GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers

R.F. Missio[1], E.T. Caixeta[2,3], E.M. Zambolim[2], G.F. Pena[2], L. Zambolim[2], L.A.S. Dias[4] and N.S. Sakiyama[2,4]

[1]Universidade Federal do Paraná, Campus Palotina, Palotina, PR, Brasil
[2]Laboratório de Biotecnologia do Cafeeiro (BioCafé), Universidade Federal de Viçosa, Viçosa, MG, Brasil
[3]Empresa Brasileira de Pesquisa Agropecuária, Viçosa, MG, Brasil
[4]Departamento de Fitotecnia, Universidade Federal de Viçosa, Viçosa, MG, Brasil

Corresponding author: L.A.S. Dias
E-mail: lasdias@ufv

Genet. Mol. Res. 10 (4): 2366-2381 (2011)
Received November 8, 2010
Accepted June 25, 2011
Published October 6, 2011
DOI http://dx.doi.org/10.4238/2011.October.6.2

ABSTRACT. Coffee is one of the main agrifood commodities traded worldwide. In 2009, coffee accounted for 6.1% of the value of Brazilian agricultural production, generating a revenue of US$6 billion. Despite the importance of coffee production in Brazil, it is supported by a narrow genetic base, with few accessions. Molecular differentiation and diversity of a coffee breeding program were assessed with gSSR and EST-SSR markers. The study comprised 24 coffee accessions according to their genetic origin: arabica accessions (six traditional genotypes of *C. arabica*), resistant arabica (six leaf rust-resistant *C. arabica* genotypes with introgression of Híbrido de Timor), robusta (five *C. canephora* genotypes), Híbrido de Timor (three *C. arabica* x *C. canephora*), triploids (three *C. arabica* x *C. racemosa*), and racemosa (one *C. racemosa*). Allele and polymorphism analysis, AMOVA, the Student *t*-test, Jaccard's dissimilarity coefficient, cluster analysis,

correlation of genetic distances, and discriminant analysis, were performed. EST-SSR markers gave 25 exclusive alleles per genetic group, while gSSR showed 47, which will be useful for differentiating accessions and for fingerprinting varieties. The gSSR markers detected a higher percentage of polymorphism among (35% higher on average) and within (42.9% higher on average) the genetic groups, compared to EST-SSR markers. The highest percentage of polymorphism within the genetic groups was found with gSSR markers for robusta (89.2%) and for resistant arabica (39.5%). It was possible to differentiate all genotypes including the arabica-related accessions. Nevertheless, combined use of gSSR and EST-SSR markers is recommended for coffee molecular characterization, because EST-SSRs can provide complementary information.

Key words: Microsatellite marker; *Coffea*; Discriminant analysis; Genetic diversity

INTRODUCTION

Coffee is one of the main agrifood commodities marketed worldwide. According to the International Coffee Organization (http://www.ico.org), the world's production is estimated between 133 and 135 million 60-kg bags, in the 2010/11 crop. Brazil is the leading producer (47.2 million bags) and exporter (31 million bags) and the second largest consumer (16.2 million bags). In 2009, coffee accounted for 6.1% of the value of Brazilian agricultural production, generating a revenue of US$6 billion. Despite the importance of coffee production, it is supported by a narrow genetic base, built on a few accessions.

The genetic structure and diversity of spontaneous, sub-spontaneous, and cultivated accessions of *Coffea arabica* were recently assessed with simple sequence repeat (SSR) markers (Silvestrini et al., 2007). The study included 72 samples of spontaneous and sub-spontaneous accessions from Ethiopia, one sub-spontaneous accession from Eritrea, and 27 cultivated accessions (13 from Yemen, one from Malawi, and 13 Brazilian cultivars). Additionally, five accessions of *C. eugenioides*, four of *C. racemosa*, and six of *C. canephora* were also sampled. By cluster analysis based on Jaccard's coefficient, all species were distinguished, and cultivated *C. arabica* accessions were distinguished from spontaneous and sub-spontaneous ones. The Brazilian cultivars were distinguished from Yemen-cultivated accessions; however, both groups exhibited a very low genetic diversity, in agreement with the already well-known narrow genetic base of cultivated *C. arabica*. Nevertheless, the low genetic diversity of cultivated *C. arabica* has been well exploited by breeders, who have achieved success in obtaining improved cultivars (see list in http://www.agricultura.gov.br).

As the molecular mapping and scan of the genomes increased the possibility to unlock the genetic potential of plant germplasm resources (Tanksley and McCouch, 1997), the use of the spontaneous and sub-spontaneous plants will likely play an important rule in the future of *C. arabica* breeding programs. Meanwhile, understanding the genetic structure and diversity within the cultivated gene pool cannot be postponed, because it has been shown to be useful for breeding purposes even though small. Assessing the diversity of cultivated *C.*

arabica with DNA markers may be possible, but it still remains unsolved due to the lack of an appropriate number of polymorphic markers. SSR markers have been used in many plant genetic studies because they normally show good level of polymorphism among genotypes, with co-dominance of a variable number of alleles, and wide distribution throughout the genome. The easy polymerase chain reaction (PCR) detection of a specific SSR locus, by using specific primers designed for the conservative DNA sequences flanking such a locus, is another known advantage. However, the availability of polymorphic SSR markers is still a constraint in *C. arabica*.

Since SSRs are present in the coding and non-coding regions of the genome (Gur-Arie et al., 2000) they can be observed in cDNA or gDNA fragments by different methods. For instance, an SSR-enriched genomic library can be obtained by selective hybridization of gDNA fragments with complementary SSR probes (Hamilton et al., 1999). Another example is the identification of SSRs by *in silico* data mining of sequenced cDNA (e.g., expressed sequence tag - EST) databanks (Scott et al., 2000). These two methods are used to develop gSSR and EST-SSR markers, respectively.

Enriched genomic libraries of *C. arabica* were constructed with the probes $(GT)_{15}$ and $(AGG)_{10}$, and specific primers were designed to obtain gSSR markers. The PCR test with two *C. arabica* genotypes validated new gSSR markers, which were available for further genetic studies of this species (Missio et al., 2009a). A sampled gSSR evaluation showed mean polymorphic information content values ranging from 0.22 for *C. arabica* to 0.46 for *C. canephora* (Missio et al., 2010).

EST-SSR markers developed from the Brazilian Coffee Genome Project databank were used to assess the genetic diversity of coffee populations. The level of polymorphism was 88.2% within *C. canephora* accessions and 11.8% within *C. arabica* accessions. The known narrow genetic base and autogamy were probably responsible for the low level of polymorphism in *C. arabica* (Missio et al., 2009b). It was previously reported that because EST-SSR markers are originated from highly conserved genomic regions, they may show a lower level of polymorphism compared to gSSR markers, which have a wider distribution in the genome (Varshney et al., 2005).

We present here a case study of genetic diversity of valuable accessions of our *C. arabica* breeding program using gSSR and EST-SSR markers. We also evaluated and compared the discrimination capacity of these two types of SSR and their effectiveness in establishing genetic relationships in *C. arabica*.

MATERIAL AND METHODS

Genetic material

The study comprised 24 coffee tree accessions of the EPAMIG/UFV Coffee Breeding Program germplasm (Table 1). The accessions were grouped according to their genetic origin: arabica accessions (six traditional genotypes of *C. arabica*), resistant arabica (six leaf rust-resistant *C. arabica* genotypes with introgression of Híbrido de Timor), robusta (five *C. canephora* genotypes), Híbrido de Timor (three *C. arabica* x *C. canephora*), triploids (three *C. arabica* x *C. racemosa*), and racemosa (one *C. racemosa*). The DNA of each accession was extracted from young leaves of a single plant, according to the protocol described by Diniz et al. (2005).

Table 1. List of coffee accessions evaluated.

Group	Accession	Species (ploidy)
Arabica	UFV 2144 (Catuaí Vermelho IAC 44)	*Coffea arabica* (2n = 4x = 44)
Arabica	Típica UFV 2945	*Coffea arabica* (2n = 4x = 44)
Arabica	Bourbon UFV 2952	*Coffea arabica* (2n = 4x = 44)
Arabica	Bourbon Amarelo UFV 535-1	*Coffea arabica* (2n = 4x = 44)
Arabica	Arabica UFV 10832	*Coffea arabica* (2n = 4x = 44)
Arabica	Bourbon Amarelo UFV 10745	*Coffea arabica* (2n = 4x = 44)
Robusta	T 3751 (Robusta)	*Coffea canephora* (2n = 2x = 22)
Robusta	T 3580 (Robusta)	*Coffea canephora* (2n = 2x = 22)
Robusta	Conillon UFV 513 (Conillon)	*Coffea canephora* (2n = 2x = 22)
Robusta	Guarini UFV 514 (Robusta)	*Coffea canephora* (2n = 2x = 22)
Robusta	Apoatã IAC 2258 (Robusta)	*Coffea canephora* (2n = 2x = 22)
Híbrido de Timor	Híbrido de Timor CIFC 832/2	*C. arabica* x *C. canephora* (2n = 4x = 44)
Híbrido de Timor	Híbrido de Timor CIFC 4106	*C. arabica* x *C. canephora* (2n = 4x = 44)
Híbrido de Timor	Híbrido de Timor CIFC 1343/269	*C. arabica* x *C. canephora* (2n = 4x = 44)
Triploid	UFV 557-2	Triploid (*C. arabica* x *C. racemosa*) (2n = 3x = 33)
Triploid	UFV 557-3	Triploid (*C. arabica* x *C. racemosa*) (2n = 3x = 33)
Triploid	UFV 557-4	Triploid (*C. arabica* x *C. racemosa*) (2n = 3x = 33)
Racemosa	*Coffea racemosa*	*Coffea racemosa* (2n = 2x = 22)
Resistant arabica	Catiguá MG2	Commercial variety (*C. arabica* x HT) (2n = 4x = 44)
Resistant arabica	IAPAR 59	Commercial variety (*C. arabica* x HT) (2n = 4x = 44)
Resistant arabica	Oeiras MG6851	Commercial variety (*C. arabica* x HT) (2n = 4x = 44)
Resistant arabica	Sacramento MG1	Commercial variety (*C. arabica* x HT) (2n = 4x = 44)
Resistant arabica	Catucai Amarelo 2SL	Commercial variety (*C. arabica* x Icatu Vermelho) (2n = 4x = 44)
Resistant arabica	Obatã Amarelo IAC 4932	Commercial variety (*C. arabica* x HT) (2n = 4x = 44)

HT = Híbrido de Timor (*C. arabica* x *C. canephora*).

Microsatellite markers

Two sets of SSR markers with well-defined amplification products were used (Table 2). The first, containing 17 primer pairs, was obtained from *C. arabica* ESTs of the Brazilian Coffee Genome Project (Missio et al., 2009b). The second, containing 18 primer pairs, was obtained from an enriched genomic library of the *C. arabica* genotype Bourbon Amarelo UFV 570 (Missio et al., 2009a). PCR was done on a total volume of 20 µL containing 50 ng genomic DNA, 0.6 U Taq DNA polymerase and 1X buffer (Promega), 1 mM $MgCl_2$, 150 µM of each dNTP and 0.1 µM of each primer. The touchdown-PCR procedure was performed, which consisted of initial denaturation at 94°C for 2 min, followed by 13 cycles at 94°C for 30 s, 67° to 55°C for 30 s, decreasing by 1°C at each cycle, and 72°C for 30 s. Another 30 cycles were done at 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s, followed by a final extension at 72°C for 8 min. Polymorphism was visualized on a silver-stained 6% denaturing polyacrylamide gel (Brito et al., 2010).

Allelic analysis and percentage of polymorphism

The number of alleles per primer, the maximum number of alleles per accession, the number of exclusive alleles per genetic group, and the presence of null alleles (alleles with no amplification product in one or more accessions) were analyzed. The polymorphism among and within *Coffea* groups were evaluated with EST-SSR and gSSR markers. The percentage of polymorphism was calculated by dividing the number of polymorphic bands by the total number of amplified bands for each group.

AMOVA, *t*-test and cluster analysis

The allelic data of EST-SSR and gSSR markers were scored in a binary matrix, where 1 represented presence and 0 (zero) absence of bands. The genetic distances for all the pairwise combinations were calculated using Jaccard's dissimilarity coefficient (Jaccard, 1908). Analysis of molecular variance (AMOVA) among and within *Coffea* genetic groups was performed according to Excoffier et al. (1992), using the Genes software (Cruz, 2007). The mean genetic distance coefficients within genetic groups were compared by the Student *t*-test, $P < 0.05$, using the SAS software (SAS Institute, 1996). A simplified representation of the genetic distances by Jaccard was based on a dendrogram obtained with the unweighted pair group method with arithmetic averages (UPGMA), using the Genes software (Cruz, 2007). Three dendrograms were obtained with: a) EST-SSR markers only, b) gSSR markers only, and c) both EST-SSR and gSSR markers. Bootstrap iterations were calculated to test the robustness of the nodes of the dendrograms, using the Treecon software (Van de Peer and De Wachter, 1994).

Correlation of genetic distances

The effect of the number of sampled alleles on the results of Jaccard's genetic distance was analyzed. The correlation and the stress were calculated between each Jaccard's genetic distance matrix of M_i sampled alleles ($5 \leq M_i \leq 183$) and the original genetic distance matrix with 183 alleles (total number of alleles from gSSR and EST-SSR markers). The sampling was started with five alleles ($M_i = 5$), with 30 random replications. The new sampling was incremented every five alleles until all 183 alleles were sampled ($M_i = 183$). The Genes software (Cruz, 2007) was used in these analyses. The dendrograms obtained by cluster analyses with a different number of alleles were compared with the dendrogram obtained with all 183 alleles. As a result of this comparison, the minimum number of alleles was estimated to obtain a similar dendrogram for this case study.

Discriminant analysis

Stepwise discriminant analysis was applied to identify the best marker alleles to differentiate the genetic groups. The Proc Stepdisc procedure of the SAS software was used (SAS Institute, 1996, Fahima et al., 1999, Beharav and Nevo, 2003). This procedure was applied to the EST-SSR and gSSR markers, separately.

RESULTS

Allelic analysis and percentage of polymorphism

The 17 EST-SSR and 18 gSSR markers, when analyzed in 24 coffee trees of the EPAMIG/UFV Coffee Breeding Program germplasm, generated 87 and 96 alleles; the mean numbers of alleles per primer were 5.1 and 5.3, respectively. The number of alleles per primer ranged from 3 to 11 for the EST-SSR markers, and from 3 to 13 for the gSSR (Table 2). Additionally, 14 (82%) EST-SSR and 7 (39%) gSSR markers showed three or four alleles in tetraploid individuals. No SSR marker showed more than two alleles per individual in the diploid accessions *C. canephora* and *C. racemosa*.

Table 2. Description and comparison of EST-SSR and gSSR markers applied to 24 coffee accessions for a total of 183 alleles.

Marker	Primer sequence	Repeat	Size of fragment (bp)	Tm (ºC)	Number of alleles	Null alleles	Maximum number of alleles/individual
EST-SSR							
EST-SSR 005	F: GGTCCCTGATACCAAACCTC R: CAAGATCACAGGAAGGCACA	$(CA)_7$	129	58	3	N	3
EST-SSR 007	F: AGTGGCTGGGAACAAAGAGA R: TTCTCCTCCCGCAAACAGAG	$(GTC)_5$	117	58	10	N	3
EST-SSR 010	F: CTTCTTCATCCAACAACACG R: TGCCATTCCACTGTGTCACT	$(TG)_7(TA)_4$ // $(GT)_9$	152	51	11	N	3
EST-SSR 012	F: CGCGTCTGCAACAAAGGTA R: GGGTGCTAGTCAGAGCCATTT	$(CA)_7(AC)_3$	118	53	4	N	3
EST-SSR 013	F: GCCTTGCTCATATCTGCTGTCT R: GATCCTTCAACTGAGCCAAA	$(CA)_9$	141	52	3	N	3
EST-SSR 023	F: GCCATTTCACAATCTCACCTC R: AGACCCAGCAGACAACAACA	$(TA)_{15}$	207	58	5	N	3
EST-SSR 025	F: AGATACCCACCGCCTAATCCT R: GCAACAACTTCTGCTCATCC	$(AT)_7$	108	53	3	N	3
EST-SSR 027	F: ATGGAAGTGTCCTTGTCGTG R: ATGTCGGTGGGTCGGTCAAA	$(AT)_3(AC)_4(AT)_9$	259	53	6	N	4
EST-SSR 029	F: TTAACCTCCTGCCACACA R: GCCCAAATAAATCCCTCCA	$(AT)_{10}$	169	48	6	N	4
EST-SSR 047	F: GCGTCAATTAAGCCTCATCATC R: CAGCCGCTTGCAAAGTAATC	$(CAA)_5N_3(CTG)_6$	178	52	7	N	4
EST-SSR 048	F: TCCTCCTCGTGCTTCTCAAC R: GGCAGCATTCTCCTGATCCT	$(AAG)_6N_9(TGA)_4$	126	54	4	N	2
EST-SSR 054	F: GTTAGCCGTTGGTGATGGAA R: TTGGTCGAGGGAGGAAGAAC	$(TAT)_7$	184	53	4	N	2
EST-SSR 055	F: TACCACCAGCATCCAGACCA R: TGGGAGGAAATCAAGAGCAA	$(TAT)_5$	167	52	4	N	2
EST-SSR 057	F: TTGTGTTCTTTCTCCACCTC R: CAGGAGTCGTATAACGCTGAA	$(TTA)_6$	90	51	4	N	3
EST-SSR 058	F: CACACTTGATTCCGCTCACA R: GGATGCTTGCTGCTGCTATT	$(AAT)_4$	201	58	4	N	3
EST-SSR 069	F: TGAGCTAACCAAGACCAGTTCC R: CAACAGGAAATCACCGCCTA	$(GCG)_6$	101	53	3	N	3
EST-SSR 073	F: GAGGTCTTCCCACCACAACA R: GGATACGAGAGTCCCTTCCA	$(CGC)_2(GGC)_5$	160	54	6	N	4
Subtotal					87		
Mean					5.1		

Continued on next page

Table 2. Continued.

Marker	Primer sequence	Repeat	Size offragment (bp)	Tm (ºC)	Number of alleles	Null alleles	Maximum number of alleles/individual
gSSR							
SSRCa 003	F: ATGATTCGTAGGTGGAGTGG R: CTAAGCCGCAAATGACAGA	$(GT)_{12}$	196	57	4	Y	2
SSRCa 016	F: AGCAGATTCCATCCTTATCCT R: CCACTAATCCATTCCATTCC	$(GAA)_3$ // $(GGAAAG)_3$	172	56	3	N	2
SSRCa 018	F: GTCTCGTTTCACGCTCTCTC R: ATTTTTGGCACGGTATGTTC	$(GT)_{18}(GA)_{10}$	115	57	13	N	3
SSRCa 019	F: GGGTTAGATAGAGCAAGAATGA R: CTGTGAAGGTGTGGAGTTTT	$(GA)_{11}$	329	55	4	N	2
SSRCa 020	F: GGTAGGCGAAGGACAGATAA R: TGGGGCAGAGTGAAGATAAG	$(AGA)G(AGA)_3$ // $(TG)_4$// $(ATT)_6$	264	57	3	N	2
SSRCa 026	F: GAATCTGGTGGGCTTTGA R: AAGGAGAGGGGAAGAAAATG	$(T)_{16}N_{12}(TC)_7$ // $(CAC)_4$	289	57	5	N	2
SSRCa 052	F: GATGGAAACCCAGAAAGTTG R: TAGAAGGGCTTTGACTGGAC	$(TTG)_7$	129	57	4	Y	3
SSRCa 062	F: AAGTTATTAGGGCAAGAGTGGA R: AAGCTCCAAGACCAAAGATG	$(CAA)_2G(AGAA)_2$ // $(AG)_4N_8(GA)_4$	275	57	3	N	3
SSRCa 068	F: ATGTTGTTGGAGGCATTTTC R: AGGAGCAGTTGTTGTTTTCC	$(AGG)_7$ // $(GAA)_4$	236	57	4	Y	4
SSRCa 081	F: ACCGTTGTTGGATATCTTTG R: GGTTGAACCTAGACCTTATTT	$(CT)_{38}$	229	53	4	Y	2
SSRCa 082	F: GCTTGTTTCCATCGCTAAA R: TTACACGTCAACCCACAAAC	$(CT)_{17}CG(CT)_6$	178	56	7	N	2
SSRCa 083	F: TCCAACAACATTAAGCGTATTC R: GACAAACCTGAGGGAAAAGA	$(TC)_{32}$	223	56	5	N	2
SSRCa 087	F: TCACTCTCGCAGACACACTAC R: GCAGAGATGATCACAAGTCC	$(TC)_{22}$	143	56	8	N	3
SSRCa 088	F: TACCTCTCCTCCTCCTTCCT R: ATTTCTATGGACCGGCAAC	$(TTTTCT)_3$	180	57	5	N	3
SSRCa 091	F: CGTCTCGTATCACGCTCTC R: TGTTCCTCGTTCCTCTCTCT	$(GT)_8(GA)_{10}$	110	56	10	N	2
SSRCa 092	F: ATAGCCTGAGCCGTAACCA R: GGGTAATTATGACGAGGGACA	$(CCA)_7CT(TCCACC)_5$	142	58	6	N	3
SSRCa 094	F: GTGTCCTAGGGAAGGGTAAG R: GAGTGCTAGGAGAGGGAGAG	$(TC)4(TTCT)_3$ // $(TTTCCT)_3(TTTC)_5$	195	55	4	N	2
SSRCa 095	F: GAGAGAGCCGAGTGAAGAGA R: GAGAGAGAAGCCATGATTTGA	$(TG)_{11}$	185	57	4	N	2
Subtotal					96		
Mean					5.3		

Tm = melting temperature; Y (yes) = presence of null alleles; N (no) = alleles with no amplification product in one or more *C. arabica* accessions.

EST-SSR markers displayed a total of 25 exclusive alleles per genetic group, while gSSR showed 47 (Table 3). The difference was mainly due to the exclusive alleles found in racemosa, robusta, and triploid genetic groups. Null alleles were observed in approximately 22% of gSSR markers and not observed in EST-SSR markers. All SSR markers showed polymorphisms among the 24 *Coffea* accessions. The gSSR markers detected a higher percentage of polymorphism among (35% higher on average) and within (42.9% higher on average) the genetic groups, compared to EST-SSR markers (Table 3). The polymorphism of gSSR markers was approximately 170% higher than EST-SSR within the Híbrido de Timor and 158% higher than EST-SSR within the arabica group (Table 3). The highest percentage of polymorphism within the genetic groups was found with gSSR markers for robusta (89.2%) and for resistant arabica (39.5%). The lowest percentage of polymorphism within the genetic groups was found with EST-SSR markers for Híbrido de Timor (10.6%) and arabica (11.1%).

The highest percentages of polymorphism among genetic groups involved the robusta group, with both EST-SSR and gSSR. The highest polymorphisms were found between the robusta and triploid groups (97.5%), and between the robusta and racemosa groups (97.3%), with gSSR markers. The lowest percentages of polymorphism among genetic groups involved the arabica group, both with EST-SSR and gSSR. The lowest polymorphism was found between the arabica and Híbrido de Timor groups (22%), with EST-SSR. With gSSR, the lowest polymorphisms were found between Híbrido de Timor and resistant arabica groups (41.3 %), and between arabica and triploid group (43.6 %) (Table 3).

Table 3. Percentage of polymorphism among and within *Coffea* groups.

Groups	Polymorphism (%)		Superiority (%)	Mean number of alleles (exclusive alleles)	
	EST-SSR	gSSR		EST-SSR	gSSR
Within					
Arabica (Ca)	11.1	28.6	157.7	4.0 (0)	1.9 (1)
Robusta (Cc)	77.8	89.2	14.7	3.3 (19)	3.6 (29)
Híbrido de Timor (HT)	10.6	28.6	169.8	3.2 (0)	1.9 (3)
Triploids (T)	12.3	14.3	16.3	3.7 (4)	1.6 (5)
Leaf rust-resistant arabica (LRCa)	28.0	39.5	41.1	3.7 (1)	2.0 (0)
Racemosa (Cr)	-	-	-	2.0 (1)	1.3 (9)
Mean	28.0	40.0	42.9		
Among					
Ca x Cc	84.5	94.0	11.2		
Ca x HT	22.0	47.8	116.4		
Ca x T	30.4	43.6	43.4		
Ca x LRCa	33.3	43.6	30.9		
Ca x Cr	30.8	91.1	195.8		
Cc x HT	83.1	94.0	13.1		
Cc x T	86.1	97.5	13.2		
Cc x LRCa	86.5	96.3	11.3		
Cc x Cr	90.1	97.3	8.0		
HT x T	36.7	48.0	30.8		
HT x LRCa	37.7	41.3	9.5		
HT x Cr	55.4	88.5	59.7		
T x LRCa a	36.1	50.0	38.5		
T x Cr	44.8	85.1	90.0		
LRCa x Cr	61.4	87.8	43.0		
Mean	54.6	73.7	35.0		

Comparison of the efficiency of EST-SSR and gSSR markers. - = not evaluated (contained only one accession); Superiority = percentage of superiority of gSSR markers relative to EST-SSR markers in relation to polymorphism (example: [(28.6-11.1) x 100] / 11.1 = 157.7%).

Genetic diversity with AMOVA, Student *t*-test and cluster analysis

AMOVA (Table 4) revealed approximately 64% of the genetic variation among groups and 36% within groups with EST-SSR, while approximately 54% of the genetic variation was observed among groups and 46% within groups with gSSR. Apparently, gSSR were more efficient than EST-SSR in assessing the genetic variation within groups.

Table 4. Analysis of molecular variance (AMOVA) among and within *Coffea* genetic groups.

Source of variation	d.f.	Sum of squares		Variation component		Variation (%)	
		EST-SSR	gSSR	EST-SSR	gSSR	EST-SSR	gSSR
Among groups	5	151.04	148.03	6.88	6.33	64.22	54.34
Within groups	18	60.00	95.08	3.83	5.32	35.78	45.66
Total	23	220.04	243.83	10.71	11.66	100.00	100.00

d.f. = degrees of freedom.

The Student *t*-test was applied for analyses within each genetic group (Figure 1). The highest genetic distance coefficients were found within the robusta group, followed by Híbrido de Timor and leaf rust-resistant arabica. On the other hand, the lowest genetic diversity coefficients were found within the arabica and triploid groups (Figure 1). gSSR markers yielded the highest genetic distance coefficients in all groups; they were statistically superior compared to EST-SSR for accessing diversity within the arabica ($t = 4.48$; $P < 0.01$), robusta ($t = 3.62$; $P < 0.05$) and Híbrido de Timor ($t = 3.25$; $P < 0.05$). There was no difference within the triploid and resistant arabica groups (Figure 1).

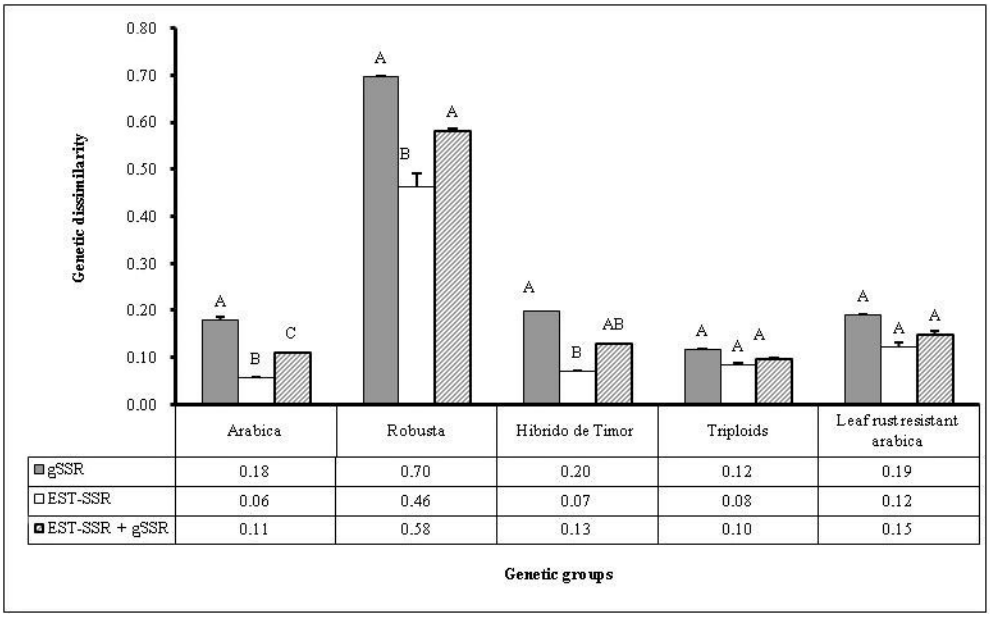

Figure 1. Mean genetic distance coefficient within each genetic group (vertical bars above columns are the standard deviation; different letters compare distance means within groups by the *t*-test at 5 %).

Figure 2 shows three dendrograms based on: 17 EST-SSR markers (87 alleles) (Figure 2A), 18 gSSR markers (96 alleles) (Figure 2B), and both EST-SSR and gSSR markers (35 markers, 183 alleles) (Figure 2C). All three dendrograms clearly demonstrated that racemosa and robusta accessions were genetically distant from arabica-related accessions (arabica, resistant arabica, Híbrido de Timor, and triploids). Racemosa showed a smaller genetic distance from triploids when EST-SSR markers were considered, compared to gSSR markers, which is consistent with the triploids' origin (*C. arabica* x *C. racemosa*). The inclusion of gSSR markers allowed us to differentiate all arabica-related accessions (arabica, resistant arabica, Híbrido de Timor, and triploids), which was not possible with EST-SSR markers. The dendrogram based on both EST-SSR and gSSR markers (35 markers, 183 alleles) was more consistent with the genetic origin of the accessions.

Correlation of genetic distances

Correlation and stress were calculated between each Jaccard's genetic distance matrix of M_i sampled alleles ($5 \leq M_i \leq 183$) and the original genetic distance matrix with 183 alleles (Figure 3). A correlation of 0.90 ($t = 109.11$; $P < 0.001$) corresponded to 120 alleles, which resulted in a similar dendrogram (data not shown), compared with the dendrogram obtained with all 183 alleles. The same order of accessions within population groups was also observed with 120 alleles. With less than 120 alleles, the order and number of groups were altered in dendrograms (data not shown). As a result of these comparisons, the minimum number of 120 alleles was estimated to obtain a similar dendrogram for this case study.

Discriminant analysis

Discriminant analysis, based on the presence and absence of alleles, revealed the best SSR loci for differentiating the genetic groups (Table 5). By this analysis, 17 alleles from 10 gSSR loci and eight alleles from five EST-SSR loci were identified as the most important alleles to differentiate the groups. For some loci, discriminant analysis identified more than one important allele, for example SSRCa 087 and EST-SSR 007, with five and four alleles, respectively. Approximately 17.7% of the gSSR and 9.2% of the EST-SSR alleles were essential for differentiating between groups. gSSR markers were approximately 52.9% more efficient than EST-SSR for the discriminant analysis of groups.

DISCUSSION

The scope of a genetic diversity study and genotype differentiation are normally limited by the number and representativeness of the sampled accessions and by the genetic characteristics that are measured. The genetic diversity of an important coffee breeding program was successfully assessed here with gSSR and EST-SSR markers, and it was possible to differentiate all genotypes including the arabica-related accessions. For this, allele and polymorphism analyses, AMOVA, the Student *t*-test, Jaccard's dissimilarity coefficient, cluster analysis, correlation of genetic distances, and discriminant analysis were performed.

The mean number of alleles per primer was almost the same for EST-SSR and

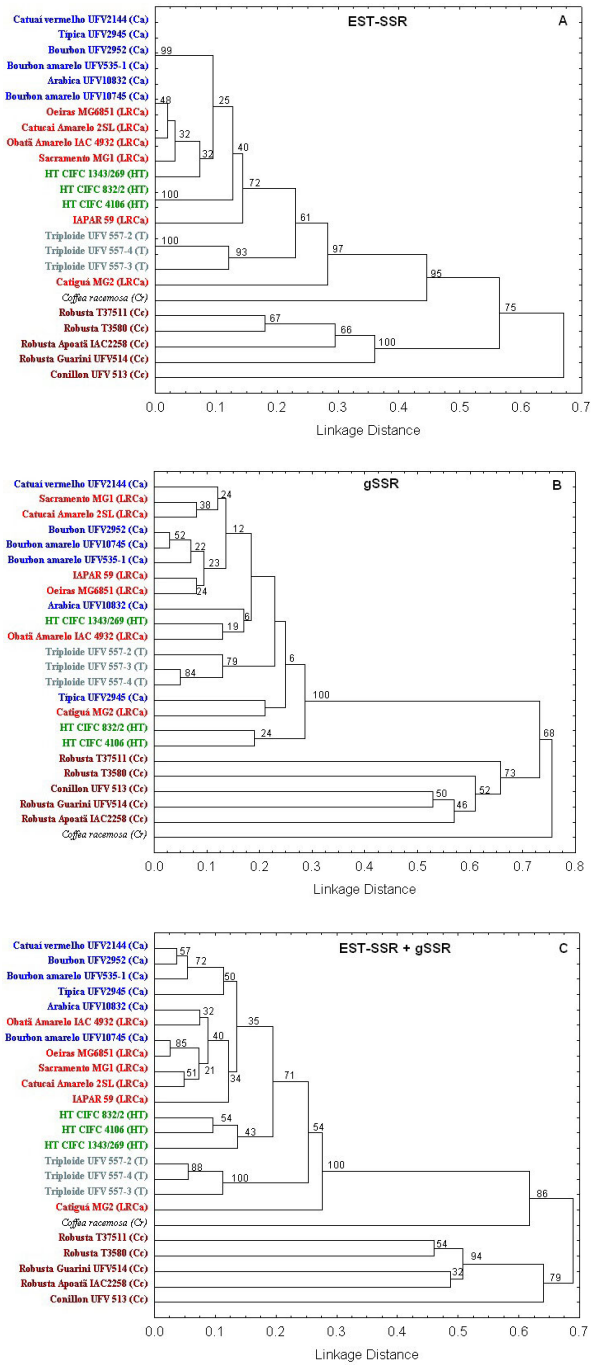

Figure 2. Dendrogram obtained using UPGMA based on genetic distances expressed as Jaccard complements, estimated for 24 coffee tree accessions. **A.** 17 EST-SSR (cophenetic correlation = 0.98). **B.** 18 gSSR markers (cophenetic correlation = 0.99). **C.** EST-SSR + gSSR (cophenetic correlation = 0.99).

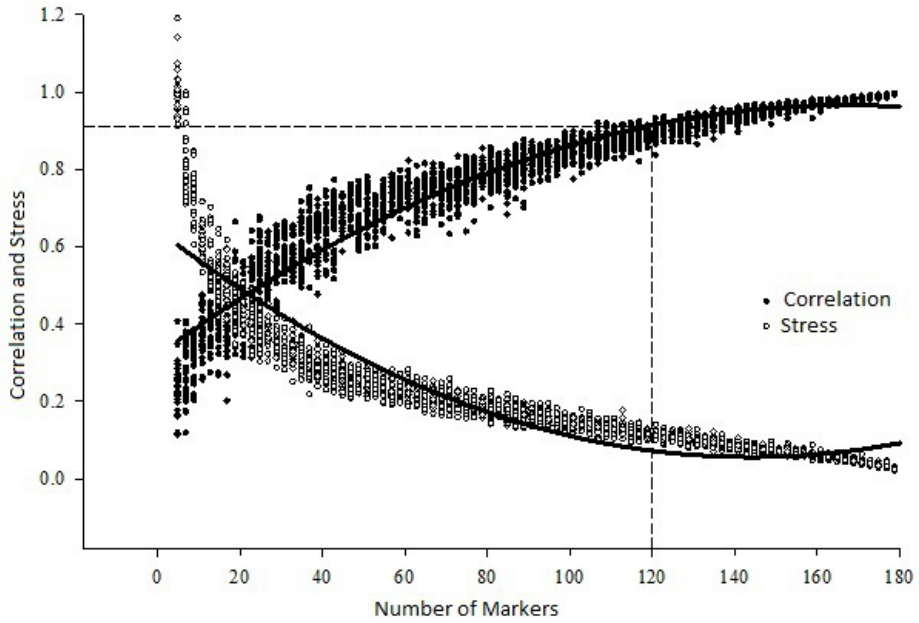

Figure 3. Optimal number of SSR alleles for assessing the genetic diversity among 24 *Coffea* accessions. The points are the projected correlations and estimated stress between the sampled alleles M_i ($5 \leq M_i \leq 183$) and the original matrix with M alleles (M = 183).

Table 5. Summary of the best differentiating loci obtained by stepwise discriminant analysis, for both classes of SSR markers.

Order	Locus	Allele	R^2	Prob >F
gSSR				
1	SSRCa 003	a2	1.00	<0.0001
2	SSRCa 003	a1	1.00	<0.0001
3	SSRCa 020	a2	1.00	<0.0001
4	SSRCa 083	a2	0.62	0.0002
5	SSRCa 087	a5	0.51	0.0024
6	SSRCa 088	a3	0.44	0.0096
7	SSRCa 087	a1	0.40	0.0089
8	SSRCa 016	a1	1.00	<0.0001
9	SSRCa 087	a7	0.20	0.0324
10	SSRCa 087	a4	0.33	0.0242
11	SSRCa 082	a2	0.30	0.0305
12	SSRCa 094	a4	0.40	0.0204
13	SSRCa 091	a9	0.40	0.0273
14	SSRCa 094	a2	0.39	0.0393
15	SSRCa 087	a3	0.73	0.0018
16	SSRCa 068	a2	0.40	0.0200
17	SSRCa 091	a8	1.00	<0.0001
EST-SSR				
1	EST-SSR 007	a5	1.00	<0.0001
2	EST-SSR 005	a2	1.00	<0.0001
3	EST-SSR 007	a2	1.00	<0.0001
4	EST-SSR 025	a1	0.67	<0.0001
5	EST-SSR 047	a6	1.00	<0.0001
6	EST-SSR 007	a8	0.50	0.0007
7	EST-SSR 007	a9	0.30	0.0450
8	EST-SSR 010	a3	0.44	0.0035

gSSR (5.1 with 17 EST-SSR primers and 5.3 with 18 gSSR primers). However, different mean num bers of alleles in *Coffea* spp were previously reported: 10.5 alleles per primer with 25 EST-SSR by Poncet et al. (2006), and 2.87 alleles with 23 gSSR by Maluf et al. (2005). In other crops, such as rice (Cho et al., 2000), beans (Blair et al., 2006), and sugar cane (Cordeiro et al., 2001), gSSR showed a higher mean number of alleles than did the EST-SSR markers. The number of SSR alleles assessed in a population depends highly on the size and genetic constitution of the population, and the number of primers that are assayed. Besides, the variation in the number of alleles in an SSR locus results from the differences in mutation rates and in the selection pressure on each locus (Metais et al., 2002); which are influenced by the structure, type and length of the tandem repeats.

In this study, we found some exclusive and null alleles. Although the exclusive alleles may only be inferred for the sampled gene pool, the results presented here may be useful for differentiating accessions and fingerprinting varieties of our coffee breeding program, as suggested in other species (Donini et al., 1998, Eujayl et al., 2001, He et al., 2005). Null alleles were observed here with gSSR primers, but not with EST-SSR primers. Null alleles of SSR markers were previously reported in coffee (Poncet et al., 2004, 2006), wheat (Gupta and Varshney, 2000; Leigh et al., 2003), and rice (Cho et al., 2000). They possibly result from mutations that occurred in the genome region for which the primers are designed, or genomic rearrangements, including insertion-deletion events within the region to be amplified (Leigh et al., 2003). The eventual negative selection of mutations at the functional regions of the genome may explain the higher chance to find null alleles with gSSR than with EST-SSR. The observation of null alleles, however, is also dependent on the sampling of the accessions.

The gSSR markers detected higher percentages of polymorphism among and within the genetic groups, compared to EST-SSR markers. Higher percentages of polymorphism with gSSR markers were also reported for grape (Scott et al., 2000), rice (Cho et al., 2000), wheat (Eujayl et al., 2001), sugar cane (Cordeiro et al., 2001), and barley (Chabane et al., 2006). It may occur due to a high conservation of DNA sequences in transcribed regions of the genome (Varshney et al., 2005), such as the ESTs. The higher percentage of gSSR (56%) with dinucleotide repeats, compared to EST-SSR (47%), may also have influenced the results. It was reported that SSR markers with dinucleotide repeats were more polymorphic than with trinucleotides in wheat (Gadaleta et al., 2007) and barley (Baek et al., 2003). The polymorphisms among and within the genetic groups were clearly influenced by the genetic origin and reproductive system of the chosen accessions. For instance, a high polymorphism was found here within the robusta group, which was also observed in previous studies (Baruah et al., 2003; Poncet et al., 2004, 2006; Aggarwal et al., 2007; Silvestrini et al., 2007; Hendre et al., 2008), and it may be mainly attributed to allogamy and the diverse parentage background.

According to the percentage of polymorphism and AMOVA results, the genetic diversity among groups was higher than within groups, as expected. Assessing the genetic diversity within groups, however, is particularly important for breeding programs, because even a small diversity has been shown to be useful to develop new improved varieties. We found here that gSSR markers were more efficient than EST-SSR markers to assess the genetic diversity within groups, based on the percentage of polymorphism and AMOVA.

The cluster analysis dendrograms based on EST-SSR and gSSR markers were con-

sistent with the genetic background of the accessions. The gSSR markers were more efficient than EST-SSR markers in differentiating all closely related arabica accessions. The advantage of a cluster analysis dendrogram is the easy identification of accessions for breeding purposes. For instance, based on the dendrogram results, Catiguá MG2, one of our new productive varieties with excellent cup quality and leaf rust resistance, was chosen to participate in new hybridizations of our breeding program, because it may be an important source of favorable new genetic variability.

The correlation of genetic distances allows one to observe the effect of using a smaller number of alleles instead the total number available. The information of the minimum number of alleles (120 in our study) will be used to continue the screening of the remaining accessions of our genebank, in further genetic distance studies. The possibility of the elimination of a number of less informative alleles from the genetic diversity study was previously reported in wheat (Fahima et al., 1999) and cashew (Pessoni, 2007). In wheat, 15 of 48 RAPD loci were important for discriminating 11 populations. In cashew, 115 to 155 of 223 ISSR alleles were important for discriminating 136 populations. Our discriminant analysis results also showed that there are some alleles that are more informative than others for differentiating the genetic groups: 17 alleles from 10 gSSR loci and eight alleles from five EST-SSR loci were identified as the most important alleles. These loci are the obvious candidates to be first used in the next step study of our breeding program.

The gSSR and EST-SSR markers were successfully used for genetic diversity evaluation of valuable accessions of a Brazilian coffee breeding program. The gSSR markers were more efficient in this evaluation, especially in differentiating *C. arabica* related accessions. Nevertheless, the combined use of gSSR and EST-SSR markers is recommended because they may provide complementary information. The selection of a more informative group of gSSR and EST-SSR markers was made for further studies.

ACKNOWLEDGMENTS

Research supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (CBP & D), and Financiadora de Estudos e Projetos (FINEP).

REFERENCES

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, et al. (2007). Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor. Appl. Genet.* 114: 359-372.

Baek HJ, Beharav A and Nevo E (2003). Ecological-genomic diversity of microsatellites in wild barley, *Hordeum spontaneum*, populations in Jordan. *Theor. Appl. Genet.* 106: 397-410.

Baruah A, Naik V, Hendre PS and Rajkumar R (2003). Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. *Mol. Ecol. Notes* 3: 647-650.

Beharav A and Nevo E (2003). Predictive validity of discriminant analysis for genetic data. *Genetica* 119: 259-267.

Blair MW, Giraldo MC, Buendia HF, Tovar E, et al. (2006). Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 113: 100-109.

Brito GG, Caixeta ET, Gallina AP, Maciel-Zambolim E, et al. (2010). Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. *Euphytica* 173: 255-264.

Chabane K, Ablett GA, Cordeiro GM, Valkoun J, et al. (2006). EST *versus* genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genet. Resour. Crop Evol.* 52: 903-909.

Cho YG, Ishii T, Temnykh S, Chen X, et al. (2000). Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 100: 713-722.

Cordeiro GM, Casu R, McIntyre CL, Manners JM, et al. (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci.* 160: 1115-1123.

Cruz CD (2007). Genes versão 2007.0.0, Programa GENES versão Windows. Available at [www.ufv.br/dbg/genes/genes.htm]. Accessed October 21, 2010.

Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET, et al. (2005). Analysis of AFLP markers associated to the *Mex*-1 resistance locus in Icatu progenies. *Crop Breed. Appl. Biotechnol.* 5: 387-393.

Donini P, Stephenson P, Bryan GJ and Koebner RMD (1998). The potential of microsatellite for high throughput genetic diversity assessment in wheat and barley. *Genet. Resour. Crop Evol.* 45: 415-421.

Eujayl I, Sorrells M, Baum M, Wolters P, et al. (2001). Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. *Euphytica* 119: 39-43.

Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Fahima T, Sun GL, Beharav A, Krugman T, et al. (1999). RAPD polymorphism of wild emmer wheat populations, *Triticum dicoccoides*, in Israel. *Theor. Appl. Genet.* 98: 434-447.

Gadaleta A, Mangini G, Mulè G and Blanco A (2007). Characterization of dinucleotide and trinucleotide EST-derived microsatellites in the wheat genome. *Euphytica* 153: 73-85.

Gupta PK and Varshney RK (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113: 163-185.

Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, et al. (2000). Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. *Genome Res.* 10: 62-71.

Hamilton MB, Pincus EL, Di FA and Fleischer RC (1999). Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* 27: 500-507.

He G, Meng R, Gao H, Guo B, et al. (2005). Simple sequence repeat markers for botanical varieties of cultivated peanut (*Arachis hypogaea* L.). *Euphytica* 142: 131-136.

Hendre PS, Phanindranath R, Annapurna V, Lalremruata A, et al. (2008). Development of new genomic microsatellite markers from robusta coffee (*Coffea canephora* Pierre ex A. Froehner) showing broad cross-species transferability and utility in genetic studies. *BMC Plant Biol.* 8: 51.

Jaccard P (1908). Nouvelles Recherches sur la Distribution Florale. *Bull. Soc. Vdoise Sci. Nat.* 44: 223-270.

Leigh F, Lea V, Law J and Wolteret P (2003). Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica* 133: 359-366.

Maluf MP, Silvestrini M, Ruggiero LMC and Guerreiro Filho O (2005). Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Sci. Agric.* 62: 366-373.

Metais I, Hamon B, Jalouzot R and Peltier D (2002). Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. *Theor. Appl. Genet.* 104: 1346-1352.

Missio RF, Caixeta ET, Zambolim EM and Zambolim L (2009a). Development and validation of SSR markers for *Coffea arabica* L. *Crop Breed. Appl. Biotechnol.* 9: 361-371.

Missio RF, Caixeta ET, Zambolim EM and Pena GF (2009b). Assessment of EST-SSR markers for genetic analysis on coffee. *Bragantia* 68: 576-581.

Missio RF, Caixeta ET, Zambolim EM and Zambolim L (2010). Polymorphic information content of SSR markers for *Coffea* spp. *Crop Breed. Appl. Biotechnol.* 10: 89-94.

Pessoni LA (2007). Estratégias de Análise da Diversidade em Germoplasmas de Cajueiro (*Anacardium* spp. L.). Doctoral thesis, Universidade Federal de Viçosa, Viçosa.

Poncet V, Hamon P, Minier J, Carasco C, et al. (2004). SSR cross-amplification and variation within coffee trees (*Coffea* spp.). *Genome* 47: 1071-1081.

Poncet V, Rondeau M, Tranchant C, Cayrel A, et al. (2006). SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Mol. Genet. Genomics* 276: 436-449.

SAS Institute (1996). SAS User Guide: Statistics, Version 609. SAS Institute Inc., Cary.

Scott KD, Eggler P, Seaton G and Rossetto M (2000). Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* 100: 723-726.

Silvestrini M, Junqueira MG, Favarin AC and Guerreiro-Filho O (2007). Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. *Genet. Resour. Crop Evol.* 54:

1367-1379.
Tanksley SD and McCouch SR (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277: 1063-1066.
Van de Peer Y and De Wachter R (1994). TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comp. Appl. Biosci.* 10: 569-570.
Varshney RK, Graner A and Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23: 48-55.