Molecular Ecology Notes (2003) **3**, 647–650 doi: 10.1046/j.1471-8286.2003.00544.x

PRIMER NOTE

# Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications

A. BARUAH, V. NAIK, P. S. HENDRE, R. RAJKUMAR, P. RAJENDRAKUMAR and R. K. AGGARWAL
*Centre for Cellular and Molecular Biology, Uppal Road, Tarnaka, Hyderabad-500007, India*

**Abstract**

**Genetic improvement of coffee (*Coffea arabica* L.) is constrained by low genetic diversity and lack of genetic markers, suitable screening tools, information on the genetic make-up of available gene pool and long generation time. In this context, use of DNA markers such as microsatellites that provide high genetic-resolution becomes highly desirable. Here, we report the development of nine new microsatellite markers from partial genomic library of an elite variety of *Coffea arabica*. The developed microsatellites revealed robust cross-species amplifications in 17 related species of coffee, and their Polymorphic Information Content varied from 0 to 0.6, 0 to 0.78 and 0.67 to 0.90 for the arabica, robusta genotypes and species representatives, respectively. The data thus suggest their potential use as genetic markers for assessment of germplasm diversity and linkage analysis of coffee.**

*Keywords*: *Coffea*, genetic diversity, genomic library, microsatellites, short-sequence repeat

*Received 27 August 2003; revision accepted 14 September 2003*

Coffee is an important plantation crop belonging to the genus *Coffea* of the family Rubiaceae. Although, > 100 species of coffee are known, its commercial cultivation relies only on two species, an amphidiploid *Coffea arabica* L. ($2n = 4x = 44$) and diploid *C. canephora* Pierre ($2n = 22$). Considerations such as sustainability over ever-changing demands of agro-climatic conditions and the commercial markets calls for continuous efforts to develop better coffee genotypes. The conventional breeding efforts to this end, especially in arabica coffee have been seriously constrained, mainly due to the narrow genetic base of the arabica genepool, lack of efficient breeding tools and linkage maps, long pre-bearing periods and difference in ploidy level. Recent advances in DNA marker technology offer possibilities to overcome some of these limitations by providing highly informative genetic markers and marker assisted breeding.

Out of various kinds of DNA markers, polymerase chain reaction (PCR)-based microsatellites are most desirable for germplasm characterization and crop improvement due to their high-information content, codominant nature, sensitivity and ease to analyse with minimal quantities of test samples (Powell *et al*. 1996). Microsatellite markers have been developed for a large number of plant species and are increasingly being used for germplasm diversity, linkage analysis and molecular breeding (Gupta & Varshney 2000). Despite these advantages, only ~150 microsatellite markers have been reported till to date for coffee (Combes *et al*. 2000; Rovelli *et al*. 2000), signifying the need for expanding the repertoire of these genetically highly informative markers for efficient management and improvement of coffee germplasm resources. In this study, we describe nine new coffee-specific microsatellite markers developed from a partial genomic library for use as potential genetic markers.

A partial small-insert genomic library was constructed using standard procedures (Sambrook *et al*. 1989) from total cellular DNA isolated from an elite arabica genotype, *C. arabica* var. HDT (Hibrido-De-Timor). DNA was isolated from fresh leaves as described by Aggarwal *et al*. (2002). Isolated DNA was purified once by treating with 1/

Correspondence: Ramesh K. Aggarwal, PhD. Fax: 00 91 40 27160591; E-mail: rameshka@ccmb.res.in

**Table 1** Details of the microsatellite markers developed in the study

| Marker | Repeat motif | Primer sequence (5′–3′)* | $T_m$ (°C) | Size range (bp) | Allelic diversity in coffee germplasm | | | | | | | | | GenBank Accession no. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Arabicas ($n$ = 45) | | | Robustas($n$ = 6) | | | Species ($n$ = 17)§ | | | |
| | | | | | $N_a$ | $H_O$ | PIC | $N_a$ | $H_O$ | PIC | $N_a$ | $H_O$ | PIC | |
| CM2 | $(AC)_{10}(AT)_9$ | F: TGTGATGCCATTAGCCTAGC<br>R: TCCAACATGTGCTGGTGATT | 60 | 177–229 | 3 | 0.120 | 0.300 | 6 | 0.750 | 0.780 | 12 | 0.450 | 0.850 | AY220270 |
| CM3 | $(GA)_2CA(GA)_8GGA(GA)_3$ | F: CCTAGGCAAACATGCATTGA<br>R: TCCATTTGTACACGGTTAGGC | 59 | 196–222 | 2 | 0.020 | 0.200 | 5 | 0.750 | 0.780 | 8 | 0.480 | 0.810 | AY220271 |
| CM5 | $(CCT)_7$ | F: GTAACCACCACCTCCTCTGC<br>R: TGGAGGTAACGGAAGCTCTG | 60 | 167–230 | 2 | 0.700 | 0.400 | 2 | 1.000 | 0.500 | 13 | 0.590 | 0.890 | AY220272 |
| CM6 | $(AC)_{11}$ | F: GCTAAGTTCAATTGCCCCTGT<br>R: GGGTTAATTTGATTGCGTGA | 59 | 210–232 | 2 | 0.000 | 0.162 | 5 | 0.330 | 0.740 | 9 | 0.500 | 0.830 | AY220273 |
| CM8 | $(GA)_7(GT)_{12}$ | F: GCCAATTGTGCAAAGTGCT<br>R: ATTCATGGGGCCTTTGTCTT | 60 | 159–183 | 3 | 0.330 | 0.600 | 3 | 0.330 | 0.490 | 8 | 0.360 | 0.860 | AY220274 |
| CM11 | $(GA)_{11}$ | F: AATCACCTTCGCAAACCAAC<br>R: CCGAACGCAATATCTTATGC | 59 | 200–254 | 2† | 0.72‡ | nc | 1 | 0.000 | 0.000 | 9 | 0.270 | 0.670 | AY220276 |
| CM16 | $(GA)_{14}$ | F: TGGGGAAAAGAAGGATATAGACAAGAG<br>R: GAGGGGGGCTAAGGGAATAACATA | 55 | 95–129 | 3 | 0.330 | 0.420 | 3 | 0.500 | 0.400 | 13 | 0.730 | 0.890 | AY220277 |
| CM17 | $(CA)_{12}$ | F: CCAGCCTTTTCACAATTCTCACCC<br>R: TGCCCCCTAGATATGGTACAAGCTTTC | 55 | 308–390 | 2 | 0.000 | 0.080 | 5 | 0.670 | 0.750 | 16 | 0.650 | 0.900 | AY220278 |
| CM48 | $(CA)_8$ | F: GGTCCCACTCTCAAGCTGAA<br>R: GGCAATTGATTCTGGAACCT | 59 | 139–157 | 2† | 0.000 | 0.000 | 2 | 0.000 | 0.440 | 8 | 0.370 | 0.820 | AY220279 |

$N_a$, number of alleles; $H_O$, obseved heterozygosity; PIC, polymorphism information content — it is equal to '$H_E$', the expected heterozygosity; nc, not calculated.
*All forward primers were labeled at the 5′-end with fluorescent dye 6-FAM.
†Indicated to be two homologous copies.
‡One fixed and the other showing quantitative segregation.
§Data pertains to 17 selected species of coffee detailed in Table 2.

 

5th vol. of Nucleon resin from the Phytopure Plant DNA isolation kit (Amersham Biosciences) for 2 min followed by centrifugation at 2500 × *g*. Approximately 10 μg of genomic DNA was digested with *Sau*3A1 restriction endonuclease (NEB) and fractionated in 1% agarose gel. Genomic fragments of 500–1500 bp were gel-excised, purified using the GeneClean kit (BIO-101, Inc.), ligated to *Bam*HI-digested, dephosphorylated pBluescriptKS$^{(+)}$ plasmid vector using T4 DNA-ligase, and cloned in *Escherichia coli* DHα10B host cells by electroporation. The transformed cells were grown overnight and recombinant white colonies were individually picked up and maintained in 96- and 384-well microtiter culture plates, and replicated onto Hybond-N$^{+}$ nylon membranes (Amersham Biosciences) to obtain high-density hybridization filters for screening. Approximately 8000 arrayed recombinant clones were Southern hybridized to $\gamma^{32}$P-labelled synthetic oligonucleotides, viz. $(CA)_{15}$, $(GA)_{15}$, $(CAA)_{10}$, $(GGT)_{10}$, $(ATT)_{10}$, $(AGG)_{10}$, $(AGA)_{10}$, $(ACT)_{10}$ and $(CATA)_{8}$, in three combinations, to isolate short-sequence repeat (SSR) rich clones.

Eighty clones were identified and plasmids extracted by the alkaline-lysis method. Coffee specific fragments were then amplified from the recombinant plasmids using M13 universal primers and sequenced for both strands on an automated ABI-PRISM 3700 DNA sequencer using the Big Dye terminator cycle sequencing kit (Applied Biosystems). The sequences having SSR motifs were identified using the online Simple Sequence Repeat Identification tool (http://www.gramene.org/gramene/searches/ssrtool). Primers were designed for 16 sequences having > 16 bp long SSR regions, using the online Primer3 program (http://www-genome.wi.mit.edu/genome_software/other/primer3.html), and synthesized on DNA/RNA Oligosynthesizer-394 (Applied Biosystems, USA).

Initially, all primer pairs were used to standardize the PCR-amplification conditions and then the working pairs were tested for their utility as potential genetic markers. For the later purpose, microsatellite markers were tested on 45 elite genotypes/accessions of *C. arabica*, six of *C. canephora* (robusta) and 17 species of *Coffea* and related genera *Psilanthus*. PCR amplifications were carried out in 15-μL reactions [containing: 10 ng genomic DNA, 1 pmole of each primer, 2 mM $MgCl_2$, 100 μM dNTPs, 1 × PCR buffer-II and 1 unit AmpliTaq Gold DNA polymerase (Applied Biosystems)] on a PTC-200 Thermal-Cycler (MJ Research), using a five-step program comprising: initial denaturation at 94 °C for 10 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing at 55–60 °C (Table 1) for 1 min, extension at 72 °C for 1.5 min and final extension at 72 °C for 5 min. All amplification products were checked on 1.5% agarose gel before resolving through GeneScan analysis on ABI-377 DNA sequencer (Applied Biosystems) as per the manufacturer's instructions. Finally, the visualized microsatellite alleles were precisely sized using the software Genotyper 2.1 (Applied Biosystems) to calculate the number, range, and distribution of amplified alleles. PIC (polymorphism information content) values were calculated (http://www.agri.huji.ac.il/~weller/Hayim/parent/PIC.htm) and cross-species specificities were tested for all the working microsatellites.

Nine of the 16 primer pairs could be validated as useful microsatellite markers. The details of these new markers, namely locus designation, primer sequences, repeat motifs, allele attributes, PIC estimates and GenBank Accession number, are summarized in Table 1, whereas their cross-species status is shown in Table 2. The data revealed very low polymorphism across 45 arabica genotypes (only two to three alleles and 0–0.08 PIC for five of the nine microsatellites) compared to the six *C. canephora* accessions (one to six alleles and 0.40–0.78 PIC for eight of the nine microsatellites) used in the study. These results are in agreement with earlier studies supporting a narrow genetic base of the autogamous *C. arabica* genepool and considerable variability in the self-incompatible robustas. Moreover, allelic data for two microsats (CM11, CM48) suggested the possibility of duplicated loci in arabica genotypes, which possibly represent the homologous copies specific to the two progenitor genomes supporting an allo-tetraploid status of the arabica. Microsat alleles at the remaining seven loci (CM2, CM3, CM5, CM6, CM8, CM16, CM17) showed significant departures from Hardy–Weinberg equilibrium (tested using Markov chain algorithm and Fisher's exact test in ARLEQUIN version 2.0; Schneider *et al*. 2000), among the 45 arabica samples. This was expected as the analysed arabica samples represented unrelated accessions/genotypes and not a random mating population.

Most interestingly, studies with related taxa revealed highly robust amplifications in most of the species of *Coffea* and *Psilanthus* (Table 2) for each of the developed markers with very high level of polymorphism (8–16 alleles and 0.67–0.90 PIC over the nine microsatellites, Table 1). These findings suggest the potential of the markers for genetic analysis of coffee, especially its secondary genepool for conservation, management, resolving taxonomic relationships, and even more importantly for their use as efficient, informative genetic landmarks on interspecific molecular linkage maps being developed for genetic improvement of coffee.

## Acknowledgements

The authors thank the Department of Biotechnology, New Delhi, India for the financial support, Dr Lalji Singh, Director, CCMB, Hyderabad for the facilities to undertake the study, Dr R Naidu, Director Research Coffee Board, Banglore and Drs CS Srinivasan, N SuryaPrakash, A Santaram, D Ganesh, HL Sreenath of the Central Coffee Research Institute, Chikamagalur for the plant materials, Dr Anju Singh for helping with the genomic library and Ms. Mehar Sultana for primer synthesis.

**Table 2** Cross-species amplification status using microsatellite markers of *Coffea arabica* developed in the study

| Related species | Coffee type | Geographical distribution | Microsatellite marker | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | CM2 | CM3 | CM5 | CM6 | CM8 | CM11 | CM16 | CM17 | CM48 |
| *C. canephora* | E | Ethiopia | 191, 205 | 204 | 194 | 216 | 183 | 232 | 101, 111 | 346, 364 | 157 |
| *C. congenesis* | E | WCA | 205 | 204, 206 | 194, 215 | 210 | 167, 173 | 234 | 99, 101, 109 | 358, 366 | 157 |
| *C. eugenoides* | M | CA | 211 | 198, 200 | 167, 191 | 216, 228 | 159 | 234 | 101, 111 | 348 | 147 |
| *C. dewevrei* | P | CA | 211 | 206, 210 | 194, 215 | 216, 222 | 169, 175 | 234 | 99, 109 | 348, 352 | 157 |
| *C. stenophylla* | Me | WA | 227 | — | 209 | 214, 226 | 167 | 234 | 113, 129 | 338 | 149 |
| *C. excelsa* | P | WA | 183, 215 | 218 | 206, 212 | 222, 230 | 169, 179 | 246 | 105 | 346, 356 | 145, 151 |
| *C. liberica* | P | WCA | 227 | 200, 204 | 194, 215 | 216, 222 | 183 | 234 | 111, 115 | 356, 362 | 157 |
| *C. abeokutae* | P | WCA | 183, 229 | 204 | 194, 206 | 230 | 169, 179 | 240, 248 | 105, 109 | 346, 354 | 145 |
| *C. recemosa* | M | EA | 211 | 198, 200 | 191 | 214, 228 | 167 | 234 | 99, 103 | 342, 348 | 149 |
| *C. salvatrix* | M | EA | 207, 211 | — | 176 | 212 | 159 | 222 | 95 | 346 | — |
| *C. arnoldiana* | P | CA | 181, 205 | 222 | 212 | 230 | 171 | 236, 254 | 125, 127 | 346 | 151 |
| *C. aruwemiensis* | P | CA | 205, 217 | 218 | 188, 212 | 226, 228 | 167, 175 | 236, 254 | 111, 113 | 342, 356 | 151 |
| *C. kapakata* | M | CA | 177, 211 | 198, 200 | 167, 212 | 218 | — | 234 | 95, 99, 105 | 342, 348 | 147, 149 |
| *P. khasiana* | Pa | India | 195 | 196 | 203, 227 | 222 | 183 | 232, 240 | 111 | 308, 390 | 139, 149, 155 |
| *P. wightiana* | Pa | India | — | — | 227, 230 | 218, 228 | — | — | — | 328, 334 | 145, 149 |
| *P. bengalensis* | Pa | India | 211 | — | 224 | — | — | — | — | 308 | 137, 139 |
| *P. travencorensis* | Pa | India | 195 | 196 | 167 | 222 | 183 | 230 | 99, 111 | 320 | 145, 149 |

One representative sample was used for each species; the allelic diversity estimates across species given in Table 1.
E, Erythrocoffea; M, Mozambicoffea; P, Pachycoffea; Me, Melanocoffea; Pa, Paracoffea.
WCA, West and central Africa; CA, Central Africa; EA, East Africa; WA, West Africa.
—, no amplification.

## References


Aggarwal RK, Shenoy VV, Ramadevi J, Rajkumar R, Singh L (2002) Molecular characterization of some Indian Basmati and other elite rice genotypes using fluorescence-AFLP. *Theoretical and Applied Genetics*, **105**, 680–690.

Combes MC, Andrzejewski S, Anthony F *et al.* (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Molecular Ecology*, **9**, 1171–1193.

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

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

Rovelli P, Mettulio R, Anthony F *et al.* (2000) Microsatellites in *Coffea arabica* L. In: *Coffee Biotechnology and Quality* (eds Sera T, Soccol CR, Pandey A, Roussos S), pp. 123–133. Kluwer Academic Publishers.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin ver. 2000: a Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory. University of Geneva, Switzerland.