Crop Breeding and Applied Biotechnology 9: 361-371, 2009
Brazilian Society of Plant Breeding. Printed in Brazil

Development and validation of SSR markers for *Coffea arabica* L.

Robson Fernando Missio[1], Eveline Teixeira Caixeta[1,2*], Eunize Maciel Zambolim[1], Laércio Zambolim[1] and Ney Sussumu Sakiyama[1]

Received 20 August 2009

Accepted 29 October 2009

ABSTRACT - *With the objective of developing new SSR markers for* Coffea arabica*, two enriched genomic libraries with probes* $(GT)_{15}$ *and* $(AGG)_{10}$ *were constructed. A total of 835 clones were sequenced and 756 presented good quality sequences. Redundant sequences were observed for 113 clones (14.94%). SSRs were found in 287 clones (38%). An estimated size of 417.5Kb of the* C. arabica *genome was sampled, with an average of one SSR per 1.46Kb. Dinucleotide repeats were more frequent than trinucleotides. Four repeat sequences,* $(AG/CT)_n$, $(AC/GT)_n$, $(AAG/CTT)_n$, *and* $(AGG/CCT)_n$ *represented 61.1% of the total observed. A total of 96 SSR primers were designed and tested by PCR for two C. arabica genotypes. Ninety new SSR markers were validated for further genetic studies of* C. arabica.

Key words: SSR marker, enriched genomic library, coffee, molecular marker.

INTRODUCTION

Microsatellites or *Simple Sequence Repeats* (SSRs) correspond to DNA sequences in that a single pair or a small number of base pairs (1-6) are repeated in tandem (Litt and Luty 1989). The SSRs are present in the coding and non-coding regions of the genome of the eukaryots and prokaryots and are characterized by the high level of polymorphism (Gur-Arie et al. 2000). The SSR markers became one of the main molecular markers for genetic studies, especially as a result of the high level of polymorphism, multialelism and high reproducibility (Zane et al. 2002).

The major disadvantages of the SSR markers are the high cost as well as the time and effort necessary for the development of the primers (Zane et al. 2002). Currently, there are different strategies for the development of SSR primers, but the enriched genomic library method with selective hybridization stands out (Zane et al. 2002). Two different strategies are frequently used in this method: SSR probes attached to nylon membranes (Armour et al. 1994), and biotinylated SSR probes (Hamilton et al. 1999). The selective hybridization method allows the selection of a high quantity of DNA fragments containing SSR regions. With this method the sampled DNA fragments are hybridized with complementary probes, thus increasing the number of clones containing SSR sequences for the design of primers.

The development of SSR primers using enriched genomic library has been widely used for many species of plants such as eucalyptus (Brondani et al. 1998), piqui (Collevatti et al. 1999), pepper (Buso et al. 2000), sugarcane (Cordeiro et al. 2000), bean (Benchimol et al. 2007), rice (Brondani et al. 2001), avocado (Ashworth et al. 2004), lychee (Viruel and Hormaza 2004), melon (Ritschel et al. 2004), hop (Stajner et al. 2005), mulberry (Zhao et al. 2005) and wheat (Song et al. 2005).

[1] Universidade Federal de Viçosa (UFV), Laboratório de Biotecnologia do Cafeeiro (Biocafé), BIOAGRO, 36570-000, Viçosa, MG, Brazil. *E-mail: eveline.caixeta@embrapa.br
[2] Empresa Brasileira de Pesquisa Agropecuária, Embrapa Café, 36570-000, Viçosa, MG, Brazil

For *Coffea* species, a small number of SSR primers have been developed and are available for genetic studies, when compared to cultures such as maize, soybean, rice, wheat and barley (http://www.gramene.org/). With soybean, Song et al. (2004) developed an integrated map containing 1015 SSR markers. With maize, 2095 of these markers have already been mapped (http://www.maizegdb.org/ssr.php). To date, there are 263 SSR markers available for coffee, of which 165 (63%) were obtained from *C. canephora* and 98 (37%) from *C. arabica* (Combes et al. 2000, Rovelli et al. 2000, Baruah et al. 2003, Moncada and McCouch 2004, Leroy et al. 2005, Bhat et al. 2005, Poncet et al. 2006, Poncet et al. 2007, Aggarwal et al. 2007, Tesfaye et al. 2007, Hendre et al. 2008, Cristancho e Gaitán 2008). Not many have been mapped and 186 out of 251 were derived from enriched genomic libraries.

New efforts for the development of SSR genomic markers are important in order to increase the availability of this class of markers for genetic studies of the *Coffea* species. The objective of this work was to develop and validate new coffee SSR markers and make them available to the scientific community.

MATERIAL AND METHODS

Construction of the enriched genomic libraries

The construction of the enriched genomic libraries with SSR probes $(GT)_{15}$ and $(AGG)_{10}$ was carried out following the protocol described by Hamilton et al. (1999) with modifications. The *C. arabica* genotype Bourbon Amarelo, access number UFV 570 from Universidade Federal de Viçosa germplasm bank, was used to obtain the libraries.

The genomic DNA (50µg) of the Bourbon Amarelo UFV 570 genotype was digested into fragments of approximately 200-1000bp using the restriction enzymes *Eco*RI, *Nhe*I, *Hae*III and *Rsa*I (New England BioLabs). The blunt-ended fragments were obtained by the *Mung Bean Nuclease* (MBN) enzyme treatment, and then dephosphorylated with Calf intestinal phosphatase (CIP), and ligated to the double-stranded SNX adaptors. Enrichment was carried out by hybridization of the DNA with two biotinylated SSR probes $(GT)_{15}$ and $(AGG)_{10}$. After washing, the fragments were amplified by PCR (*Polymerase Chain Reaction*) using the SNX_F adapter as a primer. The enriched fragments were then digested with *Nhe*I and ligated into the plasmid *pBluescript* SK+ (Stratagene), previously digested with *Xba*I (New England Biolabs). Competent *Escherichia coli* DH5á cells were transformed with the recombinant plasmids by the thermal shock procedure. PCR amplifications were carried out for the white colonies using the T3 and T7 primers (Invitrogen). The amplification products were separated by electrophoresis in 1.2% agarose gels. The colonies containing transformants with insertions greater than 400bp were selected and cultivated in LB liquid medium containing ampicillin (100µg mL^{-1}) to compose the library.

Clone sequencing and SSR analysis

The selected clones were sequenced in an automatic sequencer (MegaBACE 1000, GE). The analysis of the DNA fragment sequences was performed with CodonCode Aligner 1.6.3 (CodonCode Corporation) and SSRIT (http://www.gramene.org) programs. The CodonCode Aligner program was used to discriminate the regions of genome fragments of *C. arabica*, eliminate the plasmid sequences and verify the presence of redundant sequences. The SSRIT program was used to identify SSR repeats in the sequences. The criteria used for the SSR definition were: a minimum of four repeats of dinucleotides or three repeats of tri-, tetra-, penta-, or hexanucleotides. For imperfect repeats the maximum difference of 10bp between two motifs was adopted.

Design of SSRCa Primers

Specific flanking primers for each SSR locus was designed with the Primer3 program (Rozen e Skaletsky 2000) using the following criteria: 1) size of the primers from 18 to 24bp; 2) Tm of 55 to 60 ºC; 3) salt concentration of 50mM; 4) amplification product of 100 to 600bp; 5) GC percentage of 40 to 60%. The primers were named SSRCa followed by an order number.

Evaluation of the SSRCa primers

Two *C. arabica* genotypes, the accesses Híbrido de Timor UFV 445-46 and Catuaí UFV 2143-235 from UFV germplasm bank, were PCR tested with SSRCa primers. DNA from young leaves was extracted according to the protocol described by Diniz et al. (2005). Each DNA sample was prepared for PCR according to Missio et al. (2009) in a total volume of 20µL containing 50ηg of the genomic DNA, 0.6 units of

Taq DNA polymerase and 1x buffer (Promega), 1mM of $MgCl_2$, 150µM of each dNTP and 0,1µM of each primer. The DNA amplification was carried out in a PTC 200 (MJ Research) thermocycler using the *touchdown*-PCR procedure which involved an initial denaturation at 94 °C/2 minutes followed by 13 cycles at 94 °C/30 seconds, 67 °C to 55 °C/30 seconds, reducing by 1 °C for each cycle and 72 °C/30 seconds. The 13 cycles were followed by another 30 cycles at 94 °C/30 seconds, 55 °C/30 seconds and 72 °C/30 seconds and final extension at 72 °C/8 minutes. The electrophoretic pattern was visualized in 6% denaturing polyacrylamide gel and silver stained in accordance with the protocol described by Creste et al. (2001).

RESULTS AND DISCUSSION

Enriched genomic libraries of *C. arabica*

Two enriched genomic libraries of *C. arabica* were obtained with a total of 835 clones, which were sequenced for analysis (Table 1). The insert sequencing revealed 756 good clones, while 64 clones with sequencing problems and 15 with inserts smaller than 100bp were discarded. All clones presented sequenced inserts and none presented sequenced *Escherichia Coli* DNA. Redundant sequences were observed for 113 clones. SSRs were found in 287 (38%) out of 756 clones. The average size of the sequenced clones was 500bp. Therefore, the estimated size of 417.5Kb of the *C. arabica* genome was sampled, with an average of one SSR per 1.46Kb (417.5 Kb / 287 SSRs).

SSR markers have been developed using different methods for a wide range of species. The efficiency of each method is indicated by the proportion of clones containing SSRs in relation to the total number of clones examined (Zane et al. 2002). The efficiency of 38% in this study of *C. arabica* was high, compared to the previous studies with mulberry, 26% (Zhao et al. 2005), *Brassica*, 18.5% (Cui et al. 2008), piqui (*Caryocar brasiliense*), 14.4% (Collevatti et al. 1999), Jute (*Corchorus capsularis*), 34.5% (Mir et al. 2008) and orange tree (*Citrus*), 25% (Novelli et al. 2006). The additional evidence that the methodology was well executed was demonstrated by the fact that all clones presented sequenced inserts and none presented sequenced *Escherichia Coli* DNA contamination. The redundancy of sequences was expected, since a PCR amplification of the enriched DNA fragments were performed before the random fragment cloning, therefore increasing the possibility of cloning more than one copy of the same DNA fragment.

The 287 SSRs were classified according to their repeat compositions (Table 2). Dinucleotides represented 51% of the total SSRs, trinucleotides 33%, and tetranucleotides 8%. All mono-, penta-, and hexanucleotides together represented only 8%. (AG)n

Table 1. Analysis of the sequenced clones from enriched genomic libraries in *C. arabica*

Results of the sequenced clones	Number	%
Total of sequenced clones	835	100.0
Sequencing problems	64	7.6
Clones without inserts	0	0.0
DNA contamination from *Escherichia Coli*	0	0.0
Sequences with insufficient size for primer design	15	1.8
Total of sequences available for SSR primer design	756	90.5
Total of sequences available for SSR primer design	756	100.0
Sequences from library $(GT)_{15}$	530	70.1
Sequences from library $(AGG)_{10}$	226	29.9
Total of SSRs found	287	38.0
Redundant sequences	113	14.9
Sequences with SSR	218	28.8
Sequences with more than one SSR	69	9.1
Sequences enriched with $(GT)_{15}$	37	7.0
Sequences enriched with $(AGG)_{10}$	16	7.1
Primers designed	96	100.0
Primers validated	90	93.8

(32.8%), $(AC)_n$ (12.9%), $(AAG)_n$ (9.8%) and (AGG)n (5.6%) were more abundant. Single repeat types of SSR represented 59.2% and compound repeats represented 40.8%. Perfect types of repeats corresponded to 98.2% of the single repeats and 55.6% of the compound repeats. Considering that an estimated size of 417.5Kb of the *C. arabica* genome was sampled, the three major classes of SSR presented an average of: one dinucleotide SSR per 2.9Kb, one trinucleotide SSR per 4.4Kb, and one tetranucleotide SSR per 18.2Kb. The

Table 2. Number and frequency of SSR from enriched genomic libraries of *C. arabica*, according to their classification and the number of repeats

Classes	Number of repeats (n)									Total	%
	3	4	5	6	7	8	9	10	>10		
Mononucleotide										3	1.0
$(A/T)_n$	-	-	-	-	-	-	-	1	2	3	1.0
Dinucleotide										146	51.0
$(AC/GT)_n$	3	9	2	3	-	3	4	-	13	37	12.9
$(AG/CT)_n$	1	25	13	21	5	2	-	2	25	94	32.8
$(AT/AT)_n$	-	5	8	1	-	-	-	-	-	14	4.9
$(CG/CG)_n$	-	1	-	-	-	-	-	-	-	1	0.4
Trinucleotide										95	33.0
$(AAC/GTT)_n$	11	2	-	-	2	-	-	-	-	15	5.2
$(AAG/CTT)_n$	20	7	1	-	-	-	-	-	-	28	9.8
$(AAT/ATT)_n$	6	3	-	1	-	-	-	-	-	10	3.5
$(ACC/GGT)_n$	7	5	-	-	1	-	1	-	-	14	4.9
$(ACG/CGT)_n$	1	-	-	-	1	-	-	-	-	2	0.7
$(ACT/AGT)_n$	3	-	-	-	-	-	-	-	-	3	1.0
$(AGC/GCT)_n$	1	-	-	-	-	-	-	-	-	1	0.4
$(AGG/CCT)_n$	11	2	1	1	1	-	-	-	-	16	5.6
$(AGT/ACT)_n$	4	-	-	-	-	-	-	-	-	4	1.4
$(CCG/CGG)_n$	1	-	-	1	-	-	-	-	-	2	0.7
Tetranucleotide										23	8.0
$(AAAC/GTTT)_n$	2	-	-	-	-	-	-	-	-	2	0.7
$(AAAG/CTTT)_n$	9	-	1	-	-	-	-	-	-	10	3.5
$(AAAT/ATTT)_n$	4	-	-	-	-	-	-	-	-	4	1.4
$(AACC/GGTT)_n$	1	-	-	-	-	-	-	-	-	1	0.4
$(AACT/AGTT)_n$	1	-	-	-	-	-	-	-	-	1	0.4
$(AAGG/CCTT)_n$	2	-	-	-	-	-	-	-	-	2	0.7
$(AAGT/ACTT)_n$	1	-	-	-	-	-	-	-	-	1	0.4
$(AGGG/CCCT)_n$	-	1	-	-	-	-	-	-	-	1	0.4
$(AGGT/ACCT)_n$	1	-	-	-	-	-	-	-	-	1	0.4
$(Pentanucleotide)_n$	9	1	-	-	-	-	-	-	-	10	3.5
$(Hexanucleotide)_n$	10	-	-	-	-	-	-	-	-	10	3.5
Total	109	61	26	28	10	5	5	3	40	287	

Type of repetition	Number
Single	170
Perfect	167
Imperfect	3
Compound	117
Perfect	65
Imperfect	52

frequency of individual SSR showed strong variation also within each class of dinucleotide, trinucleotide, and tetranucleotide SSR (Figure 1).

It was previously reported that the frequency, distribution and abundance of SSRs can vary strongly among different organisms, mainly due to the different search criteria, origin of the sequences, and the size of the sampled genome (Varshney et al. 2005). Dinucleotide repeats were also the most frequent class of SSR derived from genomic DNA of quinoa (Jarvis et al. 2008), peanut (Cuc et al. 2008), melon (Ritschel et al. 2004), orange tree (Novelli et al. 2006), bean (Benchimol et al. 2007) and sugarcane (Cordeiro et al. 2000). Analyzing ESTs for the *Coffea* species Aggarwal et al. (2007) found 46% of dinucleotides and 26% of trinucleotides. However, the trinucleotide repeats were the most frequent SSR class observed in ESTs of *C. canephora* (Poncet et al. 2006), soybean and rice (Cardle et al. 2000, Gao et al. 2003), maize, tomato and cotton (Cardle et al. 2000). This higher frequency of trinucleotides class could be attributed to the lower mutation events in coding regions of the genome represented by the ESTs (Metzgar et al. 2000), while the high pressure of selection may reduce this class in the noncoding regions (Katti et al. 2001).

In this study of *C. arabica* enriched genomic libraries, the highest frequencies of SSRs were observed for $(AG/CT)_n$, $(AC/GT)_n$, $(AAG/CTT)_n$, and $(AGG/CCT)_n$. Similar results were reported for *C. canephora,* where $(AG/CT)_n$ and $(AC/GT)_n$ were the most frequent repeats (Hendre et al. 2008). Analyzing ESTs for the *Coffea* species Aggarwal et al. (2007) observed that $(AG/CT)_n$ was the most frequent dinucleotide and that $(AAG/CTT)_n$ were the most abundant trinucleotide SSR. Poncet et al. (2006) reported that ESTs of *C. canephora* $(GA)_n$ was the most frequent dinucleotide and that $(AGG/CCT)_n$ were the most abundant trinucleotide SSR.

Validation of the SSR markers for *C. arabica*

A total of 96 SSRCa primer pairs were designed, synthesized and PCR tested in *C. arabica* genotypes (Table 3). Ninety SSRCa primers produced DNA amplification products and were, therefore, validated as useful SSR markers for genetic studies of *C. arabica.* Among these, 21 (23.3%) presented polymorphism between the 'Híbrido de Timor UFV 445-46' and 'Catuaí UFV 2143-235' accesses and 69 (76.7%) were non-polymorphic. The number of alleles, considering these two accessions varied from 1 to 4, with an average of 1.86 alleles per primer (Table 3).

The proportion of SSR primers that successfully amplify the tested DNA may be used to measure the rate of conversion of SSR primer into SSR markers (Hendre et al. 2008). The conversion rate may vary among species. Garner (2002) observed that the percentage of SSR primers that do not produce PCR products is high and positively correlated to the size of the genomes. We found, however, that the conversion rate in *C. arabica* (93.7%) was higher than in *C. canephora* (75.8%, Hendre et al. 2008), nevertheless the double of the size of *C. arabica* genome, respectively, $2{,}56x10^9$bp and $1{,}38x10^9$bp (Clarindo and Carvalho 2009).

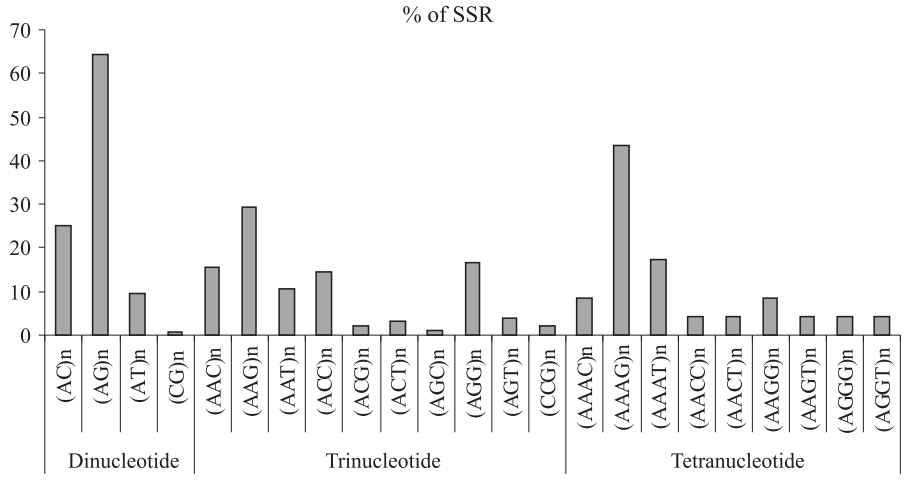

Figure 1. Frequency of individual SSR within the dinucleotide, trinucleotide, and tetranucleotide SSR classes, in two enriched genomic libraries of *C. arabica*

Table 3. Description of the 96 pairs of *primer* for SSR *loci* and the number of alleles obtained from two genotypes of *C. arabica*

Primer	Repeats	Forward primer (5'–3')	Reverse primer (5'–3')	Tm ºC	Exp. Size of frag.	Number of alleles
SSRCa 001	$(CCCTTT)_3$.....$(TC)_3$ $C(CTT)_3$	CCCACTACTCCATTCCATTC	AGCAGATTCCATCCTTATCCT	57	173	2
SSRCa 002*	$(TTCC)_3$.....$(GT)_{17}$	CTGTCCCACCAACCAAAA	CTTCAACCCCCAACACAC	57	258	2
SSRCa 003	$(GT)_{12}$	ATGATTCGTAGGTGGAGTGG	CTAAGCCGCAAATGACAGA	57	196	1
SSRCa 004	$(CT)_8CG(CT)_4$	CCATGAGCACTTGTCCATAAA	ATCAAAGAACAAACCCGACA	58	287	1
SSRCa 005	$(CT)_5$	TGTCACTTCCTTGTTGGATT	GCTTGATTGAGATGATTTGC	55	201	2
SSRCa 006	$(CT)_6$	CTTGCTCAGTGAACCATCC	TGCCTCTTATGCCACTACTAAA	56	209	4
SSRCa 007	$(GGA)_3N_3(AT)_2$	GTTCTTTCATTCCAGGTAAAGC	TAGAAGGAATCGGTGGAGAA	57	178	2
SSRCa 008	$(AG)_6$	TTACCCACTTTTTCCACCTC	TTTGGCTTCAATCTTGCTC	56	373	2
SSRCa 009	$(TTTA)_3$	CAGTTTGGAATGCTTGAGTG	CCGGAACTTAACCTTATTGG	56	352	2
SSRCa 010	$(CT)_6$	GTTGATTGGTGGAGTGATTG	AAGCATCAAGTAAGGGAGGA	56	105	2
SSRCa 011	$(CT)_6$	ATCCAACCAACCATTGAAAC	CATCCACTTTTTCCACCTTC	57	347	2
SSRCa 012*	$(CT)_4N_6(CT)_4$ ....$(TTTTC)_3$....$(AAT)_4$	TCTCCTCTATTCGCTGTTCTC	TCTGTGCTCGTTTTTTTCAC	56	595	1
SSRCa 013	$(AG)_6$	TCAAAAACAACCACACCATC	CCATTTCACTCAATCTTCCA	56	317	2
SSRCa 014	$(TA)_5$	ATTCCTCTTTCTCCCACACA	AGCGGAAAACATCCAAAAC	57	214	1
SSRCa 015	$(AT)_5$	TCGCAATAACCAATCACAAG	AGCTATTGACCCCACTGAAA	57	273	1
SSRCa 016	$(GAA)_3$….$(GGAAAG)_3$	AGCAGATTCCATCCTTATCCT	CCACTAATCCATTCCATTCC	56	172	3
SSRCa 017	$(ATTTT)_3$	TATGATTGGTTGCTTGGATG	ATCCTACAAGGCGGTGTG	57	205	2
SSRCa 018*	$(GT)_{18}(GA)_{10}$	GTCTCGTTTCACGCTCTCTC	ATTTTTGGCACGGTATGTTC	57	115	3
SSRCa 019	$(GA)_{11}$	GGGTTAGATAGAGCAAGAATGA	CTGTGAAGGTGTGGAGTTTT	55	329	2
SSRCa 020	$(AGA)G(AGA)_3$ …$(TG)_4$…$(ATT)_6$	GGTAGGCGAAGGACAGATAA	TGGGGCAGAGTGAAGATAAG	57	264	2
SSRCa 021*	$(GGA)_3N_4(AAG)_2$	GCTGAGAGTTTTGAGGGAAA	CCGACGTAGTTGATGATTGA	57	232	4
SSRCa 022	$(GA)_5$…$(AAT)_3$	GGGAGCCATTCTGTGGA	CCCCATCTGGAAACCAA	57	445	2
SSRCa 023	$(AATG)_3$	GACCCTTGCCTTTTGTTG	GCCATTCATCCATTCATTC	56	259	2
SSRCa 024	$(AG)_3(CT)_3$	CCACTTACCGCTCTACCACT	CTTGGCTTGTCTCAGTCCTT	57	299	2
SSRCa 025	$(TAA)_2(TCT)_3$	CTGCAACTTGTGAAATGGAC	ATACGGAGGATGAAGAAGCA	56	176	1
SSRCa 026	$(T)_{16}N_{12}(TC)_7$….$(CAC)_4$	GAATCTGGTGGGCTTTGA	AAGGAGAGGGGAAGAAAATG	57	289	2
SSRCa 027*	$(AC)_6$	TGACCTCTCTTTTCATTTGG	CATCACTGCCTTTCTTTTTG	55	221	1
SSRCa 028	$(AGG)_3$…$(CT)_6$	GCTTGGTTGAGGTTGAAAAA	GCCGAAATACGAAAATGTGT	57	328	2
SSRCa 029*	$(ACAA)_3$…$(AAC)_3$...$(AAG)_3$	AATGCACGAGAACAAAGATG	TAGCACCAAAATCAATCCAC	56	344	1
SSRCa 030*	$(CCAT)_3$	GAGGAATCGAGAACCAGTGT	GTTTAGGGTTGCATTTTTCC	56	189	1
SSRCa 031	$(AG)_6$	TCGGACAGATTAGGGGTTC	TGGTGGAGTTTGTTTGAAGAG	57	350	1
SSRCa 032	$(GAA)G(GAA)_3(GCA)_2$	TCACACCATCCATACATTCC	ACATCCCACATTTCAGCAC	56	328	1
SSRCa 033*	$(AAT)_3$...$(GA)_4N_5(CA)_2$ $N_3(CG)_3N_3(GC)_3$	GTTTTTACGCGCACGATTA	TTCAAAAGTCAACTCATTCTCC	57	179	2
SSRCa 034	$(CT)_3N_3(CT)_5$…$(TC)_4(AC)_2$	TGGACAAGAAATTGAAGTGG	GGGTTTAAATTATCGGGTGT	55	257	2
SSRCa 035	$(TC)_5N_3(CT)_3$	GCTTAGTGGTTCCTTCTCCA	CAAGCCATTTCTTCCTTCTC	56	192	1
SSRCa 036	$(CA)_8$	ATGTTCGTGAAACACACGTC	GGTTTGCCTTCATCTTTGTT	56	128	1
SSRCa 037	$(CT)_6$	TTTTGGCTTCAATCTTGCTC	TTACCCACTTTTTCCACCTC	57	374	1
SSRCa 038*	$(AAGA)_3$….$(A)_{18}$	CGCAGGAATCATCAAGAA	ATAAGGAAGCAGGCTAATGG	56	312	1
SSRCa 039	$(AG)_6$	GAGTCAAAGCCCCTTATTACC	AGTTTGGTGGAGTTTGTTTG	56	263	1
SSRCa 040*	$(GAG)_3A(AG)_3$	AGGGATGTAGAACCAGCAAA	CCAATAGCTCACAACAAAGG	56	283	2
SSRCa 041	$(AC)_4N_9(TC)_3$	TCCCATGATTTCTCCACTTT	TTGAGCACTGGTATGGTTTG	57	195	1
SSRCa 042	$(AG)_6$…$(AGG)_3N_4(TTC)_2$	TTGTTCACCTTTCCCACCT	AATCAGCAAAACCAACCATC	57	352	1

To be continued …

Table 3. Cont.

Primer	Repeats	Forward primer (5'–3')	Reverse primer (5'–3')	Tm ºC	Exp. Size of frag.	Number of alleles
SSRCa 043*	$(CAA)_2(TCT)_2\ldots(GAA)_4$	GCCAAAATCCTTGTCTTCAC	GTCTTCCTGTTTGCTGGTTC	57	270	1
SSRCa 044	$(CT)_7$	CCCAATCTCACAAACTAACCA	CTTCATCACCTCAACCACAA	57	248	1
SSRCa 045*	$(TTTAC)_3\ldots.(AC)_3T(CA)_3$	GACTTGTTGCATTCCCCTA	GCGCATGTGAAGAGAAAGT	56	303	2
SSRCa 046	$(AAAT)_3$	ATGAAGAGGGGTTCCATCA	CATAGACTTTTCTTGCCTCCT	57	254	1
SSRCa 047	$(AT)_5\ldots(AAAGA)_3$	TAGAGGGTCTTTCGCAGTTT	AAAACCTTTCCGTCCACTT	56	466	1
SSRCa 048	$(AAAAT)_3$	TAGTCCTACAAGGCGGTGTG	TATGATTGGTTGCTTGGATG	57	207	1
SSRCa 049	$(CTT)_4\ldots(GT)_6$	TTGCATTCTACCCAACAAAG	CCCATCCACTTCAAAATACA	56	216	2
SSRCa 050	$(CA)_4(GA)_2(CA)_2(GAGG)_4\ldots(GA)_6$	AGCAATACATGCAGAGACCA	AATGTCGTTCCAACCAGAAG	57	133	2
SSRCa 051	$(ATC)_3(ACA)_3\ldots(CAC)_3(TGC)_2$	GAACAAGAACAGCAGACACAA	GAAAAGGTTGGTGGAAGAGA	56	383	2
SSRCa 052*	$(TTG)_7$	GATGGAAACCCAGAAAGTTG	TAGAAGGGCTTTGACTGGAC	57	129	3
SSRCa 053	$(ATA)_2(TCT)_2(CT)_2\ldots(AAGA)_3N_3(AAAG)_2(GGT)_4$	ACCACTTGACCACCATTTTT	TTTTCCTCCTTGATGCTCTC	56	259	1
SSRCa 054*	$(AAAG)_3$	CCGAACCCAACTAACATCTC	GCAGGTCTTCCATTGTCTGT	57	354	2
SSRCa 055	$(ATC)_3\ldots(AAGG)_2N_9(CT)_4N_6(CT)_3$	AAGGAAAACAACACCCAAGA	CGAGACAAGAGAGGGGAAA	57	294	4
SSRCa 056	$(GGT)_3(TTGG)_2\ldots.(GTTT)_2(GTT)_2(GAT)_2T(GAT)(GTAAAA)_2\ldots.(CGGAG)_3$	CGTATTGATGGCTGATGGT	AGGTCTGGTCCCTTTCTTCT	56	412	3
SSRCa 057	$(TTTTC)_2N_3(TG)_3\ldots(TTG)_3N_7(TTG)_2$	GCGGGCTAGATGAAAACTC	ATCTCACGCGACAGCAAC	57	169	2
SSRCa 058*	$(CATC)_2(AT)_3\ldots(CA)_5$	ATCATTACCTTGCCCAAATC	ACCCTTGACTGCCATAAATC	56	364	1
SSRCa 059	$(GA)_3(AAG)_2\ldots.(TCT)_3$	AGTCTCATGCACGGTTTTG	ACGTTTCATGCTTGTTTGAG	56	249	1
SSRCa 060	$(CT)_6$	AGTTTGGTGGAGTTTGTTTG	GAGTCAAAGCCCCTTATTACC	56	263	2
SSRCa 061	$(CCAA)_2(CT)_5$	GCAGGTGCAAGTGATAAAAG	CGTCTTGTGATGTGTTAGGG	56	242	4
SSRCa 062	$(CAA)_2G(AGAA)_2(AG)_4N_8(GA)_4$	AAGTTATTAGGGCAAGAGTGGA	AAGCTCCAAGACCAAAGATG	57	275	2
SSRCa 063	$(TG)_3A(GT)_3N(TG)_4$	CTCCGCTGATTTTGTCTTTT	ACCACTTTTTCCTCCCTCTC	57	222	-
SSRCa 064	$(TTCT)_3$	TGCAGTAAGTGAGACCAACC	TGGACTATCCCATACATAACCA	56	242	2
SSRCa 065	$(AG)_2(AAG)_3$	ATCTAACAAAATCCCCGTCA	ATCGGTCGCCCTTCTAAT	57	142	4
SSRCa 066*	$(GAA)_3(AG)_3$	GTGTGTCTTGAGGGCAGTTT	TCTTGATAGGTCTCCAGCATC	57	205	1
SSRCa 067	$(GAA)_4$	TCTCCTCCCATGACCTAAAA	CGAACAAAGCTGAAGTGAAA	56	107	2
SSRCa 068	$(AGG)_7\ldots.(GAA)_4$	ATGTTGTTGGAGGCATTTTC	AGGAGCAGTTGTTGTTTTCC	57	236	1
SSRCa 069*	$(AG)_4(AT)_2(GA)_2$	GATTGGGCATAAGTTTTCCA	TGAATCCTCCAAGAATAGCC	57	146	1
SSRCa 070	$(AG)_6$	AAGCATCAAGTAAGGGAGGA	GATTGGTGGAGTGATTGGA	56	102	2
SSRCa 071	$(AC)_5$	TTCCTCCTTCCTTTCTTCTTC	GGGAGTGTTTTGGTTCATTT	56	114	2
SSRCa 072	$(AC)_6N_2(TC)_5N_3(TC)_2(CT)_3$	GCCACATTTGTCGGATTTT	GCACAACAACCATCCATCTT	58	221	-
SSRCa 073	$(TC)_2(TTC)_3(CT)_3$	GCTGTGTGAGAAGCAAAGAA	CCAACAAACCCTAAAGAAGC	56	298	1
SSRCa 074	$(AAG)_3G(GA)_3\ldots.(GGAAAG)_3$	CCACTACTCCATTCCATTCC	AGCAGATTCCATCCTTATCCT	56	172	2
SSRCa 075	$(AC)_{15}$	TTCCCATGTCAAGCAAATC	CATCGCTAGTGCAGTGAAAG	57	106	-
SSRCa 076	$(TA)_5$	GTGTGTGCAAATGAATGAAG	AGGGAAATGAGCGAGTGT	55	297	2
SSRCa 077	$(TCA)_3\ldots.(GCA)_3$	TGTTCCTGGCATACTTCATC	GTTTCATGTGGGTATCTTTCCT	56	297	1
SSRCa 078	$(TCC)_5$	AGCCTCCCTTAGTTTGTTCTC	GGAAAGTCGTCAGATTGGTT	56	210	1
SSRCa 079	$(CCCT)_2N_5(GAAAA)_3$	AAGTGGAGGAGTTTTGTGGA	CCAAGTGGATAGGTGTGAGAG	57	287	2

To be continued …

Table 3. Cont.

Primer	Repeats	Forward primer (5'–3')	Reverse primer (5'–3')	Tm ºC	Exp. Size of frag.	Number of alleles
SSRCa 080*	$(CA)_9N_8(CT)_{30}$	GTTCTTTCCGCCGTCAAT	GAGAAGAGAGAGGAAGGGAAA	57	250	3
SSRCa 081	$(CT)_{38}$	ACCGTTGTTGGATATCTTTG	GGTTGAACCTAGACCTTATTT	53	229	1
SSRCa 082	$(CT)_{17}CG(CT)_6$	GCTTGTTTCCATCGCTAAA	TTACACGTCAACCCACAAAC	56	178	2
SSRCa 083	$(TC)_{32}$	TCCAACAACATTAAGCGTATTC	GACAAACCTGAGGGAAAAGA	56	223	1
SSRCa 084	$(CCA)_4....(CAC)_7$	ATCGGAAAGATGTCAACCAT	CAAATTGAAGCCAGTGGTG	57	157	3
SSRCa 085	$(TC)_{24}$	ATGTGAAAATGGGAAGGATG	CACAGGAAAGTGACACGAAG	57	105	4
SSRCa 086	$(AC)_{11}$	AGAGAGAAGCCATGATTTGA	TCAGTCCCAGAGAATAAGGA	54	105	-
SSRCa 087	$(TC)_{22}$	TCACTCTCGCAGACACACTAC	GCAGAGATGATCACAAGTCC	56	143	4
SSRCa 088	$(TTTTCT)_3$	TACCTCTCCTCCTCCTTCCT	ATTTCTATGGACCGGCAAC	57	180	3
SSRCa 089	$(TC)_{19}$	GAAATGGTGAACTCTCTCTTGG	ATTTGCATGGCTTTGGTG	58	185	1
SSRCa 090	$(GA)_{21}$	TGACTCGATTACATCCCTAATG	GTATTTTGGTTCCCCATGTT	56	120	-
SSRCa 091*	$(GT)_8(GA)_{10}$	CGTCTCGTATCACGCTCTC	TGTTCCTCGTTCCTCTCTCT	56	110	4
SSRCa 092	$(CCA)_7CT(TCCACC)_5$	ATAGCCTGAGCCGTAACCA	GGGTAATTATGACGAGGGACA	58	142	4
SSRCa 093	$(CT)_{37}$	TTGCCTACAATACCTGTCTCC	CCCAATTCCTCTCCATTCT	56	196	-
SSRCa 094	$(TC)_4(TTCT)_3$..... $(TTTCCT)_3(TTTC)_5$	GTGTCCTAGGGAAGGGTAAG	GAGTGCTAGGAGAGGGAGAG	55	195	1
SSRCa 095*	$(TG)_{11}$	GAGAGAGCCGAGTGAAGAGA	GAGAGAGAAGCCATGATTTGA	57	185	1
SSRCa 096*	$(CT)_{18}$	GAAATGGTGAACTCTCTCTTGG	ATTTGCATGGCTTTGGTG	57	183	1

(-) Did not present amplification product in the established conditions
(*) Polymorphism observed between Híbrido de Timor UFV 445-46 and Catuaí UFV 2145-235

Conversion rates similar of *C. arabica* were reported for *Corchorus capsularis* (91.0%, Mir et al. 2008), *Lolium multiflorum* (90.4%, Hirata et al. 2006), *Cucurbitaceas* (94.0%, Gong et al. 2008), *Humulus lupulus* (92.2%, Stajner et al. 2005), and *Avena sativa* (95.5%, Li et al. 2000). One of the smallest conversion rates found was *Pinus* (4.1%, Hicks et al. 1998).

The molecular markers are useful for the construction of genetic maps and identification of markers linked to the genes that control agronomic characteristics, which may open the possibilities for marker assisted selections. Different classes of genetic markers have been developed for *Coffea* species (Pailard et al. 1996, Ky et al. 2000, Lashermes et al. 2001, Pearl et al. 2004, Teixeira-Cabral et al. 2004, Oliveira et al. 2007), however, to date, only 37 SSR markers have been mapped in *C. canephora* (Hendre et al. 2008) and none in *C. arabica*. It is therefore important to increase the availability of SSR markers for the *Coffea* species.

CONCLUSION

The enriched genomic library methodology was efficient in the development of SSR markers for *C. arabica*. As a result of this work 90 new SSR markers were developed and, therefore, facilitating the genetic studies of *C. arabica.*

Desenvolvimento e validação de marcadores microssatélites para *Coffea arabica* L.

RESUMO - *Com o objetivo de desenvolver novos marcadores microssatélites para* Coffea arabica*, duas bibliotecas genômicas enriquecidas com sondas* $(GT)_{15}$ *e* $(AGG)_{10}$ *foram construídas. Um total de 835 clones foi sequenciado e 756 apresentaram sequências de boa qualidade. Foram observados 113 clones (14,94%) contendo sequências redundantes. Microssatélites foram encontrados em 287 clones (38%). Aproximadamente 417.5Kb do genoma de* C. arabica *foi analisado, com uma média de um microssatélite a cada 1,46Kb. As repetições de dinucleotídeos foram mais freqüentes do que os de trinucleotídeos. Quatro sequências repetidas,* $(AG/CT)_n$, $(AC/GT)_n$, $(AAG/CTT)_n$, *e (AGG/CCT)n representaram 61,1% do total observado.*

Um total de 96 primers SSR foram desenhados e testados por PCR em dois genótipos de C. arabica*. Noventa novos marcadores microssatélites foram validados para futuros estudos genéticos de* C. arabica*.*

Palavras-chave: Marcador microssatélite, biblioteca genômica enriquecida, café, marcador molecular.

REFERENCES

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V and Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. **Theoretical and Applied Genetics 114**: 359-372.

Armour JA, Neumann R, Gobert S and Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridization selection. **Human Molecular Genetics 3**: 599-565.

Ashworth VETM, Kobayashi MC, De La Cruz M and Clegg MT (2004) Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. **Scientia Horticulturae 101**: 255-267.

Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P and Aggarwal RK (2003) Isolation and characterization of nine microsatellite markers from *Coffea arabica* (L.) showing wide crossspecies amplifications. **Molecular Ecology Notes 3**: 647-650.

Benchimol LL, Campos T, Carbonell SAM, Colombo CA, Chioratto AF, Formighieri EF, Gouvêa LRL and Souza AP (2007) Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. **Genetics Resources and Crop Evolution 54**: 1747-1762.

Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK and Aggarwal RK (2005) Identification and characterization of expressed sequence tags-derived simple sequence repeats markers from robusta coffee variety 'CxR' (an interspecific hybrid of *Coffea canephora* x *Coffea congensis*). **Molecular Ecology Notes 5**: 80-83.

Brondani RPV, Brondani C, Tarchini R and Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical and Applied Genetics 97**: 816-827.

Brondani C, Brondani RPV, Rangel PH and Ferreira ME (2001) Development and mapping of Oryza glumaepatula-derived microsatellite markers in the interspecific cross *Oryza glumaepatula* x *O. sativa*. **Hereditas 134**: 59-71.

Buso GSC, Brondani RV, Amaral ZPS, Reis AMM and Ferreira ME (2000) Desenvolvimento de primers SSR para análise genética de pimentas e pimentões (*Capsicum* spp.) utilizando biblioteca genômica enriquecida. **Boletim Técnico da Embrapa/Epagri 15**: 01-27.

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D and Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. **Genetics 156**: 847-854.

Clarindo WR and Carvalho CR (2009) Comparison of the *Coffea canephora* and *C. arabica* karyotype based on chromosomal DNA content. **Plant Cell Reports 28**: 73-81.

Collevatti RG, Brondani RV and Grattapaglia D (1999) Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. **Heredity 83**: 748-756.

Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G and Lashermes P (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. **Molecular Ecology 9**: 1171-1193.

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

Creste S, Tulmann-Neto A and Figueira A (2001) Detection of simple sequence repeat polymorphisms in denaturing polyacrilamide sequencing gels by silver staining. **Plant Molecular Biology Report 4**: 299-306.

Cristancho MA and Gaitán AL (2008) Isolation, characterization and amplification of simple sequence repeat loci in coffee. **Crop Breeding and Applied Biotechnology 8**: 321-329.

Cuc LM, Mace ES, Crouch JH, Quang VD, Long TD and Varshney RK (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). **BMC Plant Biology 8**: 55-65.

Cui X, Dong Y, Hou X, Cheng Y, Zhang J and Jin M (2008) Development and characterization of microsatellite markers in *Brassica rapa* ssp. *chinensis* and tansferability among related species. **Agricultural Sciences in China 7**: 19-31.

Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET and Oliveira ACB (2005) Analysis of AFLP marker associated to the *Mex*-1 resistance locus in Icatu progenies. **Crop Breeding and Applied Biotechnology 5**: 387-393.

Gao LF, Tang J, Li H and Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. **Molecular Breeding 12**: 245-261.

Garner TW (2002) Genome size and microsatellites: the effect of nuclear size on amplification potential. **Genome 45**: 212-215.

Gong L, Stift G, Kofler R, Pachner M and Lelley T (2008) Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L. **Theoretical and Applied Genetic 117**: 37-48.

Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM and Kashi Y (2000) Simple Sequence Repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. **Genome Research 10**: 62-71.

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

Hendre PS, Phanindranath R, Annapurna V, Lalremruata A and Aggarwal RK (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 Biology 8:** 51-70.

Hicks M, Adams D, O'Keefe S, Macdonald E and Hodgetts R (1998) The development of RAPD and microsatellite markers in lodgepole pine (*Pinus controrta* var. *latifolia*). **Genome 41**: 797-805.

Hirata M, Cai H, Inoue M, Yuyama N, Miura Y, Komatsu T, Takamizo T and Fujimori M (2006) Development of simple sequence repeat (SSR) markers and construction of an SSR-based linkage map in Italian ryegrass (*Lolium multiflorum* Lam.). **Theoretical and Applied Genetics 113**: 270-279.

Jarvis DE, Kopp OR, Jellen EN, Mallory MA, Pattee J, Bonifacio A, Coleman CE, Stevens MR, Fairbanks DJ and Maughan PJ (2008) Simple sequence repeat marker development and genetic mapping in quinoa (*Chenopodium quinoa* Willd.). **Journal of Genetics 87**: 39-51.

Katti MV, Rangekar PK and Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. **Molecular Biology and Evolution 18**: 1161-1167.

Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Charrier A, Hamon S and Noirot M (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). **Theoretical and Applied Genetics 101**: 669-676.

Lashermes P, Combes MC, Prakash NS, Trouslot P, Lorieux M and Charrier A (2001) Genetic linkage map of *Coffea canephora:* effect of segregation distortion and analysis of recombination rate in male and female meioses. **Genome 44**: 589-596.

Li CD, Rossnagel BG and Scoles GJ (2000) The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat cultivars. **Theoretical and Applied Genetics 101**: 1259-1268.

Litt M and Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American Journal of Human Genetics 44**: 397-401.

Metzgar D, Bytof J and Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. **Genome Research 10**: 72-80.

Mir RR, Rustgi S, Sharma S, Singh R, Goyal A, Kumar J, Gaur A, Tyagi AK, Khan H, Sinhá MK, Balyan HS and Gupta PK (2008) A preliminary genetic analysis of fibre traits and the use of new genomic SSRs for genetic diversity in jute. **Euphytica 161:** 413-427.

Missio RF, Caixeta ET, Zambolim EM, Pena GF, Ribeiro AP, Zambolim L, Pereira AA and Sakiyama NS (2009) Assessment of EST-SSR markers for genetic analysis on coffee. **Bragantia 68**: 573-581.

Moncada P and McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome 47**: 501-509.

Morgante M, Hanafey M and Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics 30**: 194-200.

Novelli VM, Cristofani M, Souza AA and Machado MA (2006) Development and characterization of polymorphic microsatellite markers for the sweet orange (*Citrus sinensis* L. Osbeck). **Genetics and Molecular Biology 29**: 90-96.

Oliveira CB, Sakiyama NS, Caixeta ET, Zambolim EM, Rufino RN and Zambolim L (2007) Mapa parcial de ligação gênica de *Coffea arabica* L. e recuperação do genitor recorrente em progênies de retrocruzamentos. **Crop Breeding and Applied Biotechnology 7**: 196-203.

Paillard M, Lashermes P and Petiard V (1996) Construction of a molecular linkage map in coffee. **Theoretical and Applied Genetics 93**: 41-47.

Pearl HM, Nagai C, Moore PH, Steiger DL, Osgood RV and Ming R (2004) Construction of a genetic map for arabica coffee. **Theoretical and Applied Genetics 108**: 829-835.

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, De Kochko A and Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. **Molecular and Genetics Genomics 276**: 436-449.

Poncet V, Dufour M, Hamon P, Hamon S, Kochko A and Leroy T (2007) Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. **Genome 50**: 1156-1161.

Ritschel PS, Lins TCL, Tristan RL, Buso GSC, Buso JA and Ferreira ME (2004) Development of microsatellite markers from an enriched genomic library for genetic analysis of melon (*Cucumis melo* L.). **BMC Plant Biology 4**: 1-14.

Rovelli P, Mettulio R and Anthony F (2000) Microsatellites in *Coffea arabica* L. In: Sera T, Soccol CR, Pandey A and Roussos S (eds) **Coffee biotechnology and quality**. Kluwer Academic Publishers, Amsterdam, p.123-133.

Rozen S and Skaletsky HJ (2000) Primer 3 on the WWW for general users and for biologist programmers. In: Krawetz S and Misener S (eds) **Bioinformatics methods and protocols: methods in molecular biology.** Humana Press, Totowa, p. 365-386.

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE and Cregan PB (2004) A new integrated genetic linkage map of the soybean. **Theoretical and Applied Genetic 109**: 122-128.

Song QJ, Shi JR, Singh S, Fickus EW, Costa JM, Lewis J, Gill BS, Ward R and Cregan PB (2005) Development and mapping of microsatellite (SSR) markers in wheat. **Theoretical and Applied Genetics 110**: 550-560.

Stajner N, Jakse J, Kozjak P and Javornik B (2005) The isolation and characterization of microsatellites in hop (*Humulus lupulus* L.). **Plant Science 168**: 213-221.

Teixeira-Cabral TA, Sakiyama NS, Zambolim L, Pereira AA and Schuster I (2004) Single-locus inheritance and partial linkage map of *Coffea arabica* L. **Crop Breeding and Applied Biotechnology 4**: 416-421.

Tesfaye K, Borsch T, Govers K and Bekele E (2007) Characterization of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugenioides* chloroplast genomes. **Genome 50**: 1112-1129.

Varshney RK, Graner A and Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. **Trends in Biotechnology 23**: 48-55.

Viruel MA and Hormaza JI (2004) Development, characterization and variability analysis of microsatellites in lychee (*Litchi chinensis* Sonn., Sapindaceae). **Theoretical and Applied Genetics 108**: 896-902.

Zane L, Bargelloni L and Patarnello T (2002) Strategies for microsatellite isolation: a review. **Molecular Ecology 11**: 1-16.

Zhao W, Miao X, Jia S, Pany and Huang Y (2005) Isolation and characterization of microsatellite loci from the mulberry, *Morus* L. **Plant Science 168**: 519-525.