

ESTRUTURA E DIVERSIDADE GENÉTICA DE UMA POPULAÇÃO DE *Coffea canephora* CONILON UTILIZANDO MARCADORES SNPs IDENTIFICADOS POR nextRAD¹

Fernanda de Araújo Carneiro²; Érica Cristina Silva Rêgo³; Sinara Oliveira de Aquino⁴; Tatiana Santos Costa⁵; Edriana Araújo de Lima⁶; Omar Cruz Rocha⁷; Gustavo Costa Rodrigues⁸; Milene Alves de Figueiredo Carvalho⁹; Pierre Marraccini¹⁰; Gabriel Ferreira Bartholo¹¹; Antônio Fernando Guerra¹²; Orzenil Bonfim da Silva Jr¹³; Dario Grattapaglia¹⁴; Alan Carvalho Andrade¹⁵

¹ Trabalho financiado pelo consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café

² Bolsista CAPES, Doutoranda, Universidade Federal de Lavras, Lavras-MG, fearca14@gmail.com

³ Bolsista Consórcio Pesquisa Café, , BS, Universidade Paulista, Brasília-DF, ecsr@gmail.com

⁴ Bolsista CAPES, Doutoranda, Universidade Federal de Lavras, Lavras-MG, sinarinha2009@gmail.com

⁵ Pesquisadora, PhD, Universidade Federal de Lavras, Lavras-MG, tatanaitase@gmail.com

⁶ Pesquisadora, PhD, Universidade de Brasília, Brasília-DF, edrianaal@gmail.com

⁷ Pesquisador, PhD, Embrapa Cerrados, Planaltina-DF, omar.rocha@embrapa.br

⁸ Pesquisador, PhD, Embrapa Informática Agropecuária, Campinas-SP, gustavo.rodrigues@embrapa.br

⁹ Pesquisador, PhD, Embrapa Café, Brasília-DF, milene.carvalho@embrapa.br

¹⁰ Pesquisador, PhD, Cirad UMR AGAP, Brasília-DF, marraccini@cirad.fr

¹¹ Pesquisador, PhD, Embrapa Café, Brasília-DF, gabriel.bartholo@embrapa.br

¹² Pesquisador, PhD, Embrapa Café, Brasília-DF, antonio.guerra@embrapa.br

¹³ Pesquisador, Doutorando, Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, orzenil.silva@embrapa.br

¹⁴ Pesquisador, PhD, Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, dario.grattapaglia@embrapa.br

¹⁵ Pesquisador, PhD, Embrapa Café, Lavras-MG, alan.andrade@embrapa.br

RESUMO: Dentre as diferentes atividades ligadas ao negócio agrícola em nível mundial, o agronegócio cafeeiro está entre as de maior importância econômica e social, sendo o principal meio de subsistência para mais de 125 milhões de pessoas e produzido em mais de 60 países. A produção cafeeira comercial baseia-se principalmente em duas espécies: *Coffea arabica* e *Coffea canephora*. A grande variabilidade genética do *C. canephora* devido sua alogamia, é um fator de grande importância para os programas de melhoramento do cafeeiro, pois fornece uma fonte de novos genes. 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. Uma ferramenta muito importante no estudo da diversidade genética é o uso dos marcadores moleculares SNPs (Single Nucleotide Polymorphism), estes consistem na variação de sequência de DNA e podem ser identificados em praticamente todos os genes. No caso de *C. canephora*, estudos de diversidade podem facilitar a orientação dos programas de melhoramento na escolha de genitores para cruzamentos ou de genótipos para composição de variedades clonais. Este trabalho objetivou (i) avaliar e caracterizar, por meio da técnica de genotipagem nextRAD (Nextera-tagmented reductively-amplified DNA), a diversidade e a estrutura genética de indivíduos de *C. canephora* Conilon, pertencentes a uma população localizada em campo experimental da Embrapa Cerrados.

PALAVRAS CHAVE: *Coffea canephora*. Diversidade genética. SNPs. Genotipagem nextRAD.

GENETIC STRUCTURE OF THE POPULATION *Coffea canephora* Conilon USING SNPs MARKERS IDENTIFIED BY nextRAD

ABSTRACT: Of all the different activities related to agricultural industry worldwide, coffee agribusiness is among the most important ones, both economically and socially, being the main livelihood for more than 125 million people in more than 60 countries. Commercial coffee production is mostly based on two species, *Coffea arabica* and *C. canephora*. The high genetic variability of *C. canephora*, due to its level of allogamy, is of great importance for breeding programs of coffee as a source of novel alleles and co-evolved genetic combinations. The molecular analysis of genetic diversity has become increasingly efficient, necessary and refined, as molecular techniques have evolved and morphological traits do not supply sufficient information to fully describe an individual. An important tool in the study of genetic diversity is the use of molecular markers SNPs (Single Nucleotide Polymorphism), these consist of DNA sequence variation and can be identified in almost all the genes. In the case of *C. canephora*, diversity studies can facilitate orientation of breeding programs in selecting parents for crosses or genotypes for composition of clonal varieties. This study aimed to (i) evaluate and characterize, by genotyping technique nextRAD (Nextera-tagmented reductively-amplified DNA), the diversity and the genetic structure of *C. canephora* Conilon individuals belonging to a population located in the experimental field of Embrapa Cerrados.

KEYWORDS: *Coffea canephora*. Genetic diversity. SNPs. Genotyping nextRAD.

INTRODUÇÃO

O melhoramento convencional, utilizado para o cafeeiro, é um processo longo, necessitando de, no mínimo, 30 anos até a obtenção de uma nova cultivar. A identificação de cultivares mais produtivas baseia-se no processo de seleção e recombinação apoiando-se em características fenotípicas cujas herdabilidades e correlações são variáveis, fazendo com que sempre exista um nível de incerteza associado à seleção direcional (Mishra e Slater, 2012).

Na condução de qualquer programa de melhoramento genético, um dos pré-requisitos é a existência e o conhecimento da diversidade genética da espécie, desta forma, são realizados os planejamentos e definidas as estratégias de trabalho, como a identificação e seleção fenotípica de indivíduos superiores em populações naturais, avaliações de indivíduos de interesse para a obtenção de variedades clonais, seleção recorrente intrapopulacional, hibridação intraespecífica, entre outras (Bered et al., 1997; Ferrão et al., 2000; Fonseca et al., 2001).

O estudo da diversidade genética fundamentado em caracteres morfológicos pode não ser confiável, devido à influência que sofrem pelo ambiente, já a análise da diversidade genética utilizando marcadores moleculares é independente de fatores ambientais, pois esses não variam de acordo com o ambiente. Devido a essa relevante característica, há um aumento de interesse na identificação de um grande número de marcadores moleculares para uma rápida aplicação na avaliação da diversidade genética e seleção de genótipos desejados (Gupta et al., 2012).

O desenvolvimento de métodos eficazes para a detecção de polimorfismos de base única (*Single Nucleotide Polymorphism*-SNPs) e pequenos eventos de inserção e deleção (Indels) levou a uma revolução no seu uso como marcadores moleculares (Batley et al., 2003). Considerado o tipo mais comum de polimorfismo de DNA, sua ocorrência e distribuição ao longo do genoma varia entre as espécies em função de diversos aspectos relacionados com o sistema preferencial de reprodução, história evolutiva da espécie e da população alvo do estudo. Podendo ser identificados em regiões gênicas ou não.

Recentemente, um novo método de genotipagem por sequenciamento (*Genotyping by Sequencing: GBS*) denominado nextRAD (*Nextera-tagmented reductively-amplified DNA*) foi lançado pela empresa SNPsaurus. Esta nova técnica de *GBS* difere das demais por utilizar um método rápido de preparação de bibliotecas de sequenciamento (Nextera) e não utilizar enzimas de restrição para a redução da complexidade genômica, mas sim se basear em PCR seletiva. Esta metodologia possibilitou, utilizando marcadores SNPs, a realização de estudos de diversidade e estrutura genética em uma população de *C. canephora* Conilon.

MATERIAL E MÉTODOS

Para verificar a estrutura e a diversidade genética da população de *C. canephora* foram selecionados 436 indivíduos, dentre eles 44 possíveis parentais advindos diretamente do BAG do INCAPER, 388 indivíduos selecionados de uma população de melhoramento da Embrapa Cerrados, originada de um pool de sementes coletadas dos genitores do INCAPER e quatro outros clones de *C. canephora* 14, 22, 73 e 120. O DNA genômico das folhas dos indivíduos selecionados foi extraído pelo método de Doyle e Doyle (1990) adaptado para extração em rack de 96 mini-tubos e enviados para genotipagem, realizada pela empresa SNPsaurus (<http://snpsaurus.com/>). A genotipagem nextRAD identificou um total de 11.230 SNPs com *call rate* acima de 80%. Destes, foram selecionados 573 marcadores que estavam em equilíbrio de Hardy – Weinberg (EHW) com *call rate* acima de 90% e frequência alélica mínima (MAF) igual ou superior a 20%.

Os dados de genotipagem nextRAD dos 573 marcadores SNPs foram analisados pelo software Cervus (Kalinowski et al., 2007), que realizou o cálculo das frequências alélicas e, por meio dessas, forneceu dados como heterozigosidade observada (H_o), heterozigosidade esperada (H_e) e conteúdo de informação de polimorfismo (PIC).

Para verificar a estrutura populacional foram realizadas análises por meio de duas abordagens, pelo método Bayesiano implementado pelo software *Structure* v.2.3.4 (Pritchard et al., 2000) e pela análise discriminante de componentes principais (DAPC) do software *adegenet* (Jombart, 2008; Jombart et al., 2010). Em ambas as abordagens, foram utilizados os mesmos marcadores e os mesmos indivíduos.

RESULTADOS

Ao observar a distribuição do número de marcadores SNPs identificados pela técnica nextRAD com o tamanho dos cromossomos (Mbp) por meio de regressão linear (Figura 1), nota-se que o número de SNPs em cada cromossomo está relacionado ao tamanho do mesmo ou seja, cromossomos maiores possuem um maior número de marcadores SNPs identificados. No cromossomo 0, com tamanho aproximado de 200 Mbp, foram identificados cerca de 2.750 marcadores, seguido pelo cromossomo 2 com mais de 1.300 SNPs identificados. O menor cromossomo de *C. canephora*, cromossomo 9, com aproximadamente 22 Mbp, um total de 497 SNPs foram identificados.

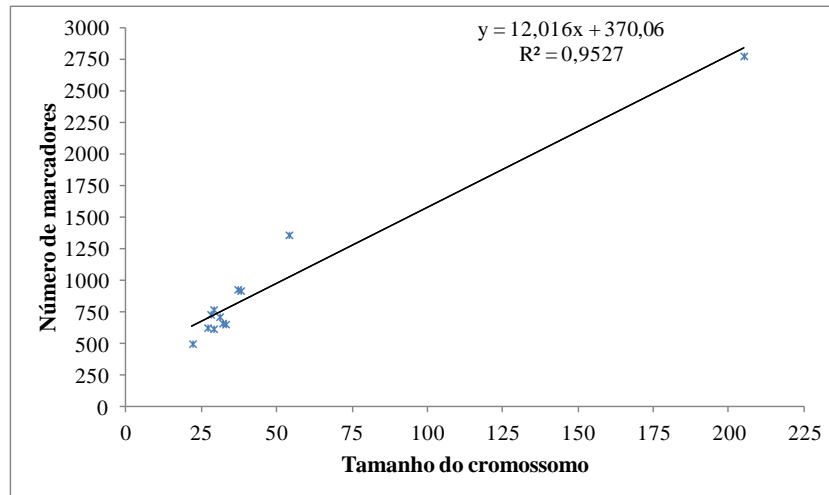


Figura 1. Correlação entre número de marcadores SNPs mapeados e tamanho dos cromossomos (Mbp) de *C. canephora*.

A média de H_o foi de 0.41 e variou de 0.237 a 0.555. Em aproximadamente 85% dos locos, a H_o estimada foi menor que 0.50. A H_e variou de 0.238 a 0.5 com média de 0.405. Neste caso, aproximadamente 41% dos SNPs apresentaram H_e maior ou igual a 0.45. Além da heteroziguidade, os valores de conteúdo de informação de polimorfismo (PIC) também permitiram quantificar o polimorfismo genético dos locos em análise, este conhecimento é importante na identificação dos marcadores mais informativos. Os valores de PIC variaram de 0.209 a 0.375 e com valor médio de 0.32. Mais de 64% dos SNPs analisados se mostraram informativos, apresentando valores de PIC superiores a 0.3. Na definição da estrutura populacional, por meio do programa *Structure*, foi possível a alocação dos 436 indivíduos da população estudada de *C. canephora* Conilon em três diferentes grupos majoritários, conforme apresentado na Figura 2.

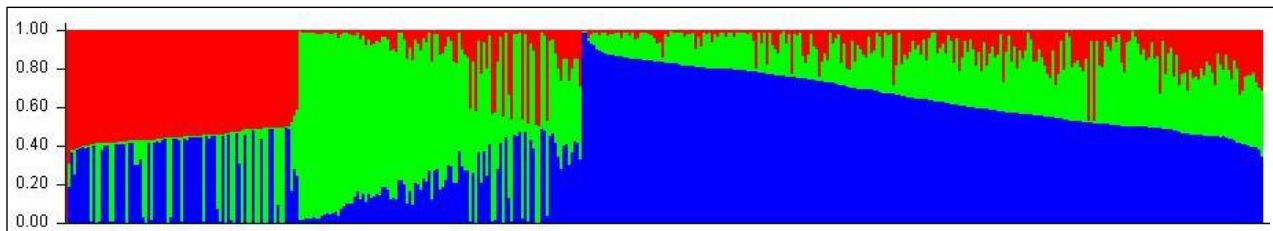


Figura 2. Resultado da atribuição dos 436 indivíduos de *C. canephora* em três grupos definidos pelo programa *Structure*. As cores vermelha, verde e azul representam os grupos 1, 2 e 3 de *C. canephora*, respectivamente.

O primeiro grupo é formado por 85 indivíduos, possuindo apenas um indivíduo parental, PAR148. O segundo grupo é composto por 103 indivíduos, incluindo 9 parentais, e o terceiro e maior grupo, inclui os 4 clones de *C. canephora* e 34 parentais, agrupando um total de 248 indivíduos.

Na análise DAPC, realizada pelo programa *adegenet*, os mesmos 436 indivíduos da população de *C. canephora* Conilon foram separados em 10 grupos, conforme demonstrado nas Figuras 3 e 4.

A Figura 3 apresenta o número de grupos definidos de acordo com o Critério de Informação Bayesiana (BIC), em que o seu menor valor representa o número provável de grupos, neste caso, 10 grupos.

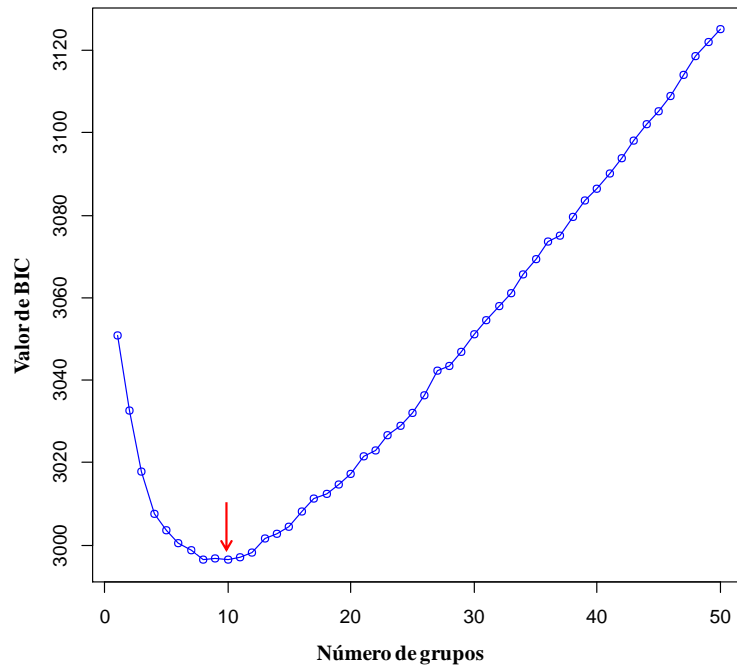


Figura 3 . Inferência do número de grupos obtidos na DAPC com 573 SNPs. A seta vermelha indica o melhor valor de K (menor valor de BIC).

Na Figura 4 estão representados por diferentes cores e números os 10 grupos estabelecidos pelo BIC, retendo 200 componentes principais (PCA eigenvalues) e 4 análises discriminantes (DA eigenvalues), onde se observa uma proximidade entre os grupos 1 (azul), 3 (verde claro), 4 (amarelo escuro), 5 (amarelo claro) e 8 (vermelho). Os grupos 7 (marrom) e 9 (verde escuro) também estão próximos entre si e os demais grupos se mantêm isolados.

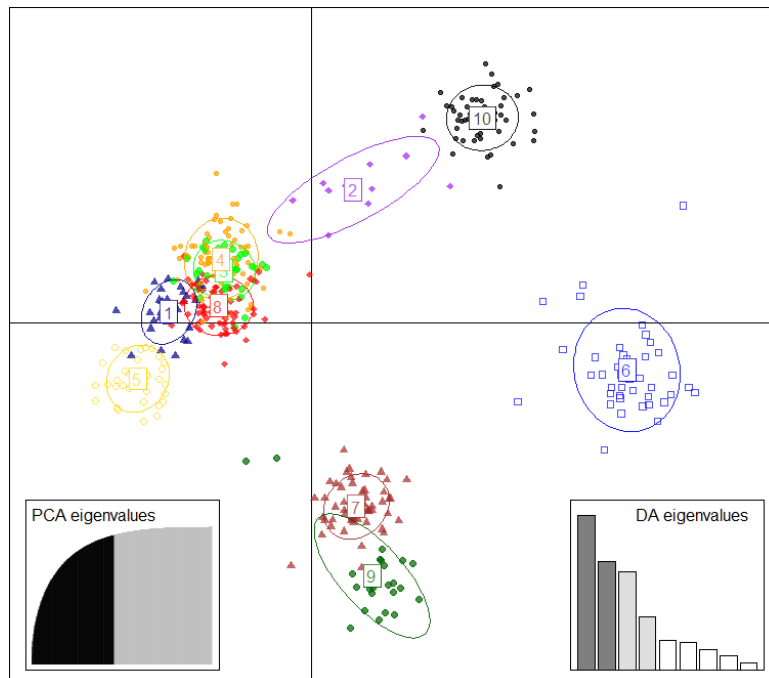


Figura 4 . Gráfico de dispersão DAPC com base na análise de 436 indivíduos da população de *C. canephora* utilizando 573 marcadores SNPs. Os grupos estão representados por cores e números diferentes, no canto esquerdo inferior o número de componentes principais retidas (PCA eigenvalues) e no canto direito inferior o número de análises discriminantes retidas (DA eigenvalues).

CONCLUSÕES

- A técnica de genotipagem nextRAD se mostrou satisfatória, identificando grande número de marcadores SNPs distribuídos em todos os cromossomos de *C. canephora* a partir de baixa concentração de DNA genômico;

- Mais da metade dos SNPs identificados se mostraram informativos, com valores de PIC iguais ou superiores a 0,3;
- Os estudos de diversidade genética mostraram que a população estudada neste trabalho contém considerável variabilidade genética, diferenciando em 3 e 10 grupos com as análises realizadas pelos programas *Structure* e *adegenet*, respectivamente.

AGRADECIMENTOS

À CAPES pelo apoio financeiro e ao Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER) pelo provimento de parte do material vegetal de *C. canephora* utilizado neste estudo.

REFERÊNCIAS

- BATLEY, J. et al. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiology*, v. 132, n. 1, p. 84-91, May 2003.
- BERED, F. et al. Marcadores moleculares e sua aplicação no melhoramento genético de plantas. *Ciência Rural*, v. 27, n. 2, p. 513-520, 1997.
- DOYLE, J. J.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. *Focus*, v. 12, p. 13-15, 1990.
- FERRÃO, R. G. et al. Banco ativo de germoplasma de *Coffea canephora*, variedade Conilon, no estado do Espírito Santo., I SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 2000, Poços de Caldas. p.405-408.
- FONSECA, A. F. A. et al. Melhoramento genético de *Coffea canephora* no estado do Espírito Santo, II SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 2001, Vitória. p.1379-1384.
- GUPTA, P. et al. Discovery and use of single nucleotide polymorphic (SNP) markers in *Jatropha curcas* L. *Molecular Breeding*, v. 30, n. 3, p. 1325-1335, Oct 2012.
- JOMBART, T. et al. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, v. 11, 94, Oct 15 2010.
- JOMBART, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, v. 24, n. 11, p. 1403-1405, Jun 1 2008.
- 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, Mar 2007.
- MISHRA, M. K.; SLATER, A. Recent advances in the genetic transformation of coffee. *Biotechnology Research International*, v. 2012, article ID, 580857, 17 p., 2012.
- PRITCHARD, J. K. et al. Inference of population structure using multilocus genotype data. *Genetics*, v. 155, n. 2, p. 945-959, Jun 2000.