GMR

The Original by FUNPEC-RP

# Genetic diversity based on AFLP markers in germplasm of the Brazilian national *Coffea arabica* trial

**C.R. Macedo[1], S.M. de Godoy[1], E.A. Ruas[2], B.D. Góes[3], C.L. Chaves[3], C.F. Ruas[1], T. Sera[4], G.H. Sera[4] and P.M. Ruas[1]**

[1] Departamento de Biologia Geral, Universidade Estadual de Londrina, Londrina, Paraná, Brasil
[2] Faculdade de Apucarana, Apucarana, Paraná, Brasil
[3] Departamento de Agronomia, Universidade Estadual de Londrina, Londrina, Paraná, Brasil
[4] Instituto de Desenvolvimento Rural do Paraná, Londrina, Paraná, Brasil

Corresponding author: P.M. Ruas
E-mail: pmruas@uel.br

Genet. Mol. Res. 20 (2): gmr18772
Received December 10, 2020
Accepted May 12, 2021
Published May 31, 2021
DOI http://dx.doi.org/10.4238/gmr18772

**ABSTRACT.** Brazil is the world largest producer and exporter of *Coffea arabica*. In this country, numerous breeding programs have generated a great multiplicity of cultivars with expressive productivity that are adapted to the different regions. Evaluating genetic diversity is critical since it provides parameters for defining breeding strategies. We estimated the genetic diversity among and within 32 cultivars of the Brazilian Coffee Trial using AFLP markers. This trial is composed of the main cultivars developed and some under development by Empresa De Pesquisa Agropecuaria do Estado de Minas Gerais (EPAMIG), Fundação PROCAFÉ, Instituto Agronômico de Campinas (IAC), and Instituto de Desenvolvimento Rural do Paraná (IDR-Paraná). A total of 982 AFLP fragments were generated, of which 97.35% were polymorphic. The percentage of polymorphic loci ranged from 22.8 to 50.5%, with genetic diversity varying from 0.06 to 0.16. Variable levels of genetic diversity observed among cultivars probably derived from the diverse germplasm sources and methods used in the genetic breeding programs, the number of advanced generations of each cultivar, as well as genetic recombination or cross-fertilization during breeding programs. Bayesian

 

cluster analysis, principal component analysis, and Neighbor-Net showed three divergent genetic groups, with a high genetic differentiation index (FST = 0.46). The pairwise FST also revealed high divergence among cultivars. IDR-Paraná had the cultivars with the highest genetic variability among these four Brazilian coffee breeding centers. We found that AFLP markers allowed us to distinguish the cultivars/progenies in the Brazilian National Trial. We conclude that Brazilian coffee germplasm still has considerable genetic variability for the development of new cultivars with high productivity, resistance to disease, superior beverage quality, and adaptation to diverse edaphoclimatic conditions in the different producing regions.

**Key words:** Genetic breeding; Coffee cultivars; Molecular markers; Genetic divergence; Germplasm evaluation

## INTRODUCTION

Agricultural coffee production began in Brazil in the first half of the 18th century with the introduction of *Coffea arabica*. This commodity has had very important economic and social impacts in this country. Currently, Brazil is considered the largest worldwide producer and exporter of coffee. International coffee production increased to 164.81 million bags of 60 kilos (increase of 5.7%) in the year 2017/2018. In Brazil coffee production is estimated at 59.9 million coffee bags of 60 kilos, with a growth of 33.2% in relation to 2017 (Conab, 2018; Ico 2018). Despite being the largest coffee producer in the world, the coffee breeding programs face great difficulties in Brazil, since cultivars of *C. arabica* released in Brazil share the same parental genotypes, which reduces substantially the basis to explore genetic variability. Several factors have contributed to the naturally low genetic variability of *C. arabica*. For instance, the predominant autogamous reproductive system of the species has produced low levels of heterozygosity from one generation to the next. Besides, the genetic base of this species was narrowed by the distribution to different producing countries of seeds and seedlings originating from only two primary cultivars (Typica and Bourbon), which generated most varieties and cultivars of coffee used all around the world (Setotaw et al., 2013).

Despite the limited genetic base, coffee research and breeding programs have generated cultivars with expressive productivity and adapted to the diverse ecological conditions in the different producing regions. One of the greatest challenges for coffee breeding programs is to improve the characteristics of the best cultivars, such as plant size, seasonal maturation uniformity, disease resistance, and beverage quality (Carvalho, 2008; Ferrão et al., 2009).

The Brazilian germplasm collection of *C. arabica* was initially set up by *ex situ* conservation and investigation, in the active germplasm banks (AGB) of the Instituto Agronômico de Campinas (IAC). This initial collection was, subsequently, expanded to several research centers in Brazil, including Empresa Brasileira de Pesquisa Agropecuária-EMBRAPA-Café, Empresa de Pesquisa Agropecuária do Estado de Minas Gerais-EPAMIG, Fundação PROCAFÉ, Instituto de Desenvolvimento Rural do Paraná-IDR in Paraná, Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural-INCAPER, Universidade Federal de Viçosa-UFV, and Universidade Federal de Lavras-UFLA

 

(Carvalho, 2008). From these research centers, many Arabica cultivars have been released in recent years, most of which are restricted to specific regions, so there is a need to better coordinated on a national level knowledge and technologies generated at each research institute.

The use of cultivars adapted to crop region and crop systems with resistance to pests and diseases, as well as to specific agronomical characteristics helps increase coffee production. Several cultivars of Arabica coffee have been released in recent years by IAC, IDR-Paraná, EPAMIG, and PROCAFÉ. However, these cultivars were restricted to the their original regions, revealing a need to outline a regional experimental network. This purpose was achieved with the establishment of the National Coffee Trial, whose objective is to estimate the stability and adaptability of the main coffee cultivars and indicate the cultivars that are best adapted to most of the coffee producing regions in Brazil and which would be best suited for commercial release.

Molecular markers are tools that enable the identification of levels of genetic variability among cultivars through the detection of DNA polymorphisms and have been used successfully in studies of *C. arabica*, to detect genetic diversity and to identify introgressive fragments in inter-specific hybrids derived from spontaneous crossings (Prakash et al., 2002). DNA markers have also allowed efficient analyses of genetic diversity within and between cultivars (Steiger, 2002), to generate linkage maps (Pearl, 2004), to evaluate relationship coefficients between cultivars (Setotaw et al., 2013), to assess the redundancies and deficiencies on levels of genetic variability of germplasm collections (Fonseca et al., 2002; Carvalho, 2008), and to examine the genetic stability in *C. arabica* plants derived from embryogenic suspensions. The use of AFLP markers showed that a genetic alteration has been limited in *C. arabica* somatic embryogenesis (Landey et al., 2013). Machado et al. (2017) employed conventional multivariate analysis to separate 16 *C. arabica* genotypes, based on morphological characteristics such as: length of the plagiotropic branch, diameter of the orthotropic branch, diameter of the crown base, root surface area ($mm^2/cm^3$), root weighted diameter (mm), and root volume ($mm^3$).

Estimates of genetic variability have been used for planning and implementation of coffee genetic breeding programs, since it allows the selection of superior genotypes and provides increased genetic gains, associated with medium to high production stability (Ferrão, 2004). In this current study we evaluated the effectiveness of the multilocus DNA markers, amplified fragment length polymorphisms (AFLP), to assess genetic diversity among and within cultivars of *C. arabica*, derived from various germplasm collections (EPAMIG, IAC, IDR - Paraná, and PROCAFÉ) belonging to the Brazilian National Coffee Trial, as a contribution to guide strategies to be implemented in further coffee breeding programs.

## MATERIAL AND METHODS

### Plant Material

The genetic material comprised 32 cultivars of *C. arabica* obtained from four Brazilian institutions, committed to developing strategies for improving coffee cultivars. These institutions are among representatives of the Brazilian Coffee Trial and they include the PROCAFÉ Foundation and EPAMIG from the State of Minas Gerais, IAC of the São

Paulo State, and IDR-Paraná. Data information of genetic material including cultivar's name, origin (Research Centers), genealogies, sampling, and year of cultivar release, are described in Table 1.

**Table 1.** General information of 32 coffee cultivars of the Brazilian Coffee Trial from active Brazilian germplasm collection maintained at four research centers. Genetic diversity indexes estimated from AFLP data of 308 plants, including: Number of plants of each cultivar (n), Percentage of polymorphic loci (PPL), Gene diversity and standard deviation (Hj ± SE), and rarity index (DW).

| Research Center/Cultivar | Genealogy | Release year | n | PPL | Hj±SE | DW |
|---|---|---|---|---|---|---|
| PROCAFÉ | | | | | | |
| 1. Catucaí Amarelo 2SL | Icatú x Catuaí | 2000 | 10 | 29.60 | 0.08±0.004 | 30.31 |
| 2. Catucaí Amarelo 24/137 | Icatú x Catuaí | 2000 | 10 | 31.50 | 0.08±0.004 | 26.69 |
| 3. Catucaí Amarelo 20/15 cv 479 | Icatú x Catuaí | 2000 | 9 | 30.20 | 0.07±0.004 | 81.10 |
| 4. Catucaí Vermelho 785/15 | Icatú x Catuaí | 2000 | 9 | 24.90 | 0.06±0.004 | 61.58 |
| 5. Catucaí Vermelho 20/15/476 | Icatú x Catuaí | 2000 | 9 | 27.50 | 0.08±0.004 | 33.20 |
| 6. Sabiá 398 | Acaiá x Catimor UFV 386 | 2000 | 9 | 33.40 | 0.09±0.004 | 32.21 |
| 7. Palma II | Catuaí Vermelho IAC 81 x Catimor UFV 353 | 2000 | 9 | 25.40 | 0.06±0.004 | 21.93 |
| 8. Acauã | Mundo Novo IAC 388-17 x Sarchimor IAC 1668 | 2000 | 10 | 28.20 | 0.07±0.004 | 39.10 |
| AVERAGE PROCAFÉ | | | | 28.84 | 0.07±0.001 | 40.77 |
| EPAMIG | | | | | | |
| 9. Oeiras M8851G | Catimor CIFC HW 26/5 (Caturra Vermelho CIF 19/1) x Timor hybrid | 1999 | 9 | 36.00 | 0.10±0.005 | 25.17 |
| 10. Catiguá MC01 | Catuaí Amarelo IAC 86 x Timor hybrid UFV 440-10 | 2004 | 9 | 26.70 | 0.07±0.004 | 15.18 |
| 11. Catiguá MG02 | Catuaí Amarelo IAC 86 x Timor hybrid UFV 440-10 | 2004 | 10 | 31.40 | 0.08±0.004 | 22.0 |
| 12. Sacramento MG1 | Catuaí Vermelho IAC 81 x Timor hybrid UF 438-52 | 2004 | 9 | 27.10 | 0.07±0.004 | 36.78 |
| 13. Araponga MG1 | Catuaí Amarelo IAC 86 x Timor hybrid UFV 446-08 | 2004 | 10 | 36.50 | 0.08±0.004 | 52.38 |
| 14. Paraíso H419-3-3-7-16-4-1 | Catuaí Amarelo IAC 30 x Timor hybrid UFV 445-46 | 2002 | 9 | 27.40 | 0.07±0.004 | 17.66 |
| 15. Paraíso H419-10-6-2-5-1 | Catuaí Amarelo IAC 30 x Timor hybrid UFV 445-46 | 2002 | 9 | 22.80 | 0.06±0.004 | 13.74 |
| 16. Paraíso H419-10-6-10-1 | Catuaí Amarelo IAC 30 x Timor hybrid UFV 445-46 | 2002 | 10 | 38.80 | 0.10±0.005 | 38.12 |
| 17. Paraíso H419-10-6-2-12-1 | Catuaí Amarelo IAC 30 x Timor hybrid UFV 445-46 | 2002 | 10 | 30.30 | 0.07±0.004 | 105.47 |
| 18. Pau Brasil MG1 | Catuaí Vermelho IAC 141 x Timor hybrid UFV 442-34 | 2004 | 10 | 31.30 | 0.08±0.004 | 46.14 |
| AVERAGE EPAMIG | | | | 30.83 | 0.08±0.001 | 37.26 |
| IDR-PARANÁ | | | | | | |
| 19. IAPAR 59 | Villa Sarchi CICF971/10 x Timor hybrid CIFC832/2 | 1992 | 10 | 41.60 | 0.13±0.005 | 60.16 |
| 20. IPR 98 | Villa Sarchi CICF971/10 x Timor hybrid CIFC832/2 | 2001 | 10 | 35.40 | 0.10±0.005 | 34.62 |
| 21 IPR 99 | Catuaí x Catuaí x H7314 (carrier of *C. liberica* genes) | 2001 | 10 | 34.80 | 0.10±0.005 | 20.75 |
| 22. IPR 100 | Catuaí x "BA-10" genotype (carrier of *C. liberica* genes) | 2001 | 10 | 49.60 | 0.16±0.006 | 40.03 |
| 23. IPR 103 | Catuaí x Icatú | 2001 | 10 | 37.00 | 0.11±0.005 | 43.46 |
| 24. IPR 104 | Villa Sarchi CICF971/10 x Timor hybrid CIFC832/2 | 2001 | 10 | 50.50 | 0.16±0.006 | 78.35 |

 

| Research Center/Cultivar | Genealogy | Release year | n | PPL | Hj±SE | DW |
|---|---|---|---|---|---|---|
| AVERAGE IDR-PARANÁ | | | | 41.48 | 0.13±0.003 | 46.23 |
| IAC 25. Tupi IAC 1669-33 | Villa Sarchi x Timor hybrid CIFC 832/2 | 1996 | 10 | 37.6 | 0.10±0.005 | 25.28 |
| 26. Bourbon Amarelo | Bourbon Vermelho mutation or natural recombination between Bourbon Vermelho and Amarelo de Botucatú | 1952 | 10 | 39.50 | 0.11±0.005 | 33.69 |
| 27. Catuaí Vermelho IAC 144 | Recombination between Caturra Amarelo and Mundo Novo IAC 374-19 | 1972 | 9 | 41.40 | 0.12±0.005 | 12.41 |
| 28. Obatã IAC 1669-20 | Villa Sarchi x Timor hybrid CIFC 832/2 | 1996 | 10 | 39.70 | 0.10±0.005 | 93.07 |
| 29. Obatã | Obatã IAC 1669-20 x Catuaí Amarelo | 2000 | 1 | 39.60 | 0.11±0.005 | 20.33 |
| AVERAGE IAC | | | | 38.84 | 0.11±0.005 | 37.67 |
| *30. Restrict material | | | | 36.20 | 0.09±0.005 | 47.65 |
| *31. Restrict material | | | | 35.90 | 0.10±0.005 | 23.53 |
| *32. Restrict material | | | | 40.80 | 0.11±0.005 | 45.40 |

* Origin Not available: these cultivars are under registration process (Registro Nacional de Cultivares of Ministério de Agricultura, Brazil).

The restricted materials are cultivars under registration process in the Registro Nacional de Cultivares of Ministério da Agricultura, Brazil. All genetic materials are currently maintained at the experimental area of the Instituto de Desenvolvimento Rural do Paraná, (IDR-Paraná), in Londrina, Paraná State, Brazil. The cultivars Catucaí Amarelo 2SL, 20/15, 24/137, and Catucaí Vermelho 785/15 are recommended for cultivation in Minas Gerais State, in the Southern and Zona da Mata regions. Catucaí Vermelho 20/15 cv 476 is adapted to cerrado conditions of Minas Gerais State and Sabiá 398 is cultivated in all growing coffee region of this state. The cultivar Acauã is cultivated in drier regions of Alto Paranaiba, Zona da Mata (both in Minas Gerais State) and Vitória da Conquista, Bahia. Oeiras MG 6851 is adapted to high altitude and high plant density of Triangulo Mineiro, Alto Paranaiba, Southern of Minas Gerais, and Zona da Mata. The cultivars Catiguá MG 1, MG 2, Sacramento MG 1, Araponga MG 1, Paraíso H 419 (3-3-7-16-4-1, 10-6-2-5-1, 10-6-2-12-1), Pau Brasil MG 1 are recommended for all Brazilian regions that cultivated Arabica coffee. Paraiso H 419 and Paul Brasil MG 1 are adapted to organic coffee production. Iapar 59, IPRs 98, 99, 100, 104 are cultivars that growing in Paraná State. IAPAR 59 is grown in cold and rainy regions. It can be recommended for other Brazilian coffee producing regions with high altitudes and temperatures between 18°C to 21°C. IPR 98 is adapted to altitudes below 500 meters and annual temperature between 20°C to 22°C and IPR 99 is suitable for plantations with high density. IPR 100 is found in Northern State with average annual temperature of 21.5°C with *Meloidogyne paranaensis* infested soil and IPR 104 is grown in all producing regions of the state with high temperatures and sandy soils or low altitude regions with lay soils. Finally, the cultivars Tupi IAC 1669-33, Bourbon Amarelo, Catucai Vermelho IAC 144, Obatã, and Obatã IAC 1669-20 are adapted to all coffee producing regions in Brazil. Tupy IAC 1669-33 is recommended to Family coffee farming and Bourbon Amarelo is grown in Brazilian region of high altitudes and cold temperatures. Obatã and Obatã IAC 1669-33 are recommended for cultivation in Brazilian regions with high temperatures (Consórcio pesquisa café, 2011).

## DNA extraction and AFLP procedures

DNA samples were obtained from leaves collected from 9-10 individual plants for each of 32 *C. arabica* cultivars, for a total of 308 plants analyzed. Genomic DNA extraction followed the CTAB method, except that CTAB was replaced by MATAB (Mixed Alkyl Trimethyl Ammonium Bromide, Sigma) in the extraction buffer, following Diniz et al. (2005). The integrity and quality of DNA samples were evaluated in 1% agarose gel, followed by quantification with a Nano Drop 2000 (Thermo Scientific, MA, USA). The AFLP protocol was applied according to Vos et al. (1995). Approximately 700 ng DNA was digested with 5U *Eco*RI and 1 U *Mse*I at 37 °C for 4 h, followed by adapter ligation using 2 U T4 DNA ligase at 22 °C for 1 h, and incubation at 70 °C for10 min for thermal inactivation of the restriction enzymes. The pre-selective reaction was prepared using 3.5 µL GoTaq® Green Master Mix (Promega, Madison, WI, USA), 0.58 µL pre-selective primer (4.75 µM), and 3 µL restriction-ligation. The pre-selective amplification was performed in a PTC-100$^{TM}$ thermocycler (MJ Research Inc., CA, USA) under the following conditions: 72°C for 2 min; 20 cycles of 94ºC for 1 s, 56ºC for 30 s and 72ºC for 2 min; and one final cycle of 60ºC for 30 min. For the selective PCR, we initially tested 10 different primer combinations with three and four selective bases. *Eco*RI primers were labeled at the 5'– end with fluorophores (FAM, NED, PET and VIC). Four most polymorphic primer pairs (FAM-*Eco*+AGG/*Mse*+CTT; NED-*Eco*+AAC/*Mse*+CTAG; VIC-*Eco*+ACT/*Mse*+CTG; PET-*Eco*+AGC/*Mse*+CTAG) were tested for reproducibility and then applied in selective amplification. The selective reactions were done using 3.5 µL *Taq* polymerase (GoTaq® Green Master Mix, Promega, Madison, WI, USA), 0.54 µL of *Mse*I (5 µM) and 0.54 µL (1 µM) of fluorescently-labeled *Eco*RI selective primers, 2.5 µL pre-selective reaction and ultra pure water, to complete a volume of 10 uL. The amplification was performed in a PTC-100$^{TM}$ thermocycler (MJ Research Inc., CA, USA) under the following conditions: 1 cycle of 94ºC for 2 min, 65ºC for 30 s and 72ºC for 2 min; 8 cycles of 94ºC for 1 s, 64ºC for 30 s and 72ºC for 2 min; 23 cycles of 94ºC for 1 s, 56ºC for 30 s and 72ºC for 2 min and 1 final cycle of 60ºC for 30 min. Selective PCR products were resolved by capillary electrophoresis on an ABI 3500xL genetic analyzer (Applied Biosystems, Foster City, CA, USA), with internal size standard (LIZ-GS-600). For reliability of the AFLP data, samples were tested for reproducibility and those with amplification problems for one or more primer combinations were discarded. Hence, some individuals from the original sample matrix were excluded, resulting in a few cultivars with nine samples only (Table 1).

## Data Analysis

Estimation of genetic parameters was done from raw AFLP data scored for each primer combination with GeneMapper® v.4.1 software (Applied Biosystems, CA, USA). AFLP fragments with size of 50-500 bp were considered to generate the binary matrix. For estimation of percentage of polymorphic loci (PPL), genetic diversity (Hj), analogous to He, and standard deviation of genetic diversity we used the software AFLP-SURV v.1.0 (Vekemans, 2002). A Bayesian approach was used to estimate the allele frequency (AFLP-SURV), with a non-uniform prior distribution of allele frequencies and assuming genotypic proportions within Hardy-Weinberg equilibrium (Vekemans, 2002). The rarity index (DW), a measure of genetic divergence, by computing 'frequency-down-weighted marker values' per population (Schönswetter and Tribsh, 2005), were estimated using the AFLPDAT R package (Ehrich, 2006) in the R environment v.3.4.1 (R Core Team, 2018).

Analysis of molecular variance (AMOVA) among and within cultivar, among and within Research Centers, and among genetic clusters, Fixation Index (FST), and pairwise FST were estimated using ARLEQUIN 3.1.1. (Excoffier et al., 2005). Principal Coordinate Analysis (PCoA) was performed using genetic distance (Nei and Li, 1979) with the "cmd scale" function in R v.3.4.1 (RCore Team, 2018). A pairwise distance matrix among cultivars was calculated with "poppr" package (Kamvar et al., 2015), implemented in R v.3.4.1 (R Core Team, 2018) and the relationships among cultivars were visualized with Neighbor-Net (Bryant and Mouton, 2004), using the Splits Tree v.4.14.2 software (Huson and Bryan, 2006). A mixture analysis, with marginal likelihood value, was estimated with BAPS v.6.0 software (Corander et al., 2008) to identify the genetic clusters among the cultivars. Values of K = 1 to K = 35 possible clusters were tested using the "Clustering with linked loci" prior, indicated for the AFLP data (Corander et al., 2007). Admixture analysis was performed to calculate the ancestral genotype mixture of each individual from the genetic groups found. This analysis was run with "2,000 interactions", "500 reference individuals per population" and "500 interactions per individual" priors.

## RESULTS

Four AFLP primer combinations applied to study 308 genotypes, representing 32 *C. arabica* cultivars, produced 982 fragments, with a mean percentage of polymorphic loci (PPL) of 97.35%. The number of fragments per primer combination was 128 (PPL = 89.84%) for FAM-Eco+AGG/ Mse+CTT, 253 (PPL = 99.60%) for NED-Eco+AAC/ Mse+CTAG, 208 (PPL= 95.67 for VIC- Eco+ACT/ Mse+CTG and 393 (PPL = 99.24%) for PET-Eco+AGC/ Mse+CTAG. The descriptive data for all 308 genotypes (32 cultivars) of *C. arabica* identified a mean percentage of polymorphic loci (PPL) among individuals within the cultivars ranging from 22.80 to 50.50%, for the cultivars Paraíso H-419-10-6-2-5-1 and IPR104, respectively (Table 1). Gene diversity (Hj) varied from 0.06, for cultivars Catucaí Vermelho 785/15, Palma II and Paraíso H-419-10-6-2-5-1, to 0.16 for cultivars IPR100 and IPR104, with a mean gene diversity of 0.09 (Table 1). The estimate gene diversity, considered among cultivars derived from different Brazilian research centers, showed higher average on gene diversity for cultivars from IDR-Paraná (0.13), followed by IAC (0.11), EPAMIG (0.08) and PROCAFÉ (0.07) (Table 1). Analyses of Molecular Variance (AMOVA) applied to 32 cultivars of *C. arabica* revealed lower variance among (46.41%) than within (53.59%) cultivars and a high index of genetic differentiation (FST = 0.46; Table 2).

**Table 2.** Analysis of molecular Variance (AMOVA) from using AFLP data of 32 genotypes (308) of *C. arabica* cultivars.

| Source of variation | d.f. | VC | PV (%) | FST |
|---|---|---|---|---|
| **All cultivars** | | | | |
| Among cultivars | 31 | 31.35 | 46.41* | 0.46* |
| Within cultivars | 276 | 36.21 | 53.59 | |
| **Research centers**† | | | | |
| Among centers | 3 | 3.36 | 4.79* | 0.04* |
| Within centers | 304 | 64.20 | 95.03 | |
| **Genetic clusters** | | | | |
| Among clusters | 2 | 36.13 | 44.23* | 0.44* |
| Within clusters | 305 | 45.54 | 55.77 | |

d.f., degree of freedom; VC, variance components; PV, percentage of variation; FST, fixation index. *$P \leq 0.01$ (significance test using 10,100 permutations). †Research centers: EPAMIG, IAC, IDR-Paraná, PROCAFÉ. Clusters given by admixture analysis (software BAPS).

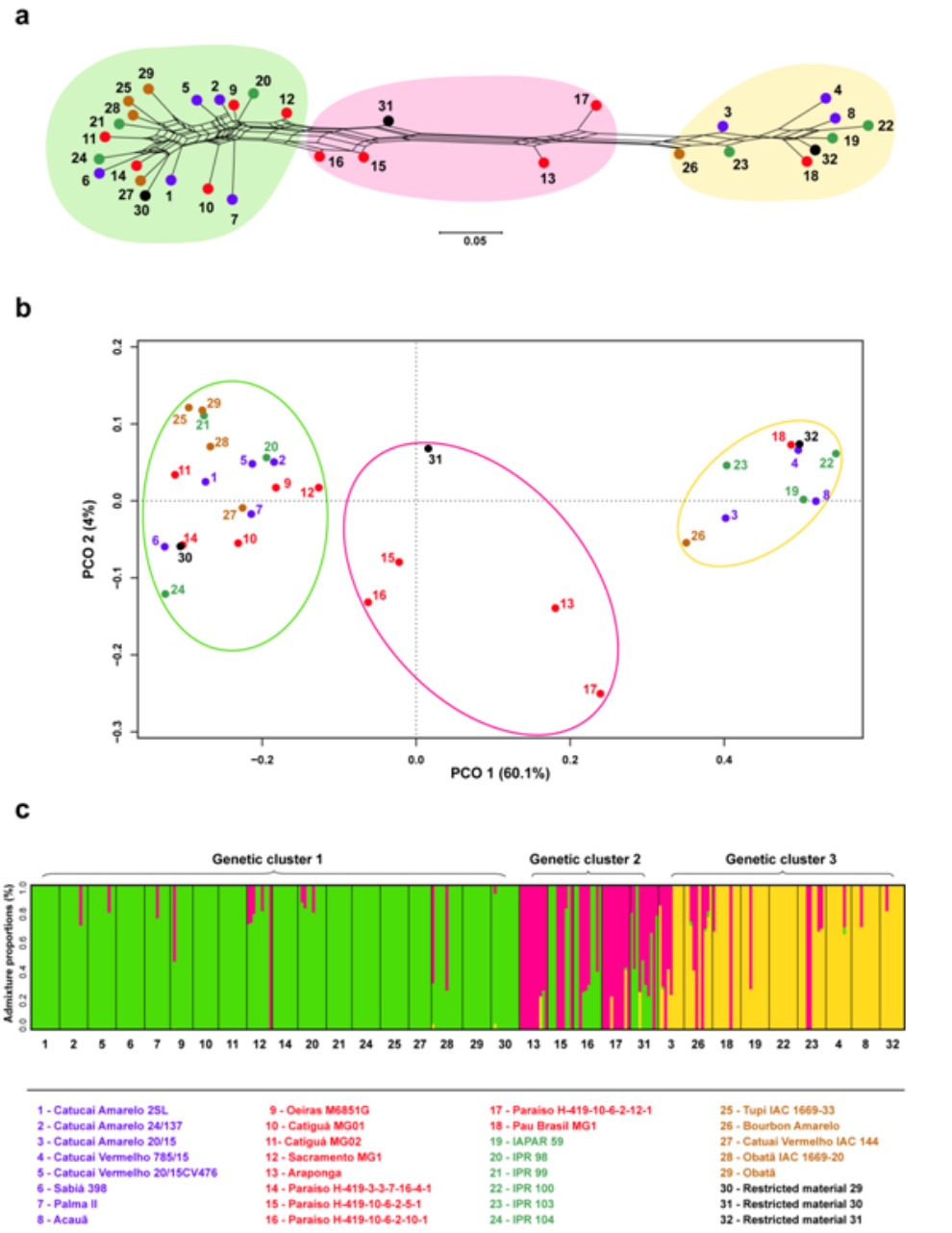


**Figure 1.** Genetic structure of 32 *C. arabica* cultivars based on 982 AFLP fragments. Dot colors represent Research Centers and the numbers are the different cultivars, as coding in Figure c. **(a)** SplitsTree Neighbor-Net inferred from Nei-Li distances based on AFLP data. The three major groups (background color) are encoded according to the number of genetic clusters found by BAPS analysis. **(b)** Principal Coordinate Analysis based on Nei-Li genetic distance. The circle colors are encoded according to BAPS. **(c)** Bayesian analysis of genetic structure of cultivars showing the three groups (K = 3) assigned by BAPS software.

 

The partitioning of genetic variation among cultivars from different research centers, revealed that most of the genetic variation (95.03%) resides within the research centers, while an extremely low value (4.97%) of genetic variance was observed among cultivars from the different research centers, with a low index of genetic differentiation (FST = 0.04; Table 2). The pairwise FST between cultivars showed that the closest (FST = 0.03) cultivars were Catucaí Vermelho 20/15CV476 and Catucaí Amarelo 24/137, both from PROCAFÉ, while the most distant (FST = 0.75) cultivars were Catucaí Vermelho 785/15 and Sabiá 398, both from PROCAFÉ (Supplementary 1).

The Neighbor-Net constructed from the Nei-Li (1979) genetic distance, showed three distinct genetic clusters, two of them are highly divergent groups, separated from each other by an intermediate small second group (Figure 1a). The first cluster grouped 18 cultivars that represent all research centers, including: Catucaí Amarelo 2SL, Catucaí Amarelo 24/137, Catuaí Vermelho 20/15 CV476, Sabiá 398 and Palma II , from PROCAFÉ; Oeiras M6851, Catiguá MG01, Catiguá MG02, Sacramento MG1 and Paraíso H-419-3-3-7-16-4-1 from EPAMIG; IPR 98, IPR 99 and IPR 104 from IDR- Paraná; Tupi IAC 1669-33, Catuaí Vermelho IAC 144, Obatã IAC1669-20, Obatã from IAC; and restricted material (Table 1, restricted material 30). The second cluster with five cultivars (four cultivars from EPAMIG) associated in two subgroups, one with the cultivars Araponga and Paraíso H-419-10-6-2-12-1, and the other with Paraíso H-419-10-6-2-5-1, H-419-10-6-2-10-1 and a restricted material (Table 1, restricted material 31). The third cluster grouped nine cultivars, also from all research centers: Catucaí Amarelo 20/15, CatucaíVermelho 785/15 and Acauã from PROCAFÉ; Pau Brasil MG1 from EPAMIG; IAPAR 59, IPR100 and IPR 103 from IDR-Paraná; Bourbon Amarelo from IAC; and one restricted material (Table 1, restricted material 32). The Neighbor-Net association of the cultivars were supported by the Principal Coordinate Analysis (PCoA) that also revealed two highly divergent groups and one intermediate group of cultivars (Figure 1b). The distribution of genetic variability detected by PCoA showed that the first two axes explained most (64.1%) of the variation found in the cultivars of coffee (Figure 1b). Likewise, the mixture model of the Bayesian analysis of population structure identified three genetic clusters, assigning for the 32 cultivars the optimal number of K = 3 clusters, with maximum posterior probabilities (p = 1) (Figure 1c). The admixture analysis applied to genetic clusters, identified K = 3 (Corander et al., 2008), and lower levels of molecular variance among (44.23%) than within clusters (55.77%), with a high index of genetic differentiation FST = 0.44 (Table 2). The genetic differentiation between clusters, as denoted by the pairwise FST values were 0.55, between cluster 1 and Cluster 3; 0.32 between cluster 2 and cluster 3; and 0.24 between cluster 1 and cluster 2.

## DISCUSSION

Molecular markers have been considered a powerful tool for DNA fingerprinting in plants, nonetheless the choice for a specific marker depends on its suitability to answer a particular question. Here we used AFLPs to assess the genetic distinctiveness of 32 *C. arabica* genotypes, belonging to the National Coffee Trial. These cultivars were released by different research centers and they are now part of germplasm collections, from where they were brought to the experimental field, at the Instituto de Desenvolvimento Rural do Paraná (IDR-Paraná), Londrina, Paraná State, Brazil. While it is worldwide accepted that *C.*

*arabica* possesses a very narrow genetic base, the analyses of 982 AFLP fragments, in a sample of 308 individual plants, revealed an overall high percentage of polymorphic loci (PPL of 97.35%) and a mean percentage of polymorphic loci within cultivars varying from 22.8%, for cultivar Paraíso H-419-10-6-2-5-1, to 50.5%, for cultivar IPR104. In other studies, different levels of polymorphic loci were found. For instance, in studies of 87 landraces of *C. arabica* from the south region of Ethiopia, Tadele et al. (2014) observed a range of 20% to 100% of polymorphic loci using ISSR markers, while Geleta et al. (2012), using SSR markers, to evaluate eight coffee cultivars, including the Catimor and *C. arabica* selections from Nicaragua, found a variation of 0.33 to 52% polymorphic loci. Sousa et al. (2017) also used SSR markers, resolved in 6% polyacrylamide gels, to analyze 34 coffee cultivars from the Brazilian National Coffee Trial, and found a range of 0.0 to 55% of polymorphic loci. The low level of polymorphism at SSR loci was attributed by the authors to the narrow genetic base of the cultivars/progenies of the coffee genotypes investigated, even though the method used for detection of alleles was, possibly, underestimated in the samples (Souza et al., 2017). While DNA markers with very similar sizes may be considered the same allele when resolved in polyacrylamide gels (Goular et al., 2001), Wang et al. (2009), using capillary electrophoresis, accurately distinguished SSR products and determined allelic size, enabling detection of alleles that differed in only one base pair. Also, capillary electrophoresis in combination with fluorescent labeled (Shi, 2003; Huang et al., 2006) primers provided higher sensitivity for detection of amplified DNA fragments. We applied AFLP to study the same sample of cultivars/progenies from the National Coffee Trial analyzed by Sousa et al. (2017) with SSR. Comparatively, we found much higher variation both within and between cultivars, hence revealing that AFLPs are much more efficient to explore the genetic variability that still exists in the Arabica coffee. The capability of AFLPs in detect high amount of polymorphism have been reported in studies of coffee. In a comparative study of 28 genotypes of *C. arabica* from Ethiopia, Dessalegn et al. (2009) demonstrated that AFLPs were more efficient over SSR markers to detect genetic variation. By performing the analyses of genetic parameters for 22 SSR and 712 AFLP fragments the authors found that the average of genetic similarity was much lower (0.560 with a range of 0.286-1.00) for SSR compared to the AFLP markers (0.982 with a range of 0.860-0.982). Dessalegn et al. (2009) also showed that with AFLPs it was possible to discriminate all genotypes, while the SSR distinguished only 64.3% of the genotypes, thus indicating that AFLPs, resolved by capillary electrophoresis, were more efficient for the characterization and evaluation of coffee genotypes.

AFLP analysis covers the entire genome, whereas SSR analysis detects variation at pre-determined loci/alleles, usually in the non-coding regions of the genome. The efficiency of AFLPs to reveal polymorphisms has been demonstrated in many plant groups. In *Dioscorea rotundata*, Mignouna et al. (2003) showed a lower level of polymorphism with SSR data than with AFLP. Reports of AFLP, RAPD, RFLP and SSR data by Garcia et al. (2004) revealed that the AFLPs provided high levels of accuracy for fingerprinting and for assessing relationships among maize inbreed lines. Gaudel et al. (2004) revealed the better accuracy of the AFLPs, compared to SSR, to discriminate individuals from neighboring populations of *Eryngium alpinum*. However, Vieira et al. (2010) using 22 polymorphic SSR loci separate twenty-five coffee varieties in two main groups, one composed of Brazilian varieties and the other with interspecific hybrids of *C. arabica*, *C. canephora* and *C. liberica*.

 

Under the current context of climatic change and disease susceptibility, genetic variability is essential for the management of cultivar vulnerability to biotic and abiotic stresses and relevant in the search for higher profitability in competitive markets (Laidò, 2013). Therefore, understanding the level and structure of the genetic diversity of a cultivated plant is important for the conservation and efficient use of the germplasms for plant breeding. Genetic parameters applied to AFLP data in the present study clearly distinguished all cultivars/progenies of the National Coffee Trial and found genetic variability within and among cultivars/progenies. For instance, while Souza et al. (2017) using SSR markers were unable to discriminate between the cultivars Catuai Vermelho IAC 144 and Paraiso H 419-3-3-7-16-4-1, the present AFLPs data assigned a genetic distance of 0.12 between these cultivars (Figures 1a, b). The cultivar Catuaí Vermelho IAC 144 arose through recombination from crossing between Caturra Amarelo IAC 476-11 x Mundo Novo IAC 374-19 cultivars and the cultivar Paraíso H 419-3-3-7-16-4-1 results from crossing between Catuaí Amarelo IAC 30 and Timor hybrid UFV 445-46. Souza et al. (2017) was also unable to determine the within genetic variability for the cultivars/progenies Catuaí Vermelho IAC 144, Catucaí 20/15, Sabiá tardio, Obatã IAC 1669-20, IPR 103 and Paraíso H 419-3-3-7-16-4-1. On the other hand, our results revealed that the percentage of polymorphic loci for these cultivars varied from 27.40% to 41.40%, with gene diversity ranging from 0.07 ± 0.005 to 0.12 ± 0.005 and rarity index from 12.41 to 105.47 (Table 1). While these values of within genetic variability varied among the cultivars, it was also possible to infer the genetic variability among the individual plants of each cultivar/progeny.

The cultivars IPR100 and IPR104, with a high genetic diversity (FST = 0.70) between them, were developed by IDR-Paraná Research Center and by crossing Catuaí x BA-10 (carrier of *C. liberica* genes) and a Icatú x Timor hybrid; both have in their constitution interspecific genotypes and, therefore, exhibit the most contrasting genetic diversity (Figure 1a, b and c). Vieira et al. (2010) using 22 SSR loci, also showed that the interspecific hybrids clustered far from most of Brazilian varieties, probably because of the presence of *C. canephora* and *C. liberica* in the genealogy of these cultivars. However, most of the cultivars in our study originated from intraspecific and interspecific hybridization presented lower genetic diversity (0.06 to 0.16) when compared to the coffee landraces studied by Tadele et al. (2014). For instance, these authors found high genetic diversity (0.07 to 0.38) in landraces sampled from Ethiopia, the original center of *C. arabica*. Geleta et al. (2012) observed gene diversity values ranging from 0.23 to 0.47 in coffee varieties from Nicaragua. Even our study detected lower genetic diversity, some cultivars still have within genetic diversity after undergoing several cycles of selection, revealing their potential for future exploitation in breeding programs.

This study showed evidence that cultivars derived from the same parental lineages eventually exhibit divergent levels of diversity and genetic structure (Catucaí and Paraíso cultivars; Figures 1a, b, c). The Catucaí cultivars, product from a natural crossing between Icatu x Catuaí, were submitted to genealogical breeding method from F3 to F6 generation. Despite the same origin, these cultivars differ on percentage of polymorphic loci (24.90 % - 31.50%), gene diversity (0.06 ± 0.004 – 0.008 ± 0.004) and rarity index (30.31 – 81.10). Also, the pairwise FST varied from 0.11, between Catucaí Amarelo 24/137 and Catucaí Amarelo 2SL, to 0.69, between Catucaí Vermelho 20/17 cv 476 and Catucaí Amarelo 2SL.

Conversely, Vieira et al. (2010) showed that SSR was unable to distinguished coffee color mutants, as Catucaí Amarelo and Catucaí Vermelho.

Our results showed that the same breeding method can result in cultivars with different levels of within and among genetic variability. It is considered that the divergence among Arabica coffee is due to the genetic origin of samples that included several species of *Coffea* and *C. arabica* varieties (Missio et al., 2011). In agreement, most of the cultivars belonging to the Brazilian National Coffee Trial have genotypes that resulted from interspecific crossing between *C. arabica*, *C. canehora* and *C. liberica*. However, the presence of cultivars with the same genealogy in different genetic clusters is possibly due to selective pressures of different methods, parental lineages used in the genetic breeding programs, variable number of advanced generations as well as gene recombination or crossed fertilization, which must have occurred at low rates in the field.

The Sarchimor cultivars, derived from crossing between Villa Sarchi CIFC 971/10 x Timor hybrid CIFC 832/2 (Table 1) such as IAPAR 59, IPR 98, IPR 104, Tupi IAC 1669-33, Obatã IAC 1669-20, also showed different levels of within genetic variability and different values of pairwise FST (Figures 1a, b; Supplementary 1). Silveira et al. (2003) studying several Sarchimor progenies found differences in within and between genetic variability. Catuaí Vermelho IAC 144 exhibit one of the highest genetic diversity, even though most of the cultivars that originated from the Catuaí genotype presented low within genetic indexes. Conversely, cultivars with distinct parental lineages presented high level of genetic variability and same genetic group, such as Bourbon Amarelo and IAPAR 59 (Figures 1a, b, c). These sources of genetic variability arise randomly and generate genotypically superior individuals, which are selected for breeding processes.

Among the sampled Brazilian germplasms (Figures 1a, b, c) the cultivars developed by IDR-Paraná constituted a group with higher genetic diversity (0.13), followed by cultivars from IAC (0.11), EPAMIG (0.08) and PROCAFÉ (0.07). IDR-Paraná cultivars are predominantly originated from Villa Sarchi, Timor hybrid, Catuaí and Icatú. IDR-Paraná and IAC have used genotypes with a wider genetic basis and higher number of parental germplasms in their programs. Thus, explaining the higher average of rarity index (DW) in IDR-Paraná cultivars and the prominent differentiation in the genetic basis of these cultivars, when compared to cultivars from other Research Centers. IPR100 is one of the cultivars with the highest genetic diversity probably due to the presence of *C. liberica* genes (BA-10 genotype). These data are consistent with those of Setotaw et al. (2013), who found a lower coefficient of parentage (COP) among cultivars of IDR-Paraná, indicating that this Research Center have released cultivars with a higher genetic variability. The progenies/cultivars from IAC showed the second highest genetic diversity. Whereas IAC is the oldest Research Institute of genetic coffee breeding, responding by 74.34% of the released cultivars of *C. arabica* in Brazil (Setotaw et al., 2013), the germplasm of IAC still presents high levels of genetic diversity among the cultivars, probably due to the use of most diversified parental (Villa Sarchi, Timor hybrid, Bourbon Vermelho, Mundo Novo and Catuaí) in the breeding programs. Conversely, PROCAFÉ and EPAMIG centers have used basically the same germplasms in the development of all cultivars, generating a narrower genetic basis in their cultivars.

Several studies have demonstrated reductions in crop genetic variability caused by the frequent use of few ancestors in the development of cultivars, been the autogamous species the most significantly affected (Pinto, 1995; Destro and Montálvan, 1999). Most of

 

the *C. arabica* cultivars released in Brazil originated from sister lineages, thus accounting its low genetic variability. Roughly 13 cultivars were used to develop the genetic material for the Brazilian cultivars, and only seven respond for 97.55% of the cultivars genetic base. Cultivar Bourbon Vermelho is the most representative between the parental cultivars, and it accounts for 52.76% of the gene pool of *C. arabica* cultivars in Brazil (Setotaw et al., 2013).

AMOVA applied to the AFLP data of the 32 cultivars revealed a lower variance within (46.41%) than among cultivars (53.59%). The genetic differentiation index (FST = 0.46) was higher than 0.25, which according to Wright (1965) can be consider an indicative of high genetic differentiation. A previous study of genetic diversity, comprising 24 accessions of *C. arabica* from Brazil, using SSR and EST-SSR markers (Missio et al., 2011), showed a molecular variance of 36% within and 64% among accessions (EST-SSR), and 46% within and 54% among accessions (SSR). Likewise, a study with RAPD markers, applied to Sarchimor cultivars (Silveira et al., 2003), revealed high differentiation (FST=0.38). The high FST values found in present study reflect the high genetic differentiation of the cultivars. The graphic distribution of the samples in the Neighbor-Net, PCoA and BAPS showed the formation of three highly divergent clusters (Figure 1a, b, c) and a high index of genetic differentiation (FST = 0.44) among them, with pairwise FST values of 0.55 between Cluster 1 and Cluster 3, 0.32 between Cluster 2 and Cluster 3 and 0.24 between Cluster 1 and Cluster 2.

The cluster 1 grouped 17 cultivars plus one cultivar that is the restricted material under number 30. The majority of cultivars of cluster 1 resulted from crossing between *C. arabica* with coffee plants containing *C. canephora* genes in different proportions. In this cluster, there were a substantial variation of pairwise FST, whose values varied from 0.03 (Catucaí Amarelo 24/137 x Catucaí Vermelho 20/15CV476) to 0.3 (Palma II x Sabiá 398), that are possible related to the origin of these cultivars. In cluster 1 64.70 % of the cultivars derived from the Catuaí genetic background, as the first parental genotype, while the Catuaí cultivars resulted from a recombination of an artificial hybridization between Caturra IAC 467-11 and Mundo Novo IAC 444-19. From F3 generation (IAC H2077-2-5) of this crossing, plants with red and yellow fruits was selected. The cultivars with red fruits were recorded (Registro Nacional de Cultivares, Ministério da Agricultura, Brazil) as Catuaí Vermelho IAC 15, IAC 24, IAC 51, IAC 72, IAC 81, IAC 99, and IAC 144. The cultivars with yellow fruits were registered as Catuaí Amarelo IAC 17, IAC 32, IAC 39, IAC 47, IAC 62, IAC 74, IAC 86, and IAC 100 (Consórcio pesquisa café, 2011). The remaining (29.40%) cultivars from cluster 1, derived from two parental lineages, the Villa Sarchi (IPR 99, IPR 104, Tupi IAC 1669-20), and cultivars or hybrids resulting from crossing with *C. canephora* (Icatú, Timor and Catimor hybrids). The Timor germplasm has 8-22% of *C. canephora* introgression, explaining the genetic diversity observed in the Timor hybrids and its derivatives (Lashermes et al. 2000), where the diversity observed in these genotypes come out as approximately the double of that seen in *C. arabica*. The exceptions are the cultivars Catuaí Vermelho IAC 144 (all parents are Arabica cultivars) and Tupi IAC 1669-33 (second parental lineage is H 7314-4 with *C. liberica* genes). For the cultivars Catucaí Amarelo and Vermelho the second parent is Icatú, which is a cultivar resulting from crossing between *C. canephora* x Bourbon Vermelho. The cultivars Sabiá 398 and Palma II have germplasm of Catimor UFV 398 (Caturra x Tymor Hybrid), as second parent. The cultivars Oeiras M6851G, Catiguá MG01, Catiguá Mg 02, Sacramento MG1, Paraíso H

419-3-3-7-16-4-1, IPR 98, IPR 104, Tupi IAC1669-33, and Obatã 1669-20 originated from the Timor hybrid lineage as second parent.

Cluster 2 grouped the cultivars Araponga MG I (resulting from the crossing Catuaí Amarelo IAC 86 x Timor hybrid UFV 446-08), Paraíso H419-10-6-2-5-1, H419-10-6-2-10-1, H419-10-6-2-12-1 and the restricted material 31. The Paraíso cultivars resulted from a seed mix of eight progenies arising from the crossing between Catuaí Amarelo IAC 30 and Timor hybrid 445-46, in the F4 generation. The origins of cultivars of cluster 2, also explain the pairwise FST between these cultivars (0.08to 0.39). The cluster 3 beyond the restricted material 32, grouped eight cultivars (Catucaí Amarelo 20/15 cv 479, Catucaí Vermelho 785/15, Acauã, Pau Brasil MG1, IAPAR 59, IPR 100. IPR 104, Bourbon Amarelo). Six of these cultivars carry *C. canephora* genes, while one has genes from *C. liberica*. Thus, the variation on the levels of pairwise FST between the cultivars within and between clusters resulted from different breeding methods and generation of selection applied in breeding program. Our results also show that, despite of the different strengths of genetic selection used by the breeding programs, some cultivars/progenies still maintain a reasonable within and among genetic variability and high potential as source of genetic variation.

## CONCLUSIONS

Using AFLP markers, it is possible to distinguish cultivars/progenies of *C. arabica* in the Brazilian National Trial. The breeding program of IDR-Paraná has the cultivars with the highest genetic variability among Brazilian coffee breeding centers. Different pairwise FST values were observed between cultivars, even between cultivars derived from the same parent plants. Our results revealed that the *Coffea* germplasm from Brazil are still a good source of genetic variability for breeding programs to search for heterosis and future development of new varieties and cultivars.

## ACKNOWLEDGMENTS

The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for a Research Fellowship to P.M.R. and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for grants to C.R.M. This study was supported by Consórcio Pesquisa Café.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

Bryant D and Moulton V (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* 21: 255-265. DOI: 10.1093/molbev/msh018.

Carvalho CHC (2008). Cultivares de café: origem, características e recomendações. Embrapa Café, Brasília, Brazil. ISBN: 978-85-61619-00-1.

Conab (Companhia Nacional de Abastecimento) (2018). Acompanhamento da safra brasileira de café, segunda estimativa. https://www.conab.gov.br. Accessed March 8, 2021.

Consórcio pesquisa café (2011). Anais do VII Simpósio de Pesquisa dos Cafés do Brasil. Embrapa Café.

 

Corander J and Tang J (2007). Bayesian analysis of population structure based on linked molecular information. *Math. Biosci*. 205: 19-31. DOI: 10.1016/j.mbs.2006.09.015.

Corander J, Marttinen P, Sirén J and Tang J (2008). Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*. 9: 539. DOI: 10.1186/1471-2105-*9-539.*

Dessalengn Y, Herselman L and Labuschagne MT (2009). Comparison of SSR and AFLP analysis for genetic diversity assessment of Ethiopian Arabica coffee genotypes. *S. Afr. J. Plant Soil.* 26: 119-125. DOI: 10.1080/02571862.2009.10639943.

Destro D and Montalván R (1999). Melhoramento Genético de Plantas. Universidade Estadual de Londrina, Londrina, Paraná, Brazil. P. 820. ISBN: 85-7216-093-0.

Diniz LEC, Ruas CF, Carvalho VP, Torres FM, et al. (2005). Genetic diversity among forty coffee varieties assessed by RAPD markers associated with restriction digestion. *Braz. Arch. Biol. Tech.* 48: 511-521. DOI: 10.1590/S1516-89132005000500002.

Ehrich D (2006). AFLP dat: a collection of R functions for convenient handling of AFLP data. *Mol. Ecol. Notes*. 6: 603-604. DOI: 10.1111/j.1471-8286.2006.01380.x.

Excoffier L, Laval G and Schneider S (2005). Arlequin ver. 3.1: an integrated software package for population genetics data analysis. *Evol. Bioinf.* 1: 47-50.

Ferrão RG (2004). Biometria aplicada ao Melhoramento Genético do Café Conilon. PhD Thesis. Universidade Federal de Viçosa, Minas Gerais, Brazil.

Ferrão MAG, Ferrão RG, Fornazier MJ, Prezotti LC, et al. (2009). Técnicas de produção de café arábica: renovação e revigoramento das lavouras no Estado do Espírito Santo, Vitória, Incaper, 3$^{th}$edition. ISSN: 1519-0528.

Fonseca AFA, Ferrão MAG and Ferrão RG (2002). A cultura do café Robusta. In Anais do VI Simpósio de pesquisa dos cafés do Brasil. Pg. 119-145.

Garcia AAF, Benchimol LL, Barbosa AMM, Geraldi IO, et al. (2004). Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. *Genet. Mol. Biol.* 27: 579-588.

Gaudel M, Till-Bottraud I, Barjon F and Manel S (2004). Genetic diversity and differentiation in *Eryngium alpinum* L. (Apiaceae): comparisonof AFLP and microsatellite markers. *Heredity*. 92: 508-518.

Geleta MI, Monzón A and 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. DOI: 10.1100/2012/939820.

Goulart IC, Merotto Junior A, Nunes AL and Bered F (2001). Otimização da utilização de marcadores moleculares microssatélites e sua aplicação em estudos de plantas daninhas. *Pl. Daninha.* 29: 1175-1181. DOI: 10.1590/S0100-83582011000500025.

Huang YF, Huang CC, HU CC and Chang HT (2006). Cappilary electrophoresis based separation technique for the analysis of proteins. *Electrophoresis*. 27: 3503-3522. DOI: 10.1002/elps.200600100.

Huson DH and Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol*. 23: 254-26. DOI: 10.1093/molbev/msj030.

Ico (Organização Internacional do Café) (2018). Relatório sobre o Mercado de cafeeiro http://www.ico.org/. Accessed November 8, 2020.

Kamvar ZN, Brooks JC and Grünwald NJ (2015). Novel R tools for analysis of genome-wide Population genetic data with emphasis on clonality. *Front Genet.* 6: 208. DOI: 10.3389/fgene.2015.00208.

Laidò GG (2013). Genetic diversity and population structure of tetraploid wheats (*Triticum turgidum* L.) estimated by SSR, DArT and pedigree data. *PLoS ONE.* 8: e67280. DOI: 10.1371/journal.pone.0067280.

Landey RB, Cenci A, Georget F, Bertrnad B, et al. (2013). High genetic and epigenetic stability in *Coffea arabica* plants derived from embryogenic suspensions and secondary embryogenesis as revealed by AFLP, MSAP and the phenotypic variation rate. *PloS ONE.* 8: e56372. DOI: 10.1371/journal.pone.0056372.

Lashersmes P, Andrzejewski S, Bertrand B and Anthony F (2000). Molecular analysis of introgressive breeding in coffee (*Coffea arabica*). *Theor. Appl. Genet.* 100: 139-146. DOI:10.1007/s0012200500.

Machado CMS, Pimentel NS, Golynsk A, Ferreira A, et al (2017). Genetic diversity among 16 genotypes of *Coffea arabica* in the Brazilian cerrado. *Genet. Mol. Res.* 21: 16(3). DOI: 10.4238/gmr16039794.

Mignouna HD, Abang MM and Fagbemi AS (2003). A comparative assessment of molecular markers assay (AFLP, RAPD, and SSR) for white yam (*Dioscorea rotundata*) germplasm characterization. *An. Appl. Biol.* 142: 269-276. DOI: 10.1111/j.1744-7348.2003.tb00250.x.

Missio RF, Caixeta EM, Pena GF, Zambolin L, et al. (2011). Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers. *Genet. Mol. Res*. 10: 2366-2381. DOI:10.4238/2011.October.6.2.

Nei M and Li WH (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Nat. Acad. Sc.* 76: 5269-5273. DOI:10.1073/pnas.76.10.5269.

Pearl HM (2004). Construction of a genetic map for Arabica coffee. *Theor. Appl. Genet.* 108: 829-835. DOI: 10.1007/s00122-003-1498-3.

Pinto R (1995). Introdução ao melhoramento genético de plantas. Universidade Estadual de Maringá, Paraná, Brazil. ISBN: 978-85-7628-134-4.

Prakash N, Combes MC, Somanna N and Lashermes P (2002). AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. *Euphytica.* 124: 265-271.

R CoreTeam R (2018): A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/. Accessed June 5, 2020.

Schönswetter P and Tribsch A (2005). Vicariance and dispersal in the alpine perennial *Bupleurum Stellatum* L. (Apiaceae). *Taxon.* 54: 725-732. DOI: 10.2307/25065429.

Setotaw TA, Caixeta ET, Pereira AA, Oliveira ACB, et al. (2013). Coefficient of parentage in *Coffea arabica* L. cultivars grown in Brazil. *Crop Sci.* 53: 1217-1227. DOI: 10.2135/cropsci2912.09.0541.

Shi L (2003). Micro preparative capillary gel electrophoresis of DNA: Rapid expressed sequence tag library construction. *Electrophoresis*. 24: 86-92. DOI: 10.1002/elps.200390035.

Silveira SR, Ruas PM, Ruas CF, Tumoru S, et al. (2003). Assessment of genetic variability within and among coffee progenies and cultivars using RAPD markers. *Genet. Mol. Biol.* 26: 329-336. DOI: 10.1590/S1415-47572003000300018.

Sousa TV, Caixeta ET, Alkimin ER, Oliveira ACB, et al. (2017). Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. *Euphytica.* 213: 75. DOI: 10.1007/s10681-017-1865-9.

Steiger DL (2002). AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theor. Appl. Gen*. 105: 209-215. DOI: 10.1007/s00122-002-0939-8.

Tadele S, Mekbib F and Tesfaye K (2014). Genetic Diversity of Coffee (*Coffea arabica* L.). Landraces from Southern Ethiopia as Revealed by Inter Simple Sequence Repeat Marker. *Gl. Adv. Res. J. Agr. Sc.* 3: 24-34.

Vekemans X (2002). AFLP-SURV version 1.0. Laboratoire de Genetique et Ecologie Vegetale, Universite Libre de Bruxelles, Bruxelles. Available at: http://www.ulb.ac.be/sciences/lagev/aflp-surv.html. Accessed November 3, 2020.

Vieira ES, Pinho EV, Carvalho MG, Esselink DG, et al. (2010). Development of microsatellite markers for identifying Brazilian *Coffea arabica* varieties. *Genet. Mol. Biol.* 33: 507-514. DOI: 10.1590/S1415-47572010005000055.

Vos P, Hogers R, Blesskers M, Reijans M, et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res*. 23: 407-4414. DOI: 10.1093/nar/23.21.4407.

Wang X, Rinehart TA, Wadl PA, Spiers JA, et al. (2009). A new electrophoresis technique to Separate microsatellite alleles. *Afr. J. Biotec.* 8: 2432-2436. DOI 10.5897/AJB08.735.

Wright S (1965). The interpretation of population structure by F-statistics with special regard to Systematic systems of mating. *Evolution.* 19: 395-420. DOI: 10.1111j.1558-5646.tb01731.x.

 