CBAB

CROP BREEDING AND
APPLIED BIOTECHNOLOGY

Isolation, characterization and amplification of simple sequence repeat loci in coffee

Marco-Aurelio Cristancho[1*], and Álvaro-León Gaitán[1]

ABSTRACT - *Simple sequence repeat (microsatellite) loci in coffee were identified in clones isolated from enriched and random genomic libraries. It was shown that coffee is a plant species with low microsatellite frequency. However, the average distance between two loci, estimated at 127kb for poly (AG), is one of the shortest of all plant genomes. In contrast, the distance between two poly (AC) loci, estimated at 769kb, is one of the largest in plant genomes. Coffee $(AC)_n$ microsatellites are frequently associated with other microsatellites, mainly $(AT)_n$ motifs, while $(AG)_n$ microsatellites are not normally associated with other microsatellites and have a higher number of perfect motifs. Dinucleotide repeats (AG) and (AC) were found in AT-rich regions in coffee. Sequence analysis of $(AC)_n$ microsatellites identified in coffee revealed the possible association of these repeated elements with miniature inverted-repeat transposable elements (MITEs). In addition, some of the evaluated SSR markers produced transposon-like amplification patterns in tetraploid genotypes. Of 12 SSR markers developed, nine were polymorphic in diploid genotypes while 5 were polymorphic in tetraploid genotypes, confirming a greater genetic diversity in diploid species.*

Key words: Coffee, *Coffea arabica*, SSRs.

INTRODUCTION

Simple sequence repeats (SSRs), also known as microsatellites, are DNA sequences composed of tandemly repeated short motifs (Tautz and Renz 1984). These motifs have extensive site-specific length polymorphism due to differing numbers of repeat units; they are also robust, easily transferable, co-dominant, chromosomally located at a single locus and can be developed as PCR-based markers (Weber and May 1989, Tautz 1989).

Microsatellite sequences constitute a high percentage of most eukaryotic genomes (Matula and Kypr 1999) and are ubiquitous in animals, plants and fungi (Tóth et al. 2000). It seems that they also played a key role in the formation and size of some genomes. For instance in *Oryza*, the amplification of these repetitive DNA sequences causes variation in the chromosome morphology and consequently in the genome size among diploid species in the genus (Uozu et al. 1997).

Polymorphism of SSR loci is thought to arise from slippage of DNA polymerase during DNA replication (Levinson and Gutman 1987). Length polymorphisms at a particular SSR locus can be evaluated on the basis of the differing electrophoretic mobility of polymerase chain reaction (PCR) products amplified by primers flanking

[1] Centro Nacional de Investigaciones de Café – CENICAFÉ, Planalto, Km. 4 via a Manizales, Chinchiná, Caldas, Colombia. *E-mail: marco.cristancho@cafedecolombia.com

the motif. SSR characterization and mapping is underway in several plant species including rice (Temnykh et al. 2000), *Arabidopsis* (Bell and Ecker 1994), barley (Liu et al. 1996), and wheat (Röder et al. 1998). The level of allelic diversity revealed at SSR loci in most of these organisms indicates that SSRs are ideal for linkage analysis, agronomic trait selection, germplasm evaluation and varietal identification (Powell et al. 1996).

Among the different repeat motifs that compose microsatellites, poly (AT) is the most frequent in most plant genomes followed by poly (AG), while poly (AAG) and poly (AAT) are the most common among trinucleotide repeats (La Rota et al. 2005, Cardle et al. 2000, Temnykh et al. 2000, Echt and May-Marquardt 1997). Several strategies to isolate and characterize microsatellites have been proposed, e.g., search of EST sequences and hybridization of total and enriched genomic libraries (for a review, see Zane et al. 2002).

The genus *Coffea* includes two cultivated species of economic importance, *C. arabica* L. and *C. canephora* Pierre. *C. arabica* ($2n = 4X = 44$) is an amphidiploid formed by a recent event of hybridization between the diploid species *C. eugenioides* and *C. canephora* (Lashermes et al. 1999); all other *Coffea* species are diploid ($2n = 2X = 22$). Microsatellites have not been developed as extensively in coffee as in other crops. There are few reports on microsatellite markers of *C. arabica* (Baruah et al. 2003, Combes et al. 2000) which have been used for partial studies of the allele number and heterozygosity level in several diploid and tetraploid coffee species, of the genetic diversity of *C. arabica* varieties (Silvestrini et al. 2007), of the origin of cultivated *C. arabica* (Anthony et al. 2002), and for introgression studies in the same species (Herrera et al. 2002). Several SSR markers have also been identified in the species *C. canephora* (Poncet et al. 2007, Hendre et al. 2008)

The purpose of this study was to establish the frequency of several microsatellite motifs in the coffee genome; the isolation and characterization of new microsatellite sequences from *C. arabica*; and the evaluation of the amplification of SSR markers developed for genetic studies in different coffee species.

MATERIAL AND METHODS

Plant material and DNA extraction

DNA for the construction of small insert libraries and microsatellite-enriched libraries were extracted from lyophilized leaves of *C. arabica* cv. Caturra using the method outlined by Tai and Tanksley (1991). The DNA was further purified by a Cesium Chloride gradient (Sambrook et al. 1989).

Construction of a small insert library

One microgram of coffee DNA was partially digested for 1h at 65°C with 0.5U of the enzyme *Tsp*509I (New England Biolabs). Size-selected fragments between 350bp and 750bp were isolated by eluting the DNA from the gel and purifying it with a Wizard Mini-Prep kit (Promega). The size-selected fragments were cloned into the *Eco*RI site of lambda ZapII (Stratagene).

Construction of microsatellite-enriched libraries

Four microsatellite-enriched genomic libraries were constructed, using DNA of *C. arabica* var. Caturra. The four microsatellites used for the enrichment process were $(AG)_n$, $(AC)_n$, $(AAT)_n$, and $(AAG)_n$, which were chosen after Southern hybridization with coffee genomic DNA with several repeated motifs. The signal of the four chosen motifs was fairly strong in Southern analysis.

The protocol described by White and Powell (1997) was used for the construction of SSR-enriched libraries. The protocol is based on the isolation of microsatellite-containing fragments by hybridization to a biotinylated SSR oligonucleotide and subsequent selection on streptavidin-coated magnetic beads. This is a modification of the enrichment protocols developed by Karagyozov et al. (1993) and Kijas et al. (1994).

In summary, coffee DNA was partially restricted with enzyme *Tsp*509I and the fragments enriched by the oligonucleotides $(AG)_n$, $(AC)_n$, $(AAT)_n$, and $(AAG)_n$. Four coffee microsatellite libraries were constructed with the enriched fractions. All four libraries were constructed using the *Eco*RI site of a lambda Zap Express vector (Stratagene).

Library screening

A small insert library of rice formerly described by Panaud et al. (1995) was used as a control, in view of the previously known relative abundance of microsatellites in the rice genome. Coffee and rice libraries were tittered, and a total of 100,000 plaques were screened for the presence of microsatellite containing plaques by plating at 10,000 pfu in ϕ 0.137m plates.

About 500 pmol of the oligonucleotides poly (AC)$_9$, poly (AG)$_9$, poly (AAT)$_8$, and poly (AAG)$_8$, was random primer-labelled with 30 µCi of 32 P[CTP] and 2U of Klenow fragment, and used to probe plaque lifts at 65°C [oligonucleotides poly (AC) and poly (AG)] or 60°C [oligonucleotides poly (AAG) and poly (AAT)] overnight according to standard protocols (Sambrook et al. 1989).

Membranes were washed twice at 65°C or 60°C in 2X SSC, 0.1% SDS, and subsequently exposed to Biomax X-ray film (Kodak). Putative positive plaques were isolated and plasmid DNA prepared by a Wizard mini-prep kit (Promega). Either plasmid DNA or PCR-amplified fragments from the positive plaques were sequenced with T3/T7 primers on an ABI 377 automated sequencer at the Cornell Sequencing Facility.

Sequence Analysis

The high quality sequences obtained were analyzed for the presence of microsatellite motifs by the programs RepeatMasker (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker) or SSRIT (http://ars-genome.cornell.edu/cgi-bin/rice/ssrtool.pl). Microsatellite loci were defined as the dinucleotide motifs with more than 7 repeat units and trinucleotide motifs containing more than 5 repeat units. Vector sequences were trimmed with the commercial software SEQUENCHER® (Gene Codes Inc., Ann Harbor MI) or LASERGENE (DNASTAR, Inc., Madison, WI). Redundant sequences were eliminated using the program Stand-alone BLAST (Altschul et al. 1990).

SSR assays

Primers were designed for SSR flanking sequences using the computer program PRIMER 0.5 (Daly et al. 1991). SSR assays were carried out exactly as described by Milbourne et al. (1998). Four different genotypes in duplicate were used for the assay of the SSR markers; the total genomic DNA of the diploid species *C. liberica* CCC1025 and *C. eugeniodes* "ofic" accessions and the tetraploids *C. arabica* var. Caturra accession 23 and Timor hybrid 1343 (CV1), a natural tetraploid hybrid, were amplified.

RESULTS AND DISCUSSION

Microsatellite frequency in coffee

In order to determine the frequency of microsatellites in the coffee genome, four motifs were radioactively labeled and used to hybridize to 100,000 clones from the random genomic library of coffee. A genomic library of rice was used as control and reference (Panaud et al. 1995). It was evidenced that coffee is among the plant species with a low frequency of the microsatellite motifs tested. The average distance between two loci, estimated at 127kb for poly (AG) (Table 1), is one of the shortest in any plant genome (coffee genome size is 1158Mb, Arumuganathan and Earle 1991). In contrast, the distance between two poly (AC) loci, estimated at 769kb, is one of the largest in plant genomes (Table 1).

The total estimated number of poly (AG) microsatellites in coffee is 9110. Among the plant genomes with a large number of poly (AG) motifs are pine with 60,000 (Echt and May-Marquardt 1997) and wheat with 36,000 (Röder et al. 1995) while the number of 1360 for rice is small (Panaud et al. 1995). The estimated number of poly (AC) microsatellites in coffee is 1505, similar to rice (1230), while pine with 120,000 and wheat with 23,000 have rather large numbers of this

Table 1. Mean distance in kpb between two microsatellites in different plant genomes

Species	AG	AC	AAG	AAT	Reference
Wheat	440	704	N/A	N/A	Röder et al. (1995)
Pinus taeda	430	520	1400	520	Echt and May-Marquardt (1997)
Pinus strobus	470	220	1300	530	Echt and May-Marquardt (1997)
Rice	330	365	N/A	N/A	Panaud et al. (1995)
Peach	100	420	N/A	N/A	Sosinski et al. (2000)
Coffee	127	769	1052	2727	This study

N/A : data not available. The mean distance between two loci was obtained by dividing the length of coffee DNA screened by the number of positive signals counted after hybridization with the corresponding oligonucleotide repeats. For the 100,000 clones screened, the corresponding length of coffee DNA was 100,000 X 0.6kbp = 60,000kbp. For each microsatellite, this value was divided by the number of signals observed

microsatellite motif. Finally, results also suggest that coffee contains a very small number of poly (AAT) microsatellites, contrary to what has been observed in other plant genomes (La Rota et al. 2005, Cardle et al. 2000, Temnykh et al. 2000, Echt and May-Marquardt 1997). A similar pattern was identified in the *C. canephora* genome (Hendre et al. 2008). It will be of interest to test the frequency of di- and trinucleotide repeats in coffee coding regions, given the fact that trinucleotide repeats are more common in coding regions of other plant genomes (Laurent et al. 2007).

Although it was shown that poly (AT) motifs are the most abundant microsatellites in plant genomes (La Rota et al. 2005, Cardle et al. 2000), the low melting temperature of poly (AT) oligonucleotides does not allow for an easy retrieval of positive clones by hybridization-based protocols. To date it is not possible to assay the frequency of poly AT motifs in the coffee genome by the commonly used database mining, due to the very small number of coffee sequences deposited in public databases. We decided to evaluate the relative frequency of other microsatellites in coffee, which were shown to be common in other plant genomes.

The frequency and number of microsatellites in the coffee genome reported here could be somehow higher since the frequency of rice microsatellites found in this study was lower than previously reported by Panaud et al. (1995) (Tables 1 and 2). However, some general conclusions are possible: As in many other plants, the frequency of poly (AG) microsatellites in coffee is higher than that of poly (AC) microsatellites. Although coffee has a much lower frequency of microsatellites than rice, the total number of microsatellites is higher, due to the larger genome. Finally, the proportion of poly (AAG) and poly (AAT) microsatellites, very common trinucleotide motifs in other plants (La Rota et al. 2005, Cardle et al. 2000, Temnykh et al. 2000, Echt and May-Marquardt 1997), is not very high in the coffee genome. It would be of interest to investigate the frequency of other microsatellite motifs in coffee.

Sequence Analysis

Identification of Microsatellite Sequences

Based on hybridization experiments using radioactive oligonucleotides, the enrichment was considered successful for the motifs $(AG)_n$ and $(AC)_n$ but not for $(AAT)_n$ and $(AAG)_n$. One of the reasons for the failure to enrich the libraries for the trinucleotides motifs could be the difficulty to optimize the hybridization temperature for these repeated sequences. The random genomic library was screened with the dinucleotides poly (AG) and poly (AC) only.

A total of 367 poly (AG) and poly (AC) microsatellite sequences were identified. The number and type of the identified coffee microsatellites is summarized in Table 2. Coffee $(AC)_n$ microsatellites are frequently associated with other microsatellites; 54 (29%) were associated to $(TC)_n$ microsatellites, 34 (18%) to $(AT)_n$ sequences, and 7 to different tetranucleotide motifs (4%). On the other hand, coffee $(AG)_n$ microsatellites are not normally associated with other microsatellites, so the number of perfect motifs is higher. Only 7 (4%) microsatellites were associated to (TG), 7 (4%) to (AT) sequences and 6 (3%) to different trinucleotide sequences. The rest of the compound sequences correspond to imperfect poly (AG) motifs.

Interestingly, the dinucleotide repeats (AG) and (AC) are found in AT-rich regions in the coffee genome (37% and 40% GC contents, respectively). A similar result was reported for the same repeats in rice [44% and 38% GC content, respectively (Temnykh et al. 2001)].

Characterization of Microsatellite Sequences

A total of 117 of the flanking regions of poly (AG) microsatellites share either a low or high portion of homology with each other (50-90%) and 32 sequences contain either direct or inverted repeats, which could be associated with transposable elements (TE). Among the flanking regions of poly (AC) microsatellites, 86 are similar to other (AC) microsatellites and 85 contain direct

Table 2. Number and type of microsatellites (repeat units >7) identified in the coffee genome

	Type of Motif	
	AG	AC
Perfect Repeat	122	56
Compound Repeat	59	130
Maximum Length of Perfect Repeat	38	30
Mean Length of Perfect Repeat	14.88	13.39
% GC Content of Flanking Regions	37.84	40.83
Total Unique Sequences	181	186

or inverted repeats. They might therefore be associated with transposable elements.

Sequence analysis of the $(AC)_n$ microsatellites identified in coffee revealed a possible association of these repeated elements with miniature inverted-repeat transposable elements or MITEs. The characteristics of coffee sequences are similar to those of the MITE family, in terms of short length (up to 400bp) and terminal inverted repeats of 10-15bp (Zhang et al. 2000). They might represent the first family of transposable elements reported in coffee. It is also notable that 2 of 12 SSR markers assayed produced transposon-like amplification patterns. The markers CFGA627 and CFGA360 amplified more than 10 alleles in tetraploid genotypes, which is close to what was expected from transposon-associated loci (Table 3).

By a BLAST search against the databases GENEBANK and RepeatMasker2 (http://repeatmasker.genome.washington.edu/), the DNA sequences were screened for interspersed repeats and low-complexity DNA sequences (Jurka 2000). Results did not indicate homology to any known sequence in these databases. However, the MITE class of elements and other transposable elements are characterized for sequence similarity within a family in a species but not among families in different species (Zhang, et al. 2000), i.e., the lack of homology

Table 3. Amplification of coffee microsatellites in diploid and tetraploid species

	Repeat Motif	PCR primer sequence (5´–> 3´)	Product size (bp)	Number of alleles/polymorphism	
GA				**Diploids**	**Tetraploids**
CFGA574[2]	$(AG)_6 TG$ $(AG)_8$	F: GTATGGCTCTGCATTCTGTCA R: GCCAACCTCTCAAATGCTTC	106	3/Yes	1 /No
CFGA627[2]	$(AG)_{16}$	F: GGGAAGGAATTCTTTCAACTCT R: CTTGGAAATTACCATGCAACC	134	2 /Yes	>10/ Yes
CFGA792b[2]	$(AG)_{12}$	F: GATCAGAACTTTGAGCTCAGCA R: AATGTGGCACGCTAGAAGTG	182	4 /Yes	6 /Yes
CFGA1122[1]	$(AG)_{13}$	F: AATGCTTCCAGTCACCCGT R: GGCTTTCAAGCATCCAGTCT	127	4 /Yes	3 /No
CFGA1255[1]	$(AG)_{11}$	F: TTGACGTTCTCGTCCGTATG R: AAAGGCTTGAAAACTTGCGG	147	4 /Yes	2 /No
CFGA1258[1]	$(AG)_{18} (N)_{20}$ $(CCA)_5$	F: CAAGACAAATGTCTCTGCTGG R: AATGTATCAAGCGCGCAAC	132	3 /Yes	1 /No
CA					
CFCA14A[2]	$(AC)_{12}$	F: CTGGTAGGAGCTCTTCAAATTG R: TGGCCAAATCTTGTCCGT	161	4 /Yes	2 /No
CFCA281[2]	$(AC)_{13}$	F: GCGTCCACGTGTTAAGTCTT R: TCAAGTGGCAGACATGTCAC	155	2 /No	10 /Yes
CFCA331[2]	$(CT)_{17} (AC)_{18}$ $(CT)_4 AT TT$ $(AC)_6$	F: TGATGGACAGGAGTTGATGG R: CACTCATTTTGCCAATCTACC	150	2 /No	10 /Yes
CFCA334[2]	$(AC)_5 AA$ $(AC)_{13}$	F: AGCCACCACAGGAAGTTTCA R: GGGAGTGAAAGACATCAGGTG	154	1 /No	1 /No
CFCA360[2]	$(AC)_{15}$	F: TTAAGACATCGGTGCATTCA R: TGTGTACTGGGTTTTTTGATGT	135	2 /Yes	>10 /Yes
CFCA530[2]	$(ACC)_8 (AC)_{11}$	F: GTGTTTGCTTGTCAATATGCC R: ACTCCTAACCTAACCGGGAA	122	5 /Yes	4 /No

[1] Microsatellite loci isolated from the random genomic library
[2] Microsatellite loci isolated from the enriched libraries

to known transposable elements was not totally unexpected.

Identification and Characterization of Microsatellite Sequences

After hybridization experiments using radioactive oligonucleotides, it was determined that the enrichment was successful for the motifs $(AC)_n$ and $(AG)_n$ but not for the motifs $(AAT)_n$ and $(AAG)_n$. One of the reasons for the failure to enrich the libraries with trinucleotide motifs could be the difficulty of optimizing the hybridization temperature for these repeated sequences. Another reason could be the very low frequency of these motifs, as shown in the screening of the coffee genomic library with the corresponding oligonucleotides. It might also be due to the very low frequency of trinucleotide repeat microsatellites found in *C. arabica*, observed in *C. canephora* as well (Hendre et al. 2008).

Therefore, for the further development of microsatellite markers in coffee it is suggested to concentrate on poly (AG) and poly (AC) motifs, to make an appropriate saturation for mapping purposes possible, due to their high frequency. The high frequency of compound microsatellites in coffee is unusual since they tend to be rare in other plant genomes they tend to be rare (Cardle et al. 2000). It is interesting that $(AC)_n (CT)_n$, the most common compound microsatellite in coffee, is also a very common compound microsatellite in humans (Bull et al. 1999).

Baruah et al. (2003) also carried out hybridization experiments with radiolabelled di- and trinucleotide repeats; they do not describe the total number of positive clones that hybridized to each oligonucleotide but identified a total of eight dinucleotide repeats and only one trinucleotide repeat microsatellite, which suggests a lower frequency of trinucleotide repeat microsatellites in the coffee genome. This pattern is different than reported for other plant genomes, where trinucleotide prevail over dinucleotide repeats (La Rota et al. 2005, Cardle et al. 2000).

The relatively low GC content of the flanking regions of the coffee microsatellite motifs could be explained by the fact that *Tsp*509I and *Eco*RI, the enzymes used to construct the libraries, recognize AT-rich sites (AATT and GAATTC respectively). Microsatellites in animals are mostly confined to AT-rich sequences (Matula and Kypr 1999), which could also be the case for coffee.

Identification of transposon sequences associated with coffee microsatellites

It was remarkable to find a high number of microsatellites in coffee associated to transposon-like sequences. The association of microsatellites with other classes of repetitive DNA has been detected in other organisms. Of the AT dinucleotide repeats in rice, 45% were found to be associated to members of a MITE family (Temnykh et al. 2001). Some barley microsatellites are frequently associated to MITE elements, mostly of the Stoneway type (Cardle et al. 2000) and 41% of 290 clones containing SSRs were associated with known repetitive elements, especially retrotransposons (Ramsay et al. 1999). Mammalian and insect microsatellites have been frequently found to be associated to short interspersed elements (SINEs). These mobile elements have played a very active role in the genesis of microsatellite repeats in these groups of organisms (Tozaki et al. 2001, Wilder and Hollocher 2001, Arcot et al. 1995). Finally, it would be of great interest to develop transposon markers for coffee from these sequences, similar to the markers developed in petunia and maize, described as the transposon-display technique (Van den Broeck et al. 1998, Casa et al. 2000); these markers will also help in the mapping of agronomically important genes of coffee.

Amplification profile of coffee microsatellites

The amplification profile of 12 coffee SSR markers in diploid and tetraploid genotypes is shown in Table 3. Nine out of 12 markers were polymorphic in diploid genotypes while 5 were polymorphic in tetraploid genotypes. It is noteworthy that in several cases the number of amplified alleles was high for tetraploid genotypes. As mentioned before, some of the markers produced transposon-like amplification patterns in the same tetraploid genotypes.

The result of the number of polymorphic microsatellites in *C. arabica* genotypes was similar to that reported by Combes et al. (2000) who found 5 polymorphic microsatellites in 11 tested. We found 5 polymorphic microsatellite markers of *C. arabica* in 12 tested. In diploid genotypes the number of polymorphic markers was higher (9 of 12), indicating a greater genetic diversity. Similar results are found when EST-SSR markers are used in both species (Aggarwal et al. 2007), although the polymorphism of SSRs in non-coding regions is greater. The reduction in polymorphism in tetraploid genotypes could be due to their autogamous nature, which may have favored a loss of genetic

diversity during geographic dispersion, contrary to the diploid varieties that are allogamous.

It is noteworthy that the number of amplified alleles in the tetraploid genotypes was high in several cases. As mentioned above, some of the markers produced transposon-like amplification patterns in the same tetraploid genotypes. It would be of interest to check for the presence of transposon-like elements in the diploid species and test if these repeated elements are as frequently associated to microsatellites in diploids as in the tetraploid *C. arabica*. After the amplification of 12 microsatellite markers, the amplification pattern of four was atypical, two had more than 10 alleles and two others more than 20, indicating the presence of transposon elements. This pattern was not observed in diploid genotypes. The number of transposable elements may therefore vary in different coffee species, depending on their ploidy level. Furthermore, the recent formation of the tetraploid *C. arabica* species from two diploid parents (Lashermes et al. 1999) could have activated transposable elements in the new genome.

Significant differences in MITE numbers have already been reported in the genus Medicago. *M. sativa* was found to contain a high number of members of the MITE family *Bigfoot* while *M. truncatula* was found to contain a very small number of members of the same family. The differences might be related to the genome size of both species, since *M. truncatula* has a very small genome in comparison with *M. sativa* (Charrier et al. 1999). The common presence of transposable elements in plant and animal genes indicates that these repetitive elements may play a central role in gene and genome evolution (Bureau et al. 1996). Evidence that transposable elements play a central role in evolutionary restructuring of genomes has grown for every organism with a reasonable amount of collected data (Fedoroff et al. 2000).

This study is a significant contribution to the genetics of coffee, a little studied but highly important crop. The microsatellites described here are valuable in view of the small number of published microsatellites for coffee; they can be used for studies of diversity, mapping and eventually for marker-assisted selection. The information about microsatellite motif frequency provides insights on the repeat type that must be searched in coffee to increase the chance of success of finding them.

The clone sequences for the markers developed here are available as GenBank accessions AY102422 to AY102427 and AY102461 to AY102466.

Isolação, caracterização e amplificação de locos de microssatélites em café

RESUMO - *Loci de seqüência simples repetidas (SSR) no genoma do café foram identificados em clones isolados de bibliotecas genômicas ao acaso. Este trabalho mostrou que o café é uma espécie com baixa freqüência de microsatélites. Entretanto, a distância média entre dois loci estimada em 127kb para poly (AG) é uma das menores de todos os genomas de plantas. Apesar da distância entre dois poly (AC) loci estimada em 169kb ser uma das maiores para plantas. (AC)n em café são frequentemente associados com outros microssatélites e possuem alto número de repetições perfeitas. Repetições de dinucleotídeos (AG) foram encontrados de regiões ricas em AT em café. Análises de seqüências de microsatélites (AC) identificadas em café revelam a possibilidade de associação entres esses elementos repetidos com miniaturas de elementos transposons de repetições-inversas (MITEs). Além disso, alguns dos marcadores SSR avaliados produziram padrões de amplificação semelhantes a transposons em genótipos tetraplóides. Dos 12 marcadores SSR desenvolvidos, 9 foram polimórficos em genótipos diplóides, enquanto 5 foram polimórficos em genótipos tetraplóides, confirmando a grande diversidade genética em espécies diplóides.*

Palavras-chave: Café, *Coffea arabica*, SSRs.

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.

Altschul, SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic Local Alignment Search Tool. **Journal of Molecular Biology 215**: 403-410.

Anthony F, Combes MC, Astorga C, Bertrand B, Graziosi G and Lashermes P (2002) The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. **Theoretical and Applied Genetics 104**: 894-900.

Arcot SS, Wang Z, Weber JL, Deininger PL and Batzer MA (1995) *Alu* Repeats: A Source for the genesis of primate microsatellites. **Genomics 29**: 136-144

Arumuganathan K, and Earle D (1991) Nuclear DNA content of some important plant species. **Plant Molecular Biology Reporter 9**: 208-218.

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 cross-species amplifications**. Molecular Ecology Notes 3**: 647–650.

Bell CJ and Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of Arabidopsis. **Genomics 19**: 137-144.

Bull LN, Pabón-Peña CR and Freimer NB (1999) Compound microsatellite repeats: practical and theoretical features. **Genome Research 9**: 830-838.

Bureau TE, Ronald PC and Wessler SR (1996) A computer based systemic survey reveals the predominance of small inverted-repeat elements in wild type rice genes. **Proceedings of the National Academy of Sciences 93**: 8524-8529.

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.

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S and Wessler SR (2000) The MITE family *Heartbreaker* (Hbr): Molecular markers in maize. **Proceedings of the National Academy of Sciences 97**: 10083-10089.

Charrier B, Foucher F, Kondorosi E, d'Auberton-Carafa Y, Thermes C, Kondorosi A and Ratet P (1999) Bigfoot: a new family of MITE elements characterized from the *Medicago* genus. **Plant Journal 18**: 431-441.

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**: 1178-1180.

Daly MJ, Lincoln SE and Lander ES (1991) "PRIMER", unpublished software, Whitehead Institute/MIT Center for Genome Research. Available at http://www-genome.wi.mit.edu/ftp/pub/software/primer0.5.

Echt CC and May-Marquardt P (1997) Survey of microsatellite DNA in pine. **Genome 40**: 9-17.

Fedoroff N (2000) Transposons and genome evolution in plants. **Proceedings of the National Academy of Sciences 97**:7002-7007.

Hendre PS, Phanindranath R, Annapurna V, Lalremruata A and Aggarwal RA (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.

Herrera JC, Combes MC, Anthony F, Charrier A and Lashermes P (2002) Introgression into the allotetraploid coffee (*Coffea arabica* L.): segregation and recombination of the *C. canephora* genome in the tetraploid interspecific hybrid (*C. arabica* x *C. canephora*). **Theoretical and Applied Genetics 104**: 661-668.

Iniguez-Luy FL, Voort MV and Osborn TC (2008) Development of a set of public SSR markers derived from genomic sequence of a rapid cycling *Brassica oleracea* L. genotype. **Theoretical and Applied Genetics 117**: 977-985.

Jurka J (2000) Repbase Update: a database and an electronic journal of repetitive elements. **Trends in Genetics 16**: 418-420.

Karagyozov L, Kalcheva I and Chapman V (1993) Construction of random small-insert libraries highly enriched for simple sequence repeats. **Nucleic Acids Research 21**: 3911-3912.

Kijas JM, Fowler JCS, Garbett CA and Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. **Biotechniques 16**: 656-662.

La Rota M, Kantety RV, Yu JK and Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. **BMC Genomics 6**: 23.

Lashermes P, Combes MC, Robert J, Trouslot P, D´Hont A, Anthony F and Charrier A (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. **Molecular and General Genetics 261**: 259-266.

Laurent V, Devaux P, Thiel T, Viard F, Mielordt S, Touzet P and Quillet MC (2007) Comparative effectiveness of sugar beet microsatellite markers isolated from genomic libraries and GenBank ESTs to map the sugar beet genome. **Theoretical and Applied Genetics 115**: 793-805.

Levinson G and Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. **Molecular Biology and Evolution 4**: 203-221.

Liu ZW, Biyashev RM and Saghai-Maroof MA (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. **Theoretical and Applied Genetics 93**: 869-876.

Matula M and Kypr J (1999) Nucleotide sequences flanking dinucleotide microsatellites in the human, mouse and Drosophila genomes. **Journal of Biomolecular Structure and Dynamics 17**: 275-280.

Milbourne DR, Meyer A, Collins J, Ramsay LD, Gebhardt CA and Waugh AR (1998) Isolation, characterization and mapping of simple sequence repeat loci in potato. **Molecular and General Genetics 259**: 233-245.

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

Panaud O, Chen X and McCouch S (1995) Frequency of microsatellite sequences in rice (*Oryza sativa* L.). **Genome 38**: 1170-1176.

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

Powell W, Machray GC and Provan J (1996) Polymorphism revealed by simple sequence repeats. **Trends Plant Sciences 1**: 215-222.

Ramsay L, Macaulay M, Cardle L, Morgante M, Ivanissevich S, Maestri E, Powell W and Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. **The Plant Journal 17**: 415-425.

Röder MS, Plaschke J, König SU, Börner A, Sorrels ME, Tanksley SD and Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. **Molecular and General Genetics 246**: 327-333.

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy, P and Ganal MW (1998) A microsatellite map of wheat. **Genetics 149**: 2007-2023.

Sambrook J, Fritsch EF and Maniatis T (1989) **Molecular cloning: a laboratory manual**. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Silvestrini M, Junqueira MG, Favarin AC, Guerreiro-Filho O, Maluf MP, Silvarolla MB and Colombo CA (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. **Genetic Resources and Crop Evolution 54**: 1367-1379.

Sosinski B, Gannavarapu M, Hager LD, Beck LE, King GJ, Ryder CD, Rajapakse S, Bard WV, Ballard RE, and Abbott AG (2000) Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. **Theoretical and Applied Genetics 101**: 421-428.

Tai A and Tanksley S (1991) A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. **Plant Molecular Biology Reporter 8**: 297-303.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Research 17**: 6463-6471.

Tautz D and Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. **Nucleic Acids Research 12**: 4127-4138.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S and McCouch S (2001) Computational and Experimental Analysis of Microsatellites in Rice (*Oryza sativa* L.): Frequency, Length Variation, Transposon Associations, and genetic marker potential. **Genome Research 11**: 1441-1452.

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T and McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa*). **Theoretical and Applied Genetics 100**: 697-712.

Tóth G, Gáspári Z and Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. **Genome Research 10**: 967-981.

Tozaki T, Mashima S, Hirota K, Miura N, Choi-Miura N.H and Tomita M (2001) Characterization of equine microsatellites and microsatellite-linked repetitive elements (eMLREs) by efficient cloning and genotyping methods. **DNA Research 8**: 33-45.

Uozu S, Ikehashi H, Ohmido N, Ohtsubo H, Ohtsubo E and Fukui K (1997) Repetitive sequences cause for variation in genome size and chromosome morphology in the genus *Oryza*. **Plant Molecular Biology 35**: 791-799.

Van den Broeck D, Maes T, Sauer M, Zethof J, De Keukeleire P, D´Hauw M, Van Montagu M and Gerats T (1998) *Transposon Display* identifies individual transposable elements in high copy number lines. **The Plant Journal 13**: 121-129.

Weber JL and May PE (1989) Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. **American Journal of Human Genetics 44**: 388-396.

White G and Powell W (1997) Isolation and characterization of microsatellite loci in *Swietenia hum*ilis (Meliaceae): an endangered tropical hardwood species. **Molecular Ecology 6**: 851-860.

Wilder J and Hollocher H (2001) Mobile elements and the genesis of microsatellites in dipterans. **Molecular Biology and Evolution 18**:384-392.

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

Zhang Q, Arbuckle J and Wessler SR (2000) Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize. **Proceedings of the National Academy of Sciences 97**: 1160-1165.