E

BMT-TWA/Sugarcane/1/2
ORIGINAL: English
DATE: September 11, 2002

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

AD HOC CROP SUBGROUP ON MOLECULAR TECHNIQUES FOR SUGARCANE

First Session
Rio de Janeiro (Brazil), September 27, 2002

ANALYSIS OF GENETIC SIMILARITY DETECTED BY AFLP AND THE COEFFICIENT OF PARENTAGE AMONG GENOTYPES OF SUGARCANE (*SACCHARUM SPP*.)

Document prepared by experts from Brazil

Theor Appl Genet (2002) 104:30–38

M.L.A. Lima · A.A.F. Garcia · K.M. Oliveira
S. Matsuoka · H. Arizono · C.L. de Souza Jr.
A.P. de Souza

Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.)

Received: 27 November 2000 / Accepted: 27 April 2001

Abstract Despite the economical importance of sugar cane, until the present-date no studies have been carried out to determine the correlation of the molecular-based genetic similarity (GS) and the coefficient of parentage (*f*)-estimates generated for cultivars. A comprehensive knowledge of the amount of genetic diversity in parental cultivars, could improve the effectiveness of breeding programmes. In this study, amplified fragment length polymorphism (AFLP) and pedigree data were used to investigate the genetic relationship in a group of 79 cultivars (interspecific hybrids), used as parents in one of the Brazilian breeding programmes, and four species of *Saccharum* (*Saccharum sinense*, *Saccharum barberi* and two of *Saccharum officinarum*) . The objectives of this study were to assess the level of genetic similarity among the sugar-cane cultivars and to investigate the correlation between the AFLP-based GS and *f*, based on pedigree information. Twenty one primer combinations were used to obtain the AFLP molecular markers, generating a total of 2,331 bands, of which 1,121 were polymorphic, with a polymorphism rate, on average, of 50% per primer combination. GSs were determined using Jaccard's similarity coefficient, and a final dendrogram was constructed using an unweighted pair-group method using arithmetic average (UPGMA). AFLP-based GS ranged from 0.28 to 0.89, with a mean of 0.47, whereas *f* ranged from 0 to 0.503, with a mean of 0.057. Cluster analysis using GS divided the genotypes into related subgroups suggesting that there is important genetic relationship among the cultivars. AFLP-based GS and *f* were significantly correlated ($r = 0.42$, $P < 0.001$), thus the significance of this *r* value suggests that the AFLP data may help to more-accurately quantify the degree of relationship among sugar-cane cultivars.

Keywords *Saccharum* spp. · AFLP · Genetic similarity · Coefficient of parentage

Communicated by P. Langridge

M.L.A. Lima · K.M. Oliveira · A.P. de Souza (✉)
Centro de Biologia Molecular e Engenharia Genética (CBMEG),
Universidade Estadual de Campinas (UNICAMP),
Cidade Universitária "Zeferino Vaz", CP 6010, CEP 13083-970,
Campinas, SP, Brasil
e-mail: anete@unicamp.br
Tel.: (55-19) 3788-1132 or 3788-1130, Fax: (55-19) 3788-1089

M.L.A. Lima · K.M. Oliveira · A.P. de Souza
Departamento Genética e Evolução – Instituto de Biologia –
Universidade Estadual de Campinas (UNICAMP),
Cidade Universitária "Zeferino Vaz", CP 6109, CEP 13083-970,
Campinas, SP, Brasil

A.A.F. Garcia · H. Arizono
Departamento de Ciências Exatas,
Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ),
Universidade de São Paulo (USP), CP 9, CEP 13400-970,
Piracicaba, SP, Brasil

S. Matsuoka
Departamento de Biotecnologia Vegetal (DBV),
Centro de Ciências Agrárias (CCA),
Universidade Federal de São Carlos (UFSCar), CEP 13600-970,
Araras, SP, Brasil

C.L. de Souza Jr.
Departamento de Genética,
Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ),
Universidade de São Paulo (USP), CP 83, CEP 13400-970,
Piracicaba, SP, Brasil

Present address:
A.P. de Souza,
Centro de Biologia Molecular e Engenharia Genética (CBMEG),
Universidade Estadual de Campinas (UNICAMP),
Cidade Universitária "Zeferino Vaz", CP 6010, CEP 13083-970,
Campinas, SP, Brasil

Introduction

Modern sugar-cane cultivars are largely the results of intercrossings from the first interspecific crosses carried out at the beginning of the century, involving essentially *Saccharum officinarum, Saccharum spontaneum* and *Saccharum barberi*. The hybridizations used a process known as "nobilization", which corresponds to a series of backcrosses with *S. officinarum*, the noble cane (Bremer 1961a). This nobilization proved to be the main step in the genetic improvement of sugar cane, solving

problems of susceptibility to diseases, while increasing production and adaptability (Roach 1972). However, as few parental cultivars participated in the initial crossings, concerns about the limited genetic base of modern sugar-cane cultivars have increased (Arceneaux 1968; Tai and Miller 1978; Roach and Daniels 1987; Lu et al. 1994a; Deren 1995).

A successful breeding program depends on the complete knowledge and understanding of the genetic diversity of the available germplasm. Several methods have been used to investigate the genetic variation of this crop. The traditional ones, which combine agronomic and morphologic characteristics, were used in the beginning (Skinner et al. 1987; Stevenson 1965). Many of the vegetative characteristics are influenced by environmental factors, presenting continuous variation and a high degree of plasticity, and which many times do not reflect the real diversity of the *Saccharum* spp. germplasm. Coefficient of parentage (*f*) (Kempthorne 1957) is an important method to estimate the genetic diversity based on pedigree. It indirectly measures the genetic diversity among cultivars by estimating, from pedigree records, the probability that alleles, in a locus, are identical by descent. However, assumptions made when calculating *f* regarding the relatedness of ancestors, selection pressure, and genetic drift are generally not met. Its use has been widespread among self-fertilization species such as soybean (Cox et al. 1985a; Vello et al. 1988), wheat (Cox et al. 1985b) and barley (Martin et al. 1991). In outcrossing species, modifications in methodology are necessary to minimize bias in the calculation of *f*, as was observed in the case of maize (Melchinger et al. 1991) and sugar cane (Chang and Lo 1993; Deren 1995).

Isoenzymes (Glaszmann et al. 1989) and molecular-marker analyses, such as RFLP using heterologous probes (D'Hont et al. 1994), ribosomal DNA (Glaszmann et al. 1990) and low-copy nuclear sequences (Burnquist et al. 1992; Lu et al. 1994a, b; Jannoo et al. 1999), are important strategies to assess sugar-cane genetic diversity. Lately, microsatellite markers from sugar cane were identified; however, they still have not yet been used for the assessment of genetic diversity (Cordeiro et al. 2000). All these methods showed a strong molecular differentiation between *S. officinarum* and *S. spontaneum*, revealing that the main part of diversity among sugar-cane cultivars is the *S. spontaneum* fraction of the genome. As sugar-cane species have a complex genome, a large number of DNA markers are necessary to reveal the exact genetic diversity. This was possible through the use of AFLP molecular markers (amplified fragment length polymorphisms), which reveal a high number of polymorphic bands in a multiplex pattern (Vos et al. 1995). AFLP markers were first used to estimate the genetic distances among 58 maize lines (Smith et al. 1993, 1994), which were followed up by studies to investigate the genetic diversity in several plant species (Hill et al. 1996; Maughan et al. 1996; Paul et al. 1997; Perera et al. 1998; Angiolillo et al. 1999; Amsellem et al. 2000), including sugar-cane (Besse et al. 1998).

Since the relationship between AFLP-based GS and *f* estimates has not yet been determined for sugar-cane cultivars, the objective of this study was to evaluate and to compare AFLP and pedigree-based genetic similarity estimates generated for 79 cultivars, most of which were either used or are now being used commercially in Brazil. They also include some newly released cultivars, and other cultivars from overseas that are important progenitors, and also four representative species.

Materials and methods

Plant material and DNA extraction

A total of 79 sugar-cane cultivars and *Saccharum* spp.-cultivars (*Saccharum sinense*, *S. barberi* and two of *S. officinarum*), were supplied by the Sugar Cane Breeding Program developed at the Federal University of São Carlos, Araras, SP. The cultivar selection for use in this study was carried out based on its previous and also recent economic importance in the Brazilian sugar-cane production areas, or on its importance as a progenitor. The identification and the geographical origin of the 83 sugar-cane genotypes used are presented in Table 1.

Young leaves were collected from each genotype, freeze-dried and ground to powder using a mechanical mill. Genomic DNA was extracted following the CTAB-method described by Hoisington et al. (1994). DNA concentration was estimated in comparison with known concentrations of lambda DNA in 0.8% agarose gel.

AFLP analysis

AFLP-analysis profiles were performed as described by Vos et al. (1995). Four-hundred nanograms of genomic DNA were double-digested using *Eco*RI and *Mse*I enzymes and linked to the specific adaptor. Pre-selective amplifications were performed with primers carrying one selective nucleotide. *Eco*RI primers carrying three selective nucleotides were end-labeled with γ-[^{33}P]-ATP (4,000Ci/mmol), and mixed with unlabelled *Mse*I primer, for hot selective amplification. All the amplifications were carried out in a PTC-100 programmable thermal controller (MJ Research, Inc., Watertown, Mass.). PCR products were mixed with an equal volume of formamide dye (98% formamide, 10 mM EDTA, 0.025% xylene cyanol w/v, 0.025% bromophenol blue w/v), of which 3.5 µl of each sample were separated by electrophoresis on a 6% polyacrylamide denaturing gel for 4 h at 75 W. After drying, the gels were exposed for 10–15 days on β max Hyperfilm MP (Armersham Life Science, UK).

Genetic-similarity estimation

Only polymorphic bands were used in the construction of a binary matrix, reflecting the presence and absence of the fragments obtained by AFLP, in the different genotypes. The estimate of genetic similarity (GS) among all the genotypes was calculated according to the Jaccard-similarity coefficient (1908): $GSij = a/(a+b+c)$, where GS*ij* is the measurement of the genetic similarity between individuals *i* and *j*, *a* is the number of polymorphic bands present in both individuals, *b* is the number of bands present in *i* and absent in *j*, and *c* is the number of bands present in *j* and absent in *i*. This definition of similarity excludes bands, which are absent in both individuals, from the calculation.

The matrix of similarity was analyzed by the unweighted pair-group method using the arithmetic average (UPGMA), as suggested by Sneath and Sokal (1973). The co-phenetic value (r_{cof}) based on GS was calculated, which is a quantitative indication of the grouping-analysis performance. All analyses were performed using NTSYS-PC software, version 2.0j (Exeter Software, N.Y.; Rohlf 1993).

Table 1 Identification of the 83 sugar-cane genotypes used in the pedigree and AFLP-based genetic similarity assessments

Clone	Origin	Clone	Origin
Badila	New Guinea (*S.officinarum*)	RB765418	Republic of Brazil
CB36–24	Campos, Brazil	RB785148	Republic of Brazil
CB40–13	Campos, Brazil	RB806043	Republic of Brazil
CB40–77	Campos, Brazil	RB825336	Republic of Brazil
CB41–75	Campos, Brazil	RB83102	Republic of Brazil
CB45–155	Campos, Brazil	RB835019	Republic of Brazil
CB45–3	Campos, Brazil	RB835089	Republic of Brazil
CB46–47	Campos, Brazil	RB835486	Republic of Brazil
CB47–355	Campos, Brazil	RB845257	Republic of Brazil
CB49–260	Campos, Brazil	RB855035	Republic of Brazil
CB53–98	Campos, Brazil	RB855113	Republic of Brazil
Co290	Coimbatore, India	RB855156	Republic of Brazil
Co331	Coimbatore, India	RB855453	Republic of Brazil
Co419	Coimbatore, India	RB855536	Republic of Brazil
Co740	Coimbatore, India	RB855563	Republic of Brazil
Co997	Coimbatore, India	SP70–1005	São Paulo, Brazil
CP51–22	Canal Point, USA	SP70–1078	São Paulo, Brazil
CP53–76	Canal Point, USA	SP70–1143	São Paulo, Brazil
Ganda Cheni	Saretha, India (*S. barberi*)	SP70–1284	São Paulo, Brazil
IAC48–65	Campinas, Brazil	SP70–1423	São Paulo, Brazil
IAC50–134	Campinas, Brazil	SP70–3370	São Paulo, Brazil
IAC51–205	Campinas, Brazil	SP71–1406	São Paulo, Brazil
IAC52–150	Campinas, Brazil	SP71–6163	São Paulo, Brazil
IAC58–480	Campinas, Brazil	SP71–6180	São Paulo, Brazil
IAC64–257	Campinas, Brazil	SP71–6949	São Paulo, Brazil
IAC82–2045	Campinas, Brazil	SP71–799	São Paulo, Brazil
IAC82–3092	Campinas, Brazil	SP72–4928	São Paulo, Brazil
IAC83–4157	Campinas, Brazil	SP79–1011	São Paulo, Brazil
IAC86–2210	Campinas, Brazil	SP79–2233	São Paulo, Brazil
IAC87–3396	Campinas, Brazil	SP79–2312	São Paulo, Brazil
IJ76–314	Iryan, Java (*S. officinarum*)	SP79–2313	São Paulo, Brazil
L60–14	Louisiana, USA	SP79–6134	São Paulo, Brazil
Maneria	Pansahi, China (*S. sinense*)	SP79–6192	São Paulo, Brazil
NA56–79	Northern Argentina	SP80–1520	São Paulo, Brazil
POJ2878	Java, Indonesian	SP80–1816	São Paulo, Brazil
RB721012	Republic of Brazil	SP80–1836	São Paulo, Brazil
RB72454	Republic of Brazil	SP80–1842	São Paulo, Brazil
RB725828	Republic of Brazil	SP80–3280	São Paulo, Brazil
RB732577	Republic of Brazil	SP81–1763	São Paulo, Brazil
RB735275	Republic of Brazil	SP81–3250	São Paulo, Brazil
RB739359	Republic of Brazil	SP83–5073	São Paulo, Brazil
RB739735	Republic of Brazil		

Bootstrap analysis

Dboot software (A. Coelho, personal communication), based on the bootstrap method (Efron 1981), was used to verify if the number of polymorphic AFLP markers, used for genetic similarity estimation, was enough to supply precise estimates among the genotypes (King et al. 1993; Tivang et al. 1994). The polymorphic markers were submitted to 100 samplings one by one, with a growing replacement of markers. Genetic similarities for each sample of markers were estimated obtaining 100 estimates of genetic similarity for each pair of genotypes. The average, the variance and the coefficient of variation were estimated for each one of these combinations. This procedure was repeated, each time, with a continuous growing number of markers, until the total number of polymorphic markers was reached. The number of markers and the mean coefficient of variation were used in the construction of a dispersion plot.

Coefficient of parentage

The calculation of the coefficient of parentage (*f*), between two genotypes, as defined by Kempthorne (1957), is carried out between two genotypes and corresponds to the probability that alleles in a locus are identical by descent to alleles in the same locus in another cultivar. The coefficient of parentage values were using the procedure "proc inbreeding" of the software SAS (version 6.12). In general, the assumptions suggested by Cox et al. (1985a) were adopted, and *f* was considered 0 among the remote ancestors. For each genotype it was assumed that the inbred coefficient was 0, due to the heterozygous character of the genotypes of this culture (Chang and Lo 1993; Deren 1995).

Correlation coefficient

To determine the correlation level among the AFLP-based genetic similarity (GS), and the coefficient of parentage (*f*), an analysis of correlation was carried out using Pearson's coefficient (*r*).

Results

Primer selection and AFLP analysis

The 64 primers combinations were evaluated, with regard to their capacity to generate polymorphic bands, using a group of 12 genotypes randomly chosen from the

Table 2 Number of polymorphic AFLP bands observed using 21 AFLP primer combinations

Primer combination	Total number of bands	Polymorphic bands	Polymorphism rate (%)
E+AAC/M+CAC	98	48	49
E+AAC/M+CAG	77	52	68
E+AAC/M+CTA	80	58	73
E+AAC/M+CTT	76	18	24
E+ACA/M+CTA	78	50	64
E+ACC/M+CAA	131	66	50
E+ACC/M+CAC	147	87	59
E+ACC/M+CAG	138	75	54
E+ACC/M+CTA	99	32	32
E+ACG/M+CAT	62	27	44
E+ACG/M+CTG	102	43	42
E+ACT/M+CAG	159	57	36
E+ACT/M+CAT	224	74	33
E+ACT/M+CTA	112	71	63
E+ACT/M+CTG	90	41	46
E+ACT/M+CTT	40	22	55
E+AGC/M+CAG	153	63	41
E+AGC/M+CAT	170	95	56
E+AGC/M+CTC	61	38	62
E+AGC/M+CTG	172	71	41
E+AGC/M+CTT	62	33	53
Totals	2331	1121	50

83 used in this study. Among the 64 primer combinations, 40 were selected, based on the presence of scorable bands and on the high number of polymorphic bands. Among the 40 primer combinations, 21 of the best combinations were used to produce the AFLP profiles in this study (Table 2). From these 21 combinations, it was possible to discriminate each one of the 83 sugarcane genotypes. An example of the obtained AFLP profiles is shown in Fig. 1.

Each selected primer combination generated a large number of bands. A total of 2331 fragments were identified, of which 1121 were polymorphic, and used to estimate the genetic similarity between the sugar-cane genotypes. An average of 50% polymorphism was obtained for each AFLP primer combination. The selected primers and the number of polymorphic fragments, revealed in each combination, are shown in Table 2.

The primer selection associated to the primer labeling using γATP^{33} and to a long exposition period, were fundamental factors for obtaining precise and reproducible AFLP profiles in sugar-cane. This procedure increased the number of polymorphic fragments revealed and reduced the time spent obtaining the same factors. Two concentrations of DNA were evaluated for AFLP reactions: 200 ng and 400 ng. The results did not present any significant variation with regard to the intensity of the AFLP bands. Primer-labeling tests with αATP^{32} and γATP^{33} were carried out. The amplifications of the samples, using the selective *Eco*RI primer with γATP^{33}, provided profiles of the bands with a far-superior definition. The visualization of the polymorphic fragments was carried out after a 10- to 15-day autoradiographic exposition. Repetitions adopting the same primer combination and the same time of exposition proved that the results were very accurate.

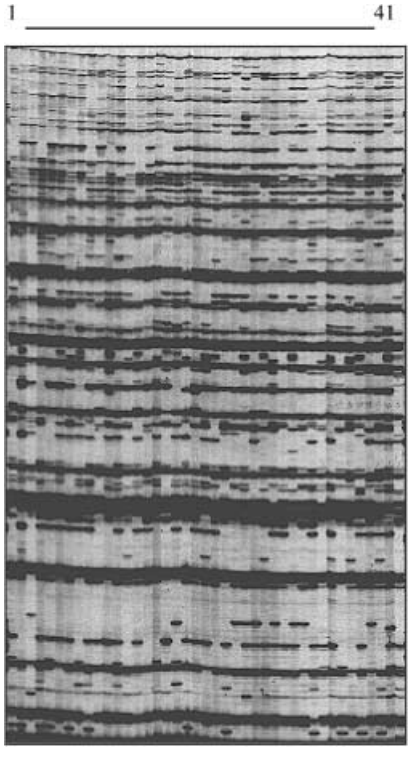

Fig. 1 AFLP profiles of 41 sugar-cane cultivars generated by the E+ACT/M+CTA primer combination

Genetic similarity, coefficient of parentage and cluster analysis

The results obtained by Jaccard's coefficient are presented in Fig. 2. The genetic-similarity estimate varied from 0.28 (between SP79–1011 and IJ76–314) to 0.89 (be-

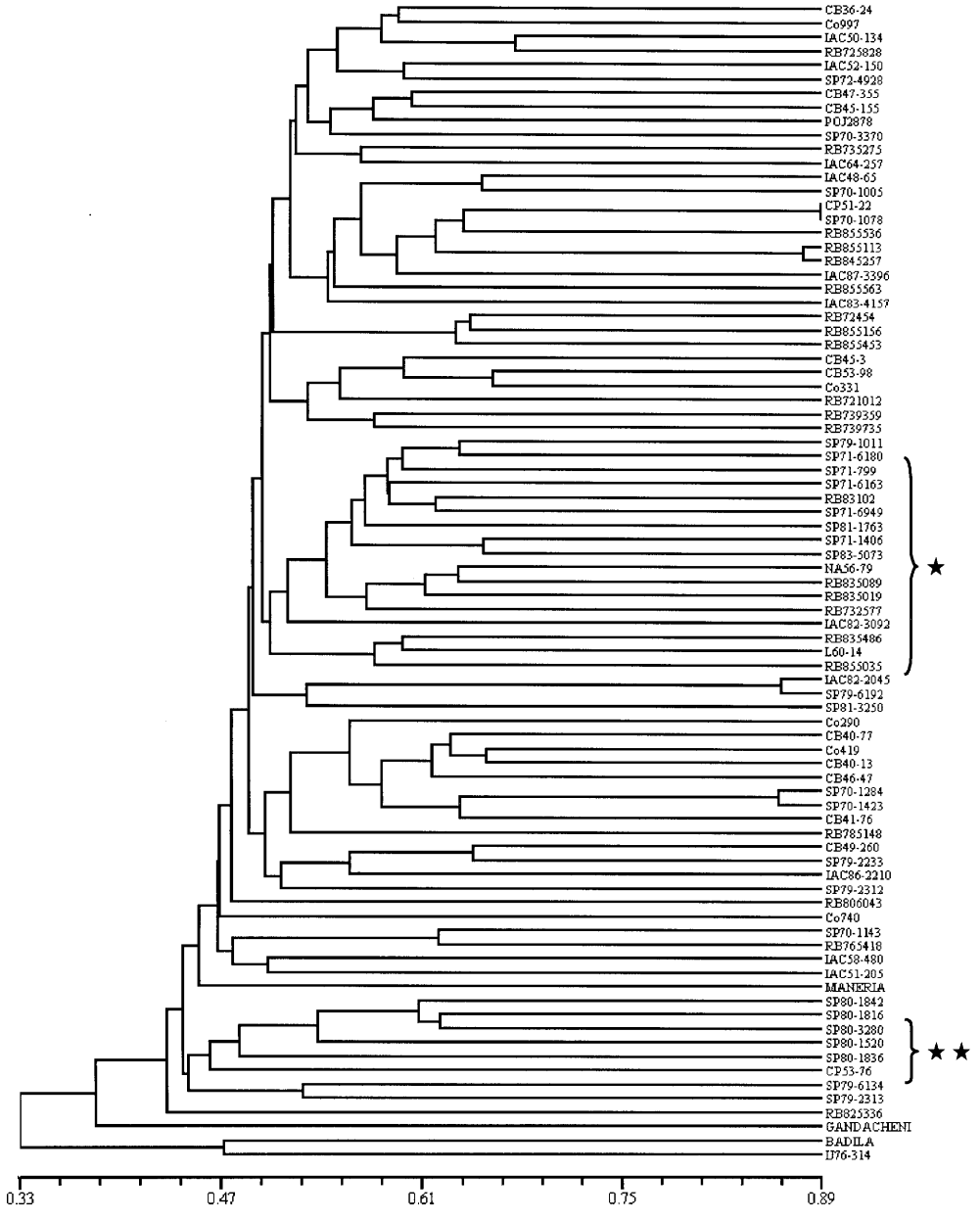

Fig. 2 Dendrogram of the 83 sugar cane revealed by UPGMA cluster analysis of AFLP-based genetic similarity (Jaccard's coefficient) estimates (r_{cof}=0.84), using 1121 AFLP polymorphic bands obtained by 21 primer combinations. *One star* indicates the group containing the NA56–79 cultivar and its offsprings; two one stars indicates the grouping of SP80 cultivars

tween CP51–22 and SP70–1078), with a mean of 0.47 in the 3,403 combinations obtained, using the data from the AFLP molecular markers. On the other hand, the coefficient of parentage varied from 0 to 0.503 (between NA56–79 and Co419), with a mean of 0.057.

The clustering of the AFLP-based GS, using the UPGMA method, formed a dendrogram showing a cophenetic coefficient of correlation (r_{cof}) of 0.84 (Fig. 2). In the dendrogram no group stood out in any special way, thus allowing all the cultivars to be in one main cluster, divided into subgroups, which presents some degree of similarity. This result indicates that there is not a high genetic diversity between the majority of the cultivars analyzed. The two clones of *S. officinarum*, Badila and IJ76–314, were placed outside the main cluster with GS = 0.47 between them. The *S. barberi* species (Ganda Cheni) was placed outside the general grouping, showing only 40% of similarity with the cultivars analyzed (GS = 0.40). The *S. sinense* (Maneria) species, although in the general group, remained separate from the complete group of cultivars. The cultivars, which presented high values of GS, were SP70–1423 and SP70–1284 (0.86); SP79–6192 and IAC82–2045 (0.86); RB845257 and RB855113 (0.87); SP70–1078 and CP51–22 (0.89).

Bootstrap analysis

The bootstrap-analysis method has shown that the accuracy of the genetic-similarity estimates, measured by the mean coefficient of variation (CV), increased according to the growth of the number of polymorphic loci ana-

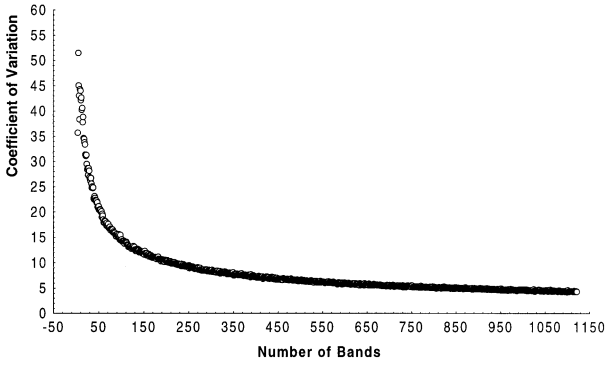

Fig. 3 Plot of the coefficient of variation of the complement of Jaccard's coefficient among all accessions estimated by bootstrap analysis for subsamples with different numbers of AFLP bands

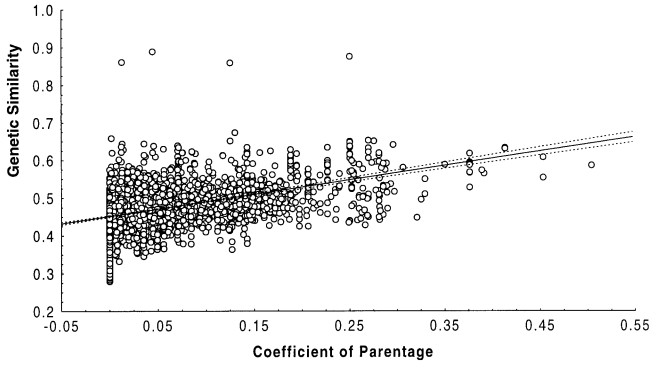

Fig. 4 Plot of AFLP-based genetic similarity (Jaccard's coefficient) and coefficient of parentage for 3403 pairs of the sugar-cane cultivars (Pearson's correlation $r = 0.42$)

lyzed. Based on the dispersion diagram, it was observed that the mean coefficient of variation decreased with an increase in the number of markers used (Fig. 3). For a mean CV close to 5.0%, which should be used as an appropriate CV value for GS estimations, an analysis of around 800 polymorphic bands would be necessary. Using all the 1121 polymorphic bands revealed in this work, the CV value was 4.29%, which is a reliable value for an appropriate GS estimation.

The methodology employed in the present work (King et al. 1993; Tivang et al. 1994) has been used by various authors to verify if the number of loci is sufficient for obtaining genetic distances with good precision (Halldén et al. 1994; Santos et al. 1994; Thormann et al. 1994). The considered value for the CV (5%) was lower than that usually recommended (10%).

Correlation between genetic similarity and the coefficient of parentage

The correlation (r) between genetic similarity and the coefficient of parentage was highly significant ($P < 0.001$); however, with a moderate to low value ($r = 0.42$). The dispersion graph (Fig. 4) showed that the points have a tendency to group together around a straight line. However, four points remained distant, representing the pairs of cultivars IAC82–2045 and SP79–6192, CP51–22 and SP70–1078, SP70–1284 and SP70–1423, RB855113 and RB845257.

Discussion

AFLP analysis

The AFLP profiles were not sensitive to the modifications carried out in the DNA concentrations, which is in agreement with the results of Vos et al. (1995). The primers, which had not been selected were discarded, due either to the high number of monomorphic bands revealed or, as observed for some primer combinations, due to the complete lack of amplification products. Qi and Lindhout (1997) showed the importance of a previous selection of primers in the variation study of AFLP profiles between barley cultivars, as well as in the setting up of a linkage map using the AFLP maize marker (Castiglioni et al. 1999).

Up to the present time there is no data of any study on sugar-cane genetic similarity using a very high number of polymorphic bands such as those obtained in this research. With the objective of evaluating the genetic diversity among sugar-cane cultivars, Glaszmann et al. (1989) used nine isozymes, evaluating a total of 114 loci; Lu et al. (1994a) used 22 probe/enzyme combinations, which identified 411 RFLP polymorphic fragments, to analyze 40 sugar-cane cultivars; and Jannoo et al. (1999) evaluated 386 RFLP polymorphic fragments in 162 sugar-cane genotypes, using 12 DNA probes.

The necessity of a high number of markers for the complete analysis of the sugar-cane genome is justified by the complex genome of this species, with an estimated size of 300 Mb (Aumuganathan and Earle 1991), as well as its polyploid structure and relatively narrow genetic base, if cultivars are taken into consideration. The results of this research showed that the AFLP technique allows the rapid obtaining of the necessary number of markers for this type of sugar-cane genome analysis. Those markers satisfactorily assessed the genetic relationship between the 83 sugar-cane genotypes. These same results have been observed in other plant species, which have a narrow genetic base, such as soybean (Maughan et al. 1996).

Genetic similarity, coefficient of parentage and correlation

The four species of *Saccharum* spp. and the 79 cultivars (interspecific hybrids) analyzed, presented moderate genetic similarity (GS) values. The two *S. officinarum* clones, Badila and IJ76–314, presented a GS value of 0.47, which is very close to the values obtained by Jannoo et al. (1999) among clones from New Guinea.

The highest values of genetic similarity observed between the cultivars (interspecific hybrids) were: 0.86 between SP70–1284 and SP70–1423; 0.86 between IAC82–2045 and SP79–6192; 0.87 between RB855113 and RB845257, and 0.88 between CP51–22 and SP70–1078.

The resulting dendrogram of the genotype grouping, based on GS (Fig. 2), revealed a general complex structure between cultivars when examined comparatively in terms of pedigree information. This is in agreement with the cultivar ploidy and its high heterozygosity, which allows the maintenance, in the genotypes, of a great number of alleles from ancestors incorporated in the initial interspecific crossings.

It was observed that there was a tendency for the cultivars to group together with others obtained from the same cross. For example, the NA56–79 cultivar was associated with its offsprings as indicated with one star in the dendrogram. This cultivar was used as a female parent of eight SP and five RB cultivars. Another example, various SP80 cultivars, all obtained from a single hybridization carried out in 1980, formed a subgroup as indicated with two stars in the dendrogram (Fig. 2). This same tendency can be observed for some cultivars identified as SP70, SP71 and RB85.

A significant correlation, with low to moderate value, was observed (r = 0.42) between the measurements of GS and f (Fig. 4). In Mantel's test, the value for the adjusted Z statistic was r = 0.37, very similar to Pearson's correlation coefficient (r = 0.42). This result was also observed in other studies of correlation between GS and f. Graner et al. (1994) compared RFLP-based GS with the coefficient of parentage estimates generated for a set of 48 barley cultivars, finding a low correlation for the winter type (r = 0.21) and a moderate one for the spring type (r = 0.42). Evaluating the correlation between isozyme-based GS and f estimates for wheat cultivars, Cox et al. (1985b) obtained a value of r = 0.27, while Tinker et al. (1993) demonstrated a moderate correlation of r = 0.61 in barley, observed between RAPD-based GS and f estimates. Similar results were found in maize (Plaschke et al. 1995) comparing estimates of GS based on microsatellite markers and f (r = 0.55). These results indicate that genetic similarity, based on molecular markers, has shown low to medium correlation with the coefficient of parentage based on pedigree data.

The dispersion graph, obtained from the correlation between the GS and f estimates between all the genotypes, showed a tendency to group values around the regression line (Fig. 4). However, the genotypes, which present the highest GS values, were those which had been placed far from the regression line. Among these are SP70–1284 and SP70–1423 (GS = 0.86 and f = 0.12), for which it is suggested that the low correlation was due to a greater genetic contribution from the female parent (CB41–76). Both cultivars come from the same female parent; however, the male parent is unknown (they are originated in a polycrossing).

The high AFLP-based GS (0.86) observed between cultivars IAC82–2045 and SP79–6192 was not expected, as the pedigree information shows a very distant relationship (f = 0.01). This result brings forth the hypothesis of an anotation error concerning the true progenitor due to polycrossing (Deren 1995).

The cultivars RB855113 and RB845257 (GS=0.87 and f = 0.25) are the result of reciprocal crossing between SP70–1143 and RB72454, and the molecular data confirms the high genetic similarity between them. The low value observed for f was probably the result of an error caused by suppositions used to calculate f.

The GS of 0.88 observed between cultivars CP51–22 and SP70–1078, by means of pedigree analysis (f = 0.44), was not expected. It is known that the female parent of SP70–1078 is IAC48–65, which resulted from a polycross. Therefore, a possible explanation for the high value of GS observed is that the male progenitor of the IAC48–65 cultivar could be CP51–22, or another clone very close to it.

A different hypothesis can be proposed to explain the low correlation between GS estimates and the coefficient of parentage. The low to moderate correlation is probably due to: unequal parental contribution (Deren 1995); gene concentration in each generation of crossing; the male parent in the polycrosses being the same or a very close cultivar, or else the mistaken anotation of the parent during the generation of one of the two estimates. In the case of the GS estimates, their consistency depends on the number and localization of the markers used in the genome. In this study, the 1121 markers used were sufficient to produce GS data with a variation below 5% probability (Fig. 3). Following the bootstrap method used, it can be observed that, with the increase in the number of molecular markers, GS variance was reduced. AFLP profile-reading errors can also result in a reduction in the correlation of GS and f. An attempt was made to avoid this type of error as much as possible in this study by eliminating the reading of any fragment that presented doubts with regards to presence or absence. Therefore, when these kinds of bands appeared they were considered "lost data", and this information was included in the standard error calculated by bootstrap.

The coefficient of parentage, although highly informative in a breeding programme, presents inherent errors during its calculation, resulting in f values with some bias. This is, in part, due to some genetic suppositions, which are assumed in the calculation of f. It is assumed that all the ancestors are not closely related, which is not always true when the history of sugar-cane cultivars is considered (Bremer 1961a). The supposition that the genotype receives the same amount of genes from each parent is questionable in the present case. Sugar-cane is polypoid and highly heterozygotic, apart from the well-known fact that when using *S. officinarum* as a female parent, its meiosis is not equivalent, resulting in one parent's alleles having an advantage over the other (Bremer 1961b). Another point to be considered during the calcu-

lation of *f* is the little-known changes in the frequency of the alleles, due to the effect of genetic drift and the selection process. Both phenomena can influence the precision of *f*. As, for example, the transmission of alleles, especially those that control qualitative characteristics with high heritability, is clearly influenced by the intensity of selection in a breeding programme. This fact results in a bias in the contribution of the parent, stressing the favorable alleles for the character in the resulting progeny (Cox et al. 1985a; Souza and Sorrells 1989).

GS estimates obtained from molecular markers will provide more information than those available through pedigree information. The results here obtained are in agreement with those presented by Barret and Kidwell (1998), using wheat cultivars, where the values of the genetic diversity obtained from AFLP markers showed themselves to be more useful than the coefficient of parentage, to identify cultivar combinations for crossings with maximum-allele variation. Thus, it can be concluded that estimates of GS based on molecular markers may provide more accurate information to plant breeders than the pedigree method, allowing breeders to more-efficiently make reliable crossings on a short-term basis or to strategically plan the breeding programme on a long-term basis.

Acknowledgements The authors thank Maria Carolina Avanci (FAPESP grant TT-2 no. 98/15205–5) for technical support during the development of this research; Prof. Alexandre S.G. Coelho, Universidade Federal de Goiás/Goiana – GO, for supplying the Dboot program and helping during bootstrap analysis; Fundação de Apoio Institutional da Universidade Federal de São Carlos/SP (FAI-UFSCar) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 98/01689–0) for financial support during the development of this research. Both M.L.A.L. and K.M.O. received TT3/MSI (98/10427–0 – 99/01347–5) and IC (98/11983–3) fellowships respectively from FAPESP. A.P.S., C.L.S.Jr. and S.M. are recipients of a research fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Amsellem L, Noyer JL, Le Bourgeois T, Hossaert-McKey M (2000) Comparison of genetic diversity of the invasive weed *Rubus alceifolius* Poir. (*Rosaceae*) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. Mol Ecol 9:443–455

Angiolillo A, Mencuccini M, Baldoni L (1999) Olive genetic diversity assessed using amplified fragment length polymorphisms. Theor Appl Genet 98:411–421

Arceneaux G (1968) Cultivated sugar canes of the world and their botanical derivation. Proc Int Soc Sugar Cane Technol 12: 844–854

Aumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Barret BA, Kidwell KK (1998) Comparison of AFLP and pedigree-based genetic diversity assessment methods using wheat cultivars from the Pacific Northwest. Crop Sci 38:1271–1278

Besse P, Taylor G, Carroll B, Berding N, Burner D, McIntyre CL (1998) Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. Genetica 104: 143–153

Bremer G (1961a) Problems in breeding and cytology of sugar cane. Euphytica 10:59–78

Bremer G (1961b) Cytological crossing research of sugar cane. Euphytica 10:229–243

Burnquist WL, Sorrells ME, Tanksley S (1992) Characterization of genetic variability in *Saccharum* germplasm by means of restriction fragment length polymorphism (RFLP) analysis. XXI Proc Int Soc Sugar Cane Technol 21:355–365

Castiglioni P, Ajmone-Marsan P, Van Wijk R, Motto M (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. Theor Appl Genet 99:425–431

Chang YS, Lo CC (1993) Genetic relationship among major sugar cane cultivars in Taiwan. Rept Taiwan Sugar Res Inst 140: 1–10

Cordeiro GM, Taylor GO, Henry RJ (2000) Characterization of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polyploid species. Plant Sci 155:161–168

Cox TS, Kiang YT, Gorman MB, Rogers DM (1985a) Relationships between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci 25:529–532

Cox TS, Lookhart GL, Walker DE, Harrell LG, Albers LD, Rogers DM (1985b) Genetic relationships among hard red winter wheat cultivars as evaluated by pedigree analysis and gliadin polyacrylamide gel-electrophoretic patterns. Crop Sci 25: 1058–1063

D'Hont A, Lu YH, Gonzales-de-Leon D, Feldmann P, Lanaud C, Glaszmann JC (1994) A molecular approach to unraveling the genetics of sugar cane, a complex polyploid of the Andropogoneae tribe. Genome 37:222–230

Deren CW (1995) Genetic base of U.S. mainland sugar cane. Crop Sci 35:1195–1199

Efron B (1981) The bootstrap, the jackknife, and other resampling plans. Society of Indiana Applied Mathematics, Philadelphia

Glaszmann JC, Fautret A, Noyer JL, Feldmann P, Lanaud C (1989) Biochemical genetic markers in sugar cane. Theor Appl Genet 78:537–543

Glaszmann JC, Lu YH, Lanaud C (1990) Variation of nuclear ribosomal DNA in sugar cane. J Genet Breed 44:191–198

Graner A, Ludwig WF, Melchinger AE (1994) Relationships among European barley germplasm. II. Comparisons of RFLP and pedigree data. Crop Sci 34:1199–1205

Halldén C, Nilsson NO, Rading IM, Sill T (1994) Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. Theor Appl Genet 88:123–128

Hill M, Witsenboer H, Zabeau M, Vos P, Kesseli R, Michelmore R (1996) PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. Theor Appl Genet 93:1202–1210

Hoisington D, Khairallah M, González-de-León D (1994) Laboratory protocols: CIMMYT applied molecular genetics laboratory. CIMMYT, Mexico, DF

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270

Jannoo N, Grivet L, Seguin M, Paulet F, Domaingue R, Rao PS, Dookun A, D'Hont A, Glaszmann JC (1999) Molecular investigation of the genetic base of sugarcane cultivars. Theor Appl Genet 99:171–184

Kempthorne O (1957) An introduction to genetic statistics. John Wiley and Sons, New York

King G; Nienhuis J, Hussey C (1993) Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphisms. Theor Appl Genet 86:1028–1032

Lu YH, D'Hont A, Paulet F, Grivet L, Arnaud M, Glaszmann JC (1994a) Molecular diversity and genome structure in modern sugar cane varieties. Euphytica 78:217–226

Lu YH, D'Hont A, Walker DIT, Rao PS, Feldmann P, Glaszmann JC (1994b) Relationships among ancestral species of sugar cane revealed with RFLP using single-copy maize nuclear probes. Euphytica 78:7–18

Martin JM, Blake TK, Hockett EA (1991) Diversity among North American spring barley cultivars based on coefficient of parentage. Crop Sci 31:1137–1149

Maughan PJ, Saghai Maroof MA, Buss GR, Huestis GM (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. Theor Appl Genet 93:392–401

Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991) Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci 31:669–678

Paul S, Wachira FN, Powell W, Waugh R (1997) Diversity and genetic differentiation among populations of India and Kenyan tea [*Camellia sinensis* (L) O. Kuntze] revealed by AFLP markers. Theor Appl Genet 94:255–263

Perera L, Russel JR, Provan J, McNicol J, Powell W (1998) Evaluating genetic relationships between indigenous coconut (*Cocos nucifera* L.) accessions from Sri Lanka by means of AFLP profiling. Theor Appl Genet 96:545–550

Plaschke J, Ganal MW, Roder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91:1001–1007

Qi X, Lindhout P (1997) Development of AFLP markers in barley. Mol Gen Genet 254:330–336

Roach BT (1972) Nobilisation of sugar cane. Proc Int Soc Sugar Cane Technol 14:206–216

Roach BT, Daniels J (1987) A review of the origin and improvement of sugar cane. Copersucar Int Sugar Cane Breeding Workshop, São Paulo, Copersucar

Rohlf FJ (1993) NTSYS-PC numerical taxonomy and multivariate analysis system. Version 2.0j. Exeter Publ, Setauket, New York

Santos JB, Nienhuis J, Skroch P, Tivang J, Slocum MK (1994) Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleraceae* L. genotypes. Theor Appl Genet 87:909–915

Skinner JC, Hogarth DM, Wu KK (1987) Selection methods, criteria, and indices. In: Heinz, DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdan. pp 409–453

Smith JSC, Zabeau M, Wright S (1993) Associations among inbred lines as revealed by RFLPs and by a thermocycling amplification methodology, amplified fragment length polymorphisms (AFLPs). Maize Genet Newslett 67:62–64

Smith S, Luk S, Sobral B, Muhawish S, Peleman J, Zabeau M (1994) Associations among inbred lines of maize using RFLP and DNA amplification technologies (AFLP and AP-PCR), and correlations with pedigree, F1 yield and heterosis. Maize Genet Newslett 68:45

Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco

Souza E, Sorrells ME (1989) Pedigree analysis of North American oat cultivars released from 1951 to 1985. Crop Sci 29:595–601

Stevenson, GC (1965) Genetics and breeding of sugar cane. Longmans, London

Tai P, Miller JD (1978) The pedigree of selected Canal Point (CP) varieties of sugar cane. Proc Am Soc Sugar Cane Technol 8:34–39

Thormann CE, Ferreira ME, Camargo LEA, Tivang JG, Osborn TC (1994) Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. Theor Appl Genet 88:973–980

Tinker NA, Fortin MG, Mather DE (1993) Random amplified polymorphic DNA and pedigree relationships in spring barley. Theor Appl Genet 85:976–984

Tivang JG, Nienhuis J, Smith OS (1994) Estimation of sampling variance of molecular marker data using the bootstrap procedure. Theor Appl Genet 89:259–264

Vello N A, Hiromoto DM, Azevedo-Filho AJBV (1988) Coefficient of parentage and breeding of Brazilian soybean germplasm. Rev Bras Genet 11:679–697

Vos P, Hogers R, Bleeker M, Reijans M, Van Der Lee T, Hornes M (1995) AFLP: a new concept for DNA fingerprinting. Nucleic Acids Res 23:4407–4414