

## MARCADORES SNP AVALIADOS EM CULTIVARES DE CAFÉ ARÁBICA<sup>1</sup>

Luiz Filipe P. Pereira<sup>2</sup>; Juliana C. M. Schenk<sup>3</sup>; Mirian P. Maluf<sup>4</sup>; Gustavo C. Sant'ana<sup>5</sup>; Livia Nogueira<sup>6</sup>; Ligia A. Pereira<sup>7</sup>; Bruna S. R. da Silva<sup>8</sup>; Paula de S. Guimarães<sup>9</sup>; Oliveiro Guerreiro Filho<sup>10</sup>; Lilian Padilha<sup>11</sup>

<sup>1</sup> Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café

<sup>2</sup> Pesquisador, PhD, Embrapa Café, Brasília-DF, [filipe.pereira@embrapa.br](mailto:filipe.pereira@embrapa.br)

<sup>3</sup> Bolsista CNPq/INCT, Embrapa Café/IAC, [julianamartinati@hotmail.com](mailto:julianamartinati@hotmail.com)

<sup>4</sup> Pesquisador, PhD, Embrapa Café, Brasília-DF, [mirian.maluf@embrapa.br](mailto:mirian.maluf@embrapa.br)

<sup>5</sup> Pesquisador, DSc, TMG-Tropical Melhoramento & Genética, [gustavocsantana80@gmail.com](mailto:gustavocsantana80@gmail.com)

<sup>6</sup> Bolsista CAPES, – Pós-Graduação Genética e Biologia Molecular -UEL, [livianogue@hotmial.com](mailto:livianogue@hotmial.com)

<sup>7</sup> Bolsista Consórcio Pesquisa Café, CPG, [ap\\_ligia@hotmail.com](mailto:ap_ligia@hotmail.com)

<sup>8</sup> Bolsista Consórcio Pesquisa Café, DSc, Embrapa Café/IAPAR [brunasilvestre@hotmial.com](mailto:brunasilvestre@hotmial.com)

<sup>9</sup> Bolsista Consórcio Pesquisa Café, CPG, [psguim@yahoo.com.br](mailto:psguim@yahoo.com.br)

<sup>10</sup> Pesquisador, PhD, Instituto Agrônomo Campinas, Bolsista CNPq DT, [oliveiro@iac.sp.gov.br](mailto:oliveiro@iac.sp.gov.br)

<sup>11</sup> Pesquisador, DSc, Embrapa Café, Brasília-DF, [lilian.padilha@embrapa.br](mailto:lilian.padilha@embrapa.br)

**RESUMO:** A demanda crescente por cafés especiais e os melhores preços alcançados por eles tem estimulado cada vez mais a participação de produtores neste segmento. Além da origem, da sustentabilidade na produção e da avaliação dos cafés diferenciados quanto aos aspectos físicos e sensoriais, a rastreabilidade da produção é uma das principais exigências para a certificação do produto. Dentro deste processo, a crescente demanda pelo atestado da pureza varietal vem exigindo o desenvolvimento de ferramentas que possam garantir a diferenciação inequívoca do produto comercial, e que viabilizem sua proteção intelectual. Assim, nosso objetivo foi desenvolver marcadores SNP, identificados em análises de genotipagem pelo sequenciamento (GBS), voltados à discriminação de cultivares de café arábica. Quarenta e oito cultivares foram genotipadas pelo sequenciamento em plataforma Illumina e 192.730 TAGs foram alinhadas ao genoma de *Coffea arabica*, gentilmente cedido pelo Consórcio Internacional de Sequenciamento do Genoma do Arabica (ACGC). Após o alinhamento ao genoma de referência de *C. arabica* e aplicação dos filtros de qualidade, 1.181 SNP foram obtidos e utilizados para identificar a relação entre os cultivares através da análise de componentes principais (PCA). Esses marcadores confirmaram a estreita base genética das variedades de *C. arabica*, e foram eficientes em separar os grupos dos Bourbons das demais cultivares, em especial, de Catuaí e Mundo Novo. Um conjunto de 18 marcadores SNP com elevado conteúdo informativo foram anotados e estão sendo validados para confirmar seu potencial para análises de pureza varietal e genética das cultivares comercializadas no país.

**PALAVRAS-CHAVE:** *coffea arabica*, GBS, discriminação de cultivares.

### SNP MARKERS ASSESSED IN ARABICA COFFEE CULTIVARS

**ABSTRACT:** The increase in specialty coffees market and the better prices achieved by them stimulate the participation of coffee producers in this segment. Besides the origin, the sustainability of production, the evaluation for physical and sensory aspects, the traceability of production is one of the main requirements for certification of differentiated coffees. In this process, the demand for varietal purity certificate requires the development of tools that allow the unequivocal differentiation of the commercial product, as well as enable its intellectual protection. Therefore, our objective was to develop SNP markers, identified with a genotyping-by-sequencing (GBS) strategy, to discriminate the arabica coffee cultivars. After alignment with the *C. arabica* reference genome, kindly provided by Consortium ACGC, and application of quality filters, 1181 SNP were obtained and used to identify the relationship between cultivars through principal component analysis (PCA). These SNP markers confirmed the narrow genetic base of *C. arabica* varieties and were efficient in separating Bourbon groups from other cultivars, especially from Catuaí and Mundo Novo. A set of 18 SNP markers of high polymorphism information content was annotated and is being validated to confirm their potential for varietal purity analysis and genetic cultivars marketed in the country.

**KEY WORDS:** *coffea arabica*, GBS, cultivar discrimination.

### INTRODUÇÃO

O Brasil é o maior produtor mundial de café arábica e, atualmente, vem se destacando na produção de cafés sustentáveis e de qualidade. Segundo o Conselho dos Exportadores de Café do Brasil (CECAFÉ), em 2019 o Brasil exportou 4,5 milhões de sacas de cafés diferenciados, aqueles com qualidade superior ou algum tipo de certificado de práticas sustentáveis. O volume foi responsável por 19,1% do total exportado de janeiro a julho deste ano e representa um crescimento de 58,8% em relação mesmo período de 2018. A receita cambial gerada com a exportação destes cafés foi de US\$ 700,3 milhões, representando 23,9% do total gerado pelo Brasil com as exportações. Os principais destinos foram: EUA (24,1%); Alemanha (13,1%); Japão (12,4%); Bélgica (10%) e Itália (8,3%) (CECAFÉ, 2019).

Para o acompanhamento de cafés diferenciados no mercado do arábica, são exigidas certificações do sistema de produção e do produto, sendo a rastreabilidade um ponto chave nesse processo. Além da origem do café, tem-se observado uma demanda crescente por atestados da pureza varietal dos lotes rastreados. Estes atestados são uma garantia, tanto para o produtor quanto para o setor comercial, da origem genética dos cafés especiais. Embora existam descritores capazes de distinguir a grande maioria das cultivares em campo, ainda não há um método disponível para o sistema de produção capaz de identificar uma cultivar a partir de suas sementes ou grãos.

Marcadores de DNA podem ser utilizados para atestar a pureza genética e varietal de cultivares e, dentre suas principais vantagens podem ser citadas a ausência de influência pelo ambiente e a detecção a partir de qualquer tecido ou órgão da planta, além de superar limitações como o baixo polimorfismo observado entre as cultivares de café arábica. A viabilização de análises em larga escala tem favorecido o desenvolvimento de marcadores capazes de identificar polimorfismos de um único nucleotídeo ou SNP (*Single Nucleotide Polymorphism*), que são modificações pontuais e herdáveis nas sequências do DNA, que ocorrem em abundância no genoma e já foram avaliados para fins de caracterização em espécies como a soja (Liu et al. 2017), o algodão (Kung et al. 2016), cerejeira (Gonopoulos et al 2013), dentre outros.

Este trabalho apresenta os resultados iniciais do desenvolvimento de marcadores SNP, identificados em análises de genotipagem pelo sequenciamento (GBS), voltados à caracterização de cultivares comerciais e de uma coleção de Bourbons - que são considerados materiais superiores para a produção de cafés de elevada qualidade de bebida.

## MATERIAL E MÉTODOS

Quarenta e oito cultivares de *C. arabica* da Coleção de Germoplasma de Café do Instituto Agrônomo (IAC) foram genotipados pelo sequenciamento, sendo: três acessos de Bourbon Vermelho, 21 de Bourbon Amarelo, Caturra Amarelo IAC 476, Caturra Vermelho IAC 477, Típica, Amarelo de Botucatu, Laurina, Ibairi, Catuaí Amarelo IAC 62 e IAC 86, Catuaí Vermelho IAC 144, IAC 81, IAC 99 e IAC 144, Catuaí IAC SH3, Mundo Novo IAC 376-4, Mundo Novo IAC 379-19, Acaiaí, Ouro Verde, Icatu Vermelho IAC 4045, Icatu Amarelo IAC 2944, Tupi IAC 1669-33, IAC 125 RN, Obatã Amarelo IAC 4739, Obatã IAC 1669-20, IPR 100. O DNA das folhas foi extraído através do protocolo CTAB adaptado em nosso laboratório e as amostras em duplicatas foram submetidas à genotipagem pelo sequenciamento em uma linha da plataforma Illumina.

Os *reads* captados no sequenciamento por síntese foram alinhados ao genoma de *Coffea arabica*, gentilmente cedido pelo Consórcio Internacional de Sequenciamento do Genoma do Arabica (ACGC), com a utilização do *pipeline-GBS* Tassel (<https://www.maizegenetics.net/tassel>) que, juntamente com as ferramentas disponibilizadas na plataforma SNIPlay Pipeline V3 ([http://sniplay.southgreen.fr/cgi-bin/analysis\\_v3.cgi](http://sniplay.southgreen.fr/cgi-bin/analysis_v3.cgi)) foram utilizados para a avaliação da qualidade das sequências, aplicação dos filtros de *Call Rate*, MAF e Heterozigosidade, assim como, na identificação dos SNP e seus perfis nos materiais estudados.

## RESULTADOS E DISCUSSÃO

A partir da GBS de cultivares de café foram obtidos 467.836.745 *reads* que produziram 7.139.655 TAGs. Desses, 192.730 TAGs foram alinhados no genoma de referência de *C. arabica*, onde 5,78% alinharam em posições únicas e 50,66% em posições múltiplas. Após a aplicação dos filtros CQ de MAF, *Call rate* 0,8, Heterozigosidade (Het) <0,9 e retirada de multialelos, um total de 1.181 SNP foram selecionados para análises posteriores. O filtro de Het < 0,9 foi utilizado para garantir que um maior número de marcadores SNP estivessem disponíveis para os estudos de diversidade e estrutura genética dessa população. Foi realizada a proporção de SNP com transição/transversão no filtro escolhido (Het < 0,9) sendo encontrados 912 SNP com transição e 268 SNP com transversão (Figura 1).

A porcentagem de dados faltantes por indivíduo dos 48 cultivares variou de 0,0008 para um Bourbon Amarelo a 0,09 para Icatu Vermelho. A frequência alélica dos SNP remanescentes foi então calculada, e os marcadores no qual o alelo menor teve frequência abaixo de 0.22 foram descartados (MAF < 0.22).

A relação entre as cultivares estudadas com 1.181 SNP, foi avaliada a partir de uma análise de componentes principais (PCA). A dispersão das cultivares no plano PC1/PC2 (Figura 2) reforçou a alta similaridade genética das cultivares de *C. arabica*, e demonstrou que os SNP obtidos na análise de GBS foram eficientes em separar os grupos dos Bourbons das demais cultivares, em especial, das cultivares Catuaí e Mundo Novo. As duas formas botânicas típica e bourbon deram origem a todas as cultivares de *C. arabica* (Antony, et al. 2001, 2002), e pela PCA pode ser observada a separação da coleção de Bourbons em relação à variedade típica. As cultivares Icatu, que possuem o *C. canephora* em sua origem de cruzamentos, foram aquelas que mais se distanciaram das demais variedades.

O trabalho está em andamento e um conjunto de 18 marcadores SNP com elevado conteúdo informativo foram anotados, os primers desenhados para a confirmação e validação destes SNP em lavouras comerciais.

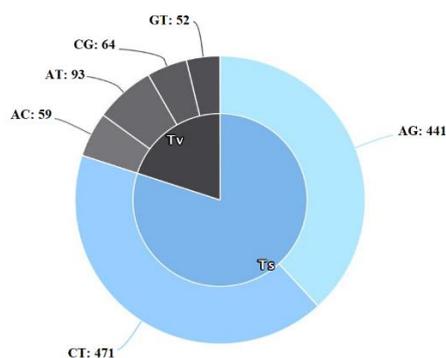


Figura 1. Proporção de SNP com transição/transversão em 48 genótipos de *C. arabica* genotipados pelo sequenciamento.

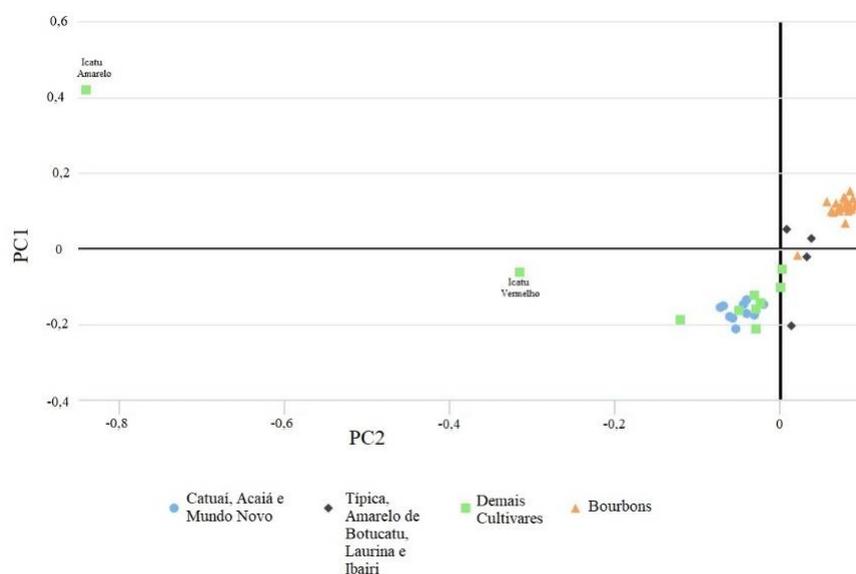


Figura 2: Análise de componentes principais-PCA para 1181 SNP selecionados em 48 cultivares de *C. arabica*.

## CONCLUSÕES

Os marcadores SNP identificados neste trabalho diferenciaram grupos de cultivares de *C. arabica* em relação à uma coleção de Bourbons.

## AGRADECIMENTOS

Os autores agradecem o apoio financeiro do Consórcio Pesquisa Café na concessão de bolsas, do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e da Coordenação de Aperfeiçoamento de Profissional de Ensino Superior - CAPES. Agradecemos ao Consórcio Internacional de Sequenciamento do Genoma do Arabica (ACGC) pelo genoma referência de *C. arabica*. Agradecemos ao Dr. Júlio César Mistro e à Masako T. Braghini do IAC-Café pelo auxílio nas amostragens do germoplasma dos Bourbons.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ANTHONY, F., BERTRAND, B., QUIROS, WILCHES, A., LASHERMES, P., BERTHAUD, J., CHARRIER, A. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. **Euphytica**, 118:53-65, 2001.
- ANTHONY, F., COMBES, M.C., ASTORGA, C. BERTRAND, B., GRAZIOSI, G, LASHERMES. The origin of cultivar *Coffea arabica* L variegates revealed by AFLP and SSR markers. **Theor Appl Genet**, 104:894–900. 2002. (Doi 10.1007/s00122-001-0798-8)

CONSELHO DOS EXPORTADORES DE CAFÉ DO BRASIL -CECAFÉ- **Relatório mensal – Julho 2019.**

Disponível em <http://cecafe.com.br> (acessado em 14/08/2019)

LIU, Z., LI, J., FAN, X., HTWE, N.M.P.S., et al. Assessing the numbers of SNP needed to establish molecular IDs and characterize the genetic diversity of soybean cultivars derived from *Tokachi nagaha*. **The Crop Journal**, 5, 326-336. 2017.

KUANG, M., WEI, S., WANG, Y., ZHOU, D., MA, L., FANG, Dan., YANG, W., MA, Z. Development of a core set of SNP markers for the identification of upland cotton cultivars in China. **Journal of Integrative Agriculture** 15(5): 954–962. 2016

GANOPOULOS, I., TSABALLA, A., XANTHOPOULOU, A., MADESIS, P., TSAFTARIS, A. Sweet cherry cultivar identification by High-Resolution-Melting (HRM) analysis using gene-based SNP markers. **Plant Mol Biol Rep.** 31:763–768. 2013. Brief Communication (doi 10.1007/s11105-012-0538-z)