

## ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA PARA IMPORTANTES CARACTERÍSTICAS AGRONÔMICAS EM *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>, Gustavo Costa Rodrigues<sup>7</sup>; Milene Alves de Figueiredo Carvalho<sup>8</sup>; Adriano Delly Veiga<sup>9</sup>, Dario Grattapaglia<sup>10</sup>, Orzenil Bonfim da Silva Júnior<sup>11</sup>, Pierre Marraccini<sup>12</sup>, Indalecio Cunha Vieira Junior<sup>13</sup>, Márcio Balestre<sup>14</sup>, Alan Carvalho Andrade<sup>15</sup>

<sup>1</sup> Trabalho financiado pelo – Consórcio Pesquisa Café e INCT-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/CAPEES, 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, Embrapa Informática Agropecuária, Campinas-SP, Brasil, [gustavo.rodrigues@embrapa.br](mailto:gustavo.rodrigues@embrapa.br)

<sup>8</sup> Pesquisadora, PhD, Embrapa Café/INOVA CAFÉ, Lavras-MG, Brasil, [milene.carvalho@embrapa.br](mailto:milene.carvalho@embrapa.br)

<sup>9</sup> Pesquisador, PhD, Embrapa Cerrados, Planaltina-DF, Brasil, [adriano.veiga@embrapa.br](mailto:adriano.veiga@embrapa.br)

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

<sup>11</sup> Pesquisador, PhD, Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, Brasil, [orzenil.silva@embrapa.br](mailto:orzenil.silva@embrapa.br)

<sup>12</sup> Pesquisador, PhD-HDR, CIRAD, UMR IPME, Instituto of Agricultural Genetics (AGI) LMI RICE2, Hanoi, Vietnã, [marraccini@cirad.fr](mailto:marraccini@cirad.fr)

<sup>13</sup> Doutorando em Genética e Melhoramento de Plantas, UFLA, Lavras-MG, Brasil, [indasjunior@hotmail.com](mailto:indasjunior@hotmail.com)

<sup>14</sup> Professor, PhD, UFLA, Lavras-MG, Brasil, [marciobalestre@ufla.br](mailto:marciobalestre@ufla.br)

<sup>15</sup> Pesquisador, PhD, Embrapa Café/INOVA CAFÉ, Lavras-MG, Brasil, [alan.andrade@embrapa.br](mailto:alan.andrade@embrapa.br)

**RESUMO:** O Estudo de Associação Genômica Ampla (*Genome-Wide Association Studies* – GWAS) surgiu com o intuito de capturar o efeito de genes na determinação de determinada característica fenotípica. Para isso, os indivíduos selecionados são genotipados em escala genômica e fenotipados para as características de interesse e então, a partir de análises estatísticas entre os polimorfismos de DNA e a variação no fenótipo, os genes que controlam essas características podem ser identificados. Por se tratar de uma espécie perene e de ciclo longo, os programas de melhoramento genético convencional do cafeeiro são demorados (vários anos) e de custo elevado. A geração de cultivares superiores via melhoramento genético ainda lida com o desafio de agregar simultaneamente diversas características quantitativas de relevância agrônoma, assim como a qualidade de bebida. Neste contexto, na tentativa de identificar genes/regiões genômicas associadas à características agrônomas importantes, 1.319 indivíduos da Embrapa Cerrados foram genotipados e fenotipados, e posteriormente, realizou-se um estudo de GWAS com os softwares TASSEL e rrBLUP. Um número diferente de marcadores associados foi identificado com as duas estratégias utilizadas. Foram identificados marcadores tanto em regiões intergênicas como em regiões gênicas, sendo que, nesta última, os marcadores estavam presentes em genes já caracterizados e descritos na literatura para outras espécies como estando relacionados às características agrônomas aqui analisadas.

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

### GENOME-WIDE ASSOCIATION STUDIES FOR IMPORTANT AGRONOMIC TRAITS IN *Coffea canephora*

**ABSTRACT:** The Genome-Wide Association Studies (GWAS) emerged to capture the effect of genes on phenotypic determination. For this, the selected individuals are genotyped and phenotyped for the traits of interest and then, from statistical analyses between DNA polymorphisms and phenotype variation, the genes controlling these traits are identified. Because it is a perennial and long-cycle species, conventional coffee breeding programs are time consuming (several years) and costly. The generation of superior cultivars via genetic improvement still deals with the challenge of simultaneously aggregating several quantitative characteristics of agronomic relevance and beverage quality. In this context, in an attempt to identify genes / genomic regions associated with important agronomic characteristics, 1,319 individuals from Embrapa Cerrados were genotyped and phenotyped, and subsequently GWAS was performed using TASSEL and rrBLUP software. A different number of associated markers were identified with both strategies used. Markers were identified in both intergenic and gene-coding regions, and at last, markers were present in genes already characterized and described in the literature for other species as being related to the agronomic characteristics analyzed here.

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

## INTRODUÇÃO

Os SNPs são mudanças em uma única base (mutações) no DNA e são unidades de variação genética que ocorrem em qualquer genoma de um organismo. Uma enorme quantidade de SNPs pode ser capturada usando tecnologias de genotipagem, como a tecnologia baseada em chips de DNA ou outras tecnologias de sequenciamento de próxima geração (NGS – *Next Generation Sequencing*). Os dados de genotipagem gerados são muito utilizados em uma abordagem denominada estudo de associação genômica ampla (GWAS – *Genome-Wide Association Studies*), em que se investiga SNPs capturados associados a fenótipos ou características de interesse ao melhoramento genético por meio de conceitos estatísticos para análise, interpretação e acompanhamento dos resultados de associação.

Com os recentes avanços na genômica do cafeeiro, como o sequenciamento completo do genoma de referência de *C. canephora* (DENOEUDE 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 NGS forneceram as ferramentas necessárias para a genotipagem de SNPs em larga escala na espécie, possibilitando o desenvolvimento e a validação de uma plataforma de genotipagem (Coffee Axiom chip – 26K), contendo SNPs informativos e de alta qualidade (ANDRADE et al., 2017).

O presente trabalho buscou identificar associações significativas entre os SNPs identificados pelo chip e dados fenotípicos, como produção, tamanho e formato de grãos, tamanho e formato de frutos, boia e peso de 100 grãos, em uma população de *C. canephora* (CARNEIRO et al., 2013). Nesse sentido, a GWAS auxiliaria na identificação marcadores para a seleção de genótipos superiores.

## MATERIAL E MÉTODOS

Um total de 1.319 indivíduos de *C. canephora* selecionados a partir de uma população de melhoramento localizada no campo experimental da Embrapa Cerrados (Planaltina, DF) foram fenotipados para diversas características de interesse agrônomo (CARNEIRO et al., 2013) e genotipados através da plataforma Coffee Axiom chip – 26K (ANDRADE et al., 2017).

### • Fenotipagem

Os dados fenotípicos utilizados nos experimentos de GWAS foram levantados a partir de avaliações realizadas para os indivíduos selecionados na população de melhoramento. Foram obtidos dados de produtividade, boia, peso de 100 grãos, peneiras e composição bioquímica por espectroscopia do infravermelho-próximo, conforme metodologias já estabelecidas. Além disso, análises fisiológicas para avaliação da tolerância à seca, foram também realizadas.

**Produtividade:** Os dados de produção anual foram obtidos para cada indivíduo selecionado na população, por um período de 3 anos consecutivos (2011/12; 2012/13 e 2013/14), sendo que cada planta foi colhida individualmente e a produção do café cereja, por planta, medida em litros (L).

**Morfologia dos frutos:** A partir da colheita, cerca de 50 frutos de café cereja, de cada planta, foram acondicionados em sacos de papel e mantidos à -20°C até a realização das análises morfológicas dos frutos. Os dados de tamanho e formato do grão foram obtidos com auxílio do software DIGI-Pro (Labomed), utilizando-se 5 frutos por amostra.

**Boia:** Após a colheita, uma amostra de 100 frutos de café cereja, de cada planta, foi colocada em um recipiente com 1L de água para obter o valor, em porcentagem, de frutos boia.

**Peso de 100 grãos:** A partir da colheita, as amostras de cada planta foram processadas em despulpador elétrico (Palini&Alves, modelo: PA-DCC/E), seguindo o processo de cereja descascado, acondicionadas em sacos de filó e colocadas para secar em terreiro, até os grãos atingirem grau de umidade de aproximadamente 12%. Posteriormente, pesou-se uma amostra de 100 grãos secos de cada indivíduo em balança eletrônica modelo AL200C (Marte).

**Formato e tamanho dos grãos:** Amostras de grãos secos (100g) de cada indivíduo foram analisadas em um conjunto de peneiras (Pinhalense) para a classificação quanto ao formato e granulometria. Conforme a Instrução Normativa do MAPA (nº8, 11/06/2003), os grãos podem ser enquadrados em duas categorias: (1) Chato – constituída de grãos com superfície dorsal convexa e a ventral plana ou ligeiramente côncava, com a ranhura central no sentido longitudinal e (2) Moca – constituída de grãos com formato ovoide, também com ranhura central no sentido longitudinal. Para cada categoria, os grãos são ainda classificados de acordo com a dimensão dos crivos das peneiras que os retêm, conforme Tabela 1.

**Composição bioquímica:** A partir dos grãos secos, conforme descrito anteriormente, amostras de 100 indivíduos (300g), foram acondicionadas em sacos de papel e enviadas para o Centro de Pesquisa R&D Tours da Nestlé (França), para a realização das análises de NIRS, conforme descrito por Vinecky et al., (2016).

**Análises fisiológicas:** Medidas do potencial hídrico foliar de antemanhã ( $\Psi_m$ ) foram realizadas no final do mês de agosto e início de setembro/2013, após um período de estresse hídrico de aproximadamente 70 dias, conforme descrito em Marraccini et al., (2012). Para esse essas análises fenotípicas, selecionou-se uma amostra de 366 indivíduos, os quais estão entre os 1.319 indivíduos da população de melhoramento selecionados para as demais características.

### • Genotipagem

Ao longo do genoma de *C. canephora* foram identificados 25.456 SNPs e após um processamento mínimo, utilizando a função *A.mat* do pacote rBLUP (ENDELMAN, 2011), no software R (R Core Team, 2017), obteve-se um conjunto de

16.688 marcadores com *call rate* acima de 90% e MAF (*Minimum Allele Frequency*) superior a 1%. Dois programas foram utilizados para as análises de GWAS, o pacote rrBLUP e o software TASSEL (BRADBURY et al., 2007), ambos tem como princípio o uso da equação de modelos mistos proposto por Yu et al. (2006).

Para se confirmar a significância das associações entre os SNPs e os fenótipos de interesse, dois métodos foram utilizados: (i) correção de Bonferroni e (ii) FDR (False Discovery Rate) (BENJAMINI; HOCHBERG, 1995), adotando-se um nível de significância de 5% ( $p$ -valor  $< 0,05$ ). A partir dos dados de  $p$ -valores significativos para cada método, o pacote CMplot, implementado no software R, gerou os gráficos do tipo Manhattan Plot.

Tabela 1: Classificação oficial de café por peneira, conforme Instrução Normativa do MAPA (nº8, 11/06/2003)

Chato Graúdo	Peneira 17 e maiores
Chato Médio	Peneiras 15 e 16
Chato Miúdo	Peneiras 12, 13 e 14
Moca Graúdo	Peneiras 11 a 13
Moca Médio	Peneira 10
Moca Miúdo	Peneiras 8 e 9

## RESULTADOS E DISCUSSÃO

Os resultados obtidos das análises realizadas por meio do rrBLUP identificaram associações para as características de peso de 100 grãos, em que nove marcadores estavam associados, sendo que seis localizam-se no cromossomo 0 e um destes, AX-168342522, também foi associado à característica de Moca Miúdo (Figura 1).

Para a característica tamanho e formato de grãos, 19 marcadores foram associados (Figura 2); onde um marcador presente no cromossomo 6 estava associado a característica Chato Médio (Figura 2A), três tiveram associação para a característica Chato Miúdo (Figura 2B), posicionados nos cromossomos 2 (AX-168348734, AX-168355324) e 6 (AX-168348505), 13 SNPs estavam associados para Moca Graúdo, podendo ser observado na Figura 2C que somente um marcador está localizado no cromossomo 2 e os demais encontram-se localizados em regiões próximas (26.401.552 – 28.859.558) do cromossomo 5, indicando a possibilidade de ser um QTL. Dois outros marcadores, associados à característica de Moca Miúdo (Figura 2D), estavam localizados nos cromossomo 3 (AX-168355983) e o no cromossomo 0 (AX-168342522).

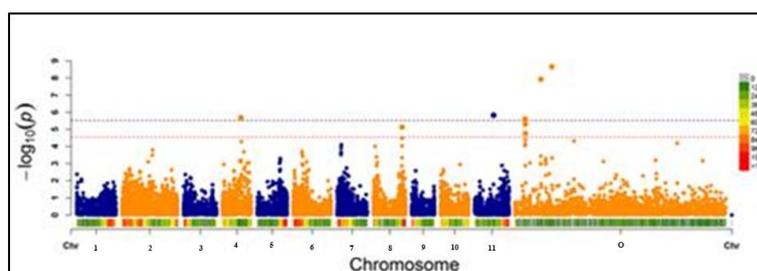


Fig. 1. Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de peso de 100 grãos. As linhas tracejadas representam o valor obtido para  $-\log_{10}(p)$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). Estão representados os 11 cromossomos de *C. canephora* e o cromossomo 0, que consiste em um conjunto não ordenado de *scaffolds*.

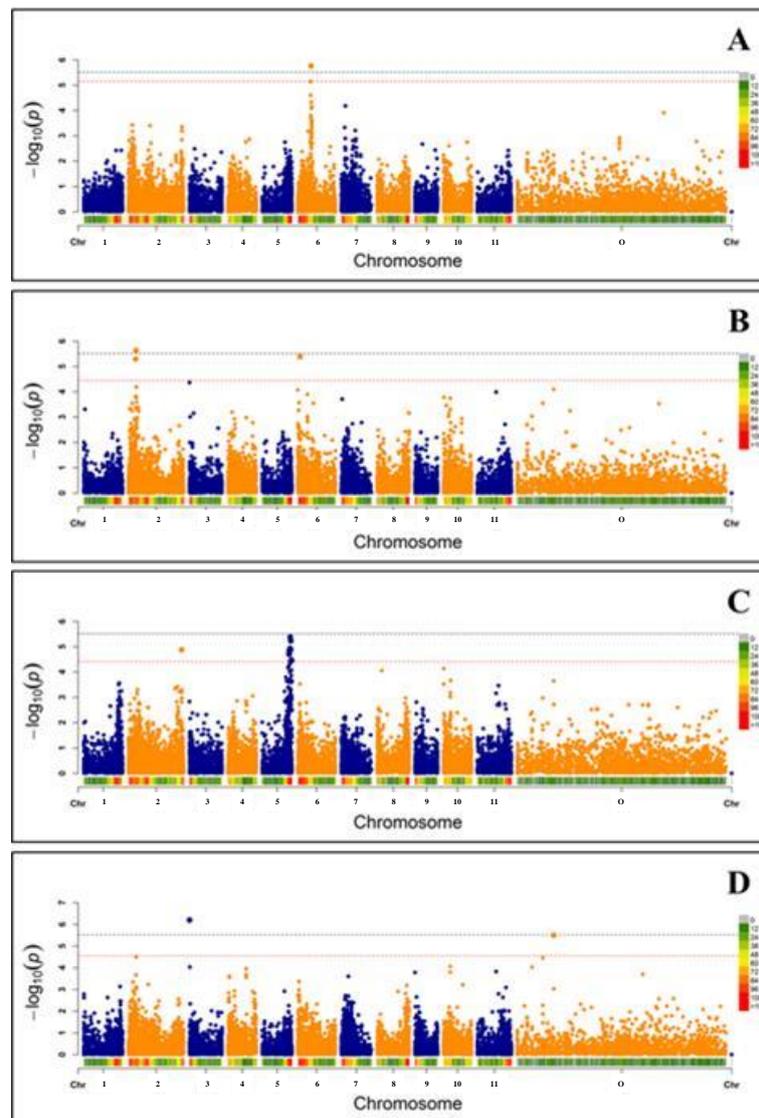


Fig. 2. Gráfico Manhattan de associação do efeito dos SNPs testados com a característica de tamanho e formato de grãos. Manhattan plot para as características de tamanho e formato do grão. SNPs associados para os grãos classificados como Chato Médio (A), Chato Miúdo (B), Moca Graúdo (C) e Moca Miúdo (D). As linhas tracejadas representam o valor obtido para  $-\log_{10}(p)$  considerando a correção de Bonferroni (linha tracejada preta) e FDR (linha tracejada vermelha). Estão representados os 11 cromossomos de *C. canephora* e o cromossomo 0, que consiste em um conjunto não ordenado de *scaffolds*.

Para as análises realizadas no software TASSEL, utilizando o mesmo conjunto de dados utilizado no rrBLUP, um marcador presente no cromossomo 11 foi associado para a característica de produção, avaliada no ano de 2013. Duas outras características que não haviam sido detectadas pelo rrBLUP foram identificadas pela análise com o TASSEL, 6 SNPs apresentaram associação para o fenótipo Boia e 23 marcadores estavam associados ao tamanho do fruto, relacionados à característica de eixo menor.

Para peso de 100 grãos, 6 SNPs, todos eles localizados no cromossomo 0, foram associados. Pode-se notar que 3 SNPs associados à característica boia, também foram detectados em associação com a característica de peso de 100 grãos (AX-168308807, AX-168355686 e AX-168358258). Já para a característica do fruto, eixo menor, a qual, somente nesta análise, marcadores associados foram detectados, pode-se observar que mais de um SNP em associação significativa foi encontrado em um mesmo gene. Cc02\_g24790 (AX-168303576 e AX-168343505), que codifica uma “*GPI-anchored protein like*”, sendo que “*Glycosylphosphatidylinositol-Anchored Proteins*” GPI proteínas são importantes na formação da parede celular e também na morfogênese (GILMOR et al, 2005) e, o gene Cc10\_g02630 que codifica uma frutokinase (“*putative pfkB family carbohydrate kinase*”), proteínas envolvidas no metabolismo de frutose (PEGO; SMEEKENS, 2001).

## CONCLUSÕES

Apesar de alguns resultados diferentes entre os dois programas estatísticos utilizados (TASSEL vs. rrBLUP) para as análises de GWAS, SNPs em associação com características agrônômicas de interesse, puderam ser confirmados por meio das duas análises realizadas. Além disto, com o suporte da caracterização funcional de vários dos genes detectados como importantes fatores na determinação fenotípica das características estudadas, já descritos na literatura, principalmente com os trabalhos pioneiros na planta modelo *Arabidopsis thaliana*, corroboram com a associação funcional desses genes encontrados neste trabalho, com as características fenotípicas análogas, como a de tamanho e formato de grãos e tamanho e formato de frutos. Esses dados corroboram com as associações significativas de GWAS encontradas em café, indicando que esses marcadores podem ter grande utilidade nos programas de melhoramento genético por seleção assistida em 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>>
- BENJAMINI, Y.;HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, v. 57, n. 1, p. 289-300, 1995.
- BRADBURY, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, v. 23, n. 19, p. 2633-2635, 2007,
- 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.
- ENDELMAN, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome, v. 4, p. 250-255, 2011.
- GILMOR, C. S. et al. Glycosylphosphatidylinositol-Anchored Proteins Are Required for Cell Wall Synthesis and Morphogenesis in Arabidopsis. The Plant Cell, v. 17, p. 1128-1140, 2005.
- PEGO, J. V.; SMEEKENS, S. Plant fructokinases: A sweet family get-together. Trends in Plant Science, v. 5, n. 12, p. 531-536, 2001.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2017 .URL <http://www.R-project.org/>.
- YU, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, v. 38, p. 203-208, 2006.