

GENETIC DIVERSITY OF CULTIVATED *Coffea arabica* INBRED LINES ASSESSED BY RAPD, AFLP AND SSR MARKER SYSTEMS

Mirian Perez Maluf^{1*}; Milene Silvestrini²; Luciana Machado de Campos Ruggiero²; Oliveira Guerreiro Filho²; Carlos Augusto Colombo³

¹*Embrapa Café, Av. Barão de Itapura, 1481 - 13001-970 - Campinas, SP - Brasil.*

²*IAC - Centro de Café "Alcides Carvalho" - 13001-970 - Campinas, SP - Brasil.*

³*IAC - Centro de Recursos Genéticos Vegetais.*

*Corresponding author <maluf@iac.sp.gov.br>

ABSTRACT: One of the greatest problems in *Coffea arabica* breeding is identifying precisely any inbred line, based only on botanical and agronomical descriptors, because of the reduced genetic variability of the species, close pedigree origin, which results in small phenotypic variation. Recently, molecular markers have been used for plant germplasm characterization and identification in several commercial species. This work evaluates the reliability of three marker systems: RAPD, AFLP and SSR, to characterize the genetic variability of commercially-used *Coffea* inbred lines developed by the Instituto Agrônômico (IAC), and their potential for cultivar identification. All methods identified polymorphisms among the cultivars. The genetic diversity recognized by the methods is very similar, although is very narrow. RAPD and SSR marker systems grouped more efficiently the evaluated cultivars according to parental origin. None of the methods allowed inbred line identification. Therefore for varietal protection, it would be necessary using a combination of botanical, agronomical and molecular markers descriptors for precise cultivar identification.

Key words: coffee, cultivar identification, molecular markers

CARACTERIZAÇÃO DA DIVERSIDADE GENÉTICA DE LINHAGENS COMERCIAIS DE *Coffea arabica* ATRAVÉS DE MARCADORES MOLECULARES DO TIPO RAPD, AFLP E SSR

RESUMO: A identificação de linhagens de *Coffea arabica* a partir de descritores botânicos e agrônômicos é um problema para o desenvolvimento de cultivares. Basicamente, a limitada variação fenotípica observada em cultivares é o resultado de uma estreita variabilidade genética em *C. arabica* associada com uma origem genealógica próxima. Recentemente, os uso de marcadores moleculares tem contribuído para a caracterização e identificação de várias espécies de interesse comercial. O objetivo deste trabalho foi comparar a confiabilidade de três tipos de marcadores moleculares, RAPD, AFLP e SSR, para a caracterização da variabilidade genética e uma possível identificação de linhagens comerciais de *Coffea* desenvolvidas pelo IAC. Os métodos avaliados permitiram identificar polimorfismos entre cultivares. A variabilidade genética detectada por eles é muito semelhante, ainda que reduzida. Marcadores do tipo RAPD e SSR foram mais eficientes em análises de parentesco, e o agrupamento das linhagens correspondeu à sua origem genealógica. No entanto, nenhum dos métodos testados permitiu a identificação individual de linhagens. Neste caso, a utilização conjunta de descritores botânicos, agrônômicos e marcadores moleculares é recomendada para a identificação precisa de linhagens, visando processos de proteção legal de cultivares de *Coffea*.

Palavras-chave: café, identificação de cultivares, marcadores moleculares

INTRODUCTION

The genus *Coffea* contains the two main commercially cultivated coffee species, *Coffea arabica* and *Coffea canephora*. Several agronomic and technological differences between these two species are responsible for their individual market value.

The *Coffea* germplasm collection of IAC includes selected commercial lines of *C. arabica* and *C. canephora* that are under continuous characterization regarding

botanic, agronomic, and technologic aspects (Carvalho et al., 1991; Aguiar, 2001). These analysis can detect significant agronomic differences among cultivars (Aguiar, 2001).

Previous studies on *Coffea* germplasm characterization, including analyses using molecular markers, reported low genetic diversity in *C. arabica*, both wild and cultivated genotypes (Lashermes et al., 1993; Orozco-Castillo et al., 1994; Cros et al., 1998; Combes et al., 2000; Anthony et al., 2001; 2002; Sera et al., 2003).

Justifications for this low variability include narrow genetic basis, reproduction system and the fact that a small number of coffee plants were introduced for cultivation in producing countries.

Even though the inbred lines selected by the IAC Breeding Program are well known and thoroughly evaluated, individual and precise identification cannot be achieved using only traditional descriptors (Aguar, 2001). New legislation on commercial exploitation of genetic resources requires correct cultivar identification for both legal protection and intellectual property requirements. Therefore, in this work, molecular markers were used as tools for germplasm evaluation, targeting the analysis of the genetic diversity of cultivated coffee lines, and the identification of cultivar specific markers. The most common methods for generating molecular markers were tested in selected coffee germplasm, and compared regarding their ability to access genetic polymorphisms and identifying genetic relationship among cultivars.

MATERIAL AND METHODS

Plant material

Inbred lines of *C. arabica* developed by the Coffee Breeding Program of the Agronomic Institute (IAC) were selected for analysis with molecular markers. All lines are commercially planted all over Brazilian coffee producing regions. A complete list of inbred coffee germplasm evaluated in this work, with respective pedigree data, is present in Table 1. Young leaves were collected from plants of each cultivar, frozen in liquid N₂, and kept at -80°C until used. *C. canephora* cultivar Apoatã was used as outgroup accession.

Genomic DNA extraction

Total genomic DNA was extracted from frozen young leaves according to Orozco-Castillo et al. (1994), using MATB (lauryl-trimethyl-ammonium bromide) as detergent. Whenever fresh-collected leaves were used, DNA was extracted using activated charcoal during tis-

Table 1 - List of *Coffea* germplasm evaluated.

Cultivar	Species	Inbred line	Origin
Acaiá	<i>Coffea arabica</i>	AC 474-4	Bourbon Vermelho × Sumatra (Typica)
Acaiá	<i>Coffea arabica</i>	AC 474-19	Bourbon Vermelho × Sumatra (Typica)
Mundo Novo	<i>Coffea arabica</i>	MN 388-17	Bourbon Vermelho × Sumatra (Typica)
Mundo Novo	<i>Coffea arabica</i>	MN 379-19	Bourbon Vermelho × Sumatra (Typica)
Mundo Novo	<i>Coffea arabica</i>	MN 505	Bourbon Vermelho × Sumatra (Typica)
Mundo Novo	<i>Coffea arabica</i>	MN 515	Bourbon Vermelho × Sumatra (Typica)
Bourbon Amarelo	<i>Coffea arabica</i>	BA 9	Bourbon Vermelho (mutation)
Catuai Vermelho	<i>Coffea arabica</i>	CV 44	Catuai Amarelo 476 × Mundo Novo 374-19
Catuai Vermelho	<i>Coffea arabica</i>	CV 46	Catuai Amarelo 476 × Mundo Novo 374-19
Catuai Vermelho	<i>Coffea arabica</i>	CV 144	Catuai Amarelo 476 × Mundo Novo 374-19
Catuai Vermelho	<i>Coffea arabica</i>	CV 99	Catuai Amarelo 476 × Mundo Novo 374-19
Catuai Vermelho	<i>Coffea arabica</i>	CV 81	Catuai Amarelo 476 × Mundo Novo 374-19
Catuai Amarelo	<i>Coffea arabica</i>	CA 86	Catuai Amarelo 476 × Mundo Novo 374-19
Catuai Amarelo	<i>Coffea arabica</i>	CA 62	Catuai Amarelo 476 × Mundo Novo 374-19
Catuai Amarelo	<i>Coffea arabica</i>	CA 74	Catuai Amarelo 476 × Mundo Novo 374-19
Catuai Amarelo	<i>Coffea arabica</i>	CA 47	Catuai Amarelo 476 × Mundo Novo 374-19
Catuai Amarelo	<i>Coffea arabica</i>	CA 100	Catuai Amarelo 476 × Mundo Novo 374-19
Icatu Precoce	<i>Coffea arabica</i>	IP 3282	Selection of Icatu Vermelho
Icatu Vermelho	<i>Coffea arabica</i>	IV 2945	(<i>C. canephora</i> cv Robusta × Bourbon Vermelho) × Mundo Novo
Icatu Vermelho	<i>Coffea arabica</i>	IV 4040	(<i>C. canephora</i> cv Robusta × Bourbon Vermelho) × Mundo Novo
Icatu Vermelho	<i>Coffea arabica</i>	IV 4042	(<i>C. canephora</i> cv Robusta × Bourbon Vermelho) × Mundo Novo
Icatu Vermelho	<i>Coffea arabica</i>	IV 4045	(<i>C. canephora</i> cv Robusta × Bourbon Vermelho) × Mundo Novo
Icatu Vermelho	<i>Coffea arabica</i>	IV 4046	(<i>C. canephora</i> cv Robusta × Bourbon Vermelho) × Mundo Novo
Icatu Amarelo	<i>Coffea arabica</i>	IA 2944	Icatu Vermelho × Bourbon Amarelo
Obatã	<i>Coffea arabica</i>	OB	(Villa Sarchi × Hybrid of Timor) × Catuai Vermelho
Tupi	<i>Coffea arabica</i>	TP	(Villa Sarchi × Hybrid of Timor) × Catuai Vermelho
Ouro Verde	<i>Coffea arabica</i>	OV	Catuai Amarelo × Mundo Novo 515
Apoatã	<i>Coffea canephora</i>	AP 2258	Selection of cultivar Robusta

sue grinding, as described by Vroh et al. (1996). All DNA samples were diluted to a final concentration of 20 ng μL^{-1} .

RAPD

A total of 40 ng of each DNA sample was used in PCR reactions for RAPD markers amplification. Random, commercial 10-mer oligonucleotides from Operon Technology, kits A, G, H and X were used. PCR reactions were set as follows in a final volume of 25 μL : 0.1 mmol L^{-1} dNTP, 2 mmol L^{-1} MgCl_2 , 0.5 $\mu\text{mol L}^{-1}$ primer, 1X reaction buffer (10 mmol L^{-1} TRIS-HCl and 50 mmol L^{-1} KCl) e 0.25 U *Taq* polymerase. Samples were submitted to 45 cycles under the following conditions: 1 min at 94°C, 45 s at 35°C and 1.5 min at 72°C. Amplified fragments were separated according to size on 1.5% agarose gel, and stained with ethidium bromide. Gel documentation and fragment size determination were performed by the software Image Master Total Lab (*Pharmacia*).

AFLP

AFLP marker analysis was performed according to Vos et al. (1995). Amplified fragments were obtained using a multi-color fluorescent primers AFLP kit (*Applied Biosystems*), which allows a total of 64 primer combinations. All combinations were tested in selected genotypes, and only those that amplified polymorphic fragments were used in all genotypes. Reaction conditions were performed as recommended by the manufacturer's manual, with some modifications. Digestion with restriction enzymes was performed using 280 ng of total DNA. After adapters ligation, digested DNA was diluted to a final concentration of 50 ng μL^{-1} . Samples were submitted to 23 cycles of pre-amplification reaction in the following conditions: 30 s at 94°C, 1 min at 56°C and 1 min at 72°C. Selective amplification reactions were: pre-amplified fragments were diluted 1:30, mixed with 1 mmol L^{-1} of *EcoRI* and *MseI* primers, 1 mmol L^{-1} of dNTP, 1X reaction buffer, 5 U of *Taq* polymerase and 2 mmol L^{-1} of MgCl_2 . Reaction conditions followed manual's recommendations amplified products were separated on 5% acrylamide gel using an ABI 377 automated sequencer (*Applied Biosystems*). Products were detected by the software Gene Scan (*Applied Biosystems*) using internal molecular size markers (GENESCAN-500 ROX). Polymorphisms were evaluated by Genotyper v 2.0 software (*Applied Biosystems*). Restriction parameters for polymorphism identification were defined based on overall amplification quality, and for final analysis only peaks that met the following pattern were selected: absorbance of 100 Abs; size between 100 – 450 bp; size tolerance of ± 5 bp.

SSR

The primers to amplify SSR locus were synthesized based on sequences described in Combes et al.

(2000) and Rovelli et al. (2000) (Table 2). The primer pair SSR1 was designed by our lab using the software Primer Designer V 2.0 (*Scientific & Educational Software*). A total of 23 microsatellite locus were evaluated, corresponding to repetitions of $(\text{TG})_n$, $(\text{TA})_n$, $(\text{CA})_n$ and $(\text{ATC})_n$. Final reaction conditions were 40 ng of genomic DNA, 1X reaction buffer, 0.2 mmol L^{-1} of dNTP, 2 mmol L^{-1} of MgCl_2 , 5 rmoles of each primer and 1.25 U of *Taq* polymerase. The complete thermal cycle program was 5 min at 95°C, followed by 30 cycles of 1 min at 95°C, 1 min at 60°C and 1 min at 72°C, and a final 5 min of elongation at 72°C, except for loci SSR24 and SSR25, for what annealing temperature was 70°C and 66°C, respectively. The amplified products were labeled by direct incorporation of fluorescent nucleotides during PCR reaction (Rhodes et al., 1997; Jouquand et al., 1999). The nucleotide [F]dNTP [Tamra] dCTP-yellow (*PE Applied Biosystems*) was used because of system compatibility. The proportion of fluorescent to non-fluorescent nucleotide was 1:100. To maximize product, non-incorporated nucleotides were eliminated by precipitation of PCR samples with isopropanol, and then resuspended in 1.25 mL of formamide and blue dextran buffer, and 0.25 μL of TAMRA 500 (*PE Applied Biosystems*). Fluorescent-amplified products were separated on 5% acrylamide gel by ABI 377 automated sequencer (*Applied Biosystems*). Products were detected by Gene Scan software (*Applied Biosystems*) using internal molecular size markers (GENESCAN-500 ROX), and amplified loci were evaluated using Genotyper v 2.0 software (*Applied Biosystems*).

Data analysis

The gels were scored for the presence or absence of amplified fragments. Although microsatellites are co-dominant markers, molecular analyses of SSR loci were performed based on presence/absence of amplified fragments because of the allotetraploid nature of *C. arabica*.

For a comparative analysis among the methods, genetic parameters such as average number of alleles per locus, allele frequency (p_i), expected heterozygosity (H), effective number of alleles per locus (ne) were calculated as described by Morgante et al. (1994): $ne = (\sum p_i^2)^{-1}$ and $H = h = 1 - \sum p_i^2$. ($N/N-1$), where N is the sample size (number of cultivars evaluated). This measure takes into account both number and frequency of alleles. An assay efficiency index (A_i) was calculated for each method, which combining the total effective number of alleles identified ($Ne = \sum ne_i$) and the number of assays performed for their detection (P), as $A_i = Ne/P$ (Morgante et al., 1994).

Genetic distance among cultivars was calculated for each method. The genetic distance matrix was estimated using the Dice similarity coefficient (Dice, 1945).

Table 2 - Access number of microsatellite loci in the Genbank and respective forward and reverse primer sequences used.

Accession number	Sequence of primers (5'- 3')	Repeated sequence	Published size (bp) ^a	Annealing temperature (°C)	Fragment size range (bp)
AJ250259	SSR1-F ^b :ATCTCCTGGACCAGTTAACC CTGCGAGGTGCTTATTGAAG	(GT) ₃ /(GT) ₇	72-103	60	128-139
AJ308738	SSR5.F= AAAAAGCTGGTCCATGTCAA GGGCGTTCAGTTATAAACA	(TG) ₈	117	60	97
AJ308746	SSR6.F= AGGCCTTCATCTCAAAAACSSR6.R= AGCGTTACTTGAGGCAAAGA	(TC) ₁₄ /(CA) ₁₁ /(CA) ₁₆	217	60	185-211
AJ308762	SSR7.F= CTGGGTTGGTTCGTATTTGSSR7.R= GGTCCCAGAGATTCTCTCC	(TG) ₁₆	341	60	325-329
AJ308763	SSR8.F= TGACATAGGGGGCTAAATTG TTAATGGTGACCTTTGATG	(CA) ₁₀	215	60	199-201
AJ308767	SSR9.F= TGCTTAGGCACCTTGATATAGGASSR9.R= CACGTGCAAGTCACATACTTTA	(CA + TA) ₃₈	150	60	101-170
-	SSR10.F ^{c1} = TGTGCTGAACTGTGTTGCT SSR10.R ^{c1} = TCCAGAGAAATGTCGGAAGT	(TG) ₁₃	198	60	148-150
AJ308785	SSR11.F= TGGCCGTGATAATAAACAGC ATGTGGCAATCTAAAGCCAA	(TG) ₂₁	212	60	178-196
AJ308765	SSR12.F= ATGCCAAGTCGGAAAAGAA GGCAAGCTCTAGCCTTTGA	(CA) ₇	136	60	116-118
AJ250258	SSR14.F= AACTCTCCATTCGCGCATT CTGGGTTTCTGTGTTCTCG	(CA) ₃ /(CA) ₃ /(CA) ₁₈	89-135	60	104-146
-	SSR15.F ^{c2} = CTCTCCCTCAGTCAATTCCA SSR15.R ^{c2} = CTTGGTCTCCCTCCTTTTTC	(ATC) ₁₄	234	60	191-219
AJ308743	SSR16.F= TTTTCTTGCTAATCTTTGAGGASSR16.R= ACTCTAATGGGGTCATGTGG	(CA) ₇	130	60	108-112
AJ308754	SSR17.F= AAGGGGAGTGGATAAGAAGG GGCTGGATTGTGCTTTAAG	(CA) ₁₂	128	60	102-108
AJ308764	SSR18.F= CACTGGCATTAGAAAGCACC GGCAAAGTCAATGATGACTC	(CA) ₁₄	198	60	177-183
AJ308766	SSR19.F= AGTGATCTTCGAGCCATT TCTTTTGTGACTGGGCTTC	(CA) ₈	175	60	137-155
AJ250251	SSR20.F= ATTCTCTCCCCTCTCTG TGTGTGCGCGTTTCTTG	(CA) ₆ /(CA) ₃ /(CA) ₃ /(CA) ₄ /(CA) ₃ /(CA) ₃ /(CA) ₃	248-258	60	274-282
AJ250252	SSR21.F= ACCCGAAAGAAAGAACCAAGSSR21.R= CCACACAACCTCCCAATC	(GT) ₄ /(GA) ₄ /(GT) ₄ /(GT) ₆	140-146	60	140-142
AJ250253	SSR22.F= CTTGTTGAGTCTGTGCGTG TTCCCTCCCAATGCTGTA	(GA) ₅ /(GT) ₈ /TT(GT) ₄ /TT(GT) ₇ /(GA) ₁₁ /(TC) ₂ /(CT) ₃ GT	240-270	60	278-311
AJ250254	SSR23.F= GGCTCGAGATACTGTTTAG TTAATGGGCATAGGGTCC	(CA) ₁₅ /(CG) ₄ /CA	132-166	60	148-165
AJ250255	SSR24.F= CCCTCCCTGCCAGAAGAAGSSR24.R= AACCACCGTCTTTTCTCTCG	(GT) ₅ /CT(GT) ₂ /(GT) ₁₂	160-170	70	175-181
AJ250256	SSR25.F=AGGAGGGAGGTGTGGTGAAGSSR25.R= AGGGGAGTGGATAAGAAGG	(GT) ₁₁	118-134	66	135-139
AJ250257	SSR26.F= GACCATTACATTCACACAC GCATTTGTGTGACACTGTA	(CTCACAA) ₄ /(CA) ₉	103-122	60	103-126
AJ250260	SSR27.F= TGATGGACAGGAGTTGATGG TGCCAATCTACCCTTACCCTT	(CT) ₉ /(CA) ₈ /(CT) ₄ /(CA) ₅	100-132	60	118-150

^aPublished size refers to the published sequence in Combes et al. (2000) and Rovelli et al. (2000).

^bF - Forward and R - Reverse.

^{c1}, ^{c2}Primers sequences were not in the Genbank, but were described in Rovelli et al. (2000).

Cluster analysis was performed on the matrix distance using the clustering UPGMA criterion; the bootstrap method (Felsenstein, 1985) was used to evaluate the tree topology reliability for 1,000 simulations. These analyses were carried out using the software Treecon (Van de Peer & Watcher, 1994). Similar cultivars were grouped based on clustering closeness among them and the outgroup cultivar. The coefficient of variation based on the assessment of the errors associated with the estimation of genetic distance of the AFLP, RAPD and SSR markers was obtained after 1,000 bootstrap samples using the dBOOT software, version 1.1 (Coelho, 2001).

RESULTS

Levels of polymorphism

The RAPD, AFLP and SSR. markers were chosen for this study based on their specific characteristics such as simplicity and low-cost (RAPD), extensive coverage of the genome (AFLP and RAPD), and reproducibility (SSR). The methods were tested in all cultivars, and repeated at least two times. However, not all methods had reliable amplification patterns and reproducibility especially the AFLP system. Overall, RAPD and SSR markers had reproducible results, while amplification patterns of AFLP markers were more difficult to reproduce.

Therefore, only AFLP results of high quality amplification were used for further analysis.

The bootstrap procedure was applied to estimate the number of loci required for stable classification of all inbred lines evaluated for each molecular marker. Results from error assessment proved that the number of loci analyzed in this study was sufficient for achieving a dendrogram stability. Values of coefficient of variation were lower than 8% in AFLP, and 3% in SSR and RAPD, with minimal rates of decrease.

Initially, 80 primers were evaluated to identify RAPD markers. For further analysis, 23 primers were selected based on their ability to amplify polymorphic fragments among lines. From a total of 157 fragments amplified, 91 were polymorphic. Amplification patterns of RAPD markers were highly conserved in all genotypes evaluated, and very few polymorphisms were identified among lines. Almost all of the observed polymorphism occurs between the *C. canephora* cultivar Apoatã and the *C. arabica* lines.

The AFLP technique allows a 64-primer combination for germplasm evaluation (Vos et al., 1995). To optimize the analysis, initially three cultivars were selected to run all primer combinations. Combinations that amplified polymorphic fragments in the selected cultivars were used in all others genotypes. Only 20 out of 64 primer combinations were able to detect polymorphism among cultivars Icatu Vermelho, Mundo Novo and Apoatã. Although 354 loci had been identified through this method, only 88 were polymorphic, and exhibit the same low diversity observed with RAPD analysis.

Early studies of microsatellites in *Coffea* identified more than 150 loci, now available in the GenBank (Combes et al., 2000). Based on these sequences, 23 primers pairs were designed and tested in coffee cultivars. All tested primers amplified polymorphic alleles, and in some loci more than two alleles were present in one genotype. However, most of polymorphisms occurred between *C. canephora* and *C. arabica* cultivars, (RAPD results).

A comparative analysis presented on Table 3 summarizes all information obtained. The AFLP method identified the highest number of loci, but the lowest amount of polymorphic fragments. Although the RAPD and SSR systems produced lower number of loci, they are more polymorphic than those identified by AFLP. The number of alleles per locus was higher in the SSR system, and also the number of effective alleles per locus. Overall, the highest assay efficiency index (Ai) was observed for AFLP and the lowest for SSR, although SSR had the highest expected heterozygosity value. All systems identified polymorphisms between cultivars, but none of them could establish an amplification pattern for each evaluated genotype.

Genetic Similarity

To estimate the genetic diversity in the evaluated germplasm, amplified data from each marker system was used for calculation of genetic distance matrices, similarity index, and drawing of inferred topology trees. The UPGMA algorithm was used for grouping all cultivars based on their genetic distances. Dendrograms representing most probable genetic relationship between cultivars are presented in Figures 1, 2 and 3. Bootstrapping values represent percentage out of 1,000 replications, and demonstrated the reliability of tree topologies, especially in those corresponding to RAPD and SSR marker systems. All analysis confirmed the low genetic diversity present in *Coffea arabica* cultivars.

Overall, trees constructed from data of RAPD (Figure 1) and SSR (Figure 2) markers displayed similar cultivars distribution, and apparently two common groups were observed. The first group includes almost all lines from *C. arabica* cultivars such as Mundo Novo, Catuaí

Table 3 - Level of polymorphism and comparative information obtained with RAPD, AFLP and SSR markers.

Parameters	Marker system		
	RAPD	AFLP	SSR
Total number of loci	157	354	23
Number of polymorphic bands	91	88	65
Average number of alleles per locus	2	2	2.87
Effective number of alleles per locus (ne)	1.13	1.35	3.63
Expected heterozygosity (H)	0.09	0.23	0.33
Number of assays (P)	23	30	23
Assay efficiency index (Ai)	7.71	15.91	3.63

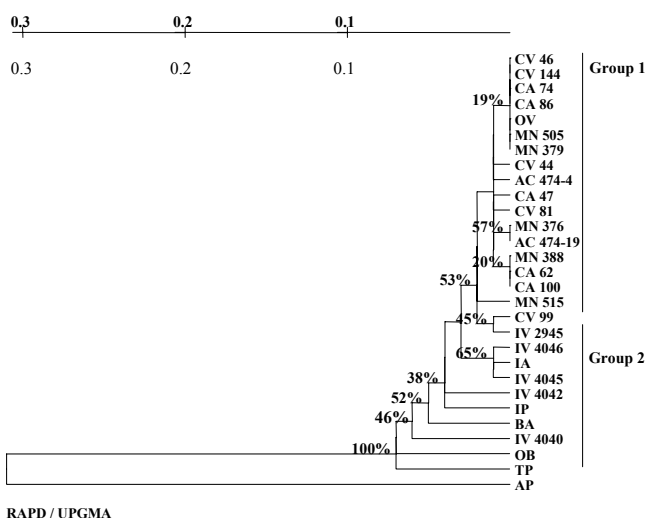
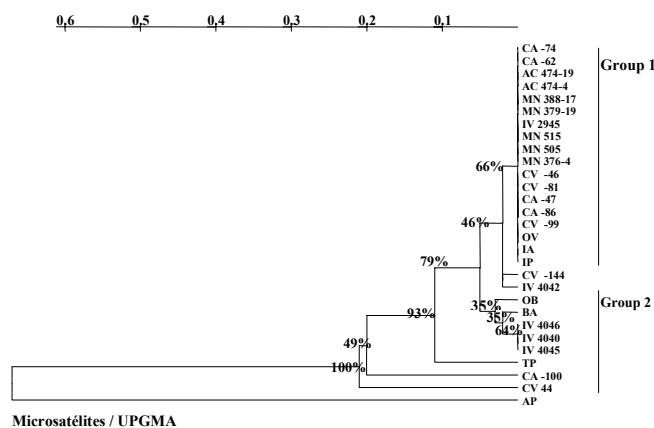
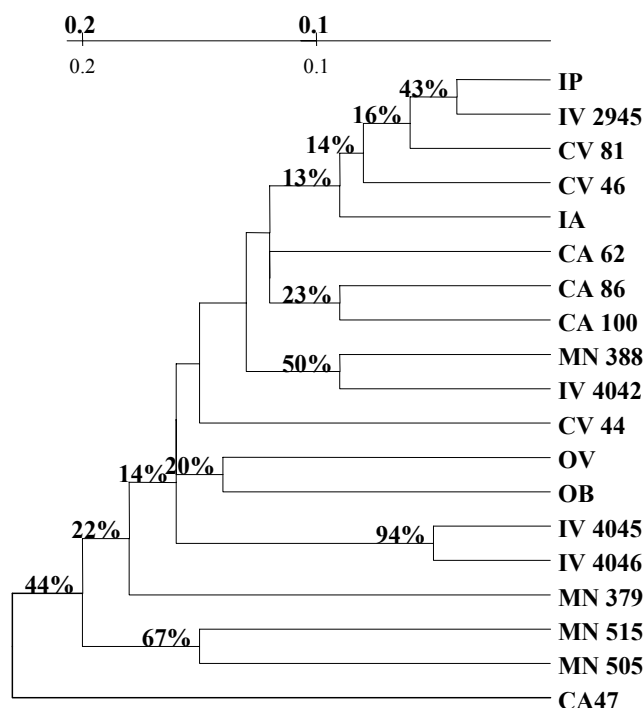


Figure 1 - Dendrogram of *Coffea* cultivars based on genetic distance obtained from RAPD markers using the UPGMA method. Numbers on branches correspond to bootstrap values (1,000 replications).



Microsatélites / UPGMA

Figure 2 - Dendrogram of *Coffea* cultivars based on genetic distance obtained from SSR markers using the UPGMA method. Numbers on branches correspond to bootstrap values (1,000 replications).



AFLP / UPGMA

Figure 3 - Dendrogram of *Coffea* cultivars based on genetic distance obtained from AFLP markers using the UPGMA method. Numbers on branches correspond to bootstrap values (1,000 replications).

and Acaia. In the second group, lines from cultivars derived from interspecific crosses, such as Icatu, Obatã and Tupi exhibit a closer relationship to the outgroup, the *C. canephora* cultivar Apoatã. However, some cultivars, such as Catuaí Vermelho 44, Catuaí Amarelo 100 and Icatu Amarelo were not observed in the same group in both analysis. Most observed polymorphisms happened between *C. arabica* and *C. canephora* cultivars.

In contrast, results of cultivars grouping using AFLP data were not satisfactory (Figure 3). Although the system was responsible for the highest number of amplified loci, the informativeness of these loci was not as accurate as expected. UPGMA grouping could not distinguish specific groups, as observed for the other marker systems. The inbred cultivars are distributed without any agreement with pedigree information.

DISCUSSION

All marker systems could identify polymorphic loci in all germplasm. However, the overall polymorphism degree detected by all systems is very low compared to similar analysis in other plant species (Lee, 1995). Most polymorphisms occur between *C. arabica* and *C. canephora*. These results agree with other studies on genetic diversity of *Coffea* commercial cultivars (Lashermes et al., 1993; Orozco-Castillo et al., 1994; Cros et al., 1998; Combes et al., 2000; Anthony et al., 2001), which in general, report low genetic diversity of *Coffea*. This is probably the result of a narrow genetic basis of the cultivated coffee, as very few accessions from African countries, such as Yemen and Ethiopia, the geographical origin of *C. arabica*, were introduced and are the bases of all breeding programs. Also, *C. arabica* is an autogamous species, what contributes to low heterozygosity levels.

To compare methods that identify polymorphisms with different basis common genetic parameters for a comparison of these methods were determined, and SSR markers were able to detect a higher average effective number of alleles (ne). This high ne results from the fact that more than two alleles were observed for some loci in individual plants. This observation could be explained by the fact that *C. arabica* is an allotetraploid species, and therefore microsatellite alleles from both parents could be detected.

The lowest efficiency index was obtained with the SSR markers. This index considered both number of loci and assays required to identify them. In this case, the AFLP method presented the highest index, for it was capable to amplify higher number of fragments. Similar relationships for molecular markers methods are also observed in other plant species (Becker et al., 1995; Pejic et al., 1998). Despite AFLP system having the highest index, the lowest degree of polymorphism was also associated with this system. Similar results were also observed for other plant species evaluated through this method (Lu et al., 1996; Powel et al., 1996; Pejic et al., 1998).

A disadvantage of the AFLP system is its technical complexity, which involves several steps of digestion, adapter ligation and selective amplification (Vos et al., 1995). The detection method chosen in this study, using fluorescent primers, is highly sensitive, and numerous peaks corresponding to amplified fragments were de-

tected. To identify real polymorphisms and to exclude non-significant amplification background, restrictive parameters were used in our analysis (see Material and Methods). These parameters were chosen based on comparisons of AFLP fragments amplified by duplicated analysis of several primers combinations. However, in this type of restrictive analysis, fragments that may be significant, but are not very frequent in the genome, are not considered for calculations. Therefore, as analysis can carry frequent mistakes, AFLP markers detection method used in this work may not be appropriate for *Coffea* germplasm characterization.

All germplasm were very close related, and no evident separation of cultivars was achieved. The trees constructed from RAPD and SSR systems were very similar. The tree corresponding to AFLP data analysis does not allow any significant grouping, and no co-relation with pedigree origin could be established with the evaluated cultivars.

Analysis of similarity trees constructed from RAPD and SSR reveal the occurrence of two groups. One group includes most of the *C. arabica* cultivars. However, no detectable genetic diversity was observed between tall cultivars, such as Mundo Novo and Acaia, and short ones, such as Catuaí and Ouro Verde. This grouping agrees with the pedigree of these cultivars, and demonstrates a common genetic origin. Actually, all Catuaí inbred lines are derived from a cross between cultivar Mundo Novo and cultivar Caturra, containing the allele *Ct* that causes shortening of branch internodes. Catuaí cultivars were developed from progenies backcrossed to Mundo Novo (Carvalho & Monaco, 1972). Therefore, genetic differences between cultivars Mundo Novo and Catuaí are expected to be very small. Similarly, the Acaia cultivar is also derived from progenies of Mundo Novo selected for larger fruit and seed size (Fazuoli, 1977).

The second group comprise most of the inter-specific cultivars represented by the tall cultivars Icatu, and medium-tall cultivars Obatã and Tupi. The pedigree origin of these cultivars is different. Icatu lines are derived from crosses of *C. canephora*, Robusta type, with *C. arabica* cultivar Bourbon Vermelho, and were specifically selected for resistance to *Hemileia vastatrix* that causes the leaf rust disease (Fazuoli, 1991). On the other hand, Tupi and Obatã were selected from crosses between *C. arabica* cultivar Villa Sarchi and Timor Hybrid, a natural *C. arabica* × *C. canephora* hybrid (Fazuoli et al., 1996). All these cultivars have in common a *C. canephora* genetic background. As this is an allogamous species, and the origin of the varieties of each cultivar is different, a greater genetic diversity among these cultivars could be expected, but was not observed. One explanation is the fact that these cultivars resulted from several backcrosses to *C. arabica*, particularly Mundo Novo. In fact, in the SSR-derived tree, cultivars Icatu Precoce,

Icatu Vermelho IAC 2945 and Icatu Amarelo were grouped together with all others *C. arabica* germplasm.

The presence of Bourbon Amarelo in the group of interspecific cultivars, in both RAPD and SSR trees, is an interesting result. The genetic origin of Bourbon Amarelo is not well known, but it is probably the result of an expontaneous cross between Arabica cultivars Bourbon Vermelho and Amarelo de Botucatu, similar to Mundo Novo and Acaia (Carvalho et al., 1957; Fazuoli, 1977). However, Bourbon Amarelo was not exposed to intense cycles of breeding, and therefore could present higher degree of genetic variability than other arabicas selected cultivars. Also, Bourbon Amarelo groups with cultivars Tupi and Obatã (Figures 1 and 2). These cultivars were selected from crosses of Villa Sarchi, a result of Sarchimor and Bourbon cross (Carvalho et al., 1991).

Even though RAPD and SSR marker systems did not had the highest A_i index, they could detect more significant degrees of polymorphism within *Coffea* germplasm, and were more efficient to determine relationships between genetic diversity and pedigree origin. However, because of the very small genetic background of commercial coffee lines, neither RAPD nor SSR is suitable for accurately clustering coffee cultivars.

The identification of individual coffee lines through the three molecular markers would be very difficult. In general, the resulting, observed groups were expected, based on previous knowledge of the genetic origin of evaluated cultivars. However, agronomic differences, such as productivity and plant architecture, are well recognized in field experiments and commercial plantations, indicating the existence of genetic differences on those cultivars. For the purpose of cultivar identification, a combination of botanical, agronomic and molecular descriptors could be complemented with gene polymorphism identification. In this case, future analysis should look for variability of genes directly related to agronomic traits.

ACKNOWLEDGEMENTS

This study was supported by Brazilian Consortium for Coffee Research and Development (CBP&D) and PRODETAB. First and second authors were supported by the CBP&D. The third author was supported by a graduate scholarship from Capes/ Brazil.

REFERENCES

- AGUIAR, A.T.E. Utilização de descritores na caracterização de cultivares e linhagens de café arábica. Campinas: Instituto Agrônômico, 2001. 98p. (Dissertação - Mestrado).
- ANTHONY, F.; BERTRAND, B.; QUIROS, O.; WILCHES, A.; LASHERMES, P.; BERTHAUD, J.; CHARRIER, A. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. *Euphytica*, v.118, p.53-65, 2001.

- ANTHONY, F.; COMBES, M.C.; ASTORGA, C.; BERTRAND, B.; GRAZIOSI, G.; LASHERMES, P. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. **Theoretical and Applied Genetics**, v.104, p.894-900, 2002.
- BECKER, J.; VOS P.; KUIPER, M.; SALAMINI, F.; HEUN, M. Combined mapping of AFLP and RFLP markers in barley. **Molecular and General Genetics**, v.249, p.65-73, 1995.
- CARVALHO, A.; MONACO, L. C. Transferência do fator Caturra para o cultivar "Mundo Novo" de *Coffea arabica*. **Bragantia**, v.31, p.379-399, 1972.
- CARVALHO, A.; ANTUNES FILHO, H.; MENDES, J.E.T.; LAZZARINI, W.; REIS, A.J.; ALOISI SOBRINHO, J.; MORAES, M.V.; NOGUEIRA, P.K.; ROCHA, T.R. Melhoramento do cafeeiro: Café Bourbon Amarelo. **Bragantia**, v.16, p.411-454, 1957.
- CARVALHO, A.; MEDINA FILHO, H.P.; FAZUOLI, L.C.; GUERREIRO-FILHO, O.; LIMA, M.M.A. Aspectos genéticos do cafeeiro (Genetic Aspects of the coffee tree). **Revista Brasileira de Genética**, v.14, p.135-183, 1991.
- COELHO, A.S.G. **Avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores**. Versão 1.1. Goiânia: UFG, Instituto de Ciências Biológicas, Departamento de Biologia Geral, 2001.
- COMBES, M.C.; ANDRZEJEWSKI, S.; ANTHONY, F.; BERTRAND, B.; ROVELLI, P.; GRAZIOSI, G.; LASHERMES, P. Characterization of microsatellites loci in *Coffea arabica* and related coffee species. **Molecular Ecology**, v.9, p.1171-1193, 2000.
- CROS, J.M.C.; COMBES, P.; TROUSLOT, F.; ANTHONY, S.; HAMON, A.; CHARRIER, A.; LASHERMES, P. Phylogenetic relationships of *Coffea* species: new evidence based on the chloroplast DNA variation analysis. **Molecular Phylogenetics and Evolution**, v.9, p.109-117, 1998.
- DICE, L.R. Measures of the amount of ecologic association between species. **Ecology**, v.26, p.297-302, 1945.
- FAZUOLI, L.C. Avaliação de progênies de café Mundo Novo (*Coffea arabica* L.). Piracicaba: USP/ESALQ, 1977. (Dissertação - Mestrado).
- FAZUOLI, L.C. Metodologias; critérios e resultados da seleção em progênies do café Icatu com resistência a *Hemileia vastatrix*. Campinas: UNICAMP, 1991. (Tese - Doutorado).
- FAZUOLI, L.C.; MEDINA FILHO H.P.; GUERREIRO FILHO O.; LIMA M.M.A.; SILVAROLLA, M.B.; GALLO P.B.; COSTA, W.M. Obatã (IAC1669-20) e Tupi (IAC1669-33); cultivares de café de porte baixo e resistentes a ferrugem. In: CONGRESSO BRASILEIRO DE PESQUISAS CAFEIRAS, 22., Águas de Lindóia. **Resumos**. Águas de Lindóia, 1996. p.149-150.
- FELSENSTEIN, J. Confidence limits on phylogenies: na approach using the bootstrap. **Evolution**, v.39, p.783-791, 1985.
- JOUQUAND, S.; CHÉRON A.; GALIBERT F. Microsatellite analysis using a two step procedure for labeling of PCR products. **Biotechniques**, v.26, p.902-905, 1999.
- LASHERMES, P.; CROS, J.; MARMEY, P.; CHARRIER, A. Use of random amplified DNA markers to analyze genetic variability and relationships of *Coffea* species. **Genetic Resources and Crop Evolution**, v.40, p.91-93, 1993.
- LEE, M. DNA markers and plant breeding programs. **Advances in Agronomy**, v.55, p.265-344, 1995.
- LU, J.; KNOX, M.R.; AMBROSE, M.J.; BROWN, J.K.M.; ELLIS, T.H.N. Comparative analysis of genetic diversity in pea assessed by RFLP and PCR-based methods. **Theoretical and Applied Genetics**, v.93, p.1103-1111, 1996.
- MORGANTE, M.; RAFALSKY, A.; BIDDLE, P.; TINGEY, S.; OLIVIERI, A.M. Genetic mapping and variability of seven soybean simple sequence repeat loci. **Genome**, v.37, p.763-769, 1994.
- NEI, M.; LI, W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. **Proceedings of the National Academy of Sciences of the USA**, v.76, p.5269-5273, 1979. Falta citação no texto.
- OROZCO-CASTILLO, C.; CHALMERS, K.J.; WAUGH, R.; POWEL, W. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. **Theoretical and Applied Genetics**, v.93, p.934-940, 1994.
- PEJIC, I.; AJMORE-MARSAN, P.; MORGANTE, M.; KOZUMPLICK, V.; CASTIGLIONI, P.; TARAMINO, G.; MOTTO, M. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs; RAPDs; SSRs; and AFLPs. **Theoretical and Applied Genetics**, v.97, p.1248-1255, 1998.
- POWELL, W.; MORGANTE, M.; ANDRE, C.; HANAFEY, M.; VOGEL, J.; TINGEY, P.; RAFALSKY A. The comparison of RFLP; RAPD; AFLP and SSR (microsatellite) markers for germplasm analysis. **Molecular Breeding**, v.2, p.225-238, 1996.
- RHODES, M.; DEARLOVE, A.; STRAW, R.; FERNANDO, S.; EVANS, A.; GREENER, M.; LACEY, T.; KELLY, M.; GIBSON, K.; BROWN, S.D.M.; MUNDY, C. High-throughput microsatellite analysis using fluorescent dUTPs for high-resolution genetic mapping of the mouse genome. **Genetical Research**, v.7, p.81-86, 1997.
- ROVELLI, P.; METTULLIO, R.; ANTHONY, F.; ANZUETO, F.; LASHERMES, P.; GRAZIOSI, G. Microsatellites in *Coffea arabica* L. In: INTERNATIONAL SEMINAR ON BIOTECHNOLOGY IN THE COFFEE AGROINDUSTRY, 3., Londrina, 2000. **Coffee biotechnology and quality**; proceedings. Dordrecht: Kluwer Academic Publishers, 2000. p.123-133.
- SERA, T.; RUAS, P.M.; RUAS, C.F.; DINIZ, L.E.C.; CARVALHO, V.P.; RAMPIM, L.; RUAS, E.A.; SILVEIRA, S.R. Genetic polymorphism among 14 elite *Coffea arabica* L. cultivars using RAPD markers associated with restriction digestion. **Genetics and Molecular Biology**, v.26, p.59-64, 2003.
- VAN DE PEER, Y.; WACHTER, R. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. **Computer Applications in the Biosciences**, v.10, p.569-570, 1994.
- VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; VAN DE LEE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.; ZABEAU, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Research**, v.23, p.4407-4414, 1995.
- VROH, L.; HARVENGT, L.; CHANDELIER, A.; MERGEAI, G.; DU JARDIN, P. Improved RAPD amplification of recalcitrant plant DNA by the use of activated charcoal during DNA extraction. **Plant Breeding**, v.115, p.205-206, 1996.

Received October 13, 2004

Accepted June 14, 2005