Genetica (2019) 147:205–216
https://doi.org/10.1007/s10709-019-00064-4

ORIGINAL PAPER

Population structure and genetic relationships between Ethiopian and Brazilian *Coffea arabica* genotypes revealed by SSR markers

Bruna Silvestre Rodrigues da Silva[1,2,9] · Gustavo César Sant'Ana[1,3,4] · Camila Lucas Chaves[5] · Leonardo Godoy Androcioli[1,6] · Rafaelle Vecchia Ferreira[1,2,9] · Gustavo Hiroshi Sera[1] · Pierre Charmetant[3,4] · Thierry Leroy[3,4] · David Pot[3,4] · Douglas Silva Domingues[7] · Luiz Filipe Protasio Pereira[1,8,9]

Received: 16 October 2018 / Accepted: 20 April 2019 / Published online: 3 May 2019

Abstract

Information about population structure and genetic relationships within and among wild and brazilian *Coffea arabica* L. genotypes is highly relevant to optimize the use of genetic resources for breeding purposes. In this study, we evaluated genetic diversity, clustering analysis based on Jaccard's coefficient and population structure in 33 genotypes of *C. arabica* and of three diploid *Coffea* species (*C. canephora*, *C. eugenioides* and *C. racemosa*) using 30 SSR markers. A total of 206 alleles were identified, with a mean of 6.9 over all loci. The set of SSR markers was able to discriminate all genotypes and revealed that Ethiopian accessions presented higher genetic diversity than commercial varieties. Population structure analysis indicated two genetic groups, one corresponding to Ethiopian accessions and another corresponding predominantly to commercial cultivars. Thirty-four private alleles were detected in the group of accessions collected from West side of Great Rift Valley. We observed a lower average genetic distance of the *C. arabica* genotypes in relation to *C. eugenioides* than *C. canephora*. Interestingly, commercial cultivars were genetically closer to *C. eugenioides* than *C. canephora* and *C. racemosa*. The great allelic richness observed in Ethiopian Arabica coffee, especially in Western group showed that these accessions can be potential source of new alleles to be explored by coffee breeding programs.

Keywords *Coffea* spp. · SSR markers · Genetic diversity · Population structure and relationships · Cultivated and wild gene pools

Introduction

Coffee is one of the most important agricultural commodities in tropical countries. More than 90% of its production occurs in developing countries providing an income for millions

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10709-019-00064-4) contains supplementary material, which is available to authorized users.

✉ Luiz Filipe Protasio Pereira
filipe.pereira@embrapa.br

[1] Instituto Agronômico do Paraná (IAPAR), Rodovia Celso Garcia Cid, 375 Km, Londrina, PR CEP 86047-902, Brazil

[2] Centro de Ciências Biológicas, Área de Genética e Biologia Molecular, Universidade Estadual de Londrina (UEL), CP 10.011, Londrina, PR CEP 86057-970, Brazil

[3] CIRAD, UMR AGAP, 34398 Montpellier, France

[4] AGAP, Univ. Montpellier, CIRAD, INRA, INRIA, Montpellier SupAgro, Montpellier, France

[5] Departamento de Agronomia, Universidade Estadual de Londrina (UEL), CP 6001, Londrina, Paraná CEP 86051-980, Brazil

[6] Instituto Agronômico do Paraná (IAPAR), Mestrado em Agricultura Conservacionista, CP 481, Londrina, PR CEP 86057-902, Brazil

[7] Instituto de Biociências de Rio Claro, Universidade Estadual Paulista (UNESP), Rio Claro, SP CEP 13506900, Brazil

[8] Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA Café), Brasília, DF CEP 70770901, Brazil

[9] Laboratório de Biotecnologia de Plantas – IAPAR, Embrapa Café, Rodovia Celso Garcia Cid, 375 km, Londrina, Paraná CP: 86047-902, Brazil

of smallholder farmess around the world that are dependent on coffee for their subsistence (Tran et al. 2016). *Coffea* L. genus belong to the Rubiaceae family and comprises approximately 124 species, but only 10 are cultivated (Davis et al. 2011). *Coffea arabica* L. and *C. canephora* P. are the two species most commercially relevant with approximately 60% and 40% of global production, respectively (ICO 2019). *C. arabica* produces a high-quality beverage, with pleasant aroma and flavor, but diseases and pests as well as abiotic stresses often affect its yield (Tran et al. 2016).

Coffee breeding programs invested intense efforts to release cultivars with high productivity, biotic and abiotic stresses tolerance and high biochemical quality of the beans (Tran et al. 2016). However, several factors are limiting the genetic gains in breeding programs (Pestana et al. 2015; Vieira et al. 2010). Commercial coffee plants are originate from a limited number of cultivars, mainly Typica and Bourbon types, and, as consequence, only narrow genetic base is available to support breeding programs. In addition, the reproductive behavior of *C. arabica* (i.e. autogamy) also contributes to the narrow genetic diversity available in this species (Anthony et al. 2002). As expected, several studies based on molecular markers demonstrated the low genetic variability available among commercial *C. arabica* varieties (Silvestrini et al. 2007; Setotaw et al. 2013; Pestana et al. 2015; Vieira et al. 2010).

The origin center of *C. arabica* is located in the highlands of southwestern Ethiopia (FAO 1968). Studies report wide agronomic diversity of arabica coffee accessions collected in this region regarding leaf size, height, biotic and abiotic stresses tolerance and yield (Bertrand et al. 2005; Pot et al. 2008; Tran et al. 2016). In addition, studies using molecular markers indicated the presence of higher genetic variability of Ethiopian accessions comparated with cultivars, demonstrating the potential of these accessions for breeding purposes (Silvestrini et al. 2007; López-Gartner et al. 2009; Teressa et al. 2010; Aerts et al. 2013; Sant'Ana et al. 2018). These accessions also showed a great variability for Caffeine, Chlorogenic acids, Lipids, Sucrose and Diterpenes contents of coffee beans (Scholz et al. 2016; Sant'Ana et al. 2018). The knowledge about population structure and genetic relationships of these Ethiopian accessions, among themselves and in relation to traditional cultivars is fundamental for efficient use of these genotypes in arabica coffee breeding programs.

Simple Sequence Repeats (SSR) markers have been used to discriminate cultivars and analyze genetic relationships among *C. arabica* cultivated and wild populations (Lashermes et al. 1995, 1999; Chaparro et al. 2004; Missio et al. 2011; Silvestrini et al. 2007; Teressa et al. 2010; Geleta et al. 2012; Aerts et al. 2013; Motta et al. 2014; Sousa et al. 2017), due to technical simplicity, speed, great resolving power, high levels of polymorphism and codominance. In addition, they are evenly dispersed across the genomes enabling accurate discrimination even between genetically related individuals (Vieira et al. 2010).

We analyzed the population structure and genetic relationships of a *C. arabica* panel, including wild genotypes from the primary center of origin of the species (Ethiopia) and commercial varieties, in order to evaluate the allelic richness of Ethiopian accessions for breeding purposes. In addition, a comparative analysis of genetic distances among *C. arabica* accessions and their two ancestral diploids, *C. canephora* and *C. eugenioides*, was carried out.

Materials and methods

Plant material and DNA extraction

A total of 36 *Coffea* genotypes were analyzed, including 33 *C. arabica* genotypes: Twenty-five from the Ethiopian collection, four cultivars (Typica, Bourbon, Iapar 59, Icatu x Catuai) and four lines developed by the breeding programs of IAPAR (IAPAR 78001-L_1C_1, IAPAR 88480-8-L_3C_3, 90-3-1, 90-8-1). Furthermore, was used one genotype of each of three different diploid *Coffea* species (*C. canephora*, *C. eugenioides* and *C. racemosa*) (Table 1).

All the plants were grown at the Londrina experimental station at Instituto Agronômico do Paraná (IAPAR), Brazil (23°23′00″S and 51°11′30″W). The FAO (Food and Agriculture Organization of the United Nations) collection at IAPAR comes from open pollinated seeds from the original collection at CATIE (Costa Rica) introduced in Brazil in 1976, and transferred from Instituto Agronômico de Campinas (IAC) to IAPAR. These accessions were collected in different Ethiopian regions. Nineteen accessions were collected from the West side of the Great Rift Valley (Kaffa, Kama, Illubador and Gojjam provinces) and six from East side of the Great Rift Valley (Sidamo and Shoa provinces).

DNA extraction and genotyping

Genomic DNA was isolated from leaves by CTAB method (Doyle and Doyle 1990) and diluted to a final concentration of 5 ng/µL. DNA quality was verified in 0.8% agarose gel, stained with ethidium bromide and quantification was estimated using spectrophotometry with absorbance at 260 and 280 nm, using Nanodrop™.

For genotyping, thirty SSR markers previously described as being polymorphic in *C. arabica* (Table 2) were used (Da Silva et al. 2013). PCR amplification reactions were performed using *Promega® Go Taq Green Master Mix* Kit with a final volume of 10 µL on *GeneAmp®* PCR System 9700 (*Applied Biosystems*) thermocycler, with the following parameters: 95 °C for 2 min, followed by 30 cycles of 95 °C

Table 1 List of 36 *Coffea* genotypes analyzed in this study

Genetic materials	Country	Region	Accession origin[a]	Species
E044	Ethiopia	Kaffa	West	*Coffea arabica*
E516	Ethiopia	Kaffa	West	*Coffea arabica*
E130	Ethiopia	Kaffa	West	*Coffea arabica*
E131	Ethiopia	Kaffa	West	*Coffea arabica*
E335	Ethiopia	Kaffa	West	*Coffea arabica*
E332	Ethiopia	Kaffa	West	*Coffea arabica*
E272	Ethiopia	Kaffa	West	*Coffea arabica*
E383	Ethiopia	Kaffa	West	*Coffea arabica*
E148	Ethiopia	Illubador	West	*Coffea arabica*
E208	Ethiopia	Illubador	West	*Coffea arabica*
E370	Ethiopia	Illubador	West	*Coffea arabica*
E363	Ethiopia	Illubador	West	*Coffea arabica*
E196	Ethiopia	Illubador	West	*Coffea arabica*
E454	Ethiopia	Illubador	West	*Coffea arabica*
E087	Ethiopia	Illubador	West	*Coffea arabica*
E464	Ethiopia	Illubador	West	*Coffea arabica*
E123A	Ethiopia	Kama	West	*Coffea arabica*
E123B	Ethiopia	Kama	West	*Coffea arabica*
E565	Ethiopia	Gojjam	West	*Coffea arabica*
E018	Ethiopia	Sidamo	East	*Coffea arabica*
E022	Ethiopia	Sidamo	East	*Coffea arabica*
E021	Ethiopia	Sidamo	East	*Coffea arabica*
E037	Ethiopia	Shoa	East	*Coffea arabica*
E237	Ethiopia	Sidamo	East	*Coffea arabica*
E238	Ethiopia	Sidamo	East	*Coffea arabica*
Typica	Amsterdam Gardens	x	Cutivar	*Coffea arabica*
Bourbon	La Réunion (Bourbon Island)	x	Cutivar	*Coffea arabica*
Iapar 59	Brazil	Paraná	Inbred line	*Coffea arabica**
Icatu x Catuai	Brazil	Paraná	Inbred line	*Coffea arabica***
H9733(90-3-1)	Brazil	Paraná	Inbred line	*Coffea arabica****
H9733(90-8-1)	Brazil	Paraná	Inbred line	*Coffea arabica****
IAPAR 78001-L_1C_1	Brazil	Paraná	Inbred line	*Coffea arabica*****
IAPAR 88480-8-L_3C_3	Brazil	Paraná	Inbred line	*Coffea arabica***
Coffea canephora	West/Central Africa/South Asia	x	Other Coffea spp.	*Coffea canephora*
Coffea eugenioides	East/Central Africa	Kenya	Other Coffea spp.	*Coffea eugenioides*
Coffea racemosa	East Africa	Mozambique	Other Coffea spp.	*Coffea racemosa*

[a]Geographical origin of the Ethiopian accessions and breeding status

xGenotypes that are not from Rift Valley region

*Villa Sarchi CIFC 971/10 × Hybrid of Timor 832/2 (Introgression of *C. canephora*)

**Introgression of *C. canephora*

***IAPAR 88480-8 L_3C_3 × (IAPAR 88480-8 L_3C_3 × C1195-5-6-2). C1195-5-6-2 = [(*C. arabica* × *C. racemosa*) × *C. arabica*] × *C. arabica*

****Crossing of the accession E335 × Catuai

for 50 s, 65 °C for 1 min, 72 °C for 30 s, and final extension of 72 °C for 5 min.

PCR products were run on 10% polyacrylamide gel electrophoresis (PAGE) and stained by ethidium bromide. Gels were visualized under ultraviolet light and captured by the Kodak ® 120 digital system. Molecular size of the amplified products was estimated using a 50 bp DNA ladder (*Ludwig biotec* ®).

Table 2 List of the 30 SSR loci used in this study

SSR loci	Primer forward	Primer reverse	Repeat motif	Nature of the repeat	Reference/contig position
CM2	TGTGATGCCATT AGCCTAGC	TCCAACATGTGC TGGTGATT	$(AC)_{10}$ $(AT)_9$	Di	Baruah et al. (2003)
CFGA792b[2]	GATCAGAACTTT GAGCTCAGCA	AATGTGGCACGC TAGAAGTG	$(AG)_{12}$	Di	Cristancho and Escobar (2008)
CFCA281[2]	GCGTCCACGTGT TAAGTCTT	TCAAGTGGCAGA CATGTCAC	$(AC)_{13}$	Di	Cristancho and Escobar (2008)
CFCA331[2]	TGATGGACAGGA GTTGATGG	CACTCATTTTGC CAATCTACC	$(CT)_{17}(AC)_{18}$	Di	Cristancho and Escobar (2008)
CFCA360[2]	TTAAGACATCGG TGCATTCA	TGTGTACTGGGT TTTTTGATGT	$(AC)_{15}$	Di	Cristancho and Escobar (2008)
CaM03[b]	CGCGCTTGCTCC CTCTGTCTCT	TGGGGGAGGGGC GGTGTT	AAC	Di	Geleta et al. (2012)
M24	GGCTCGAGATAT CTGTTTAG	TTTAATGGGCAT AGGGTCC	$(CA)_{15}(CG)_4CA$	Di	Combes et al. (2000)
M47	TGATGGACAGGA GGTGATGG	TGCCAATCTACC TACCCCTT	$(CT)_9(CA)_8/(CT)_4/(CA)_5$	Di	Combes et al. (2000)
SSRCa 002	CTGTCCCACCAA CCAAAA	CTTCAACCCCCA ACACAC	$(TTCC)_3.(GT)_{17}$	Tetra/Di	Missio et al. (2009)
SSRCa 052	GATGGAAACCCA GAAAGTTG	TAGAAGGGCTTT GACTGGAC	$(TTG)_7$	Tri	Missio et al. (2009)
SSRCa 081	ACCGTTGTTGGA TATCTTTG	GGTTGAACCTAG ACCTTATTT	$(CT)_{38}$	Di	Missio et al. (2009)
SSRCa 085	ATGTGAAAATGG GAAGGATG	CACAGGAAAGTG ACACGAAG	$(TC)_{24}$	Di	Missio et al. (2009)
SSRCa 091	CGTCTCGTATCA CGCTCTC	TGTTCCTCGTTC CTCTCTCT	$(GT)_8(GA)_{10}$	Di	Missio et al. (2009)
LEG11	CACTGAAGGCCT GGAAGAAT	AGCATCTGCAGC CTCCATAG	TGG	Tri	Pereira et al. (2011)
LEG12	CACCATAGCAAC TTCAAACACG	CACATCCAGGAA CCTTGCTC	TC	Di	Pereira et al. (2011)
LEG13	GAAGAGGAAGAA GGGGCAAG	GTGGTGGAGGAA AGGGATTC	GAA	Tri	Pereira et al. (2011)
LEG32	GGGTGATGGAAA AGCAAATG	CCAGCATCAGCA AGTAAAAGG	AGA	Tri	Pereira et al. (2011)
M32	AACTCTCCATTC CCGCATTC	CTGGGTTTTCTG TGTTCTGC	$(CA)_3/(CA)_3/(CA)_{18}$	Di	Combes et al. (2000)
No Identification	CTCTCCCTCAGT CAATTCCA	CTTGGTCTCCCT CCTTTTTC	$(ATC)_{14}$	Tri	Silvestrini et al. (2007)
AJ250253	CTTGTTTGAGTC TGTCGCTG	TTTCCCTCCCAA TGTCTGTA	$(GA)_5$ $(GT)_8TT(GT)_4TT(GT)_7/$ $(GA)_{11}/(TC)_2/(CT)_3GT$	Di	Silvestrini et al. (2007)
AJ250256	AGGAGGGAGGTG TGGGTGAAG	AGGGGAGTGGAT AAGAAGG	$(GT)_{11}$	Di	Maluf et al. (2005)
DCM01	TTTTTGGGAAAT GAAGGTGC	TGCACTTCAAGA TCCCCTTT	$(AG)_{15}$	Di	Aggarwal et al. (2007)
CaM41	CATCGTCTCCAT CGTTGCTCTATC	CCCTCCCCCTCT TTCCTATCTAAT	$(TAAA)_5$	Tetra	Hendre et al. (2008)
CFGA249	TAAGAAGCCACG TGACAAGTAAGG	TATGGCCCTTCT CGCTTTAGTT	$(AG)_{13}$	Di	Moncada and McCouch (2004)
IAPAR 14	GCGGATCTAACC AAGTAGCC	ATGATGCCGGTG ATGTTTAT	$(TTC)_4$	Tri	size37248
CHT03	GTCTCTCCGCTT TTTCTTCC	CTTGGTTGCCTG TTTCCTAA	$(CT)_8$	Di	scaffold8|size27678

Table 2 (continued)

SSR loci	Primer forward	Primer reverse	Repeat motif	Nature of the repeat	Reference/contig position
CHT12	CCGAGCATTGTG ACTCGTAT	CAGGAAAAACCA GAGACGAA	$(AT)_9$	Di	scaffold19|size38401
CHT25	CCTGTCTTGGCT CTACCTGA	TCTGTTGATCCG TGTTGATG	$(CTAT)_3$	Tetra	scaffold59|size4646
CHT28	CCGACGGGTCTC TTCTTTAT	TTCTTTACGGGA TTGCTCTG	$(AG)_5$	Di	scaffold121|size2154
CHT29	AAACCCAACCTG GCTTTTT	CATCGCCTCTCT TTCTCATC	$(TTCC)_3$	Tetra	scaffold121|size2154

Data analyses

Genetic diversity analysis

Due the allotetraploid genome of *C. arabica*, it is impossible to distinguish between the triallelic combinations of SSR loci. Therefore, although microsatellites are codominant markers, data analysis was based on presence/absence (1/0) of each allele, as performed by Aggarwal et al. (2007) and Silvestrini et al. (2007).

The analyzed genotypes were allocated to 4 genetic groups: (1) Cultivars and inbreed lines developed at IAPAR; (2) Eastern accessions (i.e. accessions from East side of the Great Rift Valley); (3) Western accessions (i.e. accessions from West side of the Great Rift Valley); and (4) others *Coffea* diploid species (*C. canephora*, *C. eugenioides* and *C. racemosa*).

The genetic diversity was estimated using the following parameters for each genetic group: Unbiased expected heterozygosity (uHe), Private alleles number, Proportion of polymorphic loci *(P %)* and Shannon's genetic index (*H'*). These analyses were performed using GenAlex software version 6.5 (Peakall and Smouse 2006). The Allelic frequency and Polimorphism Information Content (PIC) for each SSR marker were calculated according to Weir (1990).

Genetic distance matrix among all genotypes including Ethiopian accessions was estimated by the Jaccard's (1998) coefficient (Link et al. 1995). Clustering analysis was performed using the matrix distance based on the complement of Jaccard's coefficient employing the Neighbor–joining method. Bootstrap analysis (Felsenstein 1985) was performed to evaluate the tree topology reliability for 1000 simulations using FAMD software (Fingerprint Analysis with Missing Data 1.31) (Schlüter and Harris 2006). The tree was done in MEGA software (Molecular Evolutionary Genetics Analysis 6.06) (Tamura et al. 2013).

Analysis of Molecular Variance (AMOVA) was performed to estimate variation within and among genetic groups using SSR polymorphic loci. The analysis was performed using Arlequin 3.11 software (Excoffier et al. 2005) based on Weir and Cockerham method with 10,000 permutations, 100,000 steps of Markov Chain for the exact population differentiation test and 10,000 dememorisation steps, with a significance level of 0.01 (Weir and Cockerham 1984). The fixation index (Fst) among genetic groups was also estimated (Wright 1978).

Population structure analysis

The SSR profile for each genotype was used to perform the principal coordinate analysis (PCoA) and dissimilarity index between *C. arabica* genotypes in relation to *C. canephora* and *C. eugenioides* species. This analyses was performed using GenAlex software version 6.5 (Peakall and Smouse 2006).

Population structure was also estimated using Bayesian clustering method implemented in STRUCTURE software version 2.3.4 (Pritchard et al. 2000). Allele frequencies in each of the K groups (from 2 to 10) were estimated. We used a 10^5 burn-in period and 10^5 interactions MCMC (Markov Chain Monte Carlo), as these parameters resulted in relative stability of the results with 10 runs per K value. The most probable number of populations was estimated based on ΔK values (Evanno et al. 2005) using Structure Harvester software (Earl and Bridgett 2012). The level of membership that we considered to assign the genotypes to the different groups was 0.6 resulting in the assignments for 80% of the genotypes.

Results

Genetic diversity and population differentiation

We observed a total of 206 alleles across the 36 *Coffea* genotypes. The mean number of alleles over all loci was 6.9 ranging from 3 to 16 by locus and the mean of PIC values was 0.72, ranging from 0.39 to 1.00 (Supplementary Table 1).

Among Eastern, Western, Cultivars/inbred lines and species groups (*C. canephora, C. eugenioides* and *C. racemosa*)

the western group showed a highest Proportion of polymorphic markers (*P %)* and Shannon's index *(H')* (69.95% e 0.281), and the cultivars/inbred lines group presented the lowest ones (32.86% e 0.153). The number of Private alleles in Western group was also higher than in the others groups, confirming the higher genetic diversity in this group of the *C. arabica*. Regarding the uHe values, which measures the genetic diversity weighted by the sample size of each group, the species group presented the highest value (0.2), which can be explained by the fact that this group is formed by three individuals from distinct species. Considering only the *C. arabica's* groups, the Western group also had a higher value of uHe (0.183) than Eastern and cultivars/inbreed's groups (0.170 and 0.106, repectively) (Table 3).

The genetic distance to cluster analysis between all possible pairs of genotypes in this study was calculated using the polymorphic bands of all 30 SSR markers. The dendrogram using Jaccard´s coefficient reflect the higher similarity and narrow genetic diversity between genotypes in the subgroups generated, mainly in the cultivars subgroup, indicated by the low genetic distance by each cultivar/inbred line (Fig. 1).

The AMOVA shows that most of the genetic variance originates from the within-group level (74%; p = 0.05), with 26% of the total variance distributed among groups. The Fst values, which measured the magnitude of genetic differentiation between the genetic groups, was 0.32 between Ethiopian and Cultivars/inbred lines group, meaning a great differentiation between them (Wright 1978) (Table 4).

Population genetic structure

According to Evanno criterion (Evanno et al. 2005), the structure analysis with three groups (K = 3) showed a high ΔK value, but the upper-most level of the structure was in two groups (K = 2) (Fig. 1A–C). With K = 2, brazilian cultivars/inbred lines, eastern accession ^{E017} (E037) and the diploid species (*C. canephora*, *C. eugenioides* and *C. racemosa*) were allocated to the Q1 group (grey). The Q2 group (black) was formed by all accessions from west and east side of Rift Valley. In the structure with K = 3, the Q3 group was formed by *C. canephora* and *C. racemosa* species, demonstrating that these two species are more genetically distant in relation to the other groups analyzed (Fig. 2).

Principal coordinate analysis based on binary genetic distance matrix was consistent with Bayesian STRUCTURE analysis and explained 91.29% of total genetic variation of the panel (PCoA1—51.93 and PCoA2—39.36%). There was a clear division of whole panel in two groups, one (Q1) formed exclusively by accessions from Ethiopia, and another (Q2) formed by cultivars/inbred lines, two Ethiopian accessions and three diploid coffee species (i.e. *C. canephora*, *C. eugenioides* and *C. racemosa*). The second group presented a subdivision in two subgroups, in which it specifically separated the cultivars/inbred lines (Q2) from *C. canephora* and *C. racemosa* species (Q3). As in the bayesian analysis, the genotype of *C. eugenioides* presented great genetic proximity with the cultivars/inbred lines in the PCoA analysis (Fig. 3).

The comparison of binary genetic distance of *C. canephora* and *C. eugenioides* in relation to each one of *C. arabica* genotypes (Fig. 4) indicate that all *C. arabica* genotypes are genetically closer to *C. eugenioides* than to *C. canephora*. In addition, cultivars showed markedly lower genetic distances in relation to *C. eugenioides* compared to Ethiopian accessions.

Table 3 Proportion of polymorphic loci (*P %*), Shannon index of all loci (*H'*), Unbiased expected heterozygosity (uHe) and Private alleles for each genetic group over all 30 SSR loci

Genetic Group	N° of genotypes	*P* (%)	*H'*	uHe	Private Alleles
East	6	48.36	0.238	0.170	8
West	19	69.95	0.281	0.183	34
Cultivars and inbred lines	8	32.86	0.153	0.106	7
Species	3	49.06	0.249	0.200	20

Discussion

In the present study, 30 highly polymorphic SSR markers were used to genotype 36 Coffea spp. genotypes. An average of 6.9 alleles and PIC values of 0.72 were observed. Higher values were obtained for different genetic parameters compared with previous studies. Anthony et al. (2002) reported an average number over all loci of 4.7 using six SSR markers in *C. arabica* sample containing four Typica, five Bourbon and ten subspontaneous derived accessions. Using 34 SSR markers, Moncada and McCouch (2004) reported an average of 2.5 and 1.9 amplified alleles for SSR loci in 11 wild and 12 cultivated *C. arabica* genotypes, respectively, with the number of alleles per locus ranging from 1 to 8. Maluf et al. (2005) also reported an average number of 2.87 alleles in 28 cultivated *C. arabica* lines using 23 SSR markers. One reason for these differences could be due to smaller sample size and the coffee genotypes (Ethiopian vs Cultivated) used in the previous studies, mainly to the enrichment of Ethiopian *C. arabica* genotypes, as compared to the present study.

On the other hand, similar results were obtained by Teressa et al. (2010) analyzing *C. arabica* collection of 133 genotypes (78 accessions from Ethiopia and 55 cultivars) with 32 SSR markers. In this article, 209 alleles were detected

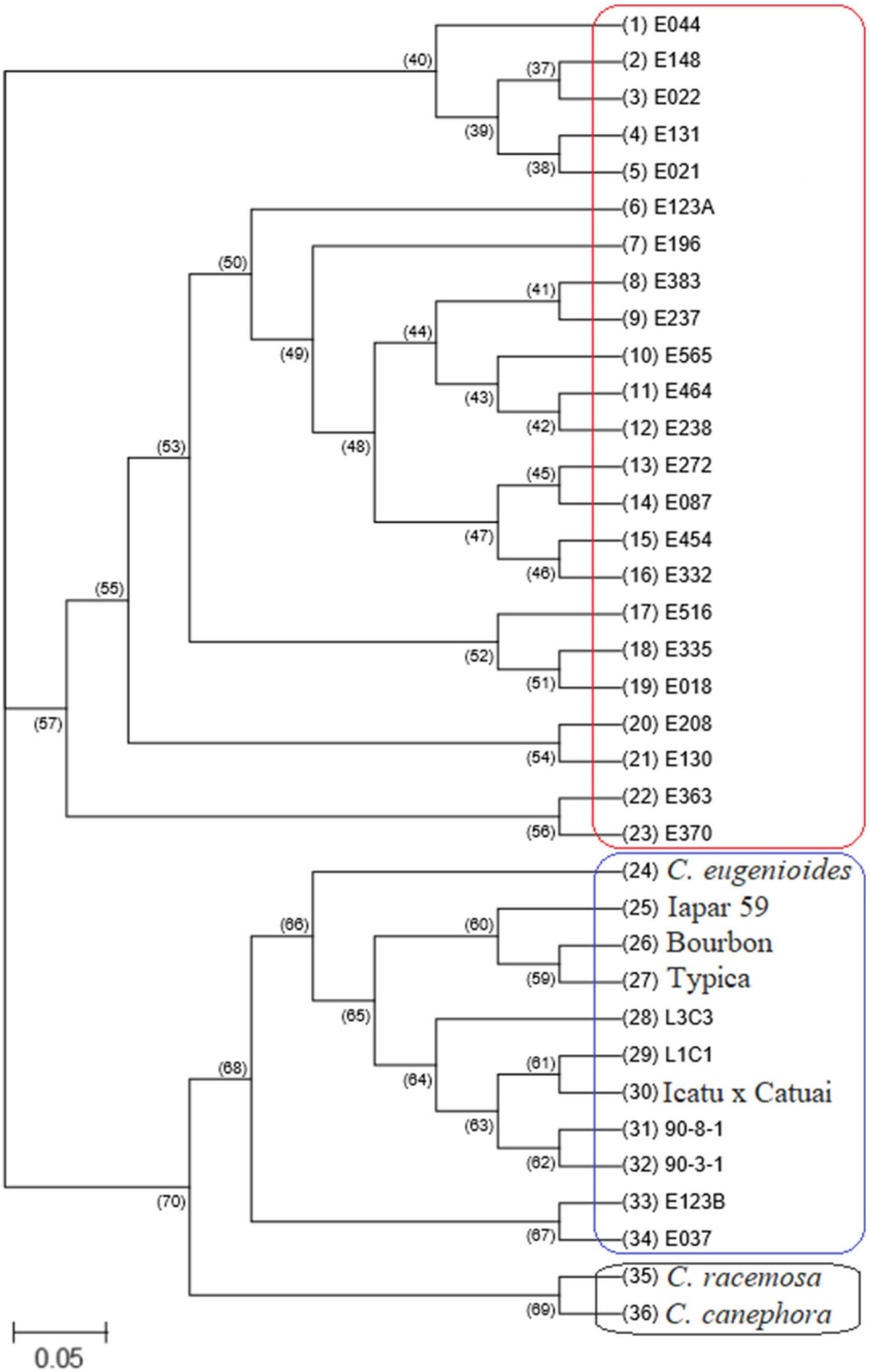

Fig. 1 Dendrogram of the 36 genotypes listed in Table 1 based on Jaccard genetic distance obtained from 30 SSR loci using Neighbor-joining method. Numbers displayed on the terminal branches indicates support of the nodes. The bootstrap used was of 1000 replications

with the number of alleles per locus ranging from 2 to 14, and an average of 6.5 alleles over all loci. Aerts et al. (2013) in a study based on populations from two coffee production systems (forest coffee and semi-forest coffee), identified 159 alleles across 703 wild accessions collected in forests from Ethiopia with 24 SSR markers. The number of alleles ranged from 2 to 19 per locus.

Our results indicated that accessions collected by the FAO mission in 1968 at the primary origin center of the species (Ethiopia), presented a high genetic diversity. On the

Table 4 Analysis of Molecular Variance (AMOVA) among and within *Coffea* genetic groups

Source of variance	df	Sum of squares	Variance component	Variation (%)
Among groups	3	232.93	7.44	26
Within groups	32	662.60	20.71	74

Fst total of the genetic groups = 0.25; Fst between West and East Ethiopian group = 0.05; Fst between Ethiopian and Cultivars/breeding lines group = 0.32; df = degrees of freedom

p = 0.05 and No. of permutations = 1000

other hand, among the cultivars, a low level of diversity was detected. This result is in agreement with the early history of *C. arabica* distribution, when the commercial cultivars have undergone successive genetic reductions (Anthony et al. 2002). Historical data indicated that the *C. arabica* populations in major producing countries were derived from few plants and/or seeds originated from Ethiopia. This could be the main factor for the low allelic richness and low polymorphism of the commercial cultivars.

The proportion of polymorphic loci and Shannon's index values estimated in our study (circa of 33% to 70% and 0.1 to 0.3) were similar to analyzes performed in populations of *Coffea arabica* using commercial cultivars and wild accessions (López-Gartner et al., 2009). These authors determined that the *P*% and *H*′ values ranged from 37 to 73% and 0.2 to 0.4 respectively. Comparing the uHe values within each genetic group (Table 3), the variability was higher in the Western group than Eastern and cultivars/inbred lines, what is consistent with *H*′ and the number of private alleles.

The high genetic richness of *Coffea arabica* Ethiopian accessions reinforce the importance of preserving the germplasm of *C. arabica* from center of origin (Ethiopia), and can help us to define which accessions are more important as source of diversity in breeding programs in the IAPAR to have a good genetic representation of the FAO collection. In the Western group 34 private alleles were identified, suggesting that particular efforts should be targeted towards the introduction of this genetic group in *C. arabica* breeding programs. The lower number of private alleles was found in brazilian cultivars/inbred lines (7 alleles), corroborating with the narrow diversity observed among cultivars in other studies (Silvestrini et al. 2007; Setotaw et al. 2013; Vieira et al. 2010; Sant'Ana et al. 2018).

The clustering analysis and the genetic structure results (genetic distance and bayesian-based approaches) indicated the presence of two main subgroups, which clearly distinguished Ethiopian accessions from others *Coffea* genotypes,

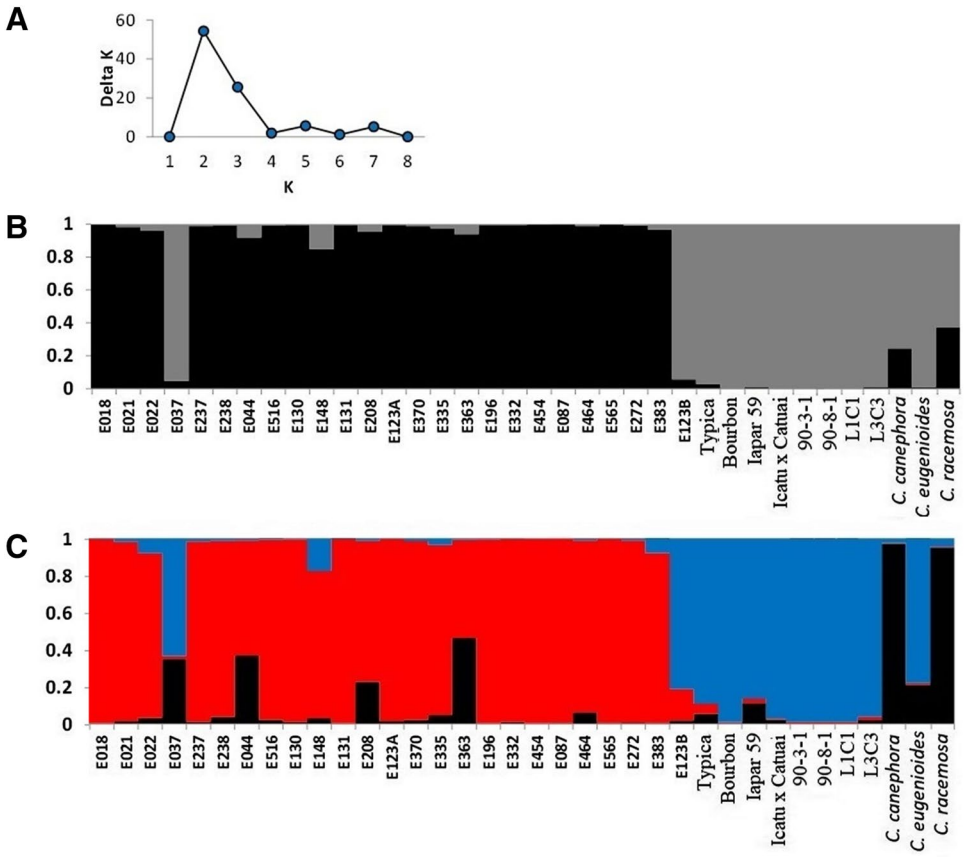

Fig. 2 Bar plot of population structure among wild accessions and cultivars/inbred lines of *Coffea arabica,* and diploid species. ΔK values in function of subgroups number (**A**) obtained from Bayesian clustering analysis considering K = 2 (**B**) and K = 3 (**C**) by STRUCTURE software version 2.3.4 from 30 SSR loci in 36 *Coffea* genotypes

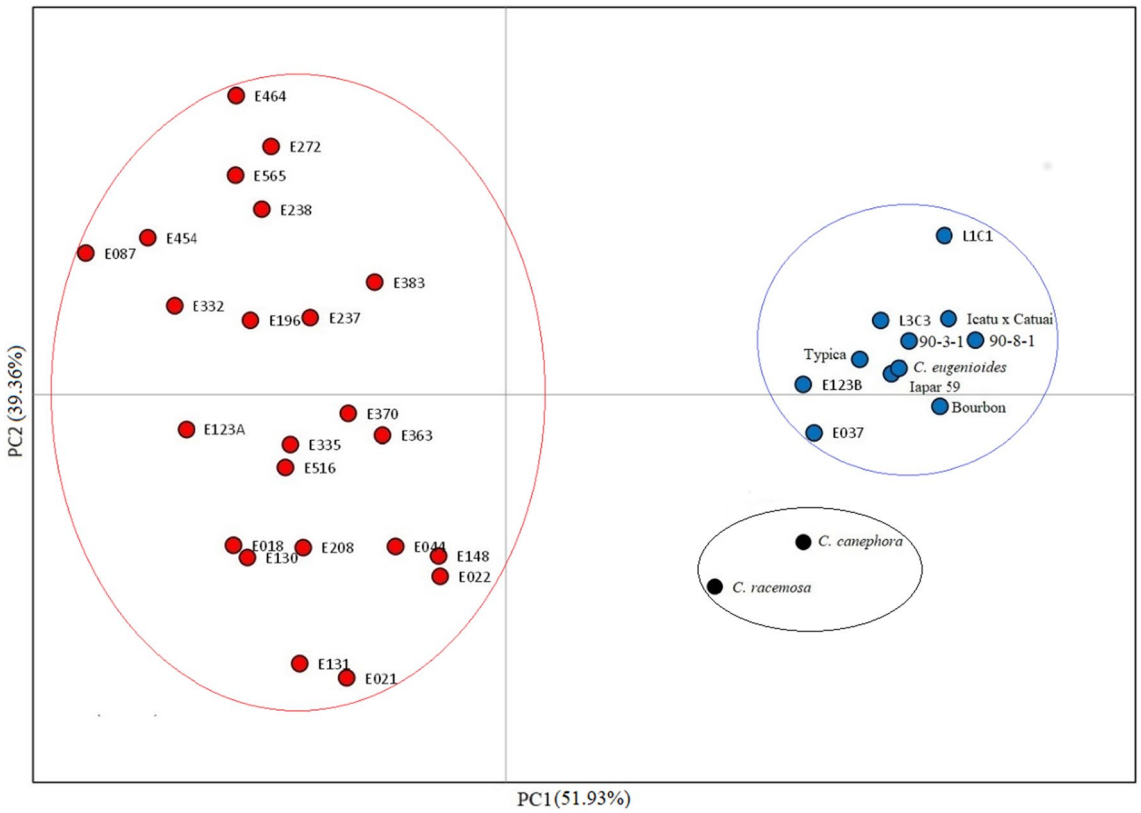

Fig. 3 Principal coordinate analysis (PCoA) based on genetic binary distance among the *Coffea* genotypes analyzed (n = 36). The dots are colored according to the colors of STRUCTURE results using K = 3

and a subgroup formed specifically by diploid species (*C. canephora* and *C. racemosa*). Similar results were observed in previous studies comparing the genetic diversity among wild and cultivated genotypes of *C. arabica* (Silvestrini et al. 2007; Lópes-Gartner et al. 2009; Teressa et al. 2010).

In the present study, binary genetic distances, clustering analysis as well as the population structure, demonstrated a closer proximity of the *C. eugenioides* genotype in relation to the cultivars group. We also observe that *C. canephora* and *C. racemosa* demonstrated high genetic dissimilarity in relation to *C. eugenioides*. Lashermes et al. (1995) studying the evolutionary history of *C. arabica* and their genetic relationships with other *Coffea* species also reported that *C. eugenioides*, followed by *C. canephora* and *C. racemosa* were the most related to *C. arabica*. Our data indicates that selection performed during the genetic improvement of *C arabica,* may have led to a decrease in genetic divergence of the breeding cultivars in relation to its diploid ancestor *C. eugenioides*. However, will be important to verify this genetic relationship in a more profound study with higher number of markers and more specimens.

C. eugenioides is the female ancestral parent of *C. arabica* and is the likely source of superior attributes for beverage quality (Medina Filho et al. 2007, 2012). Ashihara and Crozier (1999), Perrois et al. (2015) reported that the low caffeine content of *C. eugenioides* is due to the reduction of caffeine biosynthesis along with the rapid catabolism, that is regulated by specific genes. On the other side, *C. canephora* contains higher levels of the caffeine and chlorogenic acids (CGA), compounds directly related to both coffee bitterness and astringency, affecting its quality (Charrier and Berthaud 1988; Perrois et al. 2015; Jeszka-Skowron et al. 2016).

Conclusion

Our results indicate the presence of a high allelic richness in accessions from Ethiopia, especially in those collected in the West side of the Great Rift Valley, and this reinforces the importance of conserving and using germplasm of the primary center of origin of this important species. Our results indicate that *C. arabica* cultivars are genetically closer to its diploid ancestor *C. eugenioides* than wild accessions. Overall, information about genetic relationships of *Coffea* accessions estimated by SSR markers are valuable for conservation strategies and utilization of this germoplasm in breeding programs. Further analyses, including genomic comparisons of a higher number of *C. eugenioides* and *C.*

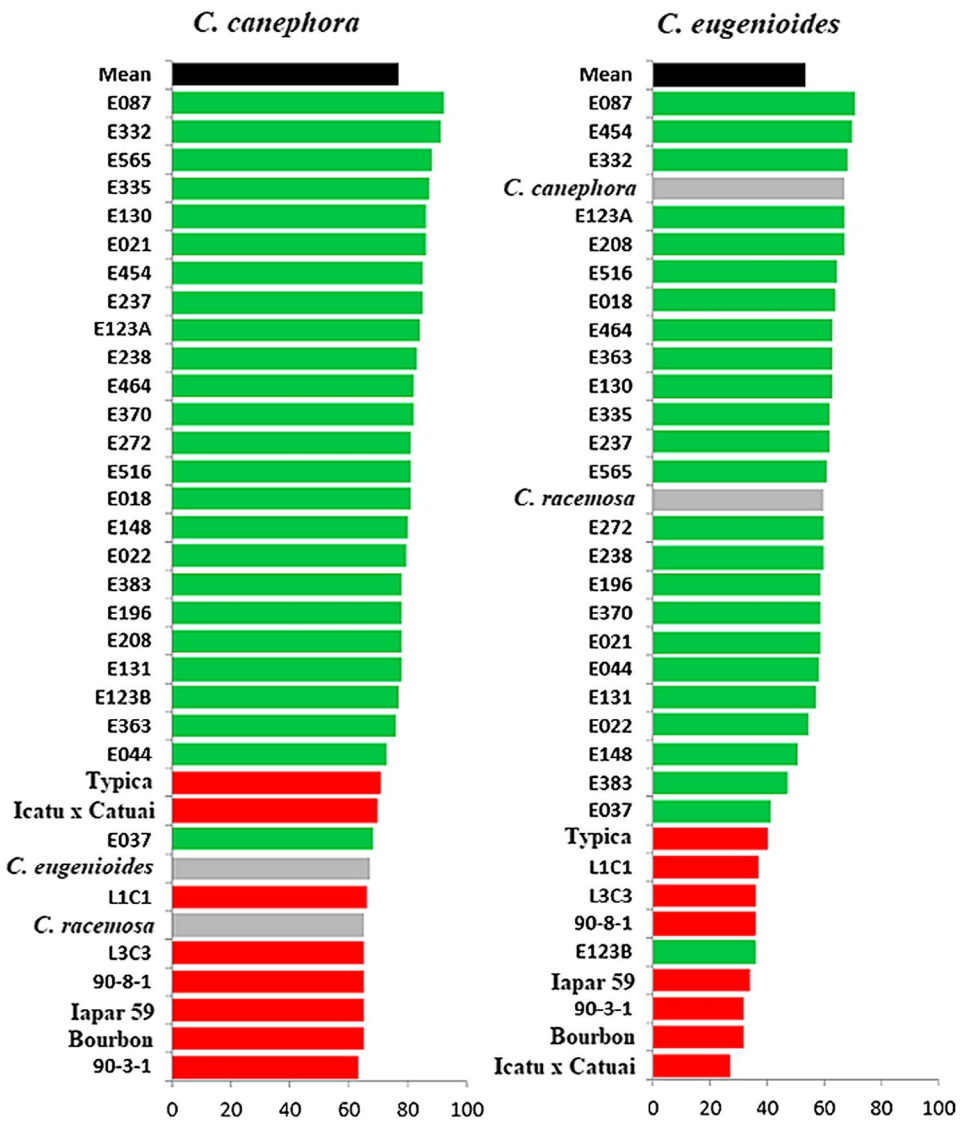

Fig. 4 Dissimilarity index between each *C. arabica* genotype in relation to *C. canephora* and *C. eugenioides* species. The bars representing genotypes from cultivars/inbred lines, Ethiopian accessions from Great Rift Valley and species are colored in red, green and grey, respectively

canephora genotypes in comparison with wild type and *C. arabica* cultivars, should provide a better understanding of the influence of the two diploid subgenomes in the domestication process of *C. arabica*.

Acknowledgements This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001, and by CAPES - Agropolis Foundation under the reference ID 1203-001 through the "Investissements d'avenir" program (Labex Agro: ANR-10-LABX-0001-01). We especially thank the Brazilian Coffee Research Consortium and INCT Café for supporting this study. DSD and LFPP acknowledge CNPq for the research fellowship. BSRS acknowledge CAPES and Brazilian Coffee Research Consortium for fellowships.

Authors contribution BSRS, GCS, LFPP conceived, designed the study and wrote initial of the manuscript draft. BSRS, GCS, CLC: performed genetic diversity and population structure analyses. BSRS, RVF, GHS and PC: selected, collected coffee plants in the field, extracted DNA and/or performed gel analyses. All authors read, edited and approved the final manuscript.

References

Aerts R, Berecha G, Gijbels P, Hundera K, van Glabeke S, Vandepitte K et al (2013) Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian montane rainforests. Evol Appl 6(2):243–252

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

Anthony F, Combs C, Astorga C, Bertrand B, Graziosi G, Lashermes P (2002) The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theor Appl Genet 104:894–900

Ashihara H, Crozier A (1999) Biosynthesis and catabolism of caffeine in low- caffeine-containing species of *Coffea*. J Agric Food Chem 47(8):3425–3431

Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P, Aggarwal RK (2003) Isolation and characterization of nine microsatellite markers from *Coffea Arabica* L., showing wide cross-species amplifications. Mol Ecol Notes 3:647–650

Bertrand B, Etienne H, Cilas C, Charrier A, Baradat P (2005) *Coffea arabica* hybrid performance for yield, fertility and bean weight. Euphytica 141(3):255–262

Chaparro AP, Cristancho MA, Cortina HA, Gaitan AL (2004) Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. Genet Resour Crop Evol 51:291–297

Charrier A, Berthaud J (1988) Principles and methods in coffee plant breeding: *Coffea canephora Pierre*. In: Clarke RJ, Macrae R (eds) Coffee, agronomy, vol 4. Elsevier, New York, pp 167–198

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

Cristancho M, Escobar C (2008) Transferability of SSR markers from related Uredinales species to the coffee rust Hemileia vastatrix. Genet Mol Res 7:1186–1192

da Silva BSR, Cação SB, Ivamoto ST, Silva JC, Domingues DS, Pereira LFP (2013) Identificação e Caracterização de Microssatélites de *Coffea arabica* a partir de dados de sequenciamento de RNA e de BACs. BBR 2(3):186–190

Davis AP, Tosh J, Ruch N, Fay MF (2011) Growing coffee : Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of Coffea. Bot J Linn Soc 167(4):357–377

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Earl DA, Bridgett M (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet 4:359–361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform 1:117693430500100003

FAO (1968) Coffee mission to Ethiopia, 1964–65. FAO, Rome

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Geleta M, Herrera I, Monzón A, Bryngelsson T (2012) Genetic diversity of arabica coffee (*Coffea arabica* L.) in Nicaragua as estimated by simple sequence repeat markers. Sci World J 2012:1–11

Hendre PS, Phanindranath R, Annapurna V, Lalremruata A, Aggarwal K (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 Biol 8:51–70

International Coffee Organization (ICO) (2019) http://www.ico.org/prices/po-production.pdf. Accessed 19 Fev 2019

Jaccard P (1998) Nouvelle recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270

Jeszka-Skowron M, Sentkowska A, Pyrzyńska K, Paz de Peña M (2016) Chlorogenic acids, caffeine content and antioxidant properties. Eur Food Res Technol 242:1403–1409

Lashermes P, Combes MC, Cros J, Trouslot P, Anthony F, Charrier A (1995) Origin and genetic diversity of *Coffea arabica* L. based on DNA molecular markers. Agronomie 528–536

Lashermes P, Combes MC, Robert J, Trouslot P, D'HONT A, Anthony F et al (1999) Molecular characterisation and origin of the *Coffea arabica* L genome. Mol Gen Genet 261(2):259–266

Link W, Dixkens C, Singh M, Schwall M, Melchinger AE (1995) Genetic diversity in European and Mediterranean faba bean germplasm revealed by RAPD markers. Theor Appl Genet 90:27–32

López-Gartner G, Cortina H, Couch MC, Susan R, Moncada MDP (2009) Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. Tree Genet Genomes 5:435–446

Maluf MP, Silvestrini M, Ruggiero LM, Guerreiro Filho O, Colombo C (2005) Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. Sci Agric 62(4):366–373

Medina-Filho HP, Maluf MP, Bordignon R, Guerreiro Filho O, Fazuoli LC (2007) Traditional breeding and modern genomics: a summary of tools and developments to exploit biodiversity for the benefit of the coffee agroindustrial chain. Acta Hortic 745:351–368

Medina-Filho HP, Bordignon R, Souza FF, Teixeira AL, Diocleciano JM, Ferro GO (2012) Arabica selections with *Coffea eugenioides* and *C.canephora* introgressions for Rondônia state in Brazilian Amazon, In: Proceedings of the 24th International Conference on Coffee Science. Costa Rica. Anais, pp. 1303–1307

Missio RF et al (2009) Development and validation of SSR markers for *Coffea arabica* L. Crop Breed Appl Biotechnol 9:361–371

Missio RF, Caixeta ET, Zambolim EM, Pena GF, Zambolim L, Dias LAS, Sakiyama NS (2011) Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers. Genet Mol Res 10(4):2366–2381

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

Motta LB, Soares TCB, Ferrao MAG et al (2014) Molecular characterization of arabica and conilon coffee plants genotypes by SSR and ISSR markers. Braz Arch Biol Technol 57:728–735

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol 6:288–295

Pereira GS, Padilha L, Pinho EVRV, Teixeira RKS, de Carvalho CHS, Maluf MP, Carvalho BL (2011) Microsatellite markers in analysis of resistance to coffee leaf miner in Arabica coffee. Pesq Agropec Bras 46(12):1650–1656

Perrois C, Strickler SR, Mathieu G, Lepelley M, Bedon L, Michaux S, Husson J, Mueller L, Privat I (2015) Differential regulation of caffeine metabolism in *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta). Planta 241(1):179–191

Pestana KN, Capucho AS, Caixeta ET et al (2015) Inheritance study and linkage mapping of resistance loci to Hemileia vastatrix in Híbrido de Timor UFV 443-03. Tree Genet Genomes 11:72

Pot D, Scholz MBS, Lannes SD, Del Grossi L, Pereira LFP, Vieira LG, Sera T (2008) Phenotypic analysis of *Coffea arabica* acessions from Ethiopia: contribution to the understanding of *Coffea arabica* diversity. In: 22nd International Conference on Coffee Science, Campinas. *Anais*. Campinas, p 165

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genet Soc Am Inference 155:945–959

Sant'ana GC, Pereira LF, Pot D, Ivamoto ST, Domingues DS, Ferreira RV, Pagiatto NF, da Silva BSR, Nogueira LM, Kitzberger CS, Scholz MB, de Oliveira FF, Sera GH, Padilha L L, Labouisse JP, Guyot R, Charmetant P, Leroy T (2018) Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. Sci Rep 8:465

Schlüter PM, Harris SA (2006) Analysis of multilocus fingerprinting data sets containing missing data. Mol Ecol Notes 6:569–572

Scholz MBS, Kitzberger CSG, Pagiatto NF, Pereira LFP, Davrieux F, Pot D et al (2016) Chemical composition in wild ethiopian Arabica coffee accessions. Euphytica 209(2):429–438

Setotaw TA, Caixeta ET, Pereira AA, Oliveira ACB, de CRUZ CD, Zambolim EM et al (2013) Coefficient of parentage in *Coffea arabica* L. cultivars grown in Brazil. Crop Sci 53(4):1237–1247

Silvestrini S, Junqueira MG, Favarin AC, Guerreiro-Filho O, Maluf MP, Silvarolla MB, Colombo CA (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. Genet Resour Crop Evol 54(6):1367–1379

Sousa TV, Caixeta ET, Alkimim ER, de Oliveira ACB, Pereira AA, Zambolim L, Sakiyama NS (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. Euphytica 213:75

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Teressa A, Crouzillat D, Petiard V, Brouhan P (2010) Genetic diversity of Arabica coffee (*Coffea arabica* L.) collections. EJAST. 1(1):63–79

Tran HT, Lee LS, Furtado A, Smyth H, Henry RJ (2016) Advances in genomics for the improvement of quality in coffee. J Sci Food Agric 96:3300–3312

Vieira ESN, von Pinho EVR, Carvalho MG, Esselink GD, Vosman B (2010) Development of microsatellite markers for the identification of brazilian *Coffea arabica* varieties. Genet Mol Biol 33:507–514

Weir BS (1990) Genetic data analysis. Methods for discrete genetic data. Sinauer Associates, Inc. Publishers, Sunderland

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358–1370

Wright S (1978) Evolution and the genetics of populations. The University of Chicago Press, London