.05 to 0.82 with an average value of 0.31. A possible reason for the low polymorphism observed in the plant material in the present study is that all evaluated material came from breeding programs and could therefore hve a narrow genetic base. On the contrary, the material used in the study of Liu et al. (2000) consisted of 97 accessions of *G. hirsutum* derived from several wild species, which explains the higher polymorphism found by these authors (5 alleles locus^{-1}). However, it must be highlighted that the PIC average value found by these authors was equal to 0.31, i.e., the polymorphism they found was low even when the PIC general mean was taken into account for all loci.

The genetic dissimilarity calculated with the SSR markers for the 435 pairs of cultivars varied from 0.097 to 0.71, presenting an average of 0.42 ± 0.01, while the genetic dissimilarity calculated based on coefficients of parentage varied from 0.25 to 1.00, presenting an average of 0.89 ± 0.01. Most cultivar pairs had dissimilarity values between 0.4 and 0.6 when using SSR markers, but values between 0.8 and 1.0 were obtained when using coefficients of parentage. This means that a higher divergence is observed among cultivars when dissimilarity is evaluated by coefficients of parentage.

Results of other studies also showed low genetic diversity in cotton cultivars by an approach through molecular markers. Multani and Lyon (1995) detected high values of genetic similarity (92.1 - 98.9%) in nine Australian cotton cultivars using RAPD markers. Iqbal et al. (1997) also found high genetic similarity (0.82 to 0.93%) in 17 *G. hirsutum* cultivars calculated based on RAPD markers. Employing SSR markers, Gutiérrez et al. (2002) detected a narrow genetic base of Australian and American cultivars.

On the other hand, when coefficient of parentage (CP) is used to estimate the genetic dissimilarity between cotton cultivars, the genetic diversity is higher than that obtained through molecular markers. Bowman et al. (1996) detected a medium value of 0.07 for coefficients of parentage estimated among 260 cotton cultivars. The low level of similarity between the cultivars evaluated in these studies suggests that the pedigree analysis overestimates the genetic diversity level in cotton cultivars. Van Esbroeck et al. (1998) who evaluated 24 cotton ancestor cultivars found a CP medium value of 0.16. Evaluating the consequence of their mistake in assuming that the cultivars were not related, Van Esbroeck et al. (1999) re-calculated the CP for these cultivars and assumed the existence of a relationship between the ancestors, considering a CP of 0.38 between them. A medium CP value of 0.46 was obtained, i.e., the CP passed from 0.16 to 0.46, expressing a reduction of the genetic diversity level in these cultivars.

According to Bowman et al. (1996), over 30% of the cotton cultivars released between 1970 and 1990 in the United States were obtained from selection within other cultivars (reselection). Although cotton is considered an autogamous plant, allogamy rates of over 50% have already been observed when pollinating insects (*Bombus spp* and *Aphis mellifera* L.) were present. In the pedigree analysis, the CP calculated for a reselection is considered to be 0.75, that is, a cultivar originated from another has a chance of 50% of being originated from self-pollination of non-homozygous plants in a non-uniform population, and 50% of being originated from cross-fertilization with a non-related individual. The frequent reselections observed in cotton have the effect of continually incorporating

Table 2. SSR primers used in the analysis of the genetic diversity of 30 cotton cultivars

SSR Locus	Chromosome location	$MgCl_2$ (mM)	Product size (pb)	Nr of alleles	Allele frequency	PIC
BNL139	-	3	150-170	3	0.08; 0.73; 0.18	0.42
BNL 946	***20Lo****	2.5	330-350	2	0.83; 0.17	0.29
BNL 1053	*3*	2	170-190	2	0.68; 0.32	0.43
BNL 1064	*6sh***	2.5	130-140	2	0.05; 0.95	0.10
BNL1231	*21**	2.5	170-200	2	0.78; 0.22	0.34
BNL1423	*9**	3	130-140	2	0.58; 0.42	0.49
BNL 1673	*12Lo**	2.5	300-360	2	0.07; 0.93	0.12
BNL 1694	*7**	2.5	230-260	2	0.48; 0.52	0.50
BNL 1721	*18Lo***	2.5	170-180	2	0.22; 0.78	0.34
BNL 2448	*5**	2.5	130-140	2	0.82; 0.18	0.30
BNL 2449	*A01**	3	140-170	3	0.70;0.02; 0.28	0.43
BNL 2495	*26Lo**	2.5	190-200	2	0.63; 0.37	0.46
BNL 2496A	-	3	110-120	2	0.75; 0.25	0.38
BNL 2590	*9Lo***	2.5	180-190	2	0.79; 0.21	0.33
BNL 2646	*15**	3	120-150	2	0.21; 0.79	0.33
BNL 2921	-	2.5	150-160	2	0.52; 0.48	0.50
BNL 2960	*10Lo***	3	140-150	2	0.53; 0.47	0.50
BNL 2986	*16Lo**	3	150-160	2	0.54; 0.46	0.50
BNL 3089	-	2.5	140-150	2	0.90; 0.10	0.18
BNL 3171	-	2.5	210-230	2	0.27; 0.73	0.39
BNL 3255	*5sh***	3	220-240	2	0.47; 0.53	0.44
BNL 3257	*8**	2.5	200-220	3	0.47; 0.37; 0.17	0.62
BNL 3408	*17Lo***	2.5	140-150	2	0.62; 0.38	0.47
BNL 3482	*26Lo***	2.5	120-130	2	0.74; 0.26	0.38
BNL 3590	*2**	2	170-190	2	0.64; 0.36	0.55
BNL 3594	*6bot**	2.5	170-190	2	0.85; 0.15	0.26
BNL 3800	-	2	180-190	2	0.90; 0.10	0.18
BNL 3838	*20**	2.5	120-130	2	0.69; 0.31	0.43
BNL 3902	*15**	2	170-200	2	0.58; 0.42	0.49
BNL 4030	*22**	2.5	110-120	2	0.27; 0.73	0.39
CNL 101	-	2.5	120-130	2	0.42; 0.58	0.49
Total				65		
Mean				2.10		0.39

sh –short arm; *Lo* – long arm; *, ** Information obtained in Lacape et al. (2003) and Liu et al. (2000), respectively. Chromosome 1 to 13 belongs to subgenome A and 14 to 26 belongs to subgenome D

Table 3. Identification, origin, relative genetic contribution (RGC), accumulated genetic contribution (AGC), accumulated number of ancestors (ANA) and frequency of ancestors in genealogy (FAG) of the 30 cotton cultivars

ANCESTOR			RGC%	AGC%	ANA	FAG%
Letter	Identification	Origin				
A	Auburn56	USA	12.70	12.7	1	50
B	Tamcot-SP37	USA	8.90	21.6	2	30
C	DP Smoothleaf	USA	7.74	29.34	3	57
D	DP 45	USA	6.07	35.41	4	43
E	CHACO 510	Argentina	5.00	40.41	5	20
F	Stoneville 213	USA	4.20	44.61	6	10
G	Reba B50	USA	3.96	48.57	7	20
H	John Cotton Polycross	USA	3.18	51.75	8	23
I	AZ 5909	USA	3.20	54.95	9	23
J	ISA 205	Africa	2.50	57.45	10	7
K	S 295	Africa	2.08	59.53	11	13
L	Mataco	Argentina	1.70	61.23	12	3
M	Siokra	Australia	1.70	61.23	13	3
N	*G. hirsutum* r. *yucatanense*	Central America	1.50	64.43	14	7
O	Tn1-Hoa	-	1.50	65.93	15	7
P	Toba/Hopicala/DP16	Triple hybrid/USA	1.46	67.39	16	10
Q	Acala 1517-70	USA	0.94	68.33	17	10
R	DPL	USA	0.70	69.03	18	3

unrelated germplasm into CP estimates and thus overestimating the diversity among cultivars.

Most cultivars released in Brazil are results of cultivar and line introductions from the USA or from reselection carried out with other previously existing cultivars. Of the 30 cultivars evaluated in the current study, 12 were obtained through reselection. The genetic contribution of the 18 ancestors to the genetic constitution of the 30 cotton cultivars evaluated in the present study is displayed in Table 3. The pedigree we took into account varied from 33.8 to 100%, with a mean of 68.74% (Table 4). Therefore, these ancestors contributed with nearly 69% to the genetic constitution of the 30 cultivars (Table 3).

Dendrograms obtained from genetic dissimilarity measures, calculated with SSR markers and coefficients of parentage for the 30 cultivars, are shown in Figure 1. Eighteen groups were obtained when using coefficients of parentage (CP) and considering a superior limit of 60% and 15 groups when using microsatellite markers (SSR), taking a superior limit of 30% into account. In a comparison of results the groups obtained by both the UPGMA and Tocher methods (data not shown) for cotton cultivars presented little agreement in clustering based on CPs or SSRs methods.

The correlation between genetic distances obtained by CPs and SSRs presented a value of 0.25 between cultivars. This value was positive and significant for a probability of 1% based on 5000 simulations. Significant correlations, but with values from low (r=0.21) to moderate (r=0.42), between genetic similarity (GS) obtained with RFLP markers or coefficient of parentage were also found by Graner et al. (1994) in wheat. On the other hand, Barbosa Neto et al. (1996) found a negative correlation (r=-0.33) between genetic similarity obtained with RFLP markers and coefficient of parentage for wheat lines. Kim and Ward (1997) found a high correlation (r=0.73) between GS calculated with RFLP

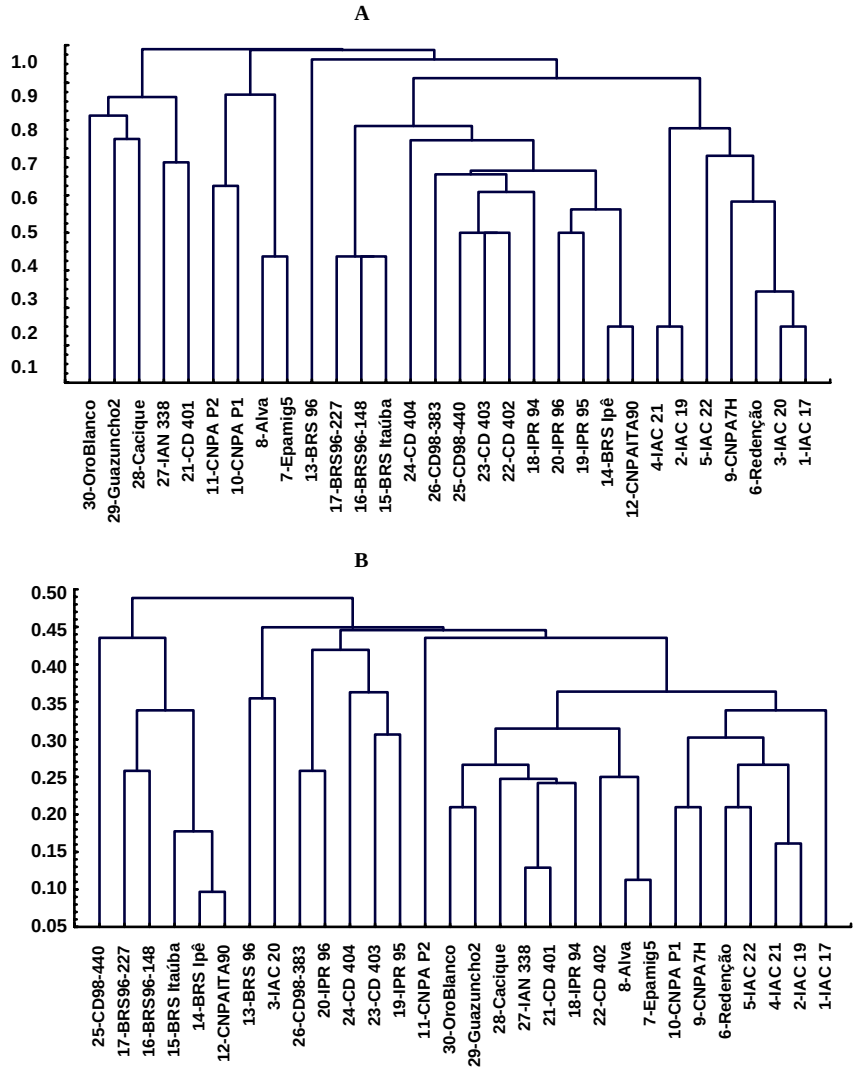

Figure 1. Genetic distance between 30 cotton cultivars obtained by coefficient of parentage (A) and microsatellite markers (B), using UPGMA (Unweigthed pair-group method using an arithmetic average) for the cluster analysis

markers and coefficient of parentage when all line pairs of winter wheat were considered, that is, red wheat (RW) and white wheat (WW). However, when RW lines and WW lines were separately considered, the correlation decreased to 0.23 for RW lines and 0.28 for WW lines.

In rice, Xu et al. (1999) found a correlation of 0.092 between genetic distances calculated with microsatellites and pedigree analysis. On the other hand, Plaschke et al. (1995), when comparing GS estimates between wheat cultivars calculated with SSR and coefficient of parentage, found a correlation of 0.55. Employing only CP values higher than 0.125 in the analysis of cultivar pairs, Tinker et al. (1993) detected a correlation of 0.61 between GS calculated with RAPD markers and CP.

According to Graner et al. (1994), the methods employed to estimate genetic similarity calculated with molecular markers and coefficient of parentage present different approaches. They are based on different kinds of underlying information and are therefore subjected to different error sources. The coefficient of parentage between two individuals *i* and *j* (f_{ij}) is defined as the probability of homologous genes taken at random, one from each parent, being identical by lineage. In contrast to this fact, the genotypic similarity between individuals is

Table 4. Genetic constitution of the 30 cotton cultivars obtained by their pedigree

Cultivars	Genetic constitution[a]	Pedigree (%)
1. IAC 17	0.562A	56.2
2. IAC 19	0.375A + 0.25N + 0.25O	87.5
3. IAC 20	0.422A	42.2
4. IAC 21	0.281A + 0.187N + 0.187O	65.5
5. IAC 22	0.211A + 0.375B	58.6
6. REDENÇÃO	0.422A	42.2
7. EPAMIG 5	0.281B + 0.375F	65.6
8. ALVA	0.281B + 0.375F	65.6
9. CNPA 7H	0.281A + 0.5B	78.1
10. CNPA P1	0.562B	56.2
11. CNPA P2	0.375B + 0.5F	87.5
12. CNPA ITA 90	0.187D + 0.187C + 0.187H + 0.187I	74.8
13. BRS 96	0.211R + 0.211A	42.2
14. BRS IPÊ	0.141D +0.141C + 0.141H + 0.141I	56.4
15. BRS ITAÚBA	0.164D + 0.164C + 0.094Q + 0.094B	51.6
16. BRS 96-148	0.164D + 0.164C + 0.094Q + 0.094B	51.6
17. BRS 96-227	0.164D + 0.164C + 0.094Q + 0.094B	51.6
18. IPR 94	0.158A + 0.125D + 0.125C + 0.125H + 0.125I	65.8
19. IPR 95	0.158A + 0.094D + 0.094C	33.8
20. IPR 96	0.158A + 0.094D + 0.094C	33.8
21. CD 401	0.25E + 0.125G + 0.125C +0.5I	100
22. CD 402	0.125D +0.125C +0.125H + 0.125I + 0.158A +0.125K	78.3
23. CD 403	0.125D +0.125C +0.125H + 0.125I + 0.1582A +0.125K	78.3
24. CD 404	0.187D + 0.125C + 0.0.125H + 0.125I + 0.187E + 0.187G + 0.062P	100
25. CD 98-440	0.125D +0.125C +0.125H + 0.125I + 0.1582A +0.125K	78.3
26. CD 98-383	0.125D + 0.125C + 0.105A + 0.25K	60.5
27. IAN 338	0.25E + 0.25J +0.25G + 0.25C	100
28. CACIQUE	0.5L + 0.25E + 0.25G	100
29. GUAZUNCHO 2	0.375E + 0.25G + 0.125C +0.25P	100
30. ORO BLANCO	0.187E + 0.125G + 0.062C + 0.125P + 0.5M	100
Mean		68.74

[a]Letters used as symbols in Table 4

based on similar genes in state, that is, genes undistinguishable in their effects. However, genes which are only similar in state but not identical by lineage are ignored in the calculation of the coefficient of parentage (Messmer et al. 1993). On the other hand, genetic similarity when estimated by molecular markers displays the similarity between genotypes as from a direct genome sample, reflecting similarity in state and lineage (Graner et al. 1994).

Therefore, this estimate may be more understandable than that obtained by coefficient of parentage that takes only the identity by lineage into account.

One great advantage of the genetic similarity analysis based on genealogy is its low cost. Disadvantages are however the need for detailed information about the evaluated cultivar genealogy and the fact that the coefficient of parentage is calculated based on some unrealistic pre-assumptions. The lack of ancestors' records hinders diversity base on genealogical studies. In fact, wherever the genealogy of some ancestors, under study here, was unknown, these ancestors were considered to be unrelated. If, on the contrary, these ancestors were related, the diversity between cultivars would be overestimated. It was assumed that a cultivar inherits 50% of its alleles from each parent. However, studies developed with molecular markers indicate that under intense selection this value may deviate up to 20% (Bernardo et al. 1996). Results obtained from morphological data suggest that the breeders have selectively favored specific traits and, as a result, favored certain genes. Therefore, deviations owing to selection and genetic drift are expected, but such deviations are not taken into account in the calculation of coefficient of parentage.

To obtain a more accurate estimate of genetic similarity based on microsatellite markers will depend on the SSR marker number used in the study and the distribution of these markers in the evaluated cultivar genome. Cotton (*G. hirsutum* L.) presents a number of chromosomes equal to 2n = 4x = 52. In the current study, the number of SSR primer pairs used in the cultivar evaluation was 31. This would correspond, in the mean, to one pair of primers per chromosome in the cotton genome.

The number of markers to be sampled is important concerning the accuracy of genetic distance estimates. In this respect, several researchers are concerned about the question of how to quantify the accuracy of genetic distance estimates and how to determine an ideal number of markers to be employed. (Tivang et al. 1994). However, there are some difficulties for setting up of variances for these estimates. Besides, information in literature is still scarce concerning the ideal number of molecular markers in the determination of genetic diversity both for natural populations as well as for improved cultivars.

CONCLUSIONS

In the current study, a significative association between the methods employed to estimate genetic diversity based on genealogical analysis and multivariate analysis obtained by microsatellite markers can be verified. The studied cotton cultivars descend from few ancestors, such as Auburn 56, Tamcot SP-37, DP Smoothleaf and DP 45, suggesting that there is the need of introducing new alleles into these cultivars' gene pool in order to avoid genetic vulnerability.

ACKNOWLEDGEMENTS

The authors are most grateful to Cooperativa Central de Pesquisa Agrícola (COODETEC), especially to researcher Jean Louis Belot, EMPRAPA, EPAMIG, IAC and IAPAR for their support and for supplying the cotton seeds analyzed in the current study. This study was financially supported by CNPq – Conselho Nacional de Pesquisa e Desenvolvimento.

Análise da diversidade genética do algodoeiro por meio de marcadores microssatélites e genealogia

RESUMO - *O objetivo geral desse trabalho foi verificar a associação entre as estimativas dos coeficientes de parentesco e de técnica multivariada como medidas de diversidade genética de cultivares de algodoeiro. Para este propósito foram utilizados 30 cultivares de algodoeiro herbáceo. A diversidade genética entre cultivares foi estimada por meio dos coeficientes de parentesco (CP) e, por meio de técnica multivariada utilizando-se marcadores microssatélites (SSR). A correlação entre as distâncias genéticas obtidas pelos CPs e SSRs para os cultivares foi positiva e significativa, com valor igual a 0,25. Os 18 ancestrais avaliados no trabalho contribuíram com 69% para a constituição genética dos 30 cultivares. A constatação de que poucos ancestrais contribuem para a constituição genética dos cultivares de algodoeiro usados no Brasil, sugere maior preocupação em introgredir novos alelos no* pool *gênico desses cultivares.*

Palavras-chave: Coeficiente de parentesco, distância genética, *Gossypium hirsutum* L., marcadores moleculares.

REFERENCES

Anderson JA, Churchill GA, Autrique JE, Tanksley SD and Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. **Genome 36**: 181-186.

Barbosa-Neto JF, Sorrells ME and Cisar G (1996) Prediction of heterosis in wheat using coefficient of parentage and RFLP-based estimates of genetic relationship. **Genome 39**: 1142-1149.

Bernardo R, Murigneux A and Karaman Z (1996) Marker-based estimates of identity by descent and alikeness in state among maize inbreds. **Theoretical and Applied Genetics 93**: 262-267.

Bowman DT, May OL and Calhoun DS (1996) Genetic base of upland cotton cultivars released between 1970 and 1990. **Crop Science 36**: 577-581.

Cruz CD (2001) **Programa Genes: versão Windows: aplicativo computacional em genética e estatística**. Viçosa, UFV, 648p.

Graner A, Ludwig WF and Melchinger AE (1994) Relationships among European barley germplasm: II, Comparison of RFLP and pedigree data. **Crop Science 34**: 1199-1205.

Gutiérrez OA, Basu S, Saha S, Jenkins JN, Shoemaker DB, Cheatham CLA and McCarty JC (2002) Genetic distances among selected cotton genotypes and its relationship with F_2 performance. **Crop Science 42**: 1841-1847.

Iqbal MJ, Aziz N, Saeed NA and Zafar Y (1997) Genetic diversity evaluation of some elite cotton varieties by RAPD analysis. **Theoretical and Applied Genetics 94**: 139-144.

Kim HS and Ward RW (1997) Genetic diversity in Eastern U.S. soft winter wheat (*Triticum aestivum* L. Thell.) based on RFLPs and coefficients of parentage. **Theoretical and Applied Genetics 94**: 472-479.

Litt M, Hauge X and Sharma V (1993) Shadows bands seen when typing polymorphic dinucleotide repeats: some causes and cures. **Biotechniques 15**: 280-284.

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

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. **Cancer Research 27**: 209-220.

McDonald MB, Elliot LJ and Sweeney PM (1994) DNA extraction from dry seeds for RAPD analyses in varietal identification studies. **Seed Science & Technology 22**: 171-176.

Messmer MM, Melchinger AE, Hermann RG and Boppenmaier J (1993) Relationships among early European maize inbreds: II. Comparison of pedigree and RFLP data. **Crop Science 33**: 944-950.

Multani DS and Lyon BR (1995) Genetic fingerprinting of Australian cotton cultivars with RAPD markers. **Genome 38**: 1005-1008.

Plaschke J, Ganal MW and Roder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. **Theoretical and Applied Genetics 91**: 1001-1007.

Sambrook J, Fritsch EF and Maniatis T (1989) **Molecular cloning: laboratory manual**. 2nd edition, CSHL, Cold Spring Harbor.

StaSoft Inc (1999) **STATISTICA for Windows [Computer program manual]**. http://www.statsoft.com

Tatineni V, Cantrell RG and Davis DD (1996) Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. **Crop Science 36**: 186-192.

Tinker NA, Fortin MG and Mather DE (1993) Random amplified polymorphic DNA and pedigree relationships in spring barley. **Theoretical and Applied Genetics 85**: 976-984.

Tivang JG, Nienhuis J and Smith OS (1994) Estimation of sampling variance of molecular marker data using the bootstrap procedure. **Theoretical and Applied Genetics 89**: 259-264.

Van Esbroeck GA, Bowman DT, Calhoun DS and May OL (1998) Changes in the genetic diversity of cotton in the USA from 1970 to 1995. **Crop Science 38**: 33-37.

Van Esbroeck GA, Bowman DT, May OL and Calhoun DS (1999) Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of modern cultivars. **Crop Science 39**: 323-328.

Wendel JF, Brubaker CL and Percival AE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. **American Journal of Botany 79**: 1291-1310.

Xu W, Virmani SS and Li ZK (1999) Association between simple sequence repeat (SSR) marker diversity, pedigree record, quantitative trait variation, and hybrid performance in rice. **International Rice Research Notes 24**: 10-11.