Research Article

Genetic relationship in *Coffea* species and parentage determination of interspecific hybrids using ISSR (Inter- Simple Sequence Repeat) markers

Paulo M. Ruas[1], Claudete F. Ruas[1], Leandro Rampim[1], Valdemar P. Carvalho[1], Eduardo A. Ruas[1] and Tumoru Sera[2]

[1]*Universidade Estadual de Londrina, Departamento de Biologia Geral, Londrina, Paraná, Brazil.*
[2]*Instituto Agronômico do Paraná (IAPAR) Londrina, Paraná, Brazil.*

Abstract

Inter-simple sequence repeat (ISSR) markers were used to evaluate genetic divergence among eight *Coffea* species and to identify the parentage of six interspecific hybrids. A total of 14 primers which contained different simple sequence repeats (SSR) were used as single primers or combined in pairs and tested for PCR amplifications. Two hundred and thirty highly reproducible fragments were amplified, which were then used to estimate the genetic similarity and to cluster the *Coffea* species and hybrids. High levels of interspecific genetic variation were revealed. The dinucleotide motif $(GA)_9T$ combined with other di- tri- and tetra-nucleotides produced a greater number of DNA fragments, mostly polymorphics, suggesting a high frequency of the poly GA microsatellite motifs in the *Coffea* genomes. The genetic similarity ranged from 0.25 between *C. racemosa* and *C. liberica* var. *dewevrei* to 0.86 between *C. arabica* var. *arabica* and Hybrid N. 2. The *C. arabica* species shared most of its markers with five of the six hybrids suggesting that it is the most likely candidate as one of the progenitors of those hybrids. These results revealed that ISSR markers could be efficiently used for genetic differentiation of the *Coffea* species and to identify the parentage of *Coffea* interspecific hybrids.

Key words: Coffea, coffee breeding, DNA markers, ISSR.

Received: November 7, 2002; Accepted: May 5, 2003.

Introduction

The genus *Coffea* subgenus *coffea* (Rubiaceae) is represented by about 80 species distributed in Continental Africa and Madagascar island (Fazuoli *et al.*, 2000). Identification of *Coffea* species is extremely important, not only from the taxonomic standpoint but also for breeding purposes. Wild species are sources of highly advantageous traits, such as resistance to diseases and insects, adaptation to many environmental conditions, and other different aspects of plant characters, for instance, root system, plant architecture, flowers, fruits, leaves, and others. *Coffea* germplasm collections have been established in many countries such as the Ivory Coast, Madagascar, Costa Rica, Colombia, and Brazil. Nevertheless, a complete *Coffea* collection is lacking since most of the available collections display only a few species, besides many *C. arabica* and *C. canephora* accessions. Additionally, in many of collections, plants were introduced as few seeds, therefore, they do not represent the original species and are actually interspecific hybrids (Fazuoli *et al.*, 2000).

Send correspondence to Paulo M. Ruas. Universidade Estadual de Londrina, Departamento de Biologia Geral, Londrina, Paraná, Brazil. E-mail: ruas@sercomtel.com.br.

While many *Coffea* species carry important characters for breeding purposes, the best quality coffee is produced from the species *C. arabica*, which represents around 70% of the world coffee production (Berthaud and Charrier, 1988). *Coffea canephora* var. *robusta* (robusta coffee), the other economically important species, responds for about 30% of the coffee produced in the world. The arabica coffee is cultivated in many developing countries representing an important source of income and employment (Anthony *et al.*, 2001). Several *C. arabica* accessions are available in the established collections. Nevertheless, breeding programs are limited by the very narrow genetic base of *C. arabica*, particularly when the aim is improvement for pest and disease resistance (van der Vossen, 1985), since most of the known cultivars have derived from the same base population (Carvalho, 1946).

Coffea arabica has an allopolyploid $(2n = 4x = 44)$ origin derived from naturally occurring interspecific hybridization (Carvalho, 1962; Charrier and Berthaud, 1985). In spite of its importance, the origin of *C. arabica* is still not very clear. Studies based on genomic in situ hybridization (GISH) and fluorescent in situ hybridization (FISH) confirm the allopolyploid nature of *C. arabica* and stated that

C. eugenioides and *C. congensis* are the diploid progenitors of *C. arabica* (Raina *et al.*, 1998). Lashermes *et al.* (1999), on the other hand, suggested that *C. arabica* arose by hybridization between *C. eugenioides,* as female parent, and *C. canephora* or related ecotypes followed by polyploidization. The mode of origin of the polyploidy remains obscure, however, it is likely to involve unreduced gametes (Lashermes *et al.*, 1999).

The knowledge of the genetic diversity and relationship available in *Coffea* collections is important for planning breeding strategies, identification of interspecific hybrids, and for germplasm conservation. It is also important to find the best choice of individuals to cross in hybrid combinations by optimizing the expression of genes of interest. In addition to morphological traits, the extent of genetic diversity has been studied in cultivated plant species using a variety of chemical and molecular characters. However, the potential to detect polymorphism depends on the method used. The analysis of six isozyme patterns in different *C. arabica* accessions from Kenya and Ethiopia, revealed absence of polymorphism contrasting with the high level of morphological variation. The results suggest that isozymes are not appropriate for the study of genetic diversity and identification of *C. arabica* accessions (Orozco-Castillo *et al.*, 1994). The development of restriction fragment length polymorphism (RFLP) analysis demonstrated the viability of studying DNA sequence polymorphisms to investigate the genetic relationship between individuals and to construct linkage maps (Botstein *et al.*, 1980). In *Coffea,* RFLP of CpDNA was used to detect genetic variation in 25 different taxa and to study the CpDNA inheritance in interspecific hybrids between *C. arabica* and *C. canephora* (Lashermes *et al.*, 1996) More recently, detection and exploitation of genetic polymorphism has been improved with the use of PCR based markers. Examples of the most commonly used PCR based techniques include random amplified polymorphic DNA (RAPD) (Williams *et al.*, 1990; Welsh and McClelland, 1990) and amplified fragment length polymorphism (AFLP) (Vos *et al.*, 1995).

Another powerful technique for DNA markers is the PCR amplification of tandemly repeated sequences referred to as Simple Sequence Repeat (SSR) or microsatellite polymorphism. Microsatellites are very polymorphic and widespread in plant genomes (Morgante and Olivieri, 1993). In SSR, the number of repeat units determines the polymorphism for fragment lengths and the heterozigote for different fragments in diploid genomes can usually be distinguished. Individual loci corresponding to a specific primer pair are then co-dominant and can be multi-allelic. The products generated have been found to be highly reproducible (Jones *et al.*, 1997).

Zietkiewics *et al.* (1994) and Kant*ety et al.* (1995) described a marker system named Inter- Simple Sequence Repeat (ISSR) amplification. The ISSR analysis involves the PCR amplification of regions between adjacent, inversely oriented microsatellites using a single simple sequence repeat (SSR)-containing primers. The technique can be applied for any species that contains a sufficient number and distribution of SSR motifs and has the advantage that genomic sequence data is not required (Gupta *et al.*, 1994; Goodwin *et al.*, 1997). The primers used in ISSR can be based on any di-, tri-, tetra, or pentanucleotide SSR motifs found at microsatellite loci, giving a wide array of possible amplification products (Blair *et al.,* 1999). The technique is more reliable than RAPD and generates larger numbers of polymorphisms per primer because variable regions in the genome are targeted (Hantula *et al.*, 1996). The potential use of ISSR markers depends on the variety and frequency of microsatellites, which changes with the species and with the targeted SSR motifs (Morgante and Olivieri, 1993). In addition, the number of bands produced by an ISSR primer with a given microsatellite repeat should reflect the relative frequency of that motif in the genome and would provide an estimate of the motif's abundance as an alternative for library hybridization (Blair *et al.,* 1999). ISSR technique has been used to assess genetic diversity in maize inbreed lines (Kant*ety, et al.*, 1995), to investigate the organization, frequency and level of polymorphism of different SSR motifs and fingerprinting in rice (Blair, *et al.*, 1999), and for fingerprinting in potato (McGregor *et al.*, 2000).

The objective of this research was the determination of the abundance and level of polymorphism of different ISSR loci in order to evaluate the genetic diversity of eight *Coffea* species, to deduce the probable origin of the *C. arabica* genomes, and to determine parentage of a further six interspecific hybrids.

Materials and Methods

Plant material

The plant material corresponding to eight *Coffea* species (one with two varieties) and six interspecific hybrids are listed in Table 1. Species and hybrids were obtained from the Germplasm Collection of the Instituto Agronômico do Paraná (IAPAR), Londrina, PR, Brazil. These *Coffea* species have been extensively used as a source to transfer important genes to the cultivated species *C. arabica* and *C. canephora*.

DNA extraction, and ISSR amplification

Genomic DNA was isolated from fresh leaf tissue following the CTAB method (Doyle and Doyle, 1987), except that CTAB was replaced by MATAB (Mixed Alyltrimethylammonium Bromide, Sigma) in the extraction buffer. The DNA concentration was estimated using a fluorometer (DyNA Quant 200, Höefer-Pharmacia), according to the manufacturer's instructions. DNA samples, obtained from at least five different plants of each species, were adjusted to 10 ng/μL and bulked (Michelmore *et al.*, 1991) for use in PCR amplifications.

Table 1 - Accessions and collection numbers of *Coffea* species and interspecific hybrids studied.

Accession	Accession number[a]
1. *C. arabica* var. *arabica*	IAPAR-75086
2. *C. canephora* var. *robusta*	IAPAR-75267
3. *C. canephora* var. *kouillou*	IAPAR-83022
4. *C. liberica* var. *dewevrei*	IAPAR-84041
5. *C. eugenioides*	IAPAR-75074
6. *C. kapakata*	IAPAR-83032
7. *C. racemosa*	IAPAR-30
8. *C. stenophylla*	IAPAR-75068
9. *C. congensis* [b]	IAC /IAPAR-83414
10. Hybrid N. 1	IAPAR-83418
11. Hybrid N. 2	IAPAR-83410
12. Hybrid N. 3	IAPAR-83419
13. Hybrid N. 4	IAPAR-36-2
14. Hybrid N. 5	IAPAR-36-1
15. Hybrid N. 6	IAPAR-35

[a]Number in the active germplasm collection at the Instituto Agronômico do Paraná (IAPAR), Londrina, Paraná, Brazil.
[b]Introduced to the IAPAR collection from the Instituto Agronômico de Campinas (IAC).

A total of 14 primers (Life Technologies) were used for ISSR amplification in the species and hybrid *Coffea* genomes. The dinucleotide motifs were all 18-mer consisting of $(GA)_9$, $(TC)_9$, and $(AT)_9$ repeats. The GA repeat was anchored at the 3' end using C or T selective nucleotide. The trinucleotide repeats were 15- and 18-mer and the tetranucleotide motifs were 16-mer in length (Table 2). For DNA amplification, the nucleotide motifs were used as single primers or combined in pairs with most of these combinations including the poly GA + T motif associated with di- tri- and tetranucleotide repeats (Table 2). The reactions were in a final volume of 15 µL containing 1.5 µL buffer 10x (75 mM Tris-/HCl) pH 9.0, 50 mM KCl and 2.0 mM $MgCl_2$, 20 mM $(NH_4)SO_4$; 0.2 mM each of dATP, dTTP, dCTP, and dGTP; 0.5 µM of each primer, 0.8 U of Taq DNA polymerase (Biotools) and 20 ng of template DNA. PCR amplification was carried out using a PTC 200 (MJ Research) thermal cycler programmed with 3 min at 94 °C for initial DNA denaturation, followed by 39 cycles of denaturation at 94 °C for 15 s; annealing at 50 °C for 1 min and 45 s; extension at 72 °C for 2 min. The final cycle was followed by a 7 min extension at 72 °C. The samples were stored at 4 °C until electrophoresis. The ISSR amplification products were resolved in 1.2 % Metaphor agarose (FMC Bioproducts) gels in 1x TAE (40 mM Tris-acetate, 1 mM EDTA pH 8.0) buffer at 120 V for 3 h and stained with ethidium bromide. The ISSR profiles were visualized under UV light, photographed with a video camera, and stored for further analysis.

Table 2 - Primer sequences with total of amplified ISSR markers and polymorphic index obtained for *Coffea* species and interspecific hybrids.

Primers repeat 5' → 3'	Number of bands	Polymorphic bands	Polymorphism index (%)
$(AG)_9$	n.a.[b]	n.a.	n.a.
$(AT)_9$	n.a.	n.a.	n.a.
$(CA)_8$	n.a.	n.a.	n.a.
$(GA)_9$	n.a.	n.a.	n.a.
$(CAG)_5$	n.a.	n.a.	n.a.
$(GAAT)_4$	n.a.	n.a.	n.a.
$(GTG)_5$	n.a.	n.a.	n.a.
$(TC)_9$	n.a.	n.a.	n.a.
$(TTA)_6$	n.a.	n.a.	n.a.
$(CCTA)_4$	5	5	100
$(TAG)_4$	8	8	100
$(GA)_9$-C	7	6	85.7
$(GACA)_4$	11	11	100
$(GGAT)_4$	11	11	100
$(GATA)_4$[a]	10	10	100
$(GGTA)_4$	3	3	100
$(GTG)_6$	2	1	50.0
$(GA)_9$-T	6	6	100
$(GA)_9 + (GACA)_4$	5	5	100
$(GA)_9$-C + $(GTG)_6$	4	4	100
$(GA)_9$-C+ $(TAG)_6$	8	8	100
$(GA)_9$-T + $(GA)_9$C	15	15	100
$(GA)_9$-T + $(TC)_9$	12	12	100
$(GA)_9$-T + $(CAG)_5$	4	4	100
$(GA)_9$-T + $(TAG)_6$	11	11	100
$(GA)_9$-T + $(GACA)_4$	7	7	100
$(GA)_9$-T + $(GGAT)_4$	20	20	100
$(GA)_9$-T + $(GGTA)_4$	13	13	100
$(TC)_9 + (GGAT)_4$	6	6	100
$(AT)_9 + (GA)_9$-T	9	9	100
$(CAG)_5 + (GACA)_4$	5	5	100
$(GTG)_6 + (GTG)_5$	8	7	87.5
$(TTA)_6 + (GTG)_6$	6	5	83.3
$(CCTA)_4 + (GA)_9$-T	13	13	100
$(GACA)_4 + (GTG)_6$	5	5	100
$(GACA)_4 + (GATA)_4$	16	16	100
Total	230	226	96.5

[a]annealing temperature of 38 °C.
[b]no amplification.

Data analysis

The ISSR products were scored for the presence (1) and absence (0) of homologous DNA bands. The similarity matrix was based on the Dice coefficient and the dendrogram was created with the UPGMA (unweighted pair-group method using arithmetical averages) method of

the NTSYS-PC (Numerical Taxonomy and Multivariate Analysis System for personal computers) software, Version 2.1 (Rohlf, 2000). The bootstrap method (Felsenstein, 1985) employed to evaluate the reliability of tree topology, was evaluated after 1000 samples. The calculations were performed with the BOOD software, version 3.1 (Coelho, 2001). The mean coefficient of variation (CV) based on the assessment of the errors associated with the estimation of the genetic similarity of the 230 ISSR markers was obtained after 1000 bootstrap samples using the Dboot software, version 1.1 (Coelho, 2001).

The matrix of genetic similarity was also used in a principal coordinate analysis (PCOORDA) to resolve the patterns of clustering among the genotypes. The cophenetic coefficient between the matrix of genetic similarity and the dendrogram were computed using the appropriate routine of the NTSYS-pc package. The significance of the cophenetic correlation was tested by the Mantel correspondence test (Mantel, 1967).

Results and Discussion

ISSR markers

The PCR amplification, performed with ISSR markers to assess the level of polymorphism in eight species (one with two varieties) and six interspecific hybrids of *Coffea* (Table 1), revealed a high percentage (96.5%) of polymorphic fragments (Table 2). This was not unexpected, since the ISSR technique amplifies microsatellite regions that are potentially polymorphic (Morgante and Olivieri, 1993). The electrophoresis pattern obtained with ISSR markers is illustrated in Figure 1. The matrix of genetic similarity, the resulting dendrogram, and the principal coordinate analysis with the graphic associations among the genotypes, are shown in Table 3 and Figures 2 and 3, respectively. The cophenetic correlation ($r = 0.91$) indicates the extent to which the clustering of genotypes demonstrated in the dendrogram accurately represents the estimates of genetic similarity among species and between species and hybrids.

According to Sneath and Sokal (1973), the number of traits required to stabilize genotype classification is very important in genetic relationship studies. The bootstrap procedure was applied to estimate the number of bands necessary to obtain a stable classification of all accessions. We observed that approximately 200 markers were sufficient for dendrogram stability and the rate of decrease was comparatively minimal beyond that (Figure 4). These results suggest that 230 markers (CV = 9.8%) are adequate for the analysis of *Coffea* species and hybrids.

In order for ISSR to be successful, pairs of simple sequence repeats must occur within a short distance (in base pairs) that can be amplified by PCR reaction, producing bands that are resolvable on standard gels (Zietkiewicz *et al.*, 1994). In this study, three dinucleotide, four
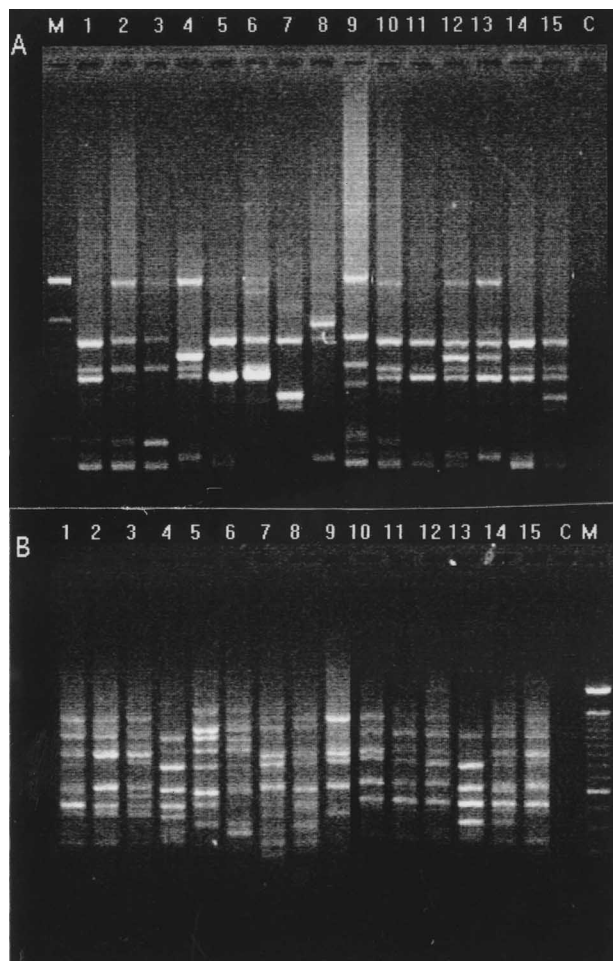

Figure 1 - Electrophoretic pattern obtained with ISSR primers. **A.** $(GA)_9T$ + $(TC)_9$ and **B.** $(GA)_9T$ + $(GA)_9C$ in eight species (one with two varieties) and six interspecific hybrids of *Coffea*. Lanes 1-9 are *Coffea* species and 10-15 are interspecific hybrids, as listed in Table 1. M is 100bp DNA ladder (Gibco).

trinucleotide, and five tetranucleotide motifs amplified a total of 230 ISSR markers, which were consistently generated in all genotypes. From all ISSR primers tested for amplification, a few produced well-defined and reproducible bands when used alone in PCR reactions. However, pair combinations of di- tri- and tetranucleotide motifs rendered a greater number of highly polymorphic bands. Best results of PCR amplification reactions were obtained with the poly (GA) dinucleotide repeat anchored at the 3' end with the T selective nucleotide (Table 2). Using this procedure varying numbers (4 to 20) of polymorphic bands were produced, depending on the SSR motif of the second primer. When used alone, the poly $(GA)_9T$ primer rendered only six bands. The combination of the poly $(GA)_9T$ with the $(GGAT)_4$ motif produced the largest number of bands (20). All these bands were polymorphic. Some significant differences were observed between the number of bands produced by other primer combinations. For example, ISSR amplifications that combined the poly $(GA)_9T$ with the

Table 3 - Genetic similarity matrix based on the Dice coefficient from 230 ISSR markers obtained for *Coffea* species and interspecific hybrids.

| Genotypes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. *C. arabica* var. arabica	1.00														
2. *C. canephora* var. robusta	0.66	1.00													
3. *C. canephora* var. kouillou	0.68	0.69	1.00												
4. *C. liberica* var. dewevrei	0.40	0.40	0.40	1.00											
5. *C. eugenioides*	0.62	0.42	0.37	0.39	1.00										
6. *C. kapakata*	0.57	0.43	0.39	0.46	0.72	1.00									
7. *C. racemosa*	0.35	0.31	0.31	0.25	0.34	0.36	1.00								
8. *C. stenophylla*	0.43	0.35	0.36	0.30	0.40	0.39	0.45	1.00							
9. *C. congensis*	0.59	0.69	0.64	0.34	0.40	0.45	0.30	0.33	1.00						
10. Hybrid N. 1	0.82	0.71	0.65	0.43	0.59	0.60	0.29	0.38	0.66	1.00					
11. Hybrid N. 2	0.86	0.60	0.59	0.41	0.67	0.63	0.36	0.39	0.54	0.79	1.00				
12. Hybrid N. 3	0.76	0.62	0.60	0.57	0.53	0.58	0.32	0.37	0.54	0.75	0.75	1.00			
13. Hybrid N. 4	0.51	0.44	0.39	0.58	0.49	0.51	0.31	0.40	0.40	0.54	0.58	0.59	1.00		
14. Hybrid N. 5	0.83	0.59	0.56	0.39	0.61	0.57	0.39	0.41	0.54	0.76	0.80	0.69	0.48	1.00	
15. Hybrid N. 6	0.78	0.61	0.61	0.37	0.53	0.53	0.50	0.41	0.52	0.72	0.71	0.69	0.44	0.74	1.00
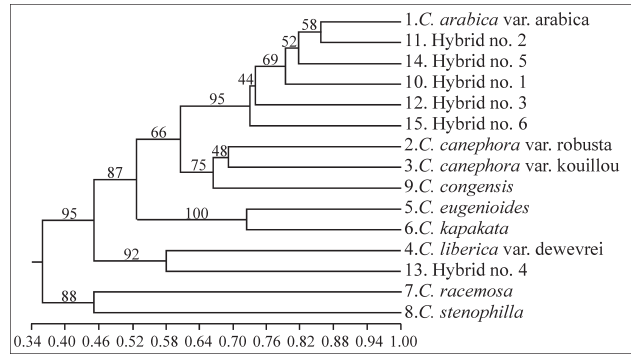
Figure 2 - Unweighted pair group method with arithmetic averages (UPGMA) dendrogram showing the relationships among eight *Coffea* species and six interspecific hybrids. The diagram was based on the Dice similarity coefficients from 230 ISSR data. Numbers at branches are bootstrap values (%) generated after 1000 replications.
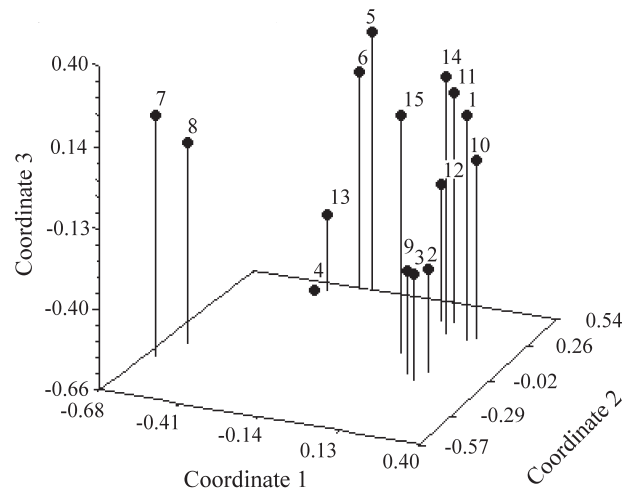

Figure 3 - Associations among *Coffea* species and interspecific hybrids obtained from a principal coordinates analyses using the Dice similarity matrix of 230 ISSR makers. The first, second and the third principal coordinates explain 19.2%, 15.3%, and 13.9% of the total variation, respectively. The numbers correspond to those listed on Table 1.

trinucleotide $(GTG)_6$ and $(CAG)_5$ motifs produced only four bands (Table 2). The effectiveness of ISSR markers in genome analysis has been reported for many plant species such as maize (Wang *et al.*, 1994; Chin *et al.*, 1996), *Arabdopsis thaliana* (Depeiges *et al.*, 1995), *Quercus macrocarpa* (Dow *et al.* 1995), rice (Akagi *et al.* 1997), wheat (Nagaoka and Ogihara, 1997), and *Vigna* (Ajibade *et al.* 2000). Nevertheless, the relative abundance of different repetitive sequences of di-, tri, and tetranucleotide motifs varied according to the plant group.

Genetic variation among *Coffea* species and parentage identification

The data matrix of genetic similarities and the dendrogram showing the relationships among the species and hybrids is presented in Table 3 and Figure 2, respectively. The principal coordinate analysis (Figure 3) was
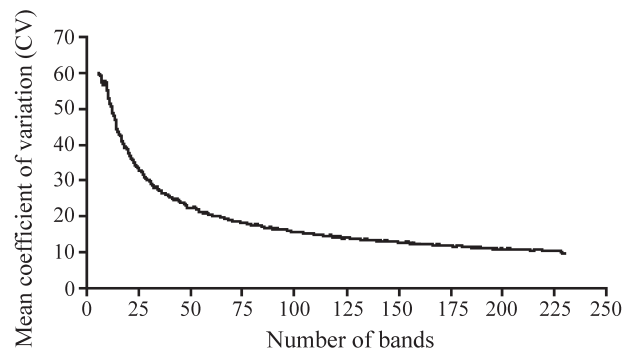
Figure 4 - Relationship between the mean coefficient of variation (CV) and sample size (number of bands). The CV value (CV = 9.8%) was obtained after 1000 bootstraps sampling using the Dice similarity coefficient.

performed to graphically display the genetic associations among species and hybrids.

The *Coffea* genus includes the sections Eucoffea and Mascarocoffea (Charrier and Berthaud, 1985). The Eucoffea section includes 12 species, which were divided into five subsections: Erythrocoffea (4 species), Pachycoffea (2 species), Melanocoffea (1 species), Mozambicoffea (3 species), and Nanocoffea with two species (Charrier, 1978). This study included the species of the Erythrocoffea subsection (*C. arabica, C. congensis, C. canephora*, and *C. eugenioides*), one of each of the Pachycoffea (*C. liberica* var. *dewevrei*) and Melanocoffea (*C. stenophylla*) subsections and two species of the Mozambicoffea subsection (*C. kapakata,* and *C. racemosa*). The associations determined with ISSR markers among *Coffea* species showed genetic similarity coefficients ranging from 0.25 (*C. racemosa* and *C. liberica* var. *dewevrei*) to 0.72 (*C. eugenioides* and *C. kapakata*) revealing high levels of genetic variation among the species studied (Table 3, Figures 2, 3).

The analysis of species relationships showed that ISSR markers placed *C. arabica* var. *arabica* closer to *C. canephora* var. *kouillou* (0.68) and *C. canephora* var. *robusta* (0.66) than to any other species (Figures 2, 3). The association of *C. canephora* and the cluster containing *C. arabica* was supported by a moderate bootstrap (BS) value of 64%. *Coffea eugenioides* was also closed to *C. arabica* with a genetic similarity coefficient of 0.62 (86% BS). These results sustain the hypothesis that this species is one of the progenitors of the arabica genomes. Although classified in the same subsection, *C. eugenioides* and *C. canephora* were placed apart with a very low genetic similarity (0.42 and 0.37), supporting the considerable variation identified by 38 isozyme markers between these two species (Berthou *et al.*, 1980). *Coffea racemosa* and *C. stenophylla* (classified in different subsections) showed a similarity coefficient of 0.45 with the cluster supported by a bootstrap value of 91%.

Coffea congensis appears associated in the same group of *C. canephora* (81% BS), showing similarity coefficients of 0.69 and 0.64 with the robusta and kouillou coffee, respectively. The coefficient of genetic similarity between *C. congensis* and *C. arabica* was 0.59. The genetic affinity between *C. canephora* and *C. congensis* was demonstrated in previous studies using RFLPs of cpDNA and mtDNA (Berthaud *et al.* 1980, 1983) and the percentage of genetic divergence (4.5%) of ITS2 sequences (Lashermes *et al.* 1997). While all these results sustain the associations shown for these species with the ISSR markers, the plants referred to as *C. congensis* used in this study may actually be interspecific hybrids between *C. canephora* and *C. congensis*. Comprehensive morphologic analysis of *C. congensis* plants that were recently introduced to the coffee gene bank of the Instituto Agronômico de Campinas (IAC) revealed noticeable differences between the genotype used in this study and the original *C. congensis* (Fazuoli *et al.,* 2000). These observations strengthen the associations obtained with ISSR markers between *C. canephora* and *C. congensis* supporting that the later could be an interspecific hybrid.

Interestingly, species classified in different subsections, such as *C. eugenioides* (Erythrocoffea) and *C. kapakata* (Mozambicoffea), were associated with higher genetic similarity (0.72) and a high bootstrap value (99%), while *C. racemosa* and *C. liberica* var. *dewevrei,* both from the subsection Mozambicoffea, were placed apart with a genetic similarity coefficient of only 0.25 (Table 3). The lowest genetic similarities were observed for *C. liberica* var. *dewevrei* with *C. stenophylla* (0.30) and with *C. racemosa* (0.25) and for *C. racemosa* with *C. congensis* and *C. canephora* var. *robusta* (0.30). Similar associations were observed with RAPD markers (Ruas *et al.*, 2000) and with comparisons of ITS2 sequences (Lashermes *et al*, 1997). The low genetic similarity values identified with the ISSR markers (Table 3) could represent the actual genetic diversity among species and clearly demonstrate the usefulness of the technique in determining the interspecific genetic relationship in *Coffea*.

Identification of *C. arabica* genomes

It is widely accepted that *C. arabica* has been derived from natural hybridization between different *Coffea* species, therefore, we attempted to use ISSR analysis to deduce the putative parents of this species, given that this knowledge is important for more efficient planning of the strategies of the coffee breeding programs. It is also well established that *C. arabica* possesses a very narrow genetic base and that it is highly susceptible to various pests and diseases (Wrigley 1995; Paillard *et al.*, 1996). The diploid species closely related to *C. arabica*, on the other hand, have wide genetic diversity and display important and useful agronomical characters (Dublin *et al.* 1991; Wrigley 1995).

The cluster analysis derived from the ISSR markers showed different coefficients of genetic similarity among the *Coffea* species (Table 3, Figures 1, 2). The highest genetic similarity of *C. arabica* var. *arabica* was with *C. canephora* var. *kouillou* (0.68), *C. canephora* var. *robusta* (0.66), and *C. eugenioides* (0.62). These associations are in agreement with previous studies about the origin of the arabica genomes. The main characteristic of the best quality coffee is the superior beverage defined by fine aroma and flavor. Besides *C. arabica*, *C. eugenioides* is the only wild species with these characteristics perhaps explaining the origin of the good flavor of arabica coffee (Fazuoli, *et al.*, 2000). Based on RFLP analysis of cpDNA and mtDNA and allozyme data, Berthou and Trouslot (1977) and Berthou *et al.* (1980, 1983) suggested that *C. eugenioides*, as maternal parent, and *C. canephora* or cytotypes with related genomes, must be considered as the progenitors of *C. arabica*. Analysis of genomic in situ hybridization (GISH) showed that the *C. eugenioides* genomic DNA preferentially hybridized with 22 chromosomes of *C. arabica* while the other 22 chromosomes hybridized with the *C. congensis* genomic DNA. Fluorescent in situ hybridization (FISH) using two ribosomal genes provided additional support to the GISH results, confirming the allopolyploid nature of *C. arabica* and showing that *C. congensis* and *C. eugenioides* are the diploid progenitors of *C. arabica* (Raina *et al*, 1998). However, Lashermes *et al* (1999) applying the RFLP and GISH clearly suggested that *C. arabica* is an amphidiploid derived from natural hybridization between *C. eugenioides* and *C. canephora*. Studies based on RAPD markers also placed *C. arabica* close to *C. eugenioides*, suggesting that this species is likely to be one of the parents of Arabic coffee (Ruas *et al*, 2000). The coefficients of genetic similarity placed *C. canephora* and *C. eugenioides* very close to *C. arabica* while *C. congensis* was less related with a similarity coefficient of 0.59, supporting previous studies and emphasizing the utility of ISSR polymorphism to deduce the origin of the arabica genome.

Parentage identification of interspecific hybrids

The ISSR markers were also used to deduce the putative parents of six interspecific hybrids (Table 1, Figures 1,2). These hybrids are important sources of genes for resistance to several coffee diseases. The ISSR markers showed high genetic affinities between *C. arabica* var. *arabica* and five interspecific hybrids, forming a group that was supported as a cluster with a bootstrap of 97% (Table 1, Figures 1, 2). ISSR data and morphological characters suggested that hybrid N. 1 derived from crossing between *C. arabica* and *C. canephora* var. *robusta*. The N. 1 hybrid was associated to *C. arabica* var. *arabica* with genetic similarity coefficient of 0.82 (supported by a 73% BS). This hybrid also showed a relatively high genetic similarity to *C. canephora* var. *robusta* (0.71) and to *C. canephora* var. *kouillou* (0.65), demonstrating that the *C. canephora* genome is well represented in this hybrid. The others species (*C. liberica* var. *dewevrei*, *C. stenophylla*, and *C. racemosa*) showed low genetic similarities with N. 1 hybrid (Table 3, Figures, 1 and 2). Results of RAPD markers were not conclusive to define the origin of N. 1 hybrid (Ruas *et al.,* 2000). The findings of ISSR markers clearly suggest that *C. arabica* and *C. canephora* are the best candidates as progenitors of N. 1 hybrid.

N. 2 and N. 5 hybrids were highly related showing a genetic similarity of 0.80 and suggesting a common origin. *Coffea arabica* shared most of its ISSR markers with N. 2 and N. 5 hybrids displaying high genetic similarity coefficients with these hybrids (0.86 and 0.83) and bootstrap values of 60% and 49%, respectively. *Coffea eugenioides* was also relatively close to both N. 2 hybrid (0.67) and N. 5 hybrid (0.61). These results are in agreement with previous studies based on RAPD data (Ruas *et al.,* 2000). Whereas N. 5 hybrid showed morphological patterns similar with *C. arabica,* N. 2 hybrid was morphologically very similar to hybrid N. 5. Thus, molecular and morphological data clearly demonstrated that both hybrids were derived from crossing between *C. eugenoides* and *C. arabica.*

The N. 3 hybrid showed genetic similarity coefficients of 0.76 with *C. arabica*, 0.57 with *C. liberica var. dewerrei*, and 0.62 with *C. canephora* var. *robusta* (Table 3, Figures 1, 2). This hybrid was very similar to *C. arabica* and *C. liberica var. dewevrei* in morphological patterns. More importantly, it carries genes for resistance to frost, leaf rust (*Hemileia vastatrix*), and nematodes (*Meloidogyne* sp), probably inherited from *C. liberica var. dewevrei.*

N. 4 hybrid, similar to N. 3 hybrid, is also an important source of genes for frost, leaf rust (*Hemileia vastatrix*), and nematode (*Meloidogyne* sp) resistance. This hybrid was less related to *C. arabica* (similarity coefficient of 0.51) and it appeared closer to *C. liberica* var. *dewevrei* showing a similarity coefficient of 0.58 and a highly supportive bootstrap value of 90%. Many morphological characters of N. 4 hybrid resemble *C. liberica* var. *dewevrei* and *C. eugenioides.* The associations based on ISSR markers are also consistent with those obtained by RAPD (Ruas *et al.,* 2000) suggesting that *C. liberica* var. *dewevrei* and *C. eugenioides* are the putative parents of the N. 4 hybrid.

N. 6 hybrid is a sterile triploid derived from crossing between tetraploid and diploid parents. The ISSR markers associated N. 6 hybrid to *C. arabica* with high genetic similarity (0.78), suggesting that they are closely related. Morphological characteristics showed that this hybrid presents narrow leaves and genes for drought, frost and leaf miner (*Perileucoptera coffeella*) resistance as observed in *C. racemosa*. The ISSR markers associated N. 6 hybrid to *C. racemosa* with a similarity coefficient of 0.50. The same conclusion was reached with RAPD data (Ruas *et al.,* 2000). Therefore, it is assumed that N. 6 hybrid derived from crossing between *C. arabica* and *C. racemosa.*

The results achieved with ISSR markers are in agreement with morphological, molecular, and crossability data suggesting that ISSR analysis could be successfully applied to study the genetic relationship, deducing putative parents of *C. arabica*, and identifying the species involved in the origin of interspecific hybrids. These results provide an essential basis for a more accurate picture of genetic diversity among *Coffea* species that may assist hybrid breeding programs.

Acknowledgements

References

Ajibade SR, Weeden NF and Chite SM (2000) Inter simple sequence repeat analysis of genetic relationship in the genus Vigna. Euphytica 111:47-55.

Akagi H, Yokozeki Y, Inagaki A and Fugimura T (1997) Highly polymorphic microsatellites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theor Appl Genet 94:61-67.

Anthony F, Bertrand B, Quiros O, Wilches A, Lachermes P, Berthaud J and Charrier A (2001) Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. Euphytica 118:53-65.

Berthaud J and Charrier A (1988) Genetic Resources of *Coffea.* In: Clarke RJ, Macrae R (eds) *Coffea* Agronomy, Elsevier Applied Science Publishers, Paris, pp 1-42.

Berthou F and Trouslot P (1977) L'analyse en électrophorère du polymorphisme enzymatique dans le genre Coffea: adaptation d'une méthod d'életrophorèse en serie. 8th Int Sci Colloquium on coffee, Paris, France.

Berthou F, Mathieu C and Vedel F (1983) Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationship in genus *Coffea* L. Theor Appl Genet 65:77-84.

Berthou F, Trouslot P, Hamon S, Vedel F and Ouetier F (1980) Analyse en électrophorèse du polymorphisme biochimique des caféiers: variation enzymatique dans dix-huit populations sauvages: *C. canephora*, *C. eugenioides et C. arabica*. Café Cacao Thé 624:313-326.

Blair MW, Panaud O and McCouch SR (1999) Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.) Theor Appl Genet 98:780-792.

Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-321.

Carvalho A (1946) Distribuição geográfica e classificação botânica do gênero *Coffea* com referência especial à espécie Arabica. Boletim da Superintendência dos Serviços do Café 21:174-180.

Carvalho A and Monaco LC (1962) Natural cross pollination in *Coffea* arabica. Proc. XVI Intern Hort Congr, Brussels, Belgium.

Charrier A (1978) La structure genetique des caféiers spontanés de la région Malgashe (Mascarocoffea), leurs relations avec les caféiers d'origine africaine (Eucoffea). Memoires ORSTOM, Paris, 224 pp.

Charrier A and Berthaud J (1985) Botanical classification of coffee. In: Clifford MN and Willson KC (eds) Coffee, Botany, Biochemistry and Production of beans and beverage. Croom Helm, London, pp 13-47.

Chin ECL, Senior ML, Shu H and Smith JSC (1996) Maize simple repetitive DNA sequences abundance and allele variation. Genome 39:886-873.

Coelho ASG (2001) BOOD - Avaliação de dendrogramas baseados em estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap, Versão 3.0. Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, GO.

Coelho ASG (2001) DBOOT - Avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores, Versão 1.1. Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, GO.

Depeiges A, Golbely C, Lenoir A, Cocherel S, Picard G, Raynal M, Grellet F and Delseny M (1995) Identification of the most represented repeated motif in *Arabidopsis thaliana* microsatellite loci. Theor Appl Genet 91:160-168.

Dow BD, Ashley MV and Hove HF (1995) Characterization of highly variable (GA/CT)n microsatellite in the burr oak, *Quercus macrocarpa*. Theor Appl Genet 91:137-141.

Doyle J J and Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytoch Bull 19:11-15.

Dublin P, Enjalrik F, Lardet L, Carron MP , Troulinder N and Pannetier C (1991) Estate crops. In: Debergh PC and Zimmerman RH (eds) Micropopagation: Technology and practice, Kluwer Academic Publ, Dordrecht, pp 337-361.

Fazuoli LC, Maluf MP, Guerreiro-Filho O, Medina-Filho HP and Silvarolla MB (2000) Breeding and biotechnology of coffee. In: Sera T, Soccol CR, Pandey A and Roussos S (eds) Coffee Biotechnology and Quality, Kluwer Academic Publishers, Dordrecht, pp 27-45.

Felsenstein J (1985) Confidence Limits on Phylogenies: an Approach Using the Bootstrap. Evolution 39(4):783-791.

Goodwin ID, Aiktken EAB and Smith LW (1997) Application of inter simple sequence repeat (ISSR) markers to plant genetics. Electrophoresis 18:1524-1528.

Gupta M, Chyi YS, Romero-Severson J and Owen JL (1994) Amplification of DNA markers from evolutionarily diverse genome using single primers of simple-sequence repeats. Theor Appl Genet 89:998-1006.

Hantula J, Dusabenygasani M and Hamelin RC (1996) Random amplified microsatellites (RAMS) - a novel method for characterizing genetic variation within fungi. Eur J For Path 26:15-166.

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, Wiel C van de, Bredemeijer G, Vosman B, Matthes M, Maly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi R, Marmiroli N, Aert R, Volckaert G, JRueda J, Linaacero R, Vazque A and Karp A (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breeding 3:381-390.

Kantety RV, Zeng X, Bebbetzen JL and Zehr BE (1995) Assessment of genetic diversity in dent and popcorn (*Zea mays* L.) inbreed lines using inter-simple sequence repeat (ISSR) amplification. Mol Breeding 1:36-373.

Lashermes P, Combes MC, Roberts J, Trouslot P, D'Hont A, Anthony F and Charrier A (1999) Molecular characterization and origin of the *Coffea arabica* L. genome. Molec Gen Genet 261:259-266.

Lashermes P, Combes MC, Trouslot P and Charrier A (1997) Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. Theor Appl Genet 94:947-955.

Lashermes P, Cros J, Combes MC, Trouslot P, Anthony F, Hamon S and Charrier A (1996) Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. Theor Appl Genet 93:626-632.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:299-220.

McGregor CE, Lambert CA, Greyling MM, Louw JH and Warnich LA (2000) comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.). Euphytica 113:135-144.

Michelmore RW, Paran I and Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828-9832.

Morgante M and Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. The Plant J 3:175-182.

Nagaoka T and Ogihara Y (1997) Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. Theor Appl Genet 94:597-602.

Orozco-Castillo C, Charmers KJ and Powell RW (1994) Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. Theor Appl Genet 87:934-940.

Paillard M, Lashermes P and Pétrard V (1996) Construction of a molecular linkage maps in coffee. Theor Appl Genet 93:41-47.

Raina SN, Mukai Y and Yamamoto M (1998) *In situ* hybridization identifies the diploid progenitor species of *Coffea arabica* (Rubiaceae). Theor Appl Genet 97:1204-1209.

Rohlf FJ (2000) NTSYS-pc Numerical Taxonomy and Multivariate Analysis System version 2.1. Owner manual.

Ruas PM, Diniz LEC, Ruas CF and Sera T (2000) Genetic Polymorphism in species and hybrids of *Coffea* revealed by RAPD. In: Sera T, Soccol CR, Pandey A, Roussos S (eds) Coffee Biotechnology and Quality. Kluwer Academic Publishers, Dordrecht, pp 187-195

Sneath PHA and Sokal RR (1973) Numerical taxonomy. Freeman. San Francisco, 573 pp.

Van der Vossen HAM (1985) Coffee selection and breeding In: Clifford MN and Willson KC (eds) Coffee: Botany, Biochemistry and Production of Beans and Beverage. Croom Helm, London, p 48-96.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M and Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucl Ac Res 23:4407-4414.

Wang, Z, Weber JL, Zhong G and Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1-6.

Welsh J and McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids 18:7213-7218.

Williams JGK, Kubelik AR, Livak KJ and Rafalski JA (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Ac Res 18:6531-6535.

Wriglei G (1995) Coffee. In: Smartt J and Simmonds NW (eds) Evolution of crop plants. Longman Scientific and Technical, London, pp 438-443.

Zietkiewics E, Rafalski A and Labuda D (1994) Genome fingerprint by sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176-183.

Editor: Marcio de Castro Silva-Filho