

## ESTUDO DA DIVERSIDADE GENÉTICA DE UMA POPULAÇÃO DE *Coffea canephora* VAR. CONILON POR MEIO DE MARCADORES MOLECULARES DO TIPO SSR<sup>1</sup>

Jean Carlos Alekcevetch<sup>2</sup>, Fernanda de Araújo Carneiro<sup>3</sup>, Érica Cristina da Silva Rêgo<sup>4</sup>, Antonio Fernando Guerra<sup>5</sup>, Gabriel Ferreira Bartholo<sup>6</sup>, Maria Amélia Gava Ferrão<sup>7</sup>, Aymbiré Francisco Almeida da Fonseca<sup>8</sup>, Pierre Marraccini<sup>8</sup>, Alan Carvalho Andrade<sup>9</sup>

<sup>1</sup> Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café e INCT-Café (CNPq/FAPEMIG).

<sup>2</sup> Bolsista CAPES, MSc, [jeancarlosembrapa@gmail.com](mailto:jeancarlosembrapa@gmail.com)

<sup>3</sup> Bolsista Consórcio Pesquisa Café, [fearca14@gmail.com](mailto:fearca14@gmail.com)

<sup>4</sup> Bolsista Consórcio Pesquisa Café, [ecsr@gmail.com](mailto:ecsr@gmail.com)

<sup>5</sup> Pesquisador, PhD, Embrapa Café, Brasília-DF, [antonio.guerra@embrapa.br](mailto:antonio.guerra@embrapa.br)

<sup>6</sup> Pesquisador, PhD, Embrapa Café, Brasília-DF, [gabriel.bartholo@embrapa.br](mailto:gabriel.bartholo@embrapa.br)

<sup>7</sup> Pesquisador, PhD, Embrapa Café, Brasília-DF, [maria.ferrao@embrapa.br](mailto:maria.ferrao@embrapa.br)

<sup>8</sup> Pesquisador, PhD, Embrapa Café, Brasília-DF, [aymbire.fonseca@embrapa.br](mailto:aymbire.fonseca@embrapa.br)

<sup>9</sup> Pesquisador, PhD, CIRAD UMR AGAP, Montpellier – FR, [marraccini@cirad.fr](mailto:marraccini@cirad.fr)

<sup>10</sup> Pesquisador, PhD, Embrapa Recursos Genéticos e Biotecnologia (LGM), Brasília-DF, [alan.andrade@embrapa.br](mailto:alan.andrade@embrapa.br)

**RESUMO:** O *Coffea canephora* é uma das duas espécies cultivadas comercialmente no mundo. A produção nacional é muito importante, com uma produção estimada a cerca de 26% do total de café beneficiado produzido no mundo para a safra de 2013. O estudo molecular da diversidade genética vem se demonstrando cada vez mais eficiente, necessário e refinado, uma vez que as técnicas moleculares estão evoluindo e os descritores morfológicos estão se tornando insuficientes para a descrição de um indivíduo. Os marcadores moleculares microssatélites ou SSR (“*Simple Sequence Repeats*”), são marcadores co-dominantes, utilizados com uma frequência cada vez maior para estudos de diversidade de populações. O objetivo desse trabalho foi avaliar a diversidade genética de cerca de 48 clones parentais de *Coffea canephora* var. Conilon presentes no Banco Ativo de Germoplasma – BAG do Incaper, bem como uma população de 266 plantas de *Coffea canephora* var. Conilon, correspondentes a uma descendência gerada por tais parentais, e mantidas no campo experimental da Embrapa Cerrados. O DNA genômico dessas plantas foi extraído, e analisado por PCR usando marcadores moleculares do tipo microssatélites (SSR). Os produtos de PCR foram analisados por meio do sequenciador automático (ABI 3130xl) para avaliar o tamanho dos fragmentos amplificados. Esses resultados permitiram demonstrar que existe uma ampla variabilidade genética entre as plantas parentais e aquelas da população descendente. Os resultados permitiram também realizar uma análise de paternidade, utilizando-se o software Cervus, sendo possível identificar o parental com maior ocorrência (24,44% de frequência) na população descendente.

**PALAVRAS-CHAVE:** *Coffea canephora*, diversidade genética, marcadores moleculares, SSRs.

### STUDY OF GENETIC DIVERSITY OF *COFFEA CANEPHORA* POPULATION USING SSR MOLECULAR MARKERS

**ABSTRACT:** *Coffea canephora* is one of the two commercially cultivated coffee species in the world. Domestic production is very important, with a production estimated at about 26% of the total coffee produced in the world from the harvest of 2013. Studies of molecular genetic diversity have been demonstrated to be increasingly efficient and refined, since molecular techniques are evolving and morphological descriptors are becoming insufficient to describe an individual. The microsatellite markers or SSRs (“*Simple Sequence Repeats*”) are co-dominant markers, used with increasing frequency for studies of population diversity. The aim of this study was to evaluate the genetic diversity of about 48 parental clones of *C. canephora* var. Conilon present in the Active Germplasm Bank - BAG of Incaper and a population of 266 plants corresponding to a seed-pool offspring generated by such parents, and maintained in the experimental field of Embrapa Cerrados. The genomic DNA of these plants was extracted and analyzed by PCR using microsatellite markers (SSR). The PCR products were analyzed using an automated sequencer (ABI 3130xl) to evaluate the size of the amplified fragments. These results allowed us to demonstrate that there is a large genetic variability among the parental plants and those of the descending population. The results also allowed to perform a paternity analysis, using the software Cervus, which could identify the parent with higher prevalence (24.44% frequency) in the descending population.

**KEY WORDS:** *Coffea canephora*, genetic diversity, molecular markers, SSRs.

## INTRODUÇÃO

A introdução de métodos biotecnológicos tem grande potencial para auxiliar os programas de melhoramento genético a obter rápidos avanços na geração de novas cultivares, principalmente em culturas perenes, como o cafeeiro. A seleção genética com base em dados fenotípicos avaliados a campo constitui o fundamento do melhoramento genético convencional do cafeeiro, tanto para a definição dos cruzamentos a serem realizados visando-se à geração de novos genótipos, como na identificação dos indivíduos superiores a serem usados comercialmente. A grande vantagem da utilização da genética molecular pelo melhoramento genético aplicado é a utilização direta das informações das sequências de DNA (particularmente/especialmente dos polimorfismos dessas sequências entre indivíduos ou populações) na seleção, propiciando alta eficiência na mesma, rapidez na obtenção de ganhos genéticos e baixos custos (após identificação dos marcadores), quando comparados com a seleção baseada em fenotipagem. Os marcadores moleculares apresentam muitas características interessantes e proveitosas no melhoramento genético, pois não interagem com o ambiente, são passíveis de aplicação em qualquer estágio de desenvolvimento e conseguem distinguir indivíduos heterozigotos (marcadores co-dominantes). Revelando-se assim um recurso muito eficiente não só na SAM, mas também na identificação de indivíduos, proteção de cultivares, análise de paternidade e de diversidade genética (REZENDE *et al.* 2012). No sentido de integrar ferramentas moleculares aos programas de melhoramento genético do cafeeiro, os objetivos deste trabalho foram de utilizar marcadores moleculares microsatélites (chamado também de “SSR” para “Simple Sequence Repeats”), para estudar a diversidade genética de 48 indivíduos presente no BAG do INCAPER de uma população (descendência) de *C. canephora* var. Conilon (266 plantas) obtida por polinização aberta dos 48 parentais e, mantida nos campos experimentais da Embrapa Cerrados e de avaliar o potencial desses marcadores na identificação dos genitores dos indivíduos analisados na população.

## MATERIAL E MÉTODOS

**Material:** Foram coletadas folhas frescas de 48 plantas de *C. canephora* var. Conilon parentais presente no banco ativo de germoplasma do INCAPER, bem como as folhas de 266 plantas de uma população ( $\pm$  3500 indivíduos) mantida na Embrapa Cerrados e obtidas a partir de sementes geradas pela polinização aberta entre as plantas parentais. As folhas foram armazenadas em a  $-20^{\circ}\text{C}$  até extrair o DNA genômico (DOYLE e DOYLE, 1990). Os DNA genômicos dos sub-grupos SG1, SG2 e Guineano de *C. canephora* foram fornecidos por o Dr T. Leroy (Cirad UMR AGAP).

### Marcadores moleculares fluorescentes SSR:

Os *primers* utilizados no presente estudo foram descritos e utilizados previamente alhos (PONCET *et al.* 2004, 2007; CUBRY *et al.* 2008, 2012). Esses marcadores foram sintetizados com marcação fluorescente na extremidade 5' com diferentes fluoróforos (6-FAN, NED e HEX) de maneira que somente o *forward* ou o *reverse* foram marcados e cada qual, somente com um fluorocromo.

Tabela 1: *primers* SSR utilizados. Os nomes dos *primers* são indicados, assim como as sequências, os *motifs* repetitivos (Rep), o fluoróforo (Flr.) utilizado na marcação e, o número de acesso da sequência no banco de dados GenBank (GB).

Nome	Sequencia	Rep.	Flr.	GB n°
Me_13_F	AGAGGGATGTCAGCATAA	CA, CT	6-FAN	AJ871892
Me_13_R	ATTGTGTTGGTAGATGTG			
Mg_257_F	GACCATTACATTTACACACAC	CA	6-FAN	AJ250257
Mg_257_R	GCATTTTGTGCACACTGTA			
Mg_360_R	ACATTTGATTGCCTCTTGACC	CA	6-FAN	AM231555
Mg_360_F	ACAGTAGTATTCATGCCACATCC			
Mg_501_F	CACCACCATCTAATGCACCT	TG	6-FAN	AM231576
Mg_501_R	CTGCACCAGCTAATTCAAGC			
Mg_753_F	GGAGACGCAGGTGGTAGAAG	CA	6-FAN	AJ308753
Mg_753_R	TCGAGAAGTCTTGGGGTGTT			
Mg_358_F	CATGCACTATTATGTTTGTGTTTT	CA	HEX	AM231554
Mg_358_R	TCTCGTCATATTTACAGGTAGGTT			
Mg_368_R	TCCTACCTACTTGCCTGTGCT	TG	HEX	AM231558
Mg_368_F	CACATCTCCATCCATAACCATTT			
Mg_445_F	CCACAGCTTGAATGACCAGA	CA	HEX	AM231567
Mg_445_R	AATTGACCAAGTAATCACCAGCT			
Mg_779_F	TCCCCATCTTTTCTTTCC	TG	HEX	AJ308779
Mg_779_R	GGGAGTGTTTTGTGTTGCTT			
Mg_461_F	CGGCTGTGACTGATGTG	AC	NED	AM231570
Mg_461_R	AATTGCTAAGGGTTCGAGAA			
Mg_477_F	CGAGGGTTGGGAAAAGGT	AC	NED	AM231574
Mg_477_R	ACCACCTGATGTTCCATTTGT			

### Condições de PCR

As reações de PCR foram realizadas em termociclador (BIO-RAD- T100™ Thermal Cycler) utilizando os reagentes do QIAGEN® Multiplex PCR Kit, seguindo as especificações de amplificação indicada pelo fabricante. As condições de

termociclagem foram realizadas integralmente conforme recomendações do fabricante (15 min 95°C, [30s 94°C, 90s 60°C, 90s 72°C] x 45 ciclos, e 10min, 72°C de extensão final).

### Análise dos dados

Em uma alíquota de 1µL de cada produto das PCRs, foram misturados 1µL de fragmentos padrões marcados com ROX (BRONDANI e GRATTAPAGLIA, 2001) (que servem como parâmetro para o software GeneMapper® V4.1 Applied Biosystem realizar os cálculos dos tamanhos dos fragmentos amplificados) e 8 µL de Formamida HI-DI Applied Biosystems, com posterior desnaturação em termociclador (95°C por 5 minutos). Em seguida, as amostras foram eletroinjetadas em um sequenciador automático modelo ABI3130 XL, (Applied Biosystems) utilizando-se capilares de sílica fundida de 36 cm de comprimento e 50µm de diâmetro preenchidos com polímero do tipo POP7 conforme recomendação do fabricante para análise de fragmentos. A visualização dos fragmentos resultantes dos eletroferogramas foi realizada no programa GeneMapper® V4.1 Applied Biosystem. As informações de distância gênica foram obtidas após análises no GenAlex - Microsoft Excel (versão 6.5) (PEAKALL e SMOUSE 2012).

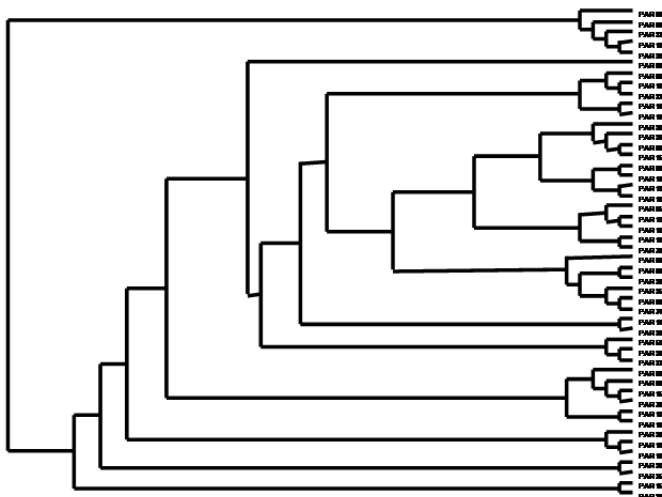
A frequência alélica, heterozigose esperada e observada, conteúdo informativo de polimorfismo (PIC), equilíbrio de Hardy-Weinberg, alelo nulo e probabilidades médias de exclusão de provável pai foram obtidos no programa Cervus (KALINOWSKI *et al.* 2007). A genotipagem obtida a partir da análise de fragmentos permite inferências sobre a distância genética entre os indivíduos. Foram realizadas diferentes comparações entre os parentais e os representantes de grupos de diversidade de *C. canephora*, previamente descritos. Todas as análises se deram a partir da submissão de dados de genotipagem à análise de distância genética do GenAlex. Para a construção dos dendrogramas, a similaridade foi calculada (1- valor da distância genética/100) e as matrizes de similaridade exportados para o programa *on-line* DendroUPGMA (GARCIA-VALLVE e PUIGBO 2013).

## RESULTADOS E DISCUSSÃO

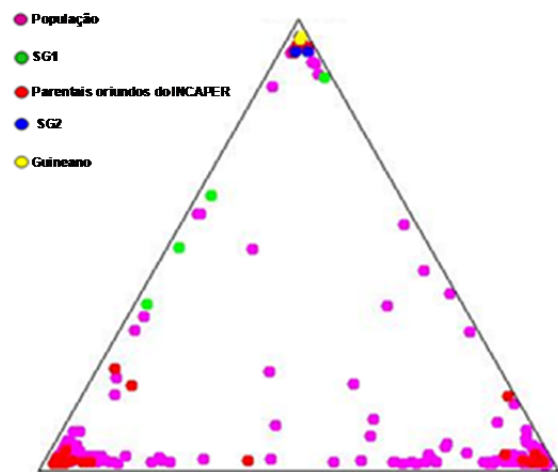
Os *primers* utilizados amplificaram fragmentos na maior parte dos indivíduos submetidos à análise sem haver problemas com as marcações o que permitiu comparar os tamanhos dos *amplicons* obtidos com os tamanhos esperados (CUBRY *et al.* 2012).

### Análise da distância gênica dos pais

Os resultados de clusterização por UPGMA, obtido pela análise da distância genética (NEI 1978), utilizando-se somente os dados dos parentais oriundos do INCAPER, mostram a formação de vários subgrupos, indicando a existência de diversidade genética nas plantas parentais (Figura 1).



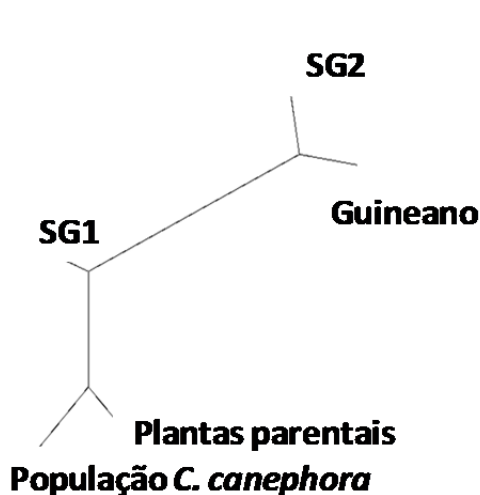
**Figura 1:** Dendrograma por agrupamento UPGMA, obtido a partir da matriz de similaridade genética dos parentais, construída com base na distância genética de NEI (1978), calculada a partir dos dados de genotipagem de 11 marcadores moleculares microsatélites.



**Figura 2:** Clusterização Bayesiana dos indivíduos da população (rosa), os parentais oriundos do INCAPER (vermelho), e os representantes dos grupos de diversidade de *C. canephora* (SG1, verde; SG2, azul e Guineano, amarelo).

Os resultados demonstram que a boa parte dos indivíduos (entre o PAR03 e o PAR32) possui coancestralidade. A planta PAR02 merece atenção especial pois ela apresenta-se isolada das outras plantas. O grupo contendo os indivíduos PAR01, PAR66, PAR333, PAR1111 e PAR 222, também se destaca por ser bastante distinto dos demais e não conter sub-agrupamentos, indicando que essas plantas são bastante próximas. Há também três sub-grupos menores, compostos apenas por dois indivíduos (PAR11 e PAR22, PAR30 e PAR35, PAR151 e PAR24). Esse comportamento indica que possivelmente sejam irmãos, ou seja, ao mesmo tempo em que há grande variabilidade genética dentro do BAG, há

também indivíduos geneticamente muito semelhantes. Foi feita também uma análise da similaridade dos parentais, com as plantas da população e plantas representantes dos grupos SG1, SG2 e Guineano de diversidade de *C. canephora* (Figura 2). Os resultados indicam claramente a formação de dois grupos distintos (vértices inferiores do triângulo) compostos dos parentais (vermelho) e indivíduos da população, além de indivíduos da população intermediários e dispersos entre os vértices da base. Além disto, os resultados mostram uma maior proximidade genética dos membros do grupo SG1 (pontos verdes), com os parentais e a população de Conilon estudada. Este fato, também pode ser claramente visualizado quando os dados são apresentados para representar a distância genética entre os clusters utilizando-se o algoritmo Neighbor-Joining (Figura 3). Estes resultados confirmam estudos anteriores de uma maior proximidade de *C. canephora* var. Conilon com os membros do grupo SG1 de Congolese (LAMBOT *et al.* 2008; CUBRY *et al.* 2008).



**Figura 3:** Representação da distância genética entre os 5 clusters identificados utilizando-se o algoritmo Neighbor-Joining (NJ), obtido por meio das análises com o software Structure 2.3.4.

Pais Candidatos	Frequência %	Pais Candidatos	Frequência %
PAR60	24,44	PAR28	2,26
PAR12	14,29	PAR70	2,26
PAR31	12,41	PAR74	2,26
PAR04	10,9	PAR10	1,88
PAR07	10,15	PAR15	1,88
PAR08_1	9,4	PAR16	1,88
PAR06	7,89	PAR29	1,88
PAR66	7,89	PAR30	1,88
PAR78	7,14	PAR24	1,5
PAR03	6,02	PAR32	1,5
PAR05	5,64	PAR59	1,5
PAR160	4,89	PAR1111	1,13
PAR151	4,51	PAR19	1,13
PAR09	4,14	PAR222	1,13
PAR13	4,14	PAR25	1,13
PAR130	4,14	PAR333	1,13
PAR148	4,14	PAR152	0,75
PAR23	3,76	PAR17	0,75
PAR121	3,38	PAR18	0,75
PAR147	3,38	PAR20	0,75
PAR11	3,01	PAR26	0,75
PAR14	3,01	PAR02	0,38
PAR08_2	2,26	PAR35	0,38
PAR22	2,26		

**Tabela 2:** Dendrograma por agrupamento UPGMA, obtido a partir da matriz de similaridade genética dos parentais, construída com base na distância genética de NEI (1978), calculada a partir dos dados de genotipagem de 11 marcadores moleculares microssatélites.

#### Análise de paternidade

Os resultados das análises de paternidade, realizada com software Cervus aos nível de confiança (80/95%), permitiram de identificar os parentais para todos os indivíduos analisados na população. Com base nesses resultados, foi possível identificar os parentais mais frequentes (Tabela 2). É evidente que os parentais possuem participação genética em quantidade bastante variada, mas ao mesmo tempo, muito discrepante, variando de 0,38% a 24,44% de recorrência. Pode-se observar que os parentais PAR 60, PAR 12, PAR 31, PAR04 e PAR07, foram os que apresentaram maior ocorrência. Isto pode indicar que estes parentais tem uma compatibilidade e/ou sincronismo de florescimento com um maior número de plantas da amostra das 48 plantas parentais.

#### CONCLUSÕES

- As estimativas de parâmetros genéticos dos marcadores moleculares utilizados nesse estudo mostraram que pelo menos oito dos onze marcadores microssatélites são informativos e adequados para os estudos realizados neste trabalho.
- Os estudos de diversidade genética mostraram que a população de *C. canephora*, contém considerável variabilidade genética.
- Os resultados das análises realizadas com o software Structure 2.3.4 confirmaram estudos prévios de uma similaridade de *C. canephora* var. conilon com membros do grupo de diversidade SG1.
- Os resultados das análises de paternidade foram eficientes em identificar os prováveis parentais de cada indivíduo da população, assim como descrever parentais com maior ocorrência.
- O parental “PAR60” é a planta mais recorrente na população, indicando que o mesmo tem elevada taxa de polinização e fecundação, sendo talvez uma planta recomendável para compor uma variedade clonal.

---

## REFERÊNCIAS BIBLIOGRÁFICAS

- BRONDANI, R. P e GRATTAPAGLIA, D. Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. **Biotechniques**. v.31, n. p.793- 800. 2001.
- CUBRY, P. *et al.* Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. **Genome**, v.51, n.1, p.50-63, 2008.
- CUBRY, P. *et al.* Global analysis of *Coffea canephora* Pierre ex Froehner (*Rubiaceae*) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. **Genetic Resources and Crop Evolution**, v.60, n.2, p.483-501, 2012.
- DOYLE, J. J.; DOYLE J. L. Isolation of plant DNA from fresh tissue. **Focus**, v.12, p.13-15, 1990.
- GARCIA-VALLVE, S; PUIGBO, P. DendroUPGMA: A dendrogram construction utility. Universitat Rovira i Virgili – URV Tarragona. Spain 2009. Disponível em: <http://genomes.urv.cat/UPGMA/index.php>
- KALINOWSKI, S.T. *et al.* Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. **Molecular Ecology**, v.16, n.5, p.1099-1106, 2007.
- LAMBOT, C. *et al.* Evaluation of Conilons for genetic diversity, cup quality and biochemical composition. 22<sup>nd</sup> International Conference on Coffee Science, **ASIC**, p.14-19, 2008.
- NEI, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics**. v.89, n.3, p.583-590, 1978.
- PEAKALL, R.; SMOUSE, P.E. GenAlex 6.5: Genetics analysis in Excel. Population genetic software for teaching and research – an update. **Bioinformatics**, v.28, n.19, p. 2537-2539, 2012;
- PONCET, V. *et al.* SSR cross-amplification and variation within coffee trees (*Coffea* ssp.). **Genome**, v.47, n.6, p.1071-1081, 2004.
- PONCET, V. *et al.* Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. **Genome**, v.50, n.12, p.1156-1161, 2007.
- REZENDE, J. C. *et al.* Biotecnologia em cafeeiro in **Biotecnologia aplicada à agropecuária**, 648p. 2012.