

## Microsatellite DNA fingerprinting of *Coffea* sp. germplasm conserved in Costa Rica through singleplex and multiplex PCR

Elodia Sánchez<sup>1</sup>, William Solano<sup>2</sup>, Andrés Gatica-Arias<sup>3</sup>, Max Chavarría<sup>1,4</sup> and Emanuel Araya-Valverde<sup>1,5\*</sup>

Crop Breeding and Applied Biotechnology  
20(1): e27812013, 2020  
Brazilian Society of Plant Breeding.  
Printed in Brazil  
<http://dx.doi.org/10.1590/1984-70332020v20n1a3>

**Abstract:** A large collection of coffee genetic resources is conserved in Costa Rica. In this study, microsatellite DNA fingerprinting of coffee through singleplex and multiplex PCR approaches coupled with capillary electrophoresis are described. To validate both methods, germplasm of *Coffea* spp. (*Arabica* and non-*Arabica*) and intraspecific F1 hybrids were analyzed using fourteen microsatellite markers. It was observed that through both PCR methods the fingerprinting profile of a subset of samples was identical. The genetic analyses revealed that non-*Arabica* coffee displayed greater genetic variation than *Arabica* coffee did. In addition, microsatellite analyses allowed the separation of *C. arabica* from other species using the principal coordinate analysis (PCoA) approach. The neighbor-joining tree clustering analysis revealed either a grouping of wild genotypes separated from cultivars of *C. arabica*, or a relation of intraspecific F1 hybrids with parental lines. The utility of our methodology for the characterization of F1 hybrids not previously analyzed through SSR (Simple Sequence Repeats) fingerprinting is demonstrated.


**Keywords:** Capillary electrophoresis, cultivars, F1 hybrids, SSR

### INTRODUCTION

Of the *Rubiaceae* family, two species are mainly responsible for the world coffee production: *Coffea arabica* L. and *Coffea canephora* Pierre ex A. Froehner, commonly known as *Arabica* and *Robusta* coffee, respectively. *C. canephora* is a diploid species ( $2n = 2x = 22$ ) and *C. arabica* is a tetraploid species ( $2n = 4x = 44$ ) derived from the hybridization of *C. canephora* and *C. eugenioides* (Lashermes et al. 1999). *C. arabica* is indigenous to Ethiopia and was first cultivated in Yemen in the seventeenth century. Until the middle of the twentieth century, cultivated coffee grown in Latin America shared the same genetic base of Yemen plants. Coffee plants introduced to Latin America are believed to be *C. arabica* var. 'Typica' and *C. arabica* var. 'Bourbon' (Anthony et al. 2002). In Costa Rica, coffee was introduced at the end of the 18th century; the first seeds are believed to have come from 'Typica' from Martinique Island (ICAFE 2015). These events of introduction and domestication, as well as the allotetraploid origin, reproductive behavior, and evolution have contributed to a narrow genetic base of coffee, as demonstrated with molecular marker analyses (Lashermes et al. 1999, Anthony et al. 2001, Zhou et al. 2016, Sousa et al. 2017).

\*Corresponding author:

E-mail: earaya@cenat.ac.cr

 ORCID: 0000-0002-0025-5420

Received: 21 November 2018

Accepted: 30 August 2019

Published: 13 February 2020

<sup>1</sup> Centro Nacional de Innovaciones Biotecnológicas, 1174-1200, San José, Costa Rica

<sup>2</sup> Centro Agronómico Tropical de Investigación y Enseñanza, 7170, Turrialba, Cartago, Costa Rica

<sup>3</sup> Laboratorio Biotecnología de Plantas, Escuela de Biología, Universidad de Costa Rica, 2060, San Pedro, Costa Rica

<sup>4</sup> Escuela de Química & Centro de Investigaciones en Productos Naturales, Universidad de Costa Rica, 11501-2060, San José, Costa Rica

<sup>5</sup> Escuela de Biología, Instituto Tecnológico de Costa Rica, 159-7050, Cartago, Costa Rica

In the period between October 2017 and June 2019, 62–65 % of the global coffee production by exporting countries came from *C. arabica* (International Coffee Organization 2017). More than half of the production of coffee in the world is supplied by Latin America, with Brazil and Colombia being the major producers. Costa Rica is responsible for about 1% of the world coffee production (International Coffee Organization 2017). Breeding efforts in Costa Rica have hence been developed through national and international participation of the Costa Rican Coffee Institute (ICAFFE), the Tropical Agricultural Research and Higher Education Center (CATIE), the Regional Cooperative Program for the Technological Development and Modernization of Coffee Cultivation (PROMECAFE), and the French Agricultural Research Centre for International Development (CIRAD). Some of this work was undertaken to obtain F1 intraspecific hybrids derived from crosses between wild Sudan-Ethiopian accessions of *C. arabica* (ET6, ET15, ET25, E41, E416, E531, Anfilo, and Rume Sudan) and American cultivars of *C. arabica* (Caturra, Catuaí, T5296). The hybrids tested in Central America showed stability, high yields and satisfactory beverage quality (Bertrand et al. 2006, Bertrand et al. 2011).

Conventional breeding efforts have included strategies such as hybridization, selection by interspecific crossings and backcrossing to transfer resistance to biotic stresses, and improve adaptation and the yield (Silva et al. 2019, Shigueoka et al. 2014). The lack of reproductive precision, the differences in ploidy levels between *C. arabica* and other diploid species, and their incompatibility are limitations associated with conventional coffee breeding. Another limitation that hinders breeding programs is the selection of genetically diverse parental lines for hybridization and the identification of hybrids at an early stage (e.g. seedlings), based on morphological characteristics (Mishra and Slater 2012). A proper identification of germplasm for breeding and conservation purposes requires the development of tools that can accelerate and provide reliability in the characterization of germplasm, to increase the efficiency of coffee breeding programs (Hendre et al. 2008). In this context, several molecular markers have been employed for coffee evaluation; they include RFLP (Lashermes et al. 1999), RAPD (Anthony et al. 2001), AFLP (Anthony et al. 2002), simple sequence repeats (SSR) (Missio et al. 2011, Ogotu et al. 2016), and single-nucleotide polymorphism approaches (SNP) (Zhou et al. 2016). Molecular markers have also served for linkage mapping and quantitative trait loci (QTL) analyses (Moncada et al. 2016), and association studies of SNP with caffeine content (Tran et al. 2018). Despite the fact that next-generation sequencing (NGS) technologies lead to an increasing use of SNP markers for several of the applications mentioned above, microsatellites are still a viable option (Flanagan and Jones 2019).

Microsatellites exhibit several advantages that make them suitable for the characterization of plant genetic resources, including coffee species. Some advantages are high polymorphism, reproducibility, relatively easy scoring, and codominance. Microsatellite-based analyses are also amenable for automation through the use of genetic analyzers with capillary electrophoresis (CE) coupled with laser-induced DNA fluorescence, which allows the combination of disparate microsatellites (Butler et al. 2004). Microsatellites are prone to true multiplexing that decreases the duration and cost of microsatellite genotyping (Guichoux et al. 2011). Regarding the use of multiplex PCR for microsatellite analysis in coffee, to the best of our knowledge, only Aerts et al. (2013) used this approach for population genetics studies in *C. arabica*; however, the method was not described in detail. Therefore, the aim of the present study was to provide a detailed procedure for microsatellite fingerprinting through both singleplex and multiplex PCR coupled with capillary electrophoresis. This analysis can be alternatively or concomitantly applied with other marker systems to characterize the genetic diversity of coffee germplasm and its genetic relationships, including F1 intraspecific hybrids that were not previously reported.

## **MATERIAL AND METHODS**

### **Plant material and DNA isolation**

Leaf samples of 46 genotypes were collected (Table 1). Cultivars and F1 intraspecific hybrids were collected from the germplasm bank of ICAFFE (lat 10° 01' 51" N, long 84° 08' 23" W, alt 1170 m asl). Genotypes of wild germplasm were collected from CATIE (lat 9° 53' 44" N, long 83° 38' 7" W, alt 602 m asl). Samples also included three non-Arabica species (*C. canephora*, *C. excelsa*, and *C. liberica*). For DNA isolation, leaves (40 mg) were desiccated in silica gel (Merck, Darmstadt, Germany), then transferred into a tube (2.0 mL), and ground with a micropistol; a CTAB (cetyl trimethylammonium bromide)-based method (Doyle and Doyle 1990) with modifications described by Pérez et al. (2017) was employed. DNA concentration was measured with a spectrophotometer (NanoDrop 2000, Thermo Scientific, Delaware, USA), and diluted (~50 ng  $\mu\text{L}^{-1}$ ) for PCR amplification.

**Table 1.** List of 46 genotypes of *Coffea* spp. used for microsatellite fingerprinting with singleplex and multiplex PCR approaches and for genetic diversity analyses

Genotype code	Cultivar Description	Genotype code	Description
BA-10	Cultivar: CATIE / T2694	E-478	Wild: CATIE / T4551
Bourbon <sup>2,4</sup>	Cultivar: <i>Coffea arabica</i> L. (Bourbon rojo)	E-531	Wild: CATIE / T4900
Catigua MG2 <sup>3,4</sup>	Cultivar: Catuaí x Timor Hybrid	ET-03	Wild: CATIE / T16691
Catuaí	Cultivar: Caturra x Mundo Novo	ET-25	Wild: CATIE / T16712
Catuaí amarillo <sup>1,4</sup>	Cultivar: Caturra x Mundo Novo	ET-41	Wild: CATIE / T16725
Caturra	Cultivar: <i>Coffea arabica</i> L. (bourbon mutation)	Rume Sudan 1	Wild: CATIE / T2724
Catuaí SH3	Cultivar: Catuaí x S288	Rume Sudan 2	Wild: CATIE / T2744
PDRY-14	Cultivar: CATIE / T21239	L02A30	Hybrid: Caturra9 x ET15
PDRY-22	Cultivar: CATIE / T21242	L04A20	Hybrid: CR 95 X Rume Sudan
S288 <sup>3,4</sup>	Cultivar: Pureline from S 26 ( <i>C. arabica</i> x <i>C. liberica</i> )	L04A42	Hybrid: Catuaí10 x ET 41
SL-28	Cultivar: CATIE / T2739	L09A22	Hybrid: Caturra7 x Amphilo
Typica	Cultivar: <i>Coffea arabica</i> L	L10A25	Hybrid: Catuaí10 x Rume Sudan
Venecia	Cultivar: <i>Coffea arabica</i> L. (bourbon selection)	L11A26	Hybrid: Caturra7 x ET41
Villa Sarchí <sup>1,4</sup>	Cultivar: <i>Coffea arabica</i> L. (bourbon mutation)	L13A12	Hybrid: T5296 x ET06
Catimor <sup>2</sup>	Cultivar: Catuaí x Timor Hybrid	L04A05	Hybrid: T5296 x ET25
IAPAR 59 <sup>1,4</sup>	Cultivar: Villa Sarchí x Timor Hybrid	L12A28	Hybrid: T5296 x Rume Sudan
Oeiras MG6851 <sup>2</sup>	Cultivar: Caturra x Timor Hybrid	L13A22	Hybrid: Caturra7 x E531
T5175 <sup>2,4</sup>	Cultivar: Caturra x Timor Hybrid	L13A44	Hybrid: T5296 x Rume Sudan
T5296 <sup>1,4</sup>	Cultivar: Villa Sarchí x Timor Hybrid	L4A34	Hybrid: Caturra x x ET41
Tupí RN <sup>2,4</sup>	Cultivar: (Villa Sarchí x Timor Hybrid)	L5A26	Hybrid: Caturra 7 x E531
Arussii S-11	Wild: CATIE / T3099	<i>Coffea canephora</i> <sup>3,4</sup>	Robusta
E-044	Wild: CATIE / T4492	<i>Coffea excelsa</i> <sup>4</sup>	<i>Coffea excelsa</i>
ET-44	Wild: CATIE / T16728	<i>Coffea liberica</i> <sup>3,4</sup>	<i>Coffea liberica</i>

<sup>1</sup> samples of DNA pool 1; <sup>2</sup> samples of DNA pool 2; <sup>3</sup> samples of DNA pool 3; <sup>4</sup> samples used in the multiplex PCR assays.

## Screening and selection of microsatellites

Twenty-seven microsatellites were initially chosen, taking into account the cross-species amplification in *Coffea* sp. and the highest polymorphic information content (PIC) as described previously (Combes et al. 2000, Baruah et al. 2003, Poncet et al. 2004, Aggarwal et al. 2007, Hendre et al. 2008, Missio et al. 2009, Missio et al. 2010, Sousa et al. 2017). PCR amplification was tested over the DNA of 14 genotypes that were distributed in three DNA pools (Table 1). PCR reaction mixes (final volume 25 mL) contained DNA (~100 ng) and 1× DreamTaq PCR Master Mix (Thermo Scientific, Delaware, USA), with MgCl<sub>2</sub> (2 mM) and primers (0.2 μM each). The amplification was performed in a thermal cycler (Veriti®, Applied Biosystems, California, USA) with an initial denaturation step at 95 °C for 5 min, followed by 35 cycles at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, and a final extension at 72 °C for 8 min. PCR products were separated by polyacrylamide gel electrophoresis (10% denatured, DCode™, Biorad, California, USA) at constant voltage (200 V) for 6 h. Fourteen of the 27 microsatellites tested showed polymorphisms between the three DNA pools, and were selected for further singleplex PCR, multiplex PCR, and capillary electrophoresis.

## Singleplex and multiplex PCR

Singleplex PCR reactions were performed over the 46 genotypes described in Table 1 with 14 microsatellites. PCR reactions and thermal cycling conditions were the same as described previously with optimized primer concentrations (Table 2). For the multiplex PCR assays, the DNA of 12 samples were used (Table 1). Four multiplex PCR reactions were performed for 11 of the 14 SSR. Each multiplex PCR reaction (final volume 25 μL) contained 1X DreamTaq PCR buffer (Thermo Scientific, Delaware, USA), MgCl<sub>2</sub> (2 μM, Thermo Scientific, Delaware, USA), dNTPs (80 μM each, Thermo Scientific, Delaware, USA), DNA polymerase (DreamTaq, 1 Unit, Thermo Scientific, Delaware, USA), bovine serum albumin (0.01 mg, Sigma, Missouri, USA), and DNA (~75 ng). The primer concentrations were different (Table 2). The thermal cycler (Veriti, Applied Biosystems, California, USA) settings were the same as for singleplex PCR. For SSR03, M32, and M764, singleplex PCR reactions were performed as previously described for singleplex PCR reactions

**Table 2.** Primer concentration of 14 microsatellites used for singleplex PCR (SiPCR), multiplex PCR (MuPCR) and the ratio of four pseudo-multiplex for capillary electrophoresis (RCE) made from singleplex PCR products

Microsatellite	Forward primer	Reverse primer	SiPCR / $\mu\text{M}$	RCE	Multiplex PCR assigned	MuPCR / $\mu\text{M}$
SSR04	NED- GGTCGCTCACTCATATCTCCAG	GCCTGGAAAGCAAACGTCTCA	0.20	1 : 1	2	0.2
SSRCa87	6FAM-TCACTCTCGCAGACACACTAC	GCAGAGATGATCACAAGTCC	0.15		1	0.1
CaM03	NED-CGCGCTTGCTCCCTCTGTCTCT	TGGGGGAGGGGCGGTGTT	0.15	1 : 1.25 : 1 : 1 : 0.75	4	0.2
M753	VIC-GGAGACGCAGGTGGTAGAAG	TCGAGAAGTCTTGGGGTGT	0.20		2	0.2
M774	PET-GCCACAAGTTTCGTGCTTTT	GGGTGTCGGGTAGGTGTATG	0.20		1	0.2
SSRCa18	6FAM-GTCTCGTTTCACGCTCTCTC	ATTTTTGGCACGGTATGTTT	0.20		3	0.2
SSRCa88	6FAM-TACCTCTCCTCCTCCTCCT	ATTTCTATGGACCGGCAAC	0.15	0.5 : 0.75 : 1 : 0.75	4	0.2
CaM16	VIC-AAGGCAGCTGAAGCGGGACAAA	TGGGGAGAGCTGCAGTTGGAGG	0.14		3	0.03
M24	VIC-GGCTCGAGATATCTGTTTAG	TTTAATGGGCATAGGGTCC	0.20		2	0.2
M764	NED-CTGGCATTAGAAAGCACCTTG	GCTTGGTCTACTGTAGGACTG	0.15		3*	---
SSR09	6FAMTTGGCTTTTGTCCCTCCTCTCTG	AGCCCATTTCCCTCTCATATTCAAG	0.14	1 : 1 : 1	4	0.1
M32	NED-AACTCTCCATTCCCGCATT	CTGGGTTTTCTGTGTCTCG	0.20		4*	---
SSR03	VIC-GGACAAAACCCGCCAAAATA	AGCGAGACAGAGGAAGGGAATATT	0.20		1*	---
SSR073	PET-GAGGTCTTCCACCACAACA	GGATACGAGAGTCCCTTCC	0.20		1	0.2

The capillary electrophoresis of singleplex PCR products were performed in four pseudo-multiplexes with optimized ratios (Table 2). For both pseudo-multiplexes and multiplex PCR products, a volume (1.5  $\mu\text{L}$ ) was combined with 0.4  $\mu\text{L}$  GeneScan 600LIZ (Applied Biosystems, California, USA) and 8.5  $\mu\text{L}$  Hi-Di formamide (Applied Biosystems, California, USA). The PCR products (1  $\mu\text{L}$ ) of SSR03, M764, and M32 (from the singleplex PCR) was added to multiplex PCR 1, 3, and 4, respectively (Table 2). Capillary electrophoresis was performed in a genetic analyzer (ABI3130xl, Applied Biosystems, California, USA).

### Data analysis

After capillary electrophoresis, allele binning and genotyping was manually inspected with the GeneMapper V4.0 software (Applied Biosystems, California, USA). Using the same software, the molecular profile of 12 samples analyzed through multiplex PCR was compared with the profile generated from singleplex PCR amplification. We employed the tetraploid data set of microsatellites in the ATETRA software (Van Puyvelde et al. 2010) to estimate the following genetic diversity parameters: number of alleles, expected heterozygosity ( $H_e$ ), expected heterozygosity corrected for sample size ( $H_{e(c)}$ ), Shannon-Wiener diversity index ( $H'$ ), and genetic differentiation coefficients ( $G_{st}$  and  $R_{st}$ ). The polymorphic information content (PIC) and the pairwise genetic distance matrix were calculated with the R package Polysat 1.7 (Clark and Jasieniuk 2011). The distance matrix was used in a principal coordinate analysis (PCoA) to visualize the grouping among the studied species. The same matrix was used in a neighbor-joining analysis to construct a tree with only *C. arabica* (cultivars and wild genotypes); a second tree was built with F1 hybrids and parental lines.

## RESULTS AND DISCUSSION

### Genetic diversity analyses comparing singleplex and multiplex PCR

The molecular profile generated using the multiplex PCR protocol was the same as that generated with the singleplex PCR protocol. The approach followed in our study was similar to that previously proposed (Hill et al. 2009), which relies on a core set of markers that amplify without major optimization; other primers are then added. To obtain the identical profiles with both PCR approaches, extensive optimization steps were performed for primer concentrations of the microsatellites as well as for the ratio of the pseudo-multiplex PCR products for capillary electrophoresis (Table 2). As suggested by Sutton et al. (2011), the fluorescent dye was invariant for each SSR in both singleplex and multiplex PCR assays, which also aided obtaining identical profiles.

The total numbers of alleles (NA) amplified in the Arabica and non-Arabica groups were 56 and 49, while the average numbers of alleles were 4.00 and 3.50, respectively (Table 3). The average number of alleles per SSR was greater in the

Arabica group than previously reported, while similar or smaller for the non-Arabica group (Combes et al. 2000, Moncada and McCouch 2004, Aggarwal et al. 2007, Hendre et al. 2008, Sousa et al. 2017). Both the total number of alleles and the average number of alleles per SSR were influenced by the number of samples, their genetic background, the number of microsatellites employed, and the polymorphism level of each microsatellite. In order to evaluate the robustness of our methodology, we included F1 intraspecific hybrids that had not been previously included in microsatellite studies, which, added to cultivars and wild genotypes, contributed to an increased average number of alleles per SSR in the Arabica group.

The average values of PIC, He, He<sub>(c)</sub>, and H' index were higher in the non-Arabica group (Table 3). Similar to the study of Baruah et al. (2003), our results demonstrate little polymorphism across the Arabica genotypes. Moncada and McCouch (2004) also found that polymorphism levels were significantly higher among diploid or wild tetraploid species of coffee than within cultivated Arabica coffee. The Shannon index of diversity (H') has also been reported to be greater in non-Arabica genotypes (Anthony et al. 2001, Moncada and McCouch 2004). Higher polymorphisms in non-Arabica coffee have been attributed to their cross-breeding process, while the Arabic coffee are self-compatible. *C. arabica* will always present less polymorphism than other coffee species due to the evolution of its genome and domestication process (Hendre et al. 2008, Missio et al. 2010).

Another parameter estimated to evaluate de suitability of the 14 microsatellites was the index of gene differentiation for multiple alleles (Gst). The average of Gst was 0.1494 with the lowest value for CaM16 (0.0416), and the highest value for M774 (0.4466). According to the classification of differentiation coefficients described by Wright (1978), only microsatellite CaM16 had a small differentiation coefficient; seven microsatellites (SSR09, M764, SSRCa88, SSR03, M32, SSR073, and M753) had moderate differentiation, and six (SSRCa18, SSR04, SSRCa87, CaM03, M24, and M774) scored high differentiation (Table 3). In addition, the Rst differentiation coefficient (Slatkin 1995) was calculated, with an average value of 0.3958. The lowest value (0.0875) was also found for CaM16, and the highest value (1.6419) was for M774 (Table 3). The values of the genetic differentiation coefficient Rst for all SSR were always greater than Gst (Table 3). The Rst coefficient treats the stepwise mutation model that is thought to reflect more accurately the mutation pattern of microsatellites, which tends to increase the Rst value (Balloux and Lugon-Moulin 2002). The 14 microsatellites analyzed for our study proved to be efficient, not only for the characterization of genetic diversity, but also for germplasm differentiation as described in the following sections.

**Table 3.** Genetic diversity parameters estimated for 14 SSR in Arabica and non-Arabica genotypes. Gst and Rst are genetic differentiation coefficients, NA: number of alleles, He: Expected heterozygosity, He(C): Corrected expected heterozygosity, H': Shannon diversity index, PIC: Polymorphic Information Content

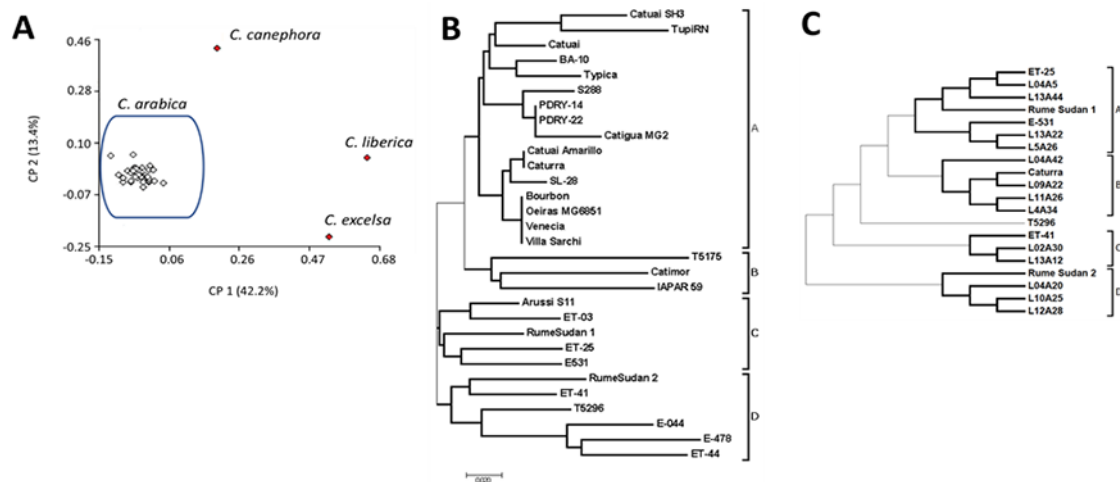
Microsatellite	Allele size range /bp	Arabica (n=43)							Non-arabica (n=3)				
		G <sub>st</sub>	R <sub>st</sub>	NA	He	He <sub>(c)</sub>	H'	PIC	NA	He	He <sub>(c)</sub>	H'	PIC
CaM03	161-188	0.1960	0.4887	10	0.689	0.693	1.486	0.698	4	0.648	0.707	1.194	0.67
CaM16	184-200	0.0416	0.0875	4	0.571	0.575	0.966	0.484	4	0.704	0.768	1.287	0.66
M24	149-169	0.2268	0.5922	7	0.402	0.404	0.894	0.5	4	0.648	0.707	1.194	0.67
M32	99-131	0.1080	0.2423	6	0.674	0.678	1.352	0.56	2	0.491	0.535	0.839	0.41
M753	283-296	0.1232	0.2845	3	0.548	0.551	0.864	0.461	5	0.548	0.551	0.864	0.75
M764	148-157	0.0748	0.1637	2	0.498	0.501	0.691	0.41	4	0.648	0.707	1.195	0.67
M774	193-224	0.4466	1.6419	2	0.498	0.501	0.691	0.41	2	0.268	0.293	0.433	0.37
SSR03	134-156	0.0969	0.2150	5	0.701	0.705	1.307	0.621	6	0.773	0.843	1.632	0.8
SSR04	116-148	0.1532	0.3626	2	0.664	0.668	1.095	0.41	2	0.602	0.657	0.992	0.39
SSR073	147-155	0.1263	0.2892	2	0.521	0.524	0.786	0.41	2	0.667	0.727	1.099	0.46
SSR09	118-126	0.0713	0.1551	3	0.552	0.555	0.901	0.422	4	0.639	0.697	1.185	0.54
SSRCa18	110-126	0.1575	0.3748	4	0.632	0.636	1.079	0.585	4	0.713	0.778	1.301	0.57
SSRCa87	116-135	0.1884	0.4646	4	0.618	0.622	1.052	0.59	3	0.667	0.727	1.099	0.46
SSRCa88	171-193	0.0813	0.1797	2	0.498	0.501	0.691	0.41	3	0.491	0.535	0.839	0.57
Average		0.1494	0.3958	4	0.576	0.580	0.99	0.498	4	0.608	0.66	1.082	0.57
Total				56					49				



### Genetic differentiation of *Coffea* spp. germplasm

A clear separation of Arabica and non-Arabica genotypes was possible employing the 14 microsatellites, as observed in the PCoA (Figure 1A). Similarly, a PCoA revealed disparate groups containing either accessions of diploid species of *Coffea* sp. or tetraploid *C. arabica* from Colombia, for which 34 microsatellites were used (Moncada and McCouch 2004). The clustering of Arabica cultivars and wild genotypes was clearer in the neighbor-joining tree, in which four groups were distinguished with most cultivars clustered in groups A and B (Fig. 1B). Wild genotypes and the commercial variety T5296 were clustered in groups C and D. Groups A and B contain genotypes that, according to World Coffee Research (2018), are classified as derived from Typica (Typica), Bourbon (Bourbon, Caturra, Villa Sarchí, Venecia, and SL-28), Bourbon-Typica (Catuai), and as introgressed germplasm (T5175, IAPAR59). Some genotypes analyzed in this study are derived from the Timor Hybrid (Catimor, Catigua MG2, T5175, Oeiras, MG6851, IAPAR 59, Tupi RN, and T5296), which grouped interspersed in groups A and B, and the genotype T5296 in group D. Our results indicate that samples derived from the Timor Hybrid share alleles with *C. arabica* from wild accessions and cultivars. The allelic diversity detected over the 14 microsatellites might indicate that alleles from the original hybrid have spread through the genotypes derived from the Timor Hybrid, which showed notable diversity and interspersed grouping (Setotaw et al. 2010). The differentiation between *Coffea* species, and between the wild and domesticated germplasm in our study demonstrate that the selected set of microsatellites seems useful for genetic characterization of coffee germplasm.

Despite cultivars grouped according to previous reports, some accessions had the same or similar genetic profile. In group A, genotypes PDRY-14 and PDRY-22 displayed the same fingerprinting profile for the 14 microsatellites, as well as Catuai Amarillo and Caturra. In addition, Bourbon, Oeiras MG6851, Villa Sarchí, and Venecia also had the same fingerprinting profile; they grouped near Catuai Amarillo and Caturra because of their highly similar DNA fingerprinting. Similarly to our results, Sousa et al. (2017) found that a set of 16 microsatellites was unable to distinguish cultivars derived from Catuai, which is included in the Bourbon-Typica group. It has also been reported that the use of 55 SNP markers did not differentiate between Caturra and Catuai (Zhou et al. 2016). These authors also found that Bourbon only differed from those of Caturra and Catuai by a single SNP, which indicated the origin of Caturra and Catuai as mutants or offspring derived from Bourbon.



**Figure 1.** Principal coordinate analysis displaying separation of (A) Arabica and non-Arabica genotypes obtained from a genetic distance matrix generated with 14 microsatellites. (B) Neighbor-joining trees for clustering of cultivars and wild germplasm of Arabica genotypes. Cultivars were clustered mainly in groups A and B. Wild genotypes and commercial variety T5296 were clustered in groups C and D. (C) Neighbor-joining tree for clustering of F1 intraspecific hybrids and parental lines. **Group A** displays the relation of parental lines (ET-25 and Rume Sudan 1) with hybrids (L04A5 and L13A44), and parental line E-531 with hybrids L13A22 and L5A26. **In group B**, hybrids L09A22, L11A26, and L4A34 are clustered with Caturra (as parental line). **Group C** comprises hybrids (L02A30 and L13A12) derived from Ethiopian accessions (ET-15 and ET-06 respectively). **Group D** displays the relation of hybrids (L04A20, L10A25, and L12A28) with Rume Sudan 2 (as a parental line).

## Grouping of F1 intraspecific hybrids

The DNA fingerprinting protocol for coffee germplasm analysis described in this study was used to produce genetic information of F1 intraspecific hybrids that had not been previously genotyped; consequently, we also enrich the existing knowledge of coffee genetic diversity. F1 intraspecific hybrids analyzed in this study are the result of crosses between a group of wild genotypes (Sudan-Ethiopian) and commercial varieties (Bertrand et al. 2006). Hybrids included in this study were vegetatively propagated and tested in Central America, demonstrating a yield earlier than and superior to traditional cultivars, greater stability across varied environments and high beverage quality provided by an intrinsic cup quality of ET parents (Bertrand et al. 2006, Bertrand et al. 2011). The NJ tree obtained from F1 intraspecific hybrids displayed four groups (Figure 1C). In group A, two subgroups were distinguished of which the parental lines ET-25 and Rume Sudan 1 are grouped with hybrids L04A5 (one parent is ET-25) and L13A44 (one parent is Rume Sudan). Another subgroup contained the parental line E-531 that was used to generate L13A22 and L5A26. The genotype Caturra is clustered in group B with hybrids L09A22, L11A26, and L4A34, which are derived from Caturra. Despite being derived from Catuaí and ET-41, L04A42 was grouped in B because it shares alleles from ET-41 that are also present in L11A26 and L4A34. Group C comprises hybrids L02A30 and L13A12, which derive from Ethiopian accessions (ET-15 and ET-06, respectively). Despite ET-41 (grouped in C) not being a parental line of L02A30 and L13A12, it might share alleles with ET-15 and ET-06, as previously described for the clustering of these three Ethiopian accessions (Anthony et al. 2001). Within group D, hybrids L04A20, L10A25, and L12A28 are derived from Rume Sudan, of which sample Rume Sudan 2 was also clustered. Sample T5296 was unclustered between groups A, B and C, D. T5296 is a parental line of hybrids in groups A (L04A5, L13A44), C (L13A12), and D (L12A28) that share alleles favored as an intermediate position in the NJ tree. These data constitute the first microsatellite analysis of F1 intraspecific hybrids of coffee. Coffee hybrids are an alternative for a profitable and sustainable production of coffee, and molecular tools might serve as a complementary descriptor for the identification, registration, and protection of new and improved coffee cultivars, particularly of those vegetatively propagated including hybrids.

In summary, our study describes a detailed microsatellite DNA fingerprinting protocol easily implementable, with low cost (mainly if multiplex PCR is used), and capable of classifying coffee germplasm. The scope of this protocol was demonstrated with the processing of F1 intraspecific hybrids that had never been genotyped by microsatellites before, so that through the results reported in this study we gain further insights into the genetic diversity of coffee.

## ACKNOWLEDGEMENTS

We thank Alejandra Robles at ICAFE for kindly providing samples for microsatellite analyses. This study was financially supported by the scholarship program of Centro Nacional de Alta Tecnología-Consejo Nacional de Rectores (CENAT-CONARE). The authors declare that they have no conflict of interest and have approved the final version of the manuscript. The genotype/marker table is available upon reasonable request to the corresponding author.

## REFERENCES

- Aerts R, Berecha G, Gijbels P, Hundera K, Van Glabeke S, Vandepitte K, Muys B, Roldán-Ruiz I and Honnay O (2013) Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian montane rainforests. **Evolutionary Applications** 6: 243-252.
- Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishakumar 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.
- Anthony F, Bertrand B, Quiros O, Wilches A, Lashermes P, Berthaud J and Charrier A (2001) Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. **Euphytica** 118: 53-65.
- 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.
- Balloux F and Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. **Molecular Ecology** 11: 155-165.
- 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.
- Bertrand B, Alpizar E, Lara L, SantaCreo R, Hidalgo M, Quijano JM, Montagnon C, Georget F and Etienne H (2011) Performance of *Coffea arabica* F1 hybrids in agroforestry and full-sun cropping systems in comparison with American pure line cultivars. **Euphytica** 181: 147-158.
- Bertrand B, Vaast P, Alpizar E, Etienne H, Davrieux F and Charmetant P (2006) Comparison of bean biochemical composition and beverage

- quality of Arabica hybrids involving Sudanese-Ethiopian origins with traditional varieties at various elevations in Central America. **Tree Physiology** **26**: 1239-1248.
- Butler JM, Buel E, Crivellente F and McCord BR (2004) Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. **Electrophoresis** **25**: 1397-1412.
- Clark LV and Jasieniuk M (2011) Polysat: An R package for polyploid microsatellite analysis. **Molecular Ecology Resources** **11**: 562-566.
- 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.
- Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. **Focus** **12**:13-15.
- Flanagan SP and Jones AG (2019) The future of parentage analysis: From microsatellites to SNPs and beyond. **Molecular Ecology** **28**: 544-567.
- Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F and Petit RJ (2011) Current trends in microsatellite genotyping. **Molecular Ecology Resources** **11**: 591-611.
- 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**: 1-19
- Hill CR, Butler JM and Vallone PM (2009) A 26plex autosomal STR assay to aid human identity testing. **Journal of Forensic Sciences** **54**: 1008-1015.
- ICAFFE - Instituto del Café de Costa Rica (2015) Historia del café de Costa Rica. ICAFFE. Available at <<http://www.icafe.go.cr/>>. Accessed on May 10, 2019.
- International Coffee Organization (2017) Trade statistics tables. International Coffee Organization. Available at <[http://www.ico.org/trade\\_statistics.asp](http://www.ico.org/trade_statistics.asp)>. Accessed on Aug 8, 2019
- 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.
- Mishra MK and Slater A (2012) Recent advances in the genetic transformation of coffee. **Biotechnology Research International** **2012**: 1-17.
- Missio R, Caixeta E, Zambolin E, Zambolin L, Cruz C and Sakiyama N (2010) Polymorphic information content of SSR markers for *Coffea* spp. **Crop Breeding and Applied Biotechnology** **10**: 89-94.
- Missio RF, Caixeta ET, Zambolim EM, Pena GF, Zambolim L, Dias LAS and Sakiyama NS (2011) Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers. **Genetics and Molecular Research** **10**: 2366-2381.
- Missio RF, Caixeta ET, Zambolim EM and Sakiyama NS (2009) Development and validation of SSR markers for *Coffea arabica* L. **Crop Breeding and Applied Biotechnology** **9**: 361-371.
- Moncada P and McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome** **47**: 501-509.
- Moncada MDP, Tovar E, Montoya JC, González A, Spindel J and McCouch S (2016) A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size. **Tree Genetics and Genomes** **12**: 1-17.
- Ogutu C, Fang T, Yan L, Wang L, Huang L, Wang X, Ma B, Deng X, Owiti A, Nyende A and Han Y (2016) Characterization and utilization of microsatellites in the *Coffea canephora* genome to assess genetic association between wild species in Kenya and cultivated coffee. **Tree Genetics and Genomes** **12**: 54.
- Pérez J, Araya-Valverde E, Garro G and Abdelnour-Esquivel A (2017) Analysis of stress indicators during cryopreservation of seeds of landrace maize (*Zea mays*). **Cryo-Letters** **38**: 445-454.
- Poncet V, Hamon P, Minier J, Carasco C, Hamon S and Noirot M (2004) SSR cross-amplification and variation within coffee trees (*Coffea* spp.). **Genome** **47**: 1071-1081.
- Setotaw TA, Caixeta ET, Pena GF, Zambolim EM, Pereira AA and Sakiyama NS (2010) Breeding potential and genetic diversity of “Híbrido do Timor” coffee evaluated by molecular markers. **Crop Breeding and Applied Biotechnology** **10**: 298-304.
- Shigueoka LF, Sera GH, Sera T, Fonseca ICB, Mariucci Junior V, Andreazi E, Carvalho FG, Gardiana CG and Carducci FC (2014) Selection of Arabic coffee progenies with rust resistance. **Crop Breeding and Applied Biotechnology** **14**: 88-93.
- Silva VA, Abrahão JCR, Lima LA, Carvalho GR, Ferrão MAG, Salgado SML, Volpato ML and Botelho CE (2019) Selection of comilion coffee clones tolerant to pests and diseases in Minas Gerais. **Crop Breeding and Applied Biotechnology** **19**: 269-276.
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. **Genetics** **139**: 457-462.
- Sousa TV, Caixeta ET, Alkimim ER, de Oliveira ACB, Pereira AA, Zambolim L and Sakiyama NS (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. **Euphytica** **213**: 75.
- Sutton JT, Robertson BC and Jamieson IG (2011) Dye shift: A neglected source of genotyping error in molecular ecology. **Molecular Ecology Resources** **11**: 514-520.
- Tran HTM, Ramaraj T, Furtado A, Lee LS and Henry RJ (2018) Use of a draft genome of coffee (*Coffea arabica*) to identify SNPs associated with caffeine content. **Plant Biotechnology Journal** **16**: 1756-1766.
- Van Puyvelde K, Van Geert A and Triest L (2010) Atetra, a new software program to analyse tetraploid microsatellite data: Comparison with tetra and tetrasat. **Molecular Ecology Resources** **10**: 331-334.
- World Coffee Research (2018) Arabica coffee varieties. Online catalog. Available at: <<https://varieties.worldcoffeeresearch.org/info/>>




## Microsatellite DNA fingerprinting of *Coffea* sp. germplasm conserved in Costa Rica through singleplex and multiplex PCR

catalog>. Accessed on March 30, 2018

Wright S (1978) **Evolution and the genetics of population, variability within and among natural populations**. The University of Chicago Press, Chicago, 580p.

Zhou L, Vega FE, Tan H, Lluch AER, Meinhardt LW, Fang W, Mischke S, Irish B and Zhang D (2016) Developing single nucleotide polymorphism (SNP) markers for the identification of Coffee germplasm. **Tropical Plant Biology** 9: 82-95.

 This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.