

## DESENVOLVIMENTO E VALIDAÇÃO DE UMA PLATAFORMA DE GENOTIPAGEM EM LARGA ESCALA PARA *Coffea canephora*<sup>1</sup>

Fernanda de Araújo Carneiro<sup>2</sup>, Sinara Oliveira de Aquino<sup>3</sup>, Nathália Gomes Mattos<sup>4</sup>, Jéssica Coelho Valeriano<sup>5</sup>, Wendel William de Jesus Carneiro<sup>6</sup>, Pierre Marraccini<sup>7</sup>, Orzenil Bonfim da Silva Júnior<sup>8</sup>, Dario Grattapaglia<sup>9</sup>, Alan Carvalho Andrade<sup>10</sup>

<sup>1</sup> Trabalho financiado pelo Consórcio Pesquisa Café

<sup>2</sup> Doutora em Biotecnologia Vegetal, UFLA, Lavras-MG, Brasil, [fearca14@gmail.com](mailto:fearca14@gmail.com)

<sup>3</sup> Bolsista Pós-Doutorado PNP/CAPEs, PhD, UFLA, Lavras-MG, Brasil, [saquinobiotec@gmail.com](mailto:saquinobiotec@gmail.com)

<sup>4</sup> Doutoranda em Biotecnologia Vegetal, UFLA, Lavras-MG, Brasil, [nagm7@hotmail.com](mailto:nagm7@hotmail.com)

<sup>5</sup> Doutoranda em Biotecnologia Vegetal, UFLA, Lavras-MG, Brasil, [jessicacoelho\\_bio@hotmail.com](mailto:jessicacoelho_bio@hotmail.com)

<sup>6</sup> Bolsista Consórcio Pesquisa Café, [wendelwill11@gmail.com](mailto:wendelwill11@gmail.com)

<sup>7</sup> Pesquisador, PhD-HDR, CIRAD, UMR IPME, Instituto of Agricultural Genetics (AGI) LMI RICE2, Hanoi, Vietnã, [marraccini@cirad.fr](mailto:marraccini@cirad.fr)

<sup>8</sup> Pesquisador, PhD, Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, Brasil, [orzenil.silva@embrapa.br](mailto:orzenil.silva@embrapa.br)

<sup>9</sup> Pesquisador, PhD, Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, Brasil, [dario.grattapaglia@embrapa.br](mailto:dario.grattapaglia@embrapa.br)

<sup>10</sup> Pesquisador, PhD, Embrapa Café/INOVACAFÊ, Lavras-MG, Brasil, [alan.andrade@embrapa.br](mailto:alan.andrade@embrapa.br)

**RESUMO:** O cultivo de espécies de café como *Coffea canephora* e *C. arabica*, são muito importantes em regiões tropicais de todo o mundo e possuem grande valor econômico e cultural. As plataformas de genotipagem de SNPs oferecem ensaios altamente multiplexados a um custo relativamente baixo e podem ser uma ferramenta válida para a identificação dos marcadores associados a características de interesse. Neste trabalho é proposto o desenvolvimento e a validação de uma plataforma de genotipagem Affymetrix Axiom (®) para *C. canephora* (Coffee Axiom chip – 26K) com SNPs identificados a partir de dados de resequenciamento de *pools* de *C. canephora*, cobrindo a maior parte da diversidade genética conhecida para a espécie, e distribuídos ao longo de todo o genoma. Mais de 1.500 indivíduos de *C. canephora* foram genotipados com o chip para se avaliar a eficácia do mesmo. A grande maioria dos SNPs (20.920 ou 82%) pertence à classe de polimorfismos de alta resolução. O chip Coffee Axiom - 26K- se constitui uma ferramenta de referência para a pesquisa de genética molecular e genética de populações em espécies de café.

**PALAVRAS-CHAVE:** *C. canephora*, SNP, diversidade genética, DNA array.

## DEVELOPMENT AND VALIDATION OF A LARGE SCALE GENOTYPING PLATFORM FOR *Coffea canephora*

**ABSTRACT:** Coffee species such as *Coffea canephora* and *C. arabica* are important crops in tropical regions around the world, and has great economic and cultural value. Single nucleotide polymorphism arrays offer highly multiplexed assays at a relatively low cost per data point and can be a valid tool for the identification of the markers associated with traits of interest. Here, we describe the development and validation of a Affymetrix Axiom(®) genotyping array for *C. canephora* (Coffee Axiom chip – 26K) whose genome – wide distributed SNP content was discovered from pooled whole-genome resequencing of *C. canephora* accessions covering most of its known genetic diversity. More than 1,500 *C. canephora* individuals have been genotyped with the chip to assess the effectiveness of the array. A large majority of SNPs (20,920 or 82%) fell in the stringent class of poly high resolution polymorphisms. The Coffee Axiom chip – 26K will likely be a reference tool for molecular breeding and population genetics investigation within coffee species.

**KEY WORDS:** *C. canephora*, SNP, genetic diversity, DNA array.

## INTRODUÇÃO

O café é uma das *commodities* agrícolas mais importantes mundialmente e é o principal meio de subsistência para mais de 125 milhões de pessoas, sendo produzido em mais de 60 países. As duas espécies *C. arabica* e *C. canephora* predominam na produção mundial, representando aproximadamente 60% e 40% do mercado cafeeiro, respectivamente (ICO, 2018).

Com os avanços recentes na genômica do cafeeiro, como o sequenciamento completo do genoma de referência de *C. canephora* (DENOEU et al., 2014), uma redução significativa em tempo e custo na seleção de plantas com características de interesse ao melhoramento pode ser alcançada. O genoma de referência juntamente com as tecnologias de sequenciamento de nova geração (NGS) forneceram as ferramentas necessárias para a genotipagem de SNPs em larga escala. As plataformas de genotipagem de SNPs em larga escala resultam em uma alta cobertura e alta precisão, características essenciais para realizar a predição genômica e a descoberta de genes em estudos de associação genômica (GWAS).

O desenvolvimento de chips de genotipagem de DNA, com milhares de SNPs cobrindo todo o genoma, tem aumentado significativamente e já foram desenvolvidos para diversas culturas (BASSIL et al., 2015; BIANCO et al., 2016;

HULSE-KEMP et al., 2015; KONING-BOUCOIRAN et al., 2015; LEE et al., 2015). Neste trabalho foi desenvolvida e validada uma plataforma de genotipagem em larga escala para *C. canephora* (Coffee Axiom chip – 26K), contendo SNPs informativos e de alta qualidade. Em resumo, espera-se que o Coffee Axiom chip – 26K forneça uma base sólida para o estabelecimento da genotipagem de alto rendimento, o que é de grande importância para a pesquisa, avançando em estudos na área genética/genômica e em aplicações de melhoramento dentro do gênero *Coffea*.

## MATERIAL E MÉTODOS

A partir dos dados gerados pelo resequenciamento genômico de pools formados por indivíduos de *C. canephora*, abrangendo a maior parte da diversidade genética presente na espécie, obteve-se uma lista de 25.456 marcadores SNPs filtrados de alta qualidade que foram utilizados na construção da plataforma de genotipagem. Após o desenvolvimento da plataforma (Coffee Axiom chip – 26K), avaliou-se a eficácia do chip a partir da validação dos SNPs incluídos no mesmo. Para a validação, o DNA de 296 indivíduos de *C. canephora* foi extraído, conforme o protocolo descrito por Russel et al. (2010), e a plataforma Affymetrix Gene Titan® foi usada para genotipar o “conjunto de validação” com o chip desenvolvido. Vale ressaltar que esses materiais utilizados como “conjunto de validação” incluem os diferentes grupos de diversidade de *C. canephora*, as principais variedades clonais resultantes dos programas de melhoramento do INCAPER e Embrapa Rondônia, bem como representantes da população proveniente da Embrapa Cerrados.

O software Axiom™ Analysis Suite (v. 1.1.1.66) foi usado para o processamento dos dados de intensidade de hibridização, agrupamento e chamada de genótipos. Amostras que apresentaram valor de Dish QC (DQC) < 0.82 e call rate < 0.97 foram removidas da análise. Executando o Axiom Best Practices Genotyping Workflow os SNPs identificados foram classificados dentro de seis tipos, de acordo com critérios de qualidade. A plataforma (Coffee Axiom chip – 26K) desenvolvida e validada para *C. canephora* foi utilizada na genotipagem em larga escala de 1.319 indivíduos selecionados na população de melhoramento da Embrapa Cerrados, a partir de avaliações fenotípicas realizadas entre os anos de 2012 e 2014, levando em consideração características como produção, tamanho e formato de grãos, tamanho e formato de frutos cereja e peso de 100 grãos (CARNEIRO et al., 2013).

## RESULTADOS E DISCUSSÃO

Do conjunto total de sondas incluídas no chip (25.456), 25.411 cobriram todos os 11 cromossomos de *C. canephora* e o “cromossomo 0”, que consiste em um conjunto de *scaffolds* não ordenados, fornecendo uma boa representação do genoma total da espécie (Figura 1), com uma média de 2.117 sondas por cromossomo e 45 sondas foram excluídas por mapearem em múltiplos locais.

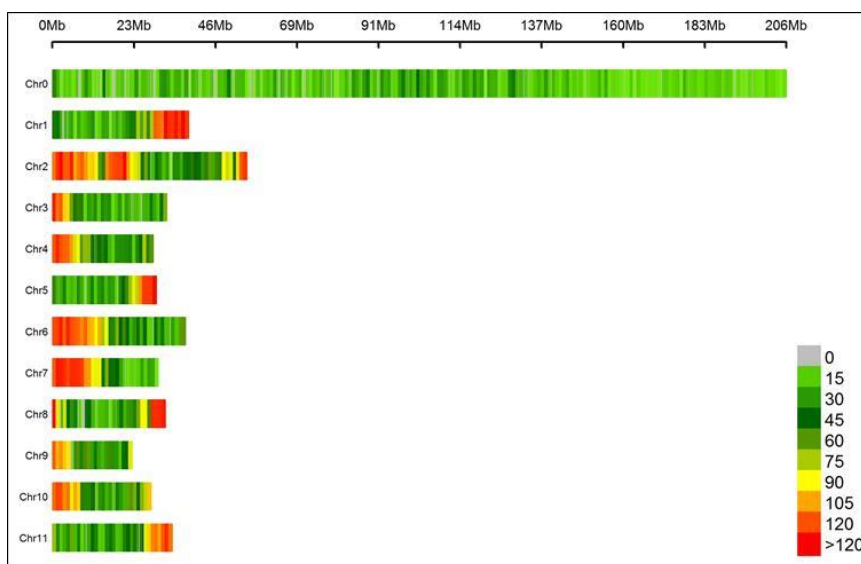


Figura 1. Distribuição dos SNPs contidos nas sondas para cada cromossomo de *C. canephora*. A quantidade de SNPs varia de 0 (em cinza claro) até >120 (em vermelho) SNPs por Mb para os 11 cromossomos de *C. canephora*, incluindo o cromossomo 0.

Verifica-se na Tabela 1 que o maior número de sondas localiza-se no cromossomo 2 (4.209), sendo este o maior cromossomo da espécie, com aproximadamente 54 Mb, enquanto que o menor número de sondas foram mapeadas no cromossomo 9 (1.257). Com relação às posições genômicas dos SNPs fixados no chip 26K de *C. canephora*, 68% estavam presentes em região gênica, sendo que desses, 42% estavam localizados em éxons (região codante – CDS) e 26% em íntrons. Os demais SNPs identificados no chip (8.348) estavam localizados em regiões intergênicas.

Tabela 1. Número de SNPs mapeados em cada cromossomo de *C. canephora*.

	Tam. cromossomo (Mb)	Nº de sondas mapeadas	Região Gênica		Região intergênica
			CDS	Íntron	
Chr 0	205.6	3.381	830	563	1.988
Chr 1	38.2	2.121	928	576	617
Chr 2	54.5	4.209	1.915	1.248	1.046
Chr 3	32.0	1.400	599	372	429
Chr 4	28.2	1.789	841	436	512
Chr 5	29.1	1.478	651	332	495
Chr 6	37.3	2.553	1.102	715	736
Chr 7	29.8	2.047	984	575	488
Chr 8	31.6	1.688	685	490	513
Chr 9	22.3	1.257	503	331	423
Chr 10	27.6	1.762	777	444	541
Chr 11	33.5	1.726	764	402	560
	<b>570</b>	<b>25.411</b>	<b>10.579</b>	<b>6.484</b>	<b>8.348</b>

A validação da plataforma de genotipagem apresentou mais de 97% de taxa de sucesso de amostragem, sendo que as 8 amostras excluídas não passaram devido ao QC *call rate*, que variou de 92,03 a 96,94 para essas amostras. As amostras incluídas como replicatas, dentro e entre placas, apresentaram uma concordância > 99,82 entre os resultados. Um total de 22.679 SNPs (89%) foram classificados como de excelente qualidade (Tabela 2; Figura 2). Destes, cerca de 90% eram polimórficos (PHR) (Figura 2A), os outros 10% eram marcadores que apresentavam somente dois *clusters* (NMH) (Figura 2B) e marcadores monomórficos (MHR) (Figura 2C). Os marcadores remanescentes (CRBT+OTV+Other) e que apresentam menor acurácia, representaram menos de 11% (2.777) do total de SNPs (Figuras 2D, 2E e 2F).

Tabela 2. Número e porcentagem de SNPs classificados em cada uma das seis categorias.

Categoria do SNP	Número de SNPs	% dos SNPs
PHR	20.920	82,18%
NMH	1.569	6,16%
MHR	190	0,74%
CRBT	571	2,24%
OTV	583	2,29%
Other	1.623	6,38%
<b>TOTAL</b>	<b>25.456</b>	

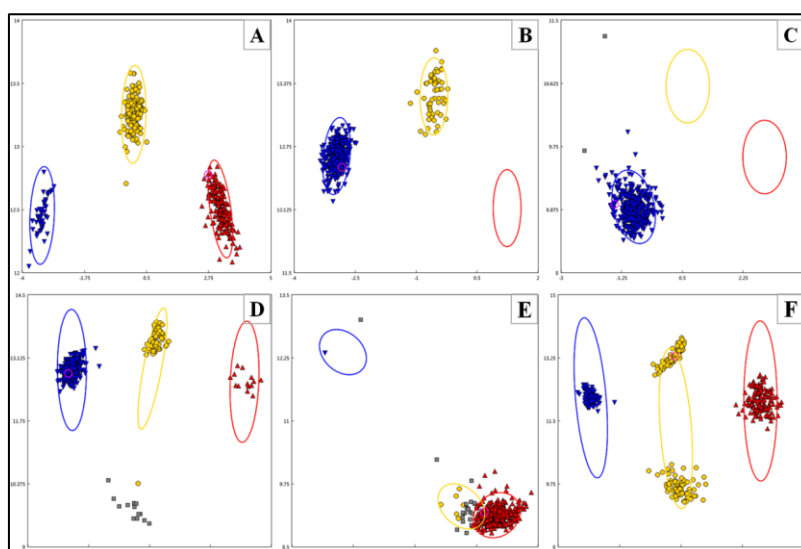


Figura 2. Classificação quanto à qualidade dos SNPs gerada pelo Axiom Best Practices Genotyping Workflow. (A) *Poly High Resolution* (PHR) - SNPs polimórficos e que passaram por todos os controles de qualidade (CQ); (B) *No MinorHomozygote* (NMH) - SNPs passaram pelo CQ, porém somente dois *clusters* são observados; (C) *Mono High Resolution* (MHR) - SNPs passaram em relação à qualidade mas são monomórficos; (D) *Call Rate Below Threshold* (CRBT) – SNPs que apresentaram *call rate* abaixo de 97%; (E) *Other* - o padrão resultante de *cluster* de SNP não se encontra em nenhuma das classes anteriores; (F) *Off-Target Variant* (OTV) - apresentam um *cluster* adicional, de baixa intensidade, resultante de desajustes entre a sonda e as sequências para esse grupo de indivíduos.

A partir da validação da plataforma de genotipagem de *C. canephora*, pode-se dizer que a genotipagem se mostrou coerente e confiável, gerando uma grande quantidade de marcadores SNPs polimórficos de qualidade. Este resultado está de acordo com outros chips publicados, melhorando as 60% e 74% das variantes PHR obtidas no chip 180K SoyaSNP (LEE et al., 2015) e Apple 480K (BIANCO et al., 2016), desenvolvidos para soja e maçã, respectivamente, mas ficando aquém dos 92% de variantes PHR apresentadas pela matriz Maize 600K (UNTERSEER et al., 2014), desenvolvida para milho.

Os marcadores moleculares têm sido amplamente utilizados para avaliação da diversidade genética, estudos evolutivos e de mapeamento. A aplicabilidade desses no melhoramento depende do custo, facilidade e precisão, tornando os SNPs uma escolha indispensável. O rápido progresso nas tecnologias de NGS durante a última década permitiu a produção de dados de sequência a baixo custo e em muito menos tempo (THUDI et al., 2012). As tecnologias baseadas em NGS foram efetivamente utilizadas para sequenciamento e resequenciamento de genomas, possibilitando a identificação de um número muito grande de marcadores SNPs. No cafeeiro, diferentes plataformas de genotipagem de SNPs foram desenvolvidas para várias aplicações (CARNEIRO et al., 2014, FERRÃO et al., 2017; SANT’ANA et al., 2018). Atualmente são os chips de DNA, com uma alta densidade de SNPs, que tem se mostrado uma ferramenta poderosa para o melhoramento molecular e a investigação genética de populações dentro de espécies de café (ANDRADE et al., 2017; MEROT-L’ANTHOENE et al., 2018), pois são de fácil utilização e econômicos para a geração e análise de dados de genotipagem. Além disso, a plataforma apresenta alta precisão e reprodutibilidade.

## CONCLUSÕES

1 - O desenvolvimento da plataforma de genotipagem (Coffee Axiom chip – 26K) para *C. canephora*, com um grande número de SNPs validados e, acima de tudo, com a alta qualidade e reprodutibilidade destes marcadores, representam uma inovação na espécie estudada e disponibiliza uma ferramenta robusta para posteriores estudos de GWAS e seleção genômica ampla (SGA), genética populacional, podendo contribuir para o aumento na eficiência dos programas de melhoramento genético do cafeeiro.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ANDRADE, A. C. et al. ‘Towards GWAS and Genomic Prediction in Coffee: Development and Validation of a 26K SNP Chip for *Coffea canephora*’, In XX INTERNATIONAL PLANT AND ANIMAL GENOME CONFERENCE, p. W173. Anais eletrônicos... Disponível em: <<https://pag.confex.com/pag/xxv/webprogram/Paper23677.html>>
- BASSIL, N. V. et al. Development and preliminary evaluation of a 90 K Axiom(R) SNP array for the allo-octoploid cultivated strawberry *Fragaria x ananassa*. BMC Genomics, v. 16, n. 1, p. 1-30, 2015.
- BIANCO, L. et al. Development and validation of the Axiom® Apple480K SNP genotyping array. The Plant Journal, v. 86, n. 1, p. 62-74, 2016.
- CARNEIRO, F. A. et al. Phenotyping and genotyping a *Coffea canephora* population, cultivated at high altitude, aiming at a GWS program for coffee. In: THE 25TH INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, 2014, Colômbia.
- CARNEIRO, F. A. et al. Phenotyping a *Coffea canephora* population, cultivated at high altitude, aiming at a GWS program for coffee. In: Workshop on Biotic and Abiotic Stress Tolerance in Plants: the Challenge for the 21st Century, 2013, Ilhéus. Book of abstracts, 2013.
- DENOEUDE, F. et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science, v. 345, n. 6201, p. 1181-1184, 2014.
- HULSE-KEMP, A. M. et al. Development of a 63K SNP Array for Cotton and High-Density Mapping of Intraspecific and Interspecific Populations of *Gossypium* spp. G3: Genes, Genomes, Genetics, v. 5, n. 6, p. 1187-1209, 2015.
- ICO – International Coffee Organization. Disponível em: <<http://www.ico.org/>>. Acesso em: 18 jan. 2019.
- KONING-BOUCOIRAN, C. F. et al. Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68k Axiom SNP array for rose (*Rosa* L.). Frontiers in Plant Science, v. 6, n. 249, p. 1-10, 2015.
- LEE, Y. G. et al. Development, validation and genetic analysis of a large soybean SNP genotyping array. The Plant Journal, v. 81, n. 4, p. 625-636, 2015.
- MEROT-L’ANTHOENE, V. et al. Development and evaluation of a genome-wide Coffee 8.5K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. Plant Biotechnology Journal, p. 1-13, 2019.
- RUSSELL, A. et al. Phylogenetics and cytology of a pantropical orchid genus *Polystachya* (Polystachyinae, Vandaeae, Orchidaceae): evidence from plastid DNA sequence data. Taxon, v. 59, n. 2, p. 389-404, 2010.
- THUDI, M. et al. Current state-of-art of sequencing technologies for plant genomics research. Briefings in Functional Genomics v. 11, n. 1, p. 3-11, 2012
- UNTERSEER, S. et al. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. BMC Genomics, v. 15, n. 1, p. 1-15. 2014.