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

Marco-Aurelio Cristancho1*, and Álvaro-León Gaitán1

Received 15 July 2008
Accepted 09 October 2008

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 Tsp509I (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 EcoRI 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 Tsp509I 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 EcoRI 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.
Isolation, characterization and amplification of simple sequence repeat loci in coffee

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 32P[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 SEQUENCER® (Gene Codes Inc., Ann Arbor 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>
<thead>
<tr>
<th>Species</th>
<th>AG</th>
<th>AC</th>
<th>AAG</th>
<th>AAT</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>440</td>
<td>704</td>
<td>N/A</td>
<td>N/A</td>
<td>Röder et al. (1995)</td>
</tr>
<tr>
<td>Pinus taeda</td>
<td>430</td>
<td>520</td>
<td>1400</td>
<td>520</td>
<td>Echt and May-Marquardt (1997)</td>
</tr>
<tr>
<td>Pinus strobus</td>
<td>470</td>
<td>220</td>
<td>1300</td>
<td>530</td>
<td>Echt and May-Marquardt (1997)</td>
</tr>
<tr>
<td>Rice</td>
<td>330</td>
<td>365</td>
<td>N/A</td>
<td>N/A</td>
<td>Panaud et al. (1995)</td>
</tr>
<tr>
<td>Peach</td>
<td>100</td>
<td>420</td>
<td>N/A</td>
<td>N/A</td>
<td>Sosinski et al. (2000)</td>
</tr>
<tr>
<td>Coffee</td>
<td>127</td>
<td>769</td>
<td>1052</td>
<td>2727</td>
<td>This study</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Type of Motif</th>
<th>AG</th>
<th>AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Repeat</td>
<td>122</td>
<td>56</td>
</tr>
<tr>
<td>Compound Repeat</td>
<td>59</td>
<td>130</td>
</tr>
<tr>
<td>Maximum Length of Perfect Repeat</td>
<td>38</td>
<td>30</td>
</tr>
<tr>
<td>Mean Length of Perfect Repeat</td>
<td>14.88</td>
<td>13.39</td>
</tr>
<tr>
<td>%GC Content of Flanking Regions</td>
<td>37.84</td>
<td>40.83</td>
</tr>
<tr>
<td>Total Unique Sequences</td>
<td>181</td>
<td>186</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>GA</th>
<th>Repeat Motif</th>
<th>PCR primer sequence (5´→3´)</th>
<th>Product size (bp)</th>
<th>Number of alleles/polymorphism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F: GTATGGCTCTGCAATTCTGTCA</td>
<td>106</td>
<td>3/Yes</td>
</tr>
<tr>
<td>CFGA574</td>
<td>(AG)₆ TG</td>
<td>R: GCCAACCTCTCAATTGCTTC</td>
<td></td>
<td>1/No</td>
</tr>
<tr>
<td></td>
<td>(AG)₈</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: GGGAGAACATTCTTTCCAAGTCAGCA</td>
<td>134</td>
<td>2 /Yes</td>
</tr>
<tr>
<td>CFGA627</td>
<td>(AG)₁₆</td>
<td>R: CTTGGAAATTACATGCAACC</td>
<td></td>
<td>&gt;10/Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: GATCACAGCTTTGAGCTCACAC</td>
<td>182</td>
<td>4 /Yes</td>
</tr>
<tr>
<td>CFGA792b</td>
<td>(AG)₁₂</td>
<td>R: AAATGTCGCAAGCTAAGTGT</td>
<td></td>
<td>6 /Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: AATGCCTTCAAGCTACCTTCT</td>
<td>127</td>
<td>3 /No</td>
</tr>
<tr>
<td>CFGA1122</td>
<td>(AG)₁₃</td>
<td>R: GGCTTTCAAGCTACCTTCT</td>
<td>147</td>
<td>2 /No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: TTGACGTCTCTGACTCGATAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFGA1255</td>
<td>(AG)₁₁</td>
<td>R: AAAGGCTTTGAAGACTCGGG</td>
<td>147</td>
<td>2 /Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: CAAAGCAGAATGTCCTTGCTTG</td>
<td>132</td>
<td>3 /Yes</td>
</tr>
<tr>
<td>CFGA1258</td>
<td>(AG)₁₈ (N)₂₀</td>
<td>R: AAATGATCAGCGCCGCAAC</td>
<td></td>
<td>1 /No</td>
</tr>
<tr>
<td></td>
<td>(CCA)₈</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: CTGTTAGGAGCTCTTCAAAATGT</td>
<td>161</td>
<td>4 /Yes</td>
</tr>
<tr>
<td>CFCA14A</td>
<td>(AC)₁₂</td>
<td>R: TGCCCAATCTTCTTGCTCT</td>
<td></td>
<td>2 /No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: WCACGGACAGACATGCAAC</td>
<td>155</td>
<td>10 /Yes</td>
</tr>
<tr>
<td>CFCA281</td>
<td>(AC)₁₃</td>
<td>R: TCAAGTGGGCAAGACATGCAAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: TGATGGAGAGGAGGTGATGGG</td>
<td>150</td>
<td>10 /Yes</td>
</tr>
<tr>
<td>CFCA331</td>
<td>(CT)₁₇ (AC)₁₈</td>
<td>R: CACTCATATGGCCATCTACC</td>
<td>154</td>
<td>1 /No</td>
</tr>
<tr>
<td></td>
<td>(CT)₁₅ ATT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(AC)₉</td>
<td>F: AGCCACAGGAGAGGTTCAC</td>
<td>154</td>
<td>1 /No</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>R: GGGATGAAAGACATCGGGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFCA334</td>
<td>(AC)₁₃</td>
<td>R: TTAGACATCCGGTCATCCA</td>
<td>135</td>
<td>2 /Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F: TTAGACATCCGGTCATCCA</td>
<td></td>
<td>&gt;10/Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: TGTGTACGGGTTTTTTGATGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFCA360</td>
<td>(AC)₁₅</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFCA530</td>
<td>(ACC)₈ (AC)₁₁</td>
<td>F: GATTTTTGGGTGGTTTAAGC</td>
<td>122</td>
<td>5 /Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R: ACTCTTAACTTAAACCGGAA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ Microsatellite loci isolated from the random genomic library
² Microsatellite loci isolated from the enriched libraries
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 Tsp509I and EcoRI, the enzymes used to construct the libraries, recognize AT-rich sites (AATT and GAATTTC 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
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 finding them.

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

**ACKNOWLEDGEMENTS**

The authors wish to thank the National Federation of Coffee Growers of Colombia for the financial support; Dr. Steve Kresovich and Dr. Sharon Mitchell for granting us the permission to use the facilities of the Institute for Genomic Diversity at Cornell University and providing assistance for the construction and evaluation of the enriched libraries; Mauricio La Rota for his valuable help in the manipulation and analysis of the microsatellite sequences; and Dr. Svieta Temnykh for sharing her knowledge on microsatellite isolation and analysis.

**Isolação, caracterização e amplificação de locos de microsatélites 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 microsatélites 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 entre 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


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


