

**RUANE ALICE DA SILVA**

**ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)  
EM *Coffea arabica***

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

Orientadora: Eveline Teixeira Caixeta

Coorientadores: Moysés Nascimento  
Tiago Vieira Sousa

**VIÇOSA- MINAS GERAIS  
2019**

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

S586e  
2019 Silva, Ruane Alice da, 1994-  
Estudo de associação genômica ampla (GWAS) em *Coffea arabica* / Ruane Alice da Silva. – Viçosa, MG, 2019.  
39 f. : il. (algumas color.) ; 29 cm.

Orientador: Eveline Teixeira Caixeta.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 36-39.

1. Café - Melhoramento genético. 2. Marcadores genéticos.  
3. Café - Seleção. I. Universidade Federal de Viçosa.  
Departamento de Fitotecnia. Programa de Pós-Graduação em  
Genética e Melhoramento. II. Título.

CDD 22. ed. 633.732

**RUANE ALICE DA SILVA**


**ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)  
EM *Coffea arabica***

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

APROVADA: 09 de agosto de 2019.



Ruane Alice da Silva  
Autora



Eveline Teixeira Caixeta  
Orientadora

**Aos meus pais Milton e Alice,  
E a toda minha família,  
Dedico.**

## AGRADECIMENTOS

A Deus pelo dom da vida, e por sempre abençoar minha trajetória, me concedendo saúde física e mental para superar os desafios encontrados.

À Universidade Federal De Viçosa (UFV) e ao Curso de Pós-Graduação em Genética e melhoramento pela oportunidade e pela excelência em ensino.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudo.

À Dra. Eveline Teixeira Caixeta pela orientação, oportunidade, paciência, conselhos e todos ensinamentos transmitidos.

Ao Dr. Moysés Nascimento por sua coorientação, atenção, paciência e generosidade durante todo o desenvolvimento desse trabalho e por aceitar o convite de participar da minha banca.

Ao meu coorientador Dr. Tiago Vieira Sousa por todo suporte, amizade, paciência e conhecimentos transmitidos.

Ao Dr. Maximiller Dal-Bianco Lamas Costa por aceitar o convite em participar da minha banca e compartilhar seus conhecimentos.

Ao Dr. Antonio Carlos Baião de Oliveira e a Dra Poliane Marcele Ribeiro Cardoso pela disponibilidade em participarem da banca.

Aos meus pais Milton e Alice, minhas maiores motivações, que mesmo sem entender sempre me apoiam como muito amor e oração.

À toda minha família, que cada um de sua forma, contribuiu nessa caminhada, em especial minha “irmã” Rondinea que sempre me impulsionou a buscar o melhor.

Ao Pedro César por todo companheirismo, motivação e ser meu suporte durante todo o mestrado.

Ao Dr. Rafael Parrella que foi um grande motivador para escolha do mestrado.

À equipe Biocafé, por toda amizade, cafés e companheirismo diário.

Ao Pedro por toda ajuda nas análises de bioinformática.

Aos amigos que construí em Viçosa em especial Danúbia (esse mestrado não seria o mesmo sem ela), Arthur, Letícia, Fumes, Fran, Laura, Dique.

Aos bons e velhos amigos Cris, Letícia e Everton, que sempre estão na torcida pelo meu sucesso.

Enfim, a todos que de alguma forma contribuíram para o desenvolvimento deste trabalho, toda minha gratidão.

*“Pensávamos que o nosso futuro estava nas estrelas,  
hoje acreditamos que está em nossos genes.”*

James Watson

## RESUMO

SILVA, Ruane Alice da, M.Sc., Universidade Federal de Viçosa, agosto de 2019. **Estudo de associação genômica ampla (GWAS) em *Coffea arabica***. Orientadora: Eveline Teixeira Caixeta. Coorientadores: Moysés Nascimento e Tiago Vieira Sousa.

Estudos de associação genômica ampla (GWAS) têm sido utilizados mundialmente para várias espécies de importância econômica. A metodologia possibilita inferir sobre as relações entre variações genéticas e características fenotípicas, em nível populacional, detectando efeitos significativos por meio de testes de hipóteses. O GWAS demonstra grande relevância nos programas de melhoramento, sobretudo de espécies perenes, tendo como principal objetivo identificar regiões cromossômicas e genes candidatos para o controle genético de determinada característica. No entanto, em *Coffea arabica*, espécie de maior importância econômica do gênero *Coffea*, existem poucos estudos publicados utilizando essa metodologia. Assim, objetivou-se a identificação de regiões cromossômicas que apresentam associações significativas entre marcadores e características fenotípicas de importância para a espécie *C. arabica*, por meio de GWAS e selecionar SNP associados para serem utilizados em seleções assistidas por marcadores moleculares. A população de trabalho foi composta por 195 indivíduos de *C. arabica*, pertencentes à 13 famílias em gerações F<sub>2</sub>, retrocruzamentos suscetíveis e retrocruzamentos resistentes à ferrugem do cafeeiro. As plantas foram fenotipadas para 18 características agrônomicas durante três anos consecutivos (2014, 2015 e 2016) e genotipadas por 21.211 marcadores moleculares SNP distribuídos em todo o genoma. O GWAS possibilitou a identificação de 110 SNP com associações significativas ( $p < 0,05$ ) para as características: altura de plantas (AP), comprimento de ramos plagiotrópicos (CRP), número de nós vegetativos (NNV), diâmetro da copa (DCo), tamanho de fruto (TFr), incidência de cercosporiose (Cer) e incidência de ferrugem (Fer). Foram analisados os efeitos de cada marcador SNP associado com as características de interesse e, dessa forma, esses marcadores poderão ser usados, posteriormente, na seleções assistidas por marcadores moleculares (SAM). Esse foi o primeiro trabalho de GWAS para essas características agrônomicas em *C. arabica* e apresentou resultados promissores para os programas de melhoramento da espécie, confirmando a eficiência da metodologia.

Palavras-chave: Melhoramento do Cafeeiro. Marcadores SNP. Seleção Assistida por Marcadores.

## ABSTRACT

SILVA, Ruane Alice da, M.Sc., Universidade Federal de Viçosa, August, 2019. **Genome-wide association Study (GWAS) in *Coffea arabica***. Adviser: Eveline Teixeira Caixeta. Co-advisers: Moysés Nascimento and Tiago Vieira Sousa.

Genomic wide association studies (GWAS) have been used worldwide for several species of economic importance. The methodology permit to infer the relationship between genetic variation and phenotypic characteristic at the population level, detecting significant effects through hypothesis testing. GWAS shows great relevance in breeding programs, especially for perennial species, aiming to identify chromosomal regions and candidate genes for the genetic control of a trait of interesting. In *Coffea arabica*, the most economically important species of the genus *Coffea*, there are few published studies using this methodology. Therefore, the objective of this work was to identify chromosomal regions that present significant associations between markers and phenotypic characteristics of importance to the species *C. arabica*, by means of GWAS. We also select associated SNP to be used in marker-assisted selection (MAS). The working population consisted of 195 *C. arabica* individuals, belonging to 13 families in F<sub>2</sub> generations, susceptible backcross and resistant backcross for coffee rust. Plants were phenotyped for 18 agronomic traits in three consecutive years (2014, 2015 and 2016) and genotyped by 21.211 SNP molecular markers distributed throughout the genome. The GWAS allowed the identification of 110 SNP with significant associations ( $p < 0.05$ ) for the characteristics: plant height (PH), plagiotropic branch length (PBL), number of vegetative nodes (NVN), crown diameter (DC), fruit size (FS), incidence of cercosporiosis (Cer) and incidence of rust (Rus). The effects of each associated SNP with the traits were obtained, allowing the future use of these markers for MAS. This is the first work of GWAS for agronomic characteristics in *C. arabica* and promising results were obtained for the breeding programs of the species, confirming the efficiency of the methodology.

Keywords: Coffee Breeding. SNP Markers. Marker-Assisted Selection.



## SUMÁRIO

Introdução.....	9
Material e métodos.....	11
Condução experimental .....	11
Material genético.....	11
Avaliações fenotípicas.....	12
Análises dos dados fenotípicos .....	12
Extração do DNA genômico .....	14
Análises de qualidade dos SNP.....	14
Identificação dos SNP em <i>Coffea arabica</i> .....	15
Estudo de associação genômica ampla.....	15
Identificação de genes .....	16
Resultados e discussão .....	16
Conclusões.....	35
Referências Bibliográficas .....	36

## Introdução

O café é a segunda bebida mais consumida mundialmente, com o Brasil no ápice do ranking de maior produtor e exportador de café do mundo (Oec, 2018). Além disso, o cenário brasileiro, diante da variedade de condições edafoclimáticas, proporciona uma diversidade de sabores da bebida, tornando o produto ainda mais diferenciado e apto para atender a demanda do mercado interno e externo (Mapa, 2018). Dentre as espécies do gênero *Coffea*, *Coffea arabica* e *Coffea canephora* são as de maior importância econômica (Combes et al., 2000), sendo ambas cultivadas no Brasil.

Diante da importância da cultura do café, se faz necessário promover pesquisas que contribuam para maior rentabilidade e sustentabilidade da sua cadeia produtiva. Nesse contexto, o melhoramento genético é um dos responsáveis por proporcionar avanços por meio do desenvolvimento de cultivares que atendam às exigências do mercado (Carvalho et al., 2011; Oliveira et al., 2010). Contudo, o processo de melhoramento é demorado, multidisciplinar e contínuo, e para torná-lo mais rápido e eficiente, recomenda-se a aplicação de ferramentas inovadoras, como o uso da biotecnologia, que podem contribuir para o progresso genético da cultura (Mishra e Slater, 2012; Ferrão et al., 2015).

A espécie *C. arabica* é mundialmente reconhecida pela elevada importância econômica. Essa espécie é alotetraploide (Clarindo e Carvalho, 2008) ( $2n = 4x = 44$  cromossomos), autógama, originada do cruzamento natural das espécies diploides *C. canephora* e *C. eugenioides* (Lashermes et al., 1999), sendo a espécie pertencente ao gênero *Coffea* mais cultivada. No entanto, *C. arabica*, devido principalmente a sua origem, tipo de reprodução e evolução, apresenta base genética estreita (Anthony et al., 2002).

Com programas de melhoramento bem consolidados, como o desenvolvido pela EPAMIG/UFV/UFLA/EMBRAPA, o Brasil se destaca pelo número de cultivares de café registradas. No entanto, segundo Setotaw et al. (2013), apenas sete ancestrais contribuem com mais de 97% da base genética dessas cultivares de *C. arabica*, o que dificulta a discriminação de genótipos nos programas de melhoramento. Essa discriminação é especialmente limitada devido ao fato de características quantitativas, governadas por muitos genes e muito afetadas pelo ambiente, não serem classificadas em grupos fenotípicos diferentes, o que torna a seleção visual dificultada (Falconer e Mackay, 1996).

Uma das ferramentas que vem auxiliando os programas de melhoramento são os marcadores moleculares. Esses podem contribuir de forma significativa, pois, permitem

detectar a diferença em nível de DNA (Mishra e Slater, 2012). Atualmente existem diferentes marcadores moleculares utilizados na espécie *C. arabica* para diversas finalidades (Mahé et al., 2008; Alvarenga et al., 2011; Sousa et al., 2017a; Sousa et al., 2017b). Contudo, os marcadores SNP, devido a sua fácil automatização permitem a genotipagem em larga escala, otimizando o tempo de melhoramento, além de oferecer uma grande quantidade de informação, pela sua abundância no genoma (He et al., 2014).

A informação obtida por meio dos marcadores SNP, junto às informações fenotípicas pode ser utilizada em estudos de associação genômica ampla (GWAS), que, por meio de testes de hipóteses, permite detectar efeitos significativos entre o marcador e características complexas (Huang e Han 2014). Esses efeitos significativos podem ser relacionados à identificação de genes que controlam determinado caráter e às suas respectivas funções biológicas (Pers et al., 2015). Além disso, o GWAS minimiza as limitações anteriormente encontradas no mapeamento de QTL tradicional, pois permite maior resolução e pode ser realizado em populações de grande variação genética, o que para mapeamento de QTL só é possível em cruzamentos biparentais (Korte e Farlow, 2013).

A eficiência do GWAS tem sido relatada em diversos trabalhos, aplicado em diferentes culturas e características de importância econômica. Para adaptação ao uso do GWAS, as culturas do milho e arroz são citadas como modelo, pois apresentam painéis compostos por um grande número de linhagens genotipadas, em vários ambientes e para várias características (Huang e Han 2014). Para a espécie *C. arabica*, poucos estudos foram realizados utilizando essa metodologia. O único trabalho disponível aborda o GWAS para características relacionadas à composição química do grão, visando à melhoria da qualidade de bebida (Sant’Ana et al., 2018). Por isso, são importantes e necessários estudos que envolvam outras características de interesse para o cafeeiro, a fim de explorar a baixa variabilidade existente nessa espécie, maximizar os ganhos genéticos e tornar o programa de melhoramento do cafeeiro mais rápido e eficiente.

Diante do exposto, objetivou-se (i) identificar regiões cromossômicas que apresentam associações significativas entre marcadores SNP e características fenotípicas de importância para a espécie *C. arabica* por meio de GWAS; e (ii) selecionar SNP associados de acordo com objetivo do programa para serem posteriormente validados e utilizados em seleções assistidas por marcadores moleculares (SAM), auxiliando os programas de melhoramento da espécie.

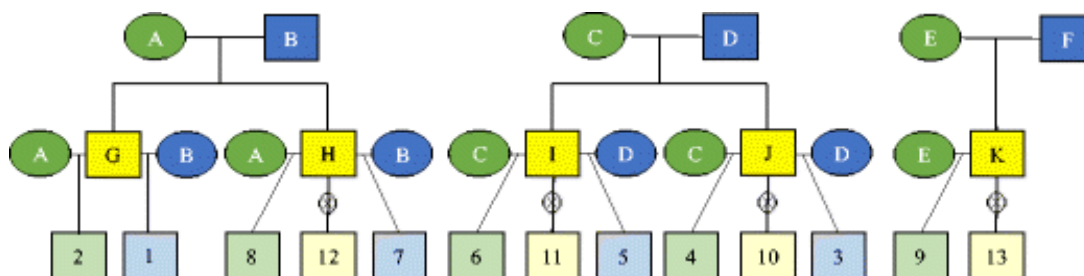
## Material e métodos

### Condução experimental

O plantio do experimento foi realizado no dia 11 de fevereiro de 2011, na estação experimental do Departamento de Fitopatologia da Universidade Federal de Viçosa, em Viçosa, Minas Gerais (20°44'25" S; 42°50'52" W). As plantas foram dispostas no espaçamento de 3,0 m entre fileiras e 0,7 m entre plantas. A calagem e adubação foram realizadas conforme as exigências da cultura, e não foram realizados métodos de controle fitossanitário para controle de ferrugem, antracnose e bicho mineiro.

### Material genético

Foram selecionados para o cruzamento, seis genótipos de dois grupos contrastantes em relação à resistência à ferrugem do cafeeiro, sendo três genitores do grupo de cultivares Catuaí (A, C e E, suscetíveis à *Hemileia vastatrix*) e três genitores do grupo Híbrido de Timor (HdT) (B, D e F, resistentes à *Hemileia vastatrix*). A partir desses cruzamentos foram obtidas 13 progênies, sendo essas correspondentes às gerações de retrocruzamentos resistentes (RCr), retrocruzamentos suscetíveis (RCs) e F<sub>2</sub> (Figura 1). Em cada progênie foram selecionados 15 genótipos, totalizando 195 indivíduos (Tabela 1).



Fonte: (Sousa et al., 2017).

**Figura 1.** Heredograma dos cruzamentos entre genitores do grupo de cultivares Catuaí e Híbrido de Timor (HdT), resultando em 13 progênies de *Coffea arabica*. A, C e E = genótipos Catuaí amarelo IAC 30, IAC 86 e IAC 64, respectivamente; B, D e F = genótipos HdT UFV 445-46, UFV 440-10 e UFV 530, respectivamente; G, H, I, J e K = híbridos oriundos de cruzamentos entre genitores Catuaí amarelos e HdT; 1, 3, 5 e 7 = progênies de geração de retrocruzamento resistente; 2, 4, 6, 8 e 9 = progênies de geração de retrocruzamento suscetível; 10, 11, 12 e 13 = progênies em geração F<sub>2</sub>.

**Tabela 1.** Progênes de *Coffea arabica* avaliadas em Viçosa (MG).

<b>Progênes</b>	<b>Genitor 1</b>	<b>Genitor 2</b>
RCr 1	H 419-1 c-17	UFV 445-46
RCs 2	H 419-1 c-17	UFV 2143-235
RCr 3	H 514-8 c-387	UFV 440-10
RCs 4	H 514-8 c-387	UFV 2154-344
RCr 5	H 514-7 c-364	UFV 440-10
RCs 6	H 514-7 c-364	UFV 2154-344
RCr 7	H 419-10 c-214	UFV 445-46
RCs 8	H 419-10 c-214	UFV 2143-235
RCs 9	UFV 2148-57	H 513-5 c-14
F <sub>2</sub> 10	H 514-8 c-387	-
F <sub>2</sub> 11	H 514-7 c-364	-
F <sub>2</sub> 12	H 419-10 c-214	-
F <sub>2</sub> 13	H 513-5 c-14	-

RCr = Retrocruzamento resistente; RCs = Retrocruzamento suscetível; F<sub>2</sub> = geração obtida pela autofecundação dos híbridos F<sub>1</sub>.

### ***Avaliações fenotípicas***

As avaliações fenotípicas foram realizadas nos anos 2014, 2015 e 2016, em que, em cada ano, se avaliou 18 características fenotípicas de importância agrônômica (Sousa et al., 2019). Dentre as características, onze foram contínuas, avaliadas a partir de mensurações, sendo essas, produção (PROD), comprimento da folha (CF), largura da folha (LF), comprimento de um ramo (CRP), números de nós vegetativos (NNV), número de nós reprodutivos (NNR), número total de frutos (NFRP), volume total de frutos (VFRP), altura de plantas (AP), diâmetro do caule (Dca) e diâmetro da copa (Dco). As demais características, tamanho dos frutos (TFr), uniformidade de maturação dos frutos (UMat), ciclo de maturação dos frutos (CMat), incidência de ferrugem (Fer), incidência de cercosporiose (Cer), infestação de bicho mineiro (BM) e vigor vegetativo (Vig), foram avaliadas de forma categórica, utilizando escala de notas. As descrições de como foram realizadas as avaliações de cada característica estão na Tabela 2.

### ***Análises dos dados fenotípicos***

Os dados fenotípicos foram corrigidos para efeitos de anos, parcelas e interação anos x parcelas. As análises foram realizadas considerando os modelos lineares mistos (procedimento REML/BLUP), por meio do *software* Selegen-REML/BLUP (Resende, 2016b). Os parâmetros genéticos foram estimados a partir da análise individual das 18 características, utilizando o seguinte modelo estatístico:

$$y = Xu + Zg + Wp + Vr + Tb + Ri + e$$

Em que:

$y$ : Equivale ao vetor de dados;

$u$ : Equivale ao vetor referente à média geral em cada ano de avaliação;

$g$ : Equivale ao vetor de efeitos de progênies (efeito aleatório);

$p$ : Equivale a variância permanente entre plantas (efeito aleatório);

$r$ : Equivale à variância entre tipos de populações (efeito aleatório);

$b$ : Equivale à variância entre parcela (efeito aleatório);

$i$ : Equivale à variância da interação progênies X anos (efeito aleatório);

$e$ : Equivale ao vetor de resíduos (efeito aleatório).

As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

**Tabela 2.** Descrição das avaliações fenotípicas realizadas nos anos de 2014, 2015 e 2016, em Viçosa (MG).

<b>Característica</b>	<b>Forma de avaliação</b>
Produção	Litros de café recém colhidos por planta
Comprimento de uma folha (cm) Largura de uma folha (cm)	Medidos em folha do 3º ou 4º par de um ramo plagiotrópicos do terço médio da planta
Comprimento de um ramo plagiotrópico (cm) Número de nós reprodutivos Número de nós vegetativos Número total de frutos Volume de frutos (litros)	Medidos no ramo plagiotrópico localizado no terço médio da planta
Altura de plantas (cm)	Medida do ramo ortotrópico (superfície do solo até o ponto final de crescimento do ramo)
Diâmetro da copa (cm)	Medido no sentido transversal à linha de plantio, medindo-se a maior projeção da "saia" do cafeeiro
Diâmetro do caule (cm)	Medido na região do coleto da planta (cerca de 5 cm da superfície do solo)
Tamanho dos frutos maduros	Avaliado por notas de 1 a 3
Uniformidade de maturação dos frutos	Avaliada por nota de 1 a 4
Ciclo de maturação dos frutos Incidência de Ferrugem Incidência de Cercosporiose Infestação de Bicho-Mineiro	Avaliada por notas de 1 a 5
Vigor vegetativo	Avaliado por notas de 1 a 10

### ***Extração do DNA genômico***

A extração do DNA genômico foi realizada por Sousa et al. (2019). Extraiu-se o DNA de folhas jovens e completamente expandidas dos 195 genótipos de *C. arabica*, utilizando a metodologia proposta por Diniz et al. 2005, no Laboratório de Biotecnologia do Cafeeiro – Biocafé/UFV. A concentração do DNA foi verificada em *NanoDrop2000* e sua qualidade avaliada em gel de agarose 1%. Após a padronização da concentração de DNA das amostras, as mesmas foram enviadas para empresa *Rapid Genomics*, localizada na Flórida/EUA, para construção das sondas, sequenciamento e identificação dos marcadores moleculares SNP. As análises de qualidade das sequências foram conduzidas com auxílio do software Rbio (Bhering, 2017).

### ***Análises de qualidade dos SNP***

Um total de 40.000 sondas de DNA foram construídas para identificação dos marcadores SNP e genotipagem dos cafeeiros. Destas, 10.000 sondas polimórficas foram selecionadas (Resende et al., 2016a). Essas sondas apresentavam tamanho de 120 pb, e foram construídas com base nas informações genômicas disponíveis das espécies *C. canephora* e *C. arabica*. Para *C. arabica* utilizou-se as sequências de ESTs (*Expressed Sequence tag*) desenvolvidas pelo Projeto Genoma Brasileiro do Café (Vieira et al., 2006), visto que, o genoma dessa espécie ainda não estava disponível. As informações genômicas de *C. canephora* foram obtidas a partir de sequências ESTs de *C. canephora* (Vieira et al., 2006) e do genoma de referência de *C. canephora* (Denoëud et al., 2014). Outras informações acerca da construção das sondas e identificação dos SNP podem ser consultadas em Sousa et al. (2017).

Os 195 cafeeiros foram genotipados usando as 10.000 sondas e a metodologia de *Capture-Seq (RAPiD genomics)*, resultando em 21.211 marcadores SNP. Os SNP foram submetidos à análise de qualidade, sendo removidos os que não atenderam aos critérios mínimos estabelecidos na análise. Os parâmetros de qualidade adotados foram CR (*Call Rate*) igual ou superior a 90% e MAF (*Minor Allele Frequency*) a 5%. Para o parâmetro MAF o nível crítico foi estabelecido pela equação  $MAF = \frac{1}{\sqrt{2N}}$ , onde N refere-se ao número de indivíduos avaliados. Outro critério de qualidade realizado foi a eliminação dos falsos SNP (Vidal et al., 2010), que pode ocorrer devido a poliploidia presente em *C. arabica*. A estratégia foi a exclusão dos SNP que não apresentaram variância entre os indivíduos que compõem a população em estudo, ou seja, a heterozigosidade foi igual a 1. Os dados foram

obtidos a partir do trabalho de Sousa et al. (2019), e esses foram utilizados para o estudo de associação genômica ampla.

### ***Identificação dos SNP em Coffea arabica***

Para a localização e verificação da distribuição dos marcadores SNP no genoma de *C. arabica*, realizou-se uma análise de similaridade *BLAST* (*Basic Local Alignment Search Tool*), entre as sequências das sondas originadas do genoma de *C. canephora* com o genoma de *C. arabica*. Por meio da localização da sequência de cada sonda, foi identificada a posição do SNP no genoma de *C. arabica*.

### ***Estudo de associação genômica ampla***

Para a realização das análises do estudo de associação genômica ampla, a estrutura genética da população foi obtida por meio das análises de componentes principais (PCA), utilizando a função *prcomp* do *software* R. A metodologia de PCA foi aplicada aos dados genotípicos para obter os componentes principais (CP) que capturam a maior parte da variação genética. Essa técnica multivariada permite a redução da massa de dados, com menor perda possível da informação (Hongyu et al., 2016). O percentual mínimo adotado para explicação da variação total foi igual a 85%, e os componentes principais obtidos foram utilizados como covariáveis no modelo de GWAS, para detectar SNP associados às características fenotípicas.

O modelo geral do GWAS é dado por:

$$Y_{ijk}^* = \mu + \text{SNP}_i + \sum_{k=1}^5 \text{PC}_{jk} + \varepsilon_{ijk}$$

Em que:

$Y_{ijk}^*$ : Equivale ao fenótipo ajustado;

$\mu$ : Equivale a média da população;

$\text{SNP}_i$ : Equivale ao efeito fixo do  $i^{\text{th}}$ SNP, ajustado como uma covariável (efeito de substituição do alelo);

$\text{PC}_{kj}$ : Equivale ao efeito fixo do  $k^{\text{th}}$ componente principal individual K;

$\varepsilon_{ijk}$ : Equivale ao erro associado ao  $Y_{ijk}^*$ .

O vetor  $\theta = [\mu, \text{SNP}_i, \text{PC}_{jk}]^T$  representa parâmetros desconhecidos. O Threshold adotado foi de  $-\log_{10}(2,54e^{-06})$ ,  $-\log(5,08e^{-07})$  calculado com base na correção do método de Bonferroni



para testes múltiplos ( $\alpha^* = \alpha / m$  onde  $m$  foi o número total de SNP usados na análise de associação) para identificar SNP significativos a 5% e 1% respectivamente.

Para quantificar o efeito dos alelos favoráveis, de acordo com o objetivo do programa de cafeeiro, estimou-se a razão entre a média fenotípica dos indivíduos que possuem um específico par de alelos e toda a população:

$$Rz = (\mu_s/\mu_g)-1*100$$

Em que:

Rz: Razão entre a média fenotípica dos indivíduos que possuem um específico par de alelos e toda a população;

$\mu_s$ : Média fenotípica dos indivíduos selecionados para 0,1,2 alelos;

$\mu_g$ : Média geral de todos os indivíduos.

O desequilíbrio de ligação foi estimado utilizando o parâmetro  $r^2$ , entre SNP localizados em uma mesma região do genoma. As análises de DL foram realizadas por meio de scripts personalizados no *software* R.

### ***Identificação de genes e anotação funcional***

Após o estudo de associação genômica ampla, os SNP com associações significativas foram analisados para a identificação de SNP localizados dentro de genes. As análises foram realizadas por meio de *scripts* personalizados no *software Python* v3.6. Essa análise permitiu a identificação dos genes em que os SNP estão inseridos. Posteriormente, foi analisada a anotação funcional dos genes contendo os SNP, utilizando como base o genoma de referência de *C. arabica* (NCBI, 2019).

## **Resultados e discussão**

### ***Análise dos marcadores SNP***

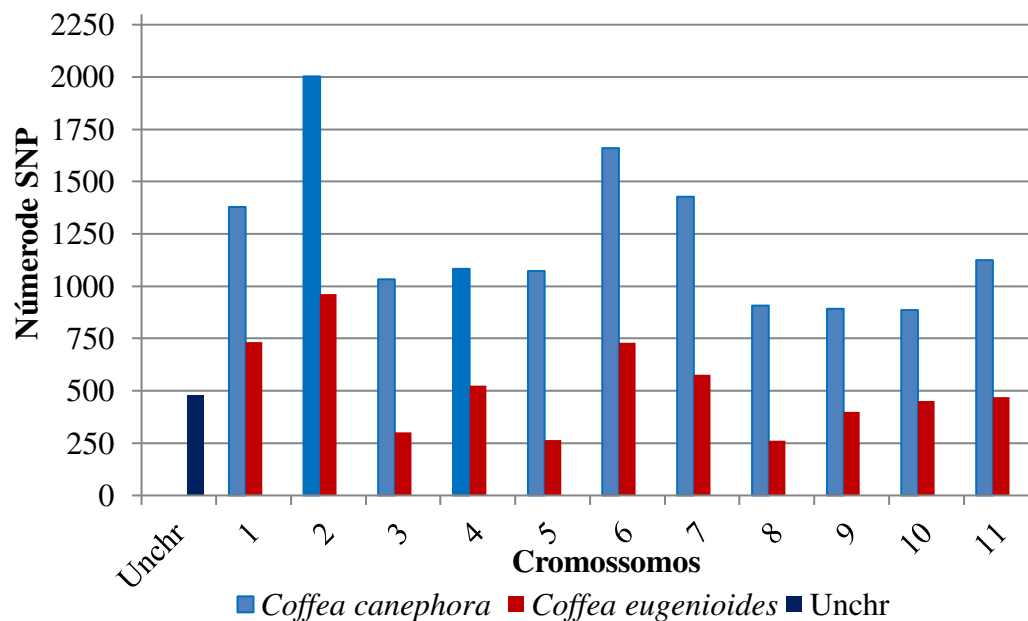
Com o intuito de garantir maior qualidade da genotipagem foram eliminados os SNP que não atenderam aos critérios mínimos de qualidade estabelecidos por meio dos parâmetros de MAF e *Call Rate*. Foram também excluídos os SNP que não apresentaram variância entre os indivíduos da população em estudo devido a poliploidia da espécie *Coffea arabica*. Essa filtragem resultou na redução de 3,46% do total de marcadores SNP, que passaram de 21.211 para 20.477, garantindo maior controle de qualidade da genotipagem, e, conseqüentemente, maior eficiência no GWAS (Turner et al., 2011; Sousa et al., 2019).

A análise de similaridade por meio do *best hit* permitiu a identificação de 19.623 SNP no genoma de *C. arabica*. Essa redução de 3,93% do total de marcadores SNP se deve ao fato de que a maioria das sondas para identificação dos SNP foram obtidas com base no genoma de referência de *C. canephora* (Resende et al., 2016a ; Sousa et al., 2017) e algumas delas não apresentaram alinhamento significativo com o genoma de *C. arabica*. E, além disso, algumas sondas apresentaram alinhamento significativo, porém, a região alinhada não era a região onde os SNP estavam localizados. A análise de similaridade garante a confiabilidade dos dados, fornecendo informações precisas sobre a posição dos SNP na espécie em estudo e demonstrando que todo o genoma de *C. arabica* foi analisado.

Após as análises de qualidade e de similaridade, foram identificados 19.623 SNP distribuídos pelo genoma de *C. arabica*. Esse número de SNP identificado é superior ao estudo de GWAS publicado por Sant'Ana et al. (2018), onde foram avaliadas características relacionadas à composição química do grão, trabalhando com um conjunto inicial de 6.696 marcadores SNP, dos quais foram selecionados 2.587 após a filtragem dos dados.

A genotipagem permitiu identificar SNP amplamente distribuídos pelo genoma de *C. arabica*, identificados em todos os cromossomos da espécie (Figura 2). Foram identificados também alguns SNP em sequências não ordenadas, isso porque, mesmo com o genoma de referência sequenciado, alguns *scaffolds* não foram agrupados em nenhum dos cromossomos. Esses *scaffolds* foram agrupados em um conjunto denominado *Uncharacterized chromosome* (Unchr) (NCBI, 2019).

O número de SNP por cromossomo variou de 479 (Unchr) a 2.965 (cromossomo 2). Como *C. arabica* é um alotetraploide (Clarindo e Carvalho, 2008), proveniente de *C. canephora* e *C. eugenioides* (Lashermes et al., 1999), o seu genoma é dividido em dois subgenomas, um de cada espécie de origem. Dessa forma, as informações genômicas estão representadas por dois conjuntos de cromossomos, um referente à *C. canephora* e o outro referente à *C. eugenioides* (Figura 2).



**Figura 2.** Marcadores moleculares SNP distribuídos ao longo dos 11 cromossomos e no *Uncharacterized chromosome* (Unchr) posteriormente ao alinhamento com genoma de *C. arabica*.

### ***Estudo de associação genômica ampla***

Para a análise do estudo de associação genômica ampla, os valores fenotípicos foram corrigidos para efeitos de anos, parcelas e interação anos x parcelas, por meio da metodologia REML/BLUP. O GWAS possibilitou a identificação de 110 SNP com associações significativas a um nível de 5% de probabilidade e 42 SNP com associações significativas a 1%, para sete das 18 características agronômicas estudadas. As características agronômicas que apresentaram associações significativas foram: altura de plantas, comprimento do ramo plagiotrópico, diâmetro da copa, número de nós vegetativos, tamanho de fruto, incidência de ferrugem e incidência de cercosporiose. Nas análises desse trabalho foram considerados os SNP com associação significativa a 5%. Ressalta-se que todas as características agronômicas avaliadas no presente trabalho são fundamentais para o desenvolvimento de cultivares de café que atendam as principais exigências do mercado, e que ainda não haviam sido trabalhadas em GWAS. O GWAS é uma ferramenta capaz de identificar locos e genes candidatos para características de caráter quantitativo (Su et al., 2019).

Para as características produção (PROD), comprimento de uma folha (CF), largura de uma folha (LF), número de nós reprodutivos (NNR), número total de frutos (NFRP), volume

total de frutos (VFRP), diâmetro do caule (Dca), uniformidade de maturação (UMat), ciclo de maturação (CMat), vigor vegetativo (VIG) e incidência de bicho-mineiro (BM) não foram identificados SNP com associações significativas com limiar (Threshold  $\log_{10} P$ ) adotado de  $2,54e^{-06}$  e  $5,08e^{-07}$  calculado com base na correção do método de Bonferroni. A correção de Bonferroni visa controlar globalmente ao nível  $\alpha$  de significância para todos os testes realizados (Miller, 1981). De acordo com Resende et al. (2017), a GWAS capta uma proporção pequena da variação genotípica, pois avalia cada alelo individualmente, por isso, para algumas características que apresentam heranças mais complexas, a metodologia pode ser limitada. Nesse caso, a característica pode ser controlada por muitos alelos raros, cada um deles apresentando grande efeito sobre o fenótipo, ou, pode apresentar muitos alelos comuns de pequenos efeitos e esses não conseguem ser detectados (Gibson, 2012).

Sousa et al. (2019) utilizaram as mesmas populações de melhoramento de *C. arabica* e as mesmas características desse trabalho para o estudo de seleção genômica. Seus resultados corroboram com os encontrados neste trabalho, pois as características que não apresentaram SNP com associações significativas PROD, CF, LF, NNR, NFRP, VFRP, Dca, UMat, CMat, VIG e BM obtiveram baixa herdabilidade genômica (0,26; 0,29; 0,32; 0,23; 0,34; 0,25; 0,16; 0,28; 0,31; 0,34; 0,30, respectivamente). Além disso, foram as características que apresentaram as maiores estimativas do número de QTL que controlam a característica (751; 3981; 17758; 1157; 1081; 1658; 14775; 1313; 440; 476, respectivamente), exceto NNR, para a qual o número de QTL não foi estimado. Esses resultados confirmam a complexidade das características, e, justifica a não captura da variação pela metodologia de GWAS para essas características.

### ***Estudo de associação genômica ampla para altura de plantas***

Cinquenta e seis SNP apresentaram associações significativas para a característica altura de plantas ( $p < 0,05$ ) (Tabela 3), distribuídos nos cromossomos 2 (chr2), 6 (chr6), 7 (chr7) e Unchr (*Uncharacterized chromosome*) (Figura 3). Visto que o Unchr é constituído por um conjunto de *scaffolds* com sequências não ordenadas, as informações encontradas nesta região não foram discutidas no presente trabalho. Portanto, considerou-se 55 SNP informativos associados à altura de plantas.

A partir dos valores da razão entre a média fenotípica corrigida dos indivíduos que possuem um dos alelos específico e toda a população, foi possível determinar, em termos percentuais, o efeito do alelo de interesse que resulta na melhoria da característica AP no contexto do melhoramento (Tabela 3). Dessa forma, esse efeito permite identificar os alelos

favoráveis, associados à redução de AP, uma vez que os programas de melhoramento do cafeeiro buscam selecionar plantas de porte baixo para facilitar os tratos culturais e principalmente a colheita.

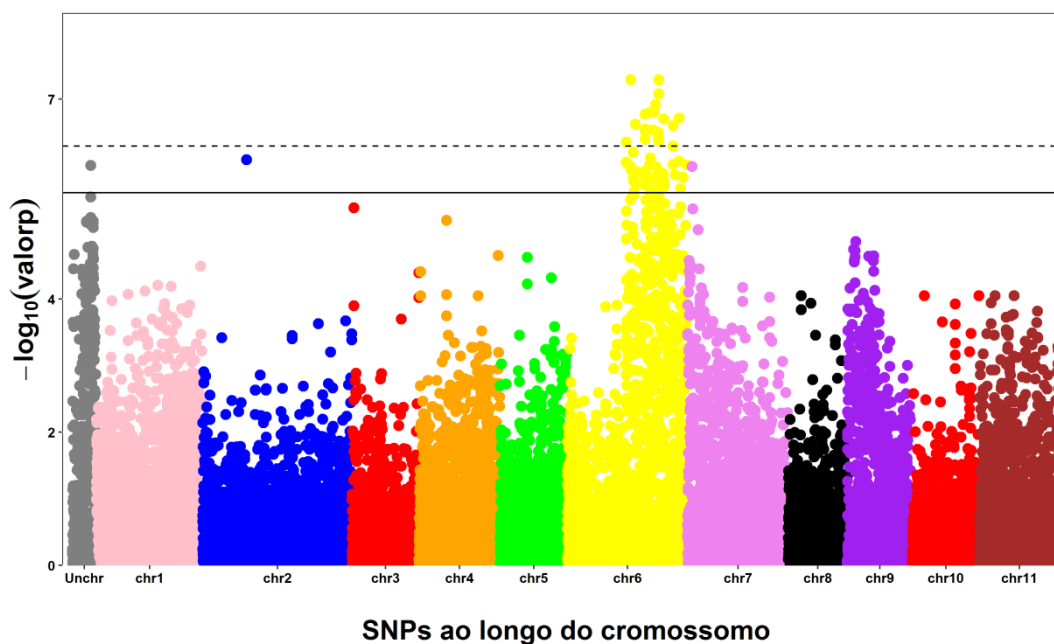
**Tabela 3.** SNP com associações significativas ( $p < 0,05$ ), posição, alelos favoráveis e efeitos associados à redução de altura de plantas.

SNP	Chr	Posição	<i>p-value</i>	R <sup>2</sup>	Alelo	Alelo Favorável	Efeito(%)
V17031	chr7_c	677592	1,03E-06	0,11360	C/T	T	4,63859
V16847*	Unchr	806813	9,95E-07	0,13497	C/G	G	7,42445
V16042	chr2_e	12528092	8,18E-07	0,13986	T/G	T	8,59216
V15899	chr6_e	16766939	8,84E-07	0,11037	T/C	T	5,81278
V15914	chr6_e	17010018	4,41E-07	0,11965	A/G	G	6,11464
V15959	chr6_c	18653343	1,02E-06	0,13359	T/G	T	7,84281
V15966	chr6_c	18683698	5,12E-08	0,16420	G/A	A	9,32675
V15967	chr6_c	18683725	1,15E-06	0,13684	C/T	T	8,03744
V16012	chr6_c	19668205	1,56E-06	0,13291	G/T	T	7,58232
V16014	chr6_c	19668268	6,28E-07	0,14141	T/C	C	8,56697
V16027	chr6_c	19711172	2,52E-06	0,12814	A/G	G	7,79241
V16029	chr6_c	19711182	1,75E-06	0,13192	A/T	T	7,81282
V16050	chr6_c	20516595	2,37E-07	0,14913	A/T	A	8,43056
V16102	chr6_e	20768840	1,93E-06	0,13382	A/T	A/T	4,70157
V16099	chr6_c	21356008	1,25E-06	0,10790	C/T	T	19,56234
V16297	chr6_c	26782635	1,93E-06	0,12073	C/A	C	5,29129
V16298	chr6_c	26782664	1,83E-06	0,11905	C/T	C	5,12150
V16293	chr6_c	26937789	3,99E-07	0,14288	C/T	T	7,50054
V16294	chr6_c	26937799	1,68E-07	0,15273	G/A	G	7,62358
V16295	chr6_c	26937808	1,22E-06	0,13603	A/G	A	7,40663
V16296	chr6_c	26937817	2,99E-07	0,14881	G/A	G	7,88765
V15785	chr6_c	30136750	9,53E-07	0,10301	T/C	C	4,60664
V15786	chr6_c	30136754	1,60E-07	0,11814	A/C	C	4,98635
V1190	chr6_c	30457537	1,53E-06	0,13143	T/C	C	21,21391
V15787	chr6_c	30968458	7,64E-07	0,13943	A/G	A	7,66001
V1094	chr6_c	31352169	2,34E-06	0,13012	C/T	T	7,18741
V16180	chr6_c	32858769	2,93E-07	0,14951	G/A	A	8,48874
V16206	chr6_c	33299570	1,57E-07	0,15237	A/G	G	8,46431
V1666	chr6_c	33759015	1,21E-07	0,15923	T/G	G	7,90736
V1504	chr6_c	34372993	1,04E-06	0,13284	G/T	G	6,92785
V16508	chr6_c	34559139	1,39E-06	0,10804	T/C	T	4,87099
V16509	chr6_c	34559142	1,79E-06	0,10844	T/C	T	4,88017
V16482	chr6_c	34836377	2,32E-06	0,12723	G/A	G	6,92015
V16481	chr6_c	34852348	3,98E-07	0,14382	A/G	A	8,26294
V16476	chr6_c	34852929	3,19E-07	0,13838	T/C	T	6,63111

Continuação...

SNP	Chr	Posição	<i>p-value</i>	R <sup>2</sup>	Alelo	Alelo Favorável	Efeito(%)
V16479	chr6_c	34852983	8,28E-08	0,15060	A/G	A	7,22469
V16480	chr6_c	34852993	4,04E-07	0,13818	A/G	A	6,83245
V1690	chr6_c	35004072	2,07E-06	0,10846	G/A	G	7,98351
V16589	chr6_c	37176686	1,31E-06	0,12377	G/A	G	6,12801
V2101	chr6_c	37696025	1,98E-07	0,15070	A/C	C	7,83073
V40	chr6_c	37767529	1,75E-06	0,12640	T/A	T	7,68267
V41	chr6_c	37767535	1,79E-06	0,12654	T/C	T	7,70794
V42	chr6_c	37767560	2,05E-06	0,12198	T/C	T	7,51621
V16391	chr6_e	43620472	5,09E-07	0,14393	T/A	A	8,11656
V16392	chr6_e	43620494	2,50E-07	0,15093	G/A	A	8,17788
V16395	chr6_e	43620540	9,61E-07	0,13135	C/T	T	7,41266
V16418	chr6_e	47341413	1,37E-06	0,12705	T/C	C	7,32441
V16762	chr6_e	51038354	8,60E-07	0,12894	C/A	A	7,62583
V16744	chr6_e	51085545	1,93E-07	0,14412	G/A	A	7,80832
V16709	chr6_e	51312825	2,21E-06	0,10674	G/C	C	5,65360
V16630	chr6_e	51939316	1,50E-06	0,12542	A/G	A	7,61825
V16336	chr6_c	54173849	9,68E-07	0,13528	A/G	G	7,77570

\* SNP localizados no Unchr.

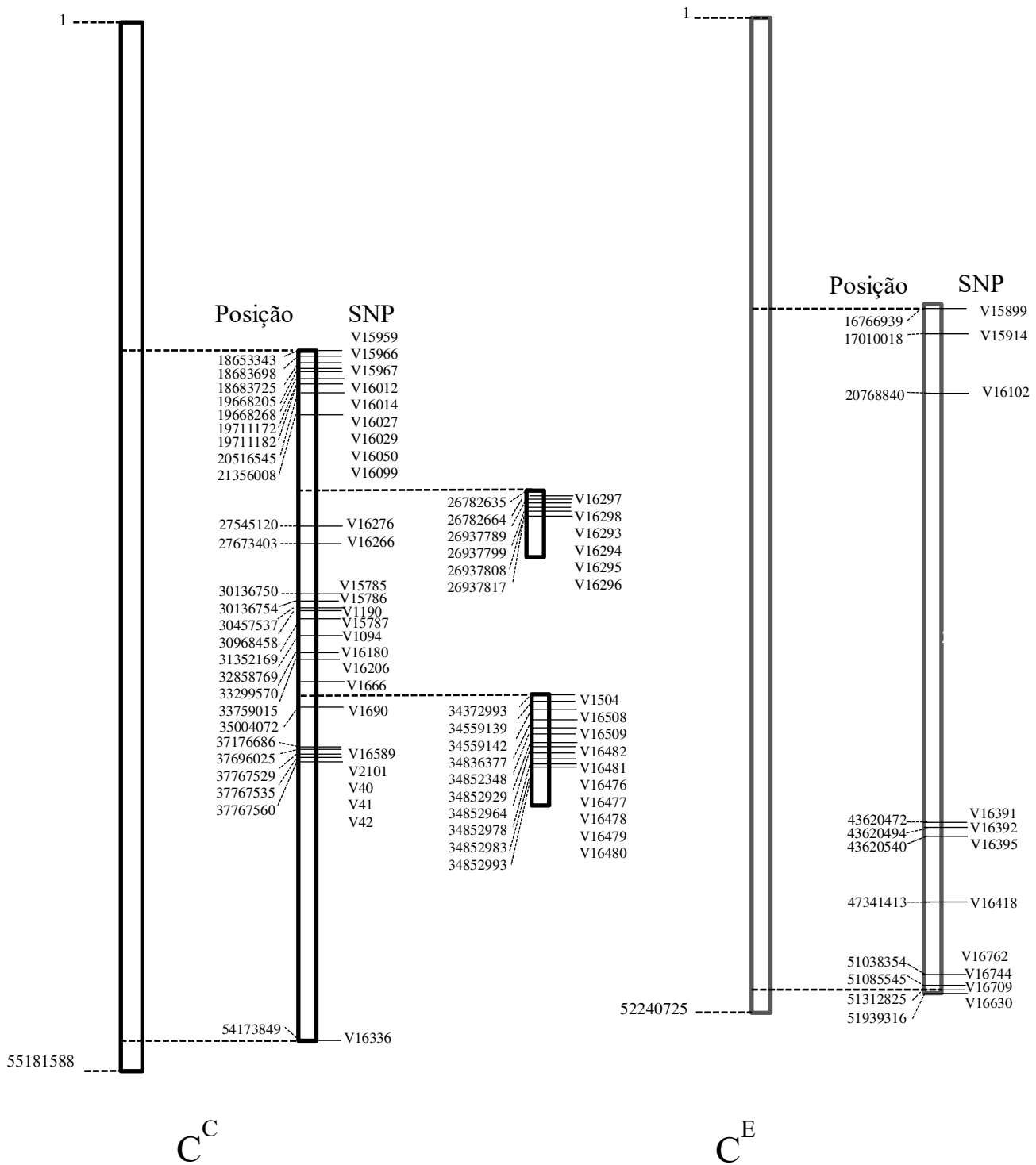


**Figura 3.** Gráfico de Manhattan para altura de plantas. Os SNP acima da linha tracejada horizontal mostram a significância com um limiar moderadamente restrito de  $-\log_{10}$  (0,01/19.623). Gráficos acima da linha contínua horizontal mostram a significância com um limiar de  $-\log_{10}$  (0,05/19.623). As diferentes cores indicam gráficos para diferentes cromossomos, que seguem a ordem: Unchr (*Un characterized chromossome*)- chr11 (Cromossomo 11).

Os alelos favoráveis e os efeitos dos SNP associados à redução de altura de plantas estão apresentados na Tabela 3. Os efeitos associados à redução de AP variaram de 4,61% (V15785) a 21,21% (V1190). Os SNP V1190 e V16099 apresentaram os maiores efeitos associados à redução de AP, 21,21% e 19,56% respectivamente. Associações com maiores efeitos geralmente são mais úteis em programas de melhoramento (Su et al., 2019).

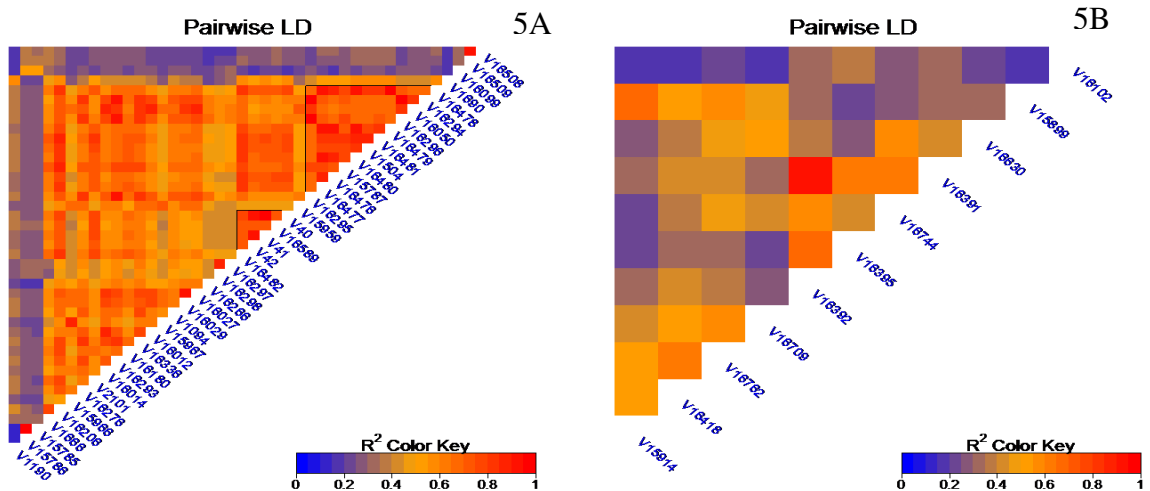
Dentre os 55 SNP associados a AP (Tabela 3), 96,36% estão localizados no cromossomo 6, e o restante, 3,64%, estão no cromossomo 2 e 7. Essa grande fração de SNP no cromossomo 6 pode sugerir um provável QTL no controle dessa característica. A posição dos 53 SNP localizados no cromossomo 6 está apresentada na Figura 4. Foram identificados 42 SNP no subgenoma de *C. canephora* e 11 SNP no subgenoma de *C. eugenioides*. Com base na posição dos 53 SNP associados com o fenótipo corrigido para altura de plantas identificados no cromossomo 6 (Figura 4), pode-se sugerir que as informações genômicas provenientes dos dois subgenomas possivelmente estão atuando no controle dessa característica. Contudo, a maior quantidade de SNP foram localizados na parte do subgenoma referente a *C. canephora*.

Diante do número elevado de SNP no cromossomo 6, realizou-se a análise de desequilíbrio de ligação (DL) com os SNP significativos para AP localizados neste cromossomo. A análise foi realizada para cada conjunto cromossômico, um referente à *C. canephora* e o outro referente à *C. eugenioides* (Figura 5). Por meio do  $r^2$  identificou-se SNP em DL. No conjunto cromossômico referente ao subgenoma de *C. eugenioides* destaca-se os SNP V16391 e V16392 em forte DL. No conjunto cromossômico referente ao subgenoma de *C. canephora* identificou-se blocos em DL. Identificou-se um bloco maior que se estende do SNP V16478 ao V16295, e outro bloco menor que se estende do SNP V40 ao V42 (Figura 5). Essas informações são de suma importância, pois podem auxiliar na seleções assistidas, permitindo selecionar os haplótipos desejáveis. Essa estratégia resulta em redução do tempo dos programas de melhoramento (Su et al., 2019).



**Figura 4.** Posição dos SNP no cromossomo 6 no genoma de *C. arabica*. O  $C^C$  representa o cromossomo referente ao subgenoma de *C. canephora* e o  $C^E$  representa a parte do subgenoma referente à *C. eugenioides*.





**Figura 5.** Bloco de desequilíbrio de ligação formado por SNP com associações significativas para AP localizados em uma região no cromossomo 6 de *C. canephora* (A) e *C. eugenioides* (B). Os valores de  $r^2$  são indicados usando a intensidade da cor na parte inferior direita.

A análise de identificação de genes com base nas informações do NCBI (2019) permitiu detectar 16 SNP associados ao fenótipo corrigido AP, inseridos em genes (Tabela 4). Identificou-se diferentes SNP inseridos no mesmo gene, como os SNP V16042, V15966 e V15967 localizados no gene denominado LOC113691295, anotado como Histidina - tRNA ligase, cloroplástico/mitocondrial. Os SNP V40, V41 e V42 foram identificados no gene denominado LOC113693412, anotado como provável proteína-quinase semelhante a receptor rico em leucina, at5g49770. Os SNP V16508, V16509, V16630 e V17031 foram localizados inseridos em genes, no entanto, os genes não apresentam anotação funcional. Os demais SNP estão inseridos nos genes LOC113692598, LOC113694028, LOC113697149, LOC113695058, LOC113695139 e LOC113696954. A anotação funcional dos mesmos está apresentada na Tabela 5. Os SNP que não foram localizados dentro de genes são também relevantes para uso como marcador genético no melhoramento. Marcador molecular não precisa estar necessariamente inserido em gene para detectar diferenças genéticas entre indivíduos, o mesmo pode estar associado ao gene e ser eficiente (Andersen e Lubberstedt, 2013). Além disso, o SNP pode estar em regiões promotoras ou reguladoras, e, portanto, envolvidos na expressão gênica.

De acordo com a anotação funcional dos genes em que os SNP estão inseridos, não foi identificado nenhum mecanismo que tenha influência direta no controle de altura de

plantas. Isso pode ser explicado devido ao fato da complexidade da característica em estudo, evidenciando a presença de mais genes atuando no controle da mesma. A característica altura de plantas é de caráter quantitativo, governada por muitos genes, o que justifica o número elevado de SNP atuando no controle dessa característica. De acordo com Sousa et al. (2019), que avaliou essa característica nas mesmas populações de *C. arabica*, altura de plantas apresentou valor moderado de herdabilidade genômica (0,46) e estimativas de número de QTL elevada (202). Esses valores confirmam a complexidade da característica.

**Tabela 4.** Anotação funcional dos SNP inseridos em genes para AP.

SNP	Chr	Posição	Gene	Anotação funcional
V16042	chr2_e	12528092	LOC113691295	Histidina - tRNA ligase, cloroplástico / mitocondrial
V15966	chr6_c	18683698	LOC113691295	Histidina - tRNA ligase, cloroplástico / mitocondrial
V15967	chr6_c	18683725	LOC113691295	Histidina - tRNA ligase, cloroplástico / mitocondrial
V16050	chr6_c	20516595	LOC113692598	Proteína de montagem ferro-enxofre IscA-like 2, mitocondrial
V16206	chr6_c	33299570	LOC113694028	Dihydroneopterin aldolase 1
V16508	chr6_c	34559139	LOC113693325	Não caracterizado
V16509	chr6_c	34559142	LOC113693325	Não caracterizado
V40	chr6_c	37767529	LOC113693412	Provável proteína-quinase semelhante a receptor rico em leucina, at5g49770
V41	chr6_c	37767535	LOC113693412	Provável proteína-quinase semelhante a receptor rico em leucina, at5g49770
V42	chr6_c	37767560	LOC113693412	Provável proteína-quinase semelhante a receptor rico em leucina, at5g49770
V15899	chr6_e	16766939	LOC113697149	Proteína estrictosidina sintase 10-like
V15914	chr6_e	17010018	LOC113695058	Álcool benzílico O-benzoiltransferase
V16744	chr6_e	51085545	LOC113695139	Putativa proteína tipo serina / treonina-proteína quinase CCR3
V16709	chr6_e	51312825	LOC113696954	Proteína contendo domínio B3, semelhante a At3g25182
V16630	chr6_e	51939316	LOC113697088	Não caracterizado
V17031	chr7_c	677592	LOC113699150	Não caracterizado

#### ***Estudo de associação genômica ampla para comprimento do ramo plagiotrópico***

Identificou-se 22 SNP com associações significativas ( $p < 0,05$ ) para comprimento do ramo plagiotrópico (CRP) (Tabela 5). Os SNP com associações significativas estão distribuídos no cromossomo 7 (chr07) e *Unchr* (Figura 6). Dentre os 22 SNP associados ao CRP, apenas cinco que estão localizados no cromossomo 7 são informativos, pois os demais

foram localizados no *Unchr*. Os efeitos associados ao aumento de CRP foram calculados e os percentuais dos SNP informativos localizados no cromossomo 7 variaram de 5,57% a 12,96%. Foi considerado alelo de interesse, aquele que resultou em aumento do CRP. Para o cafeeiro, o aumento do comprimento de ramos plagiotrópicos pode resultar no aumento da produtividade. Estudos prévios demonstraram que o comprimento do ramo plagiotrópico é uma característica correlacionada positivamente com produção (Carvalho et al., 2011; Teixeira et al., 2012) e, dessa forma, o incremento dessa característica pode acarretar em aumento da produção.

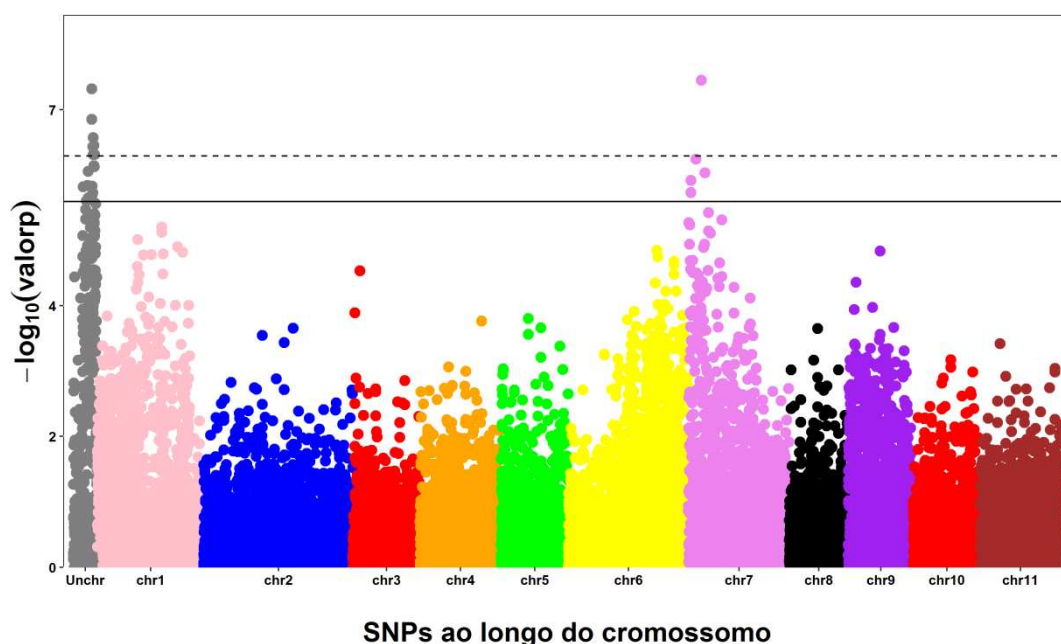
**Tabela 5.** SNP com associações significativas ( $p < 0,05$ ), posição, alelos favoráveis e efeitos associados ao aumento do comprimento do ramo plagiotrópico.

SNP	Chr	Posição	<i>p-value</i>	R <sup>2</sup>	Alelo	Alelo Favorável	Efeito(%)
V16919*	Unchr	106103	1,51E-06	0,10568	C/G	C/G	9,05564
V16900*	Unchr	218187	2,48E-06	0,11456	G/T	C	26,78280
V16888*	Unchr	331436	1,45E-06	0,11899	G/A	G/A	8,19117
V16986	chr7_c	394531	1,87E-06	0,1308	G/A	G	12,33586
V16987	chr7_c	394547	1,22E-06	0,13507	C/G	G	12,95945
V1309*	Unchr	453050	8,93E-07	0,13702	T/C	T	15,51539
V1310*	Unchr	453100	2,26E-06	0,12206	G/A	G/A	6,54548
V16847*	Unchr	806813	1,42E-07	0,15015	C/G	C	18,42508
V16848*	Unchr	806847	2,16E-06	0,1177	G/A	G	16,51437
V16849*	Unchr	806859	4,83E-08	0,14968	A/T	A	16,51437
V16832*	Unchr	844559	1,49E-06	0,13332	G/C	G/C	4,98871
V16828*	Unchr	884969	3,69E-07	0,13557	C/G	C	17,53219
V16825*	Unchr	901197	1,89E-06	0,11433	G/A	G/A	8,02920
V16816*	Unchr	950061	4,90E-07	0,14133	A/G	A	14,01326
V16817*	Unchr	950062	2,69E-07	0,14783	G/T	G	14,71262
V16798*	Unchr	1045911	3,57E-07	0,14367	C/T	C	16,24930
V16799*	Unchr	1045937	7,27E-07	0,13801	G/A	G	14,99060
V16800*	Unchr	1045944	7,54E-07	0,13683	T/G	T	16,07630
V16796*	Unchr	1056119	4,89E-07	0,11877	C/T	C/T	9,09268
V17241	chr7_e	1096941	5,73E-07	0,12744	A/C	A/C	5,56725
V17289	chr7_e	1621064	3,57E-08	0,12204	T/A	T/A	9,83961
V17181	chr7_c	2700421	9,29E-07	0,13175	T/C	T	12,44749

\* SNP localizados no *Unchr*.

Os SNP V17289 e V17241 foram localizados inseridos dentro de genes, esses SNP estão inseridos nos genes LOC113701091 e LOC113701577, respectivamente. Com base no NCBI (2019), o gene denominado LOC113701091 não apresenta função conhecida. O gene

LOC113701577, localizado no cromossomo 7 ( $C^E$ ), está descrito como proteína do tipo membrana plasmática regulada a frio. Esse gene está envolvido na resposta das plantas a baixas temperaturas, principalmente porque esse estresse à baixa temperatura pode influenciar na composição da membrana plasmática, reduzindo a fluidez e proporcionando o aumento da rigidez da membrana celular (Orvar et al., 2000; Su et al., 2018). Os SNP V17289 e V17241 apresentaram efeito para aumento de CRP quando estão em heterozigose (Tabela 6). Esse resultado sugere um possível efeito de heterose para esse loco.



**Figura 6.** Gráfico de Manhattan para comprimento de ramo (A). Os SNP acima da linha tracejada horizontal mostram a significância com um limiar moderadamente restrito de  $-\log_{10}$  (0,01/19.623). Gráficos acima da linha contínua horizontal mostram a significância com um limiar de  $-\log_{10}$  (0,05/19.623). As diferentes cores indicam gráficos para diferentes cromossomos, que seguem a ordem: Unchr (*Un caracterizad chromossome*)- chr11 (Cromossomo 11).

#### ***Estudo de associação genômica ampla para número de nós vegetativos***

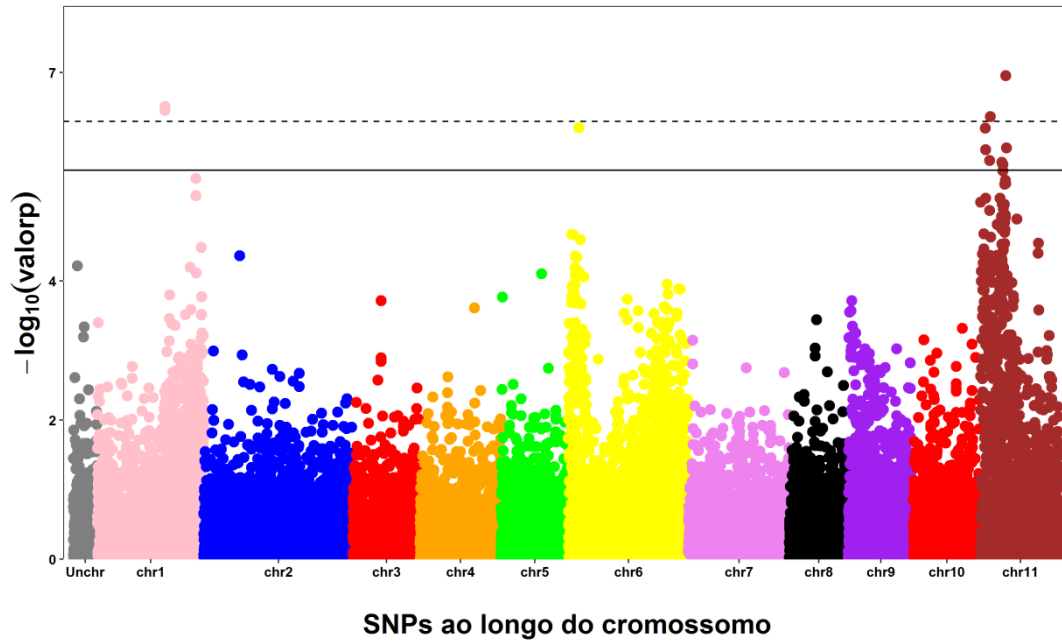
Doze SNP apresentaram associações significativas para número de nós vegetativos (NNV) ( $p < 0,05$ ) (Tabela 6). Os SNP com associações significativas estão distribuídos nos cromossomos 1 (chr01), 6 (chr06) e 11 (chr11) (Figura 7).

Os SNP V14783, V4256, V4257 apresentaram os maiores efeitos percentuais associados ao aumento de NNV (Tabela 6), 16,32%, 15,97% e 16,30%, respectivamente. Esses

mesmos SNP foram os únicos identificados inseridos em genes. Os SNP V4256 e V4257 foram identificados no mesmo gene LOC113713452. No entanto, esse gene não apresenta função conhecida. O SNP V14783 está inserido no gene LOC113697187 localizado no cromossomo 6 (C<sup>E</sup>) (Tabela 6). O gene LOC113697187 está descrito como uma proteína do tipo transportadora de auxina 2. A auxina é um fitohormônio que está diretamente relacionado com a regulação do crescimento em vegetais, se encontra em maior concentração no ápice das plantas, e pode suprir indiretamente o desenvolvimento de gemas laterais (Azizi et al., 2015; Baroni et al., 2017). Assim, a concentração correta deste fitohormônio é fundamental para desenvolvimento do número de nós vegetativos. Além disso, assim como CRP, NNV também apresenta correlação positiva com produção (Carvalho et al., 2011; Teixeira et al., 2012) e, por isso, é fundamental para os programas de melhoramento do cafeeiro.

**Tabela 6.** SNP com associações significativas ( $p < 0,05$ ), posição, alelos favoráveis e efeitos associados ao aumento de número de nós vegetativos.

SNP	Chr	Posição	<i>p-value</i>	R <sup>2</sup>	Alelo	Alelo Favorável	Efeito(%)
V14783	chr6_e	1501319	6,32E-07	0,10666	T/C	C	16,29957
V2570	chr11_e	3278100	6,35E-07	0,09448	A/G	G	13,16646
V2571	chr11_e	3278111	1,29E-06	0,094	A/C	C	11,40995
V6021	chr11_c	5295325	1,85E-06	0,09647	T/C	C	9,70242
V6011	chr11_c	5324733	4,36E-07	0,10317	A/G	A	10,14948
V6099	chr11_c	19107508	1,97E-06	0,09826	T/C	T	8,55304
V1764	chr11_c	20518053	2,11E-06	0,0983	G/A	A	9,60625
V1766	chr11_c	20518070	2,13E-06	0,09719	C/G	G	9,50464
V6225	chr11_c	22385238	1,11E-07	0,12358	T/A	T	10,28526
V6256	chr11_c	23177469	1,23E-06	0,10535	G/C	G	9,13990
V4256	chr1_e	44489730	3,50E-07	0,12602	G/A	G	16,32219
V4257	chr1_e	44489733	3,09E-07	0,1278	T/C	T	15,97100



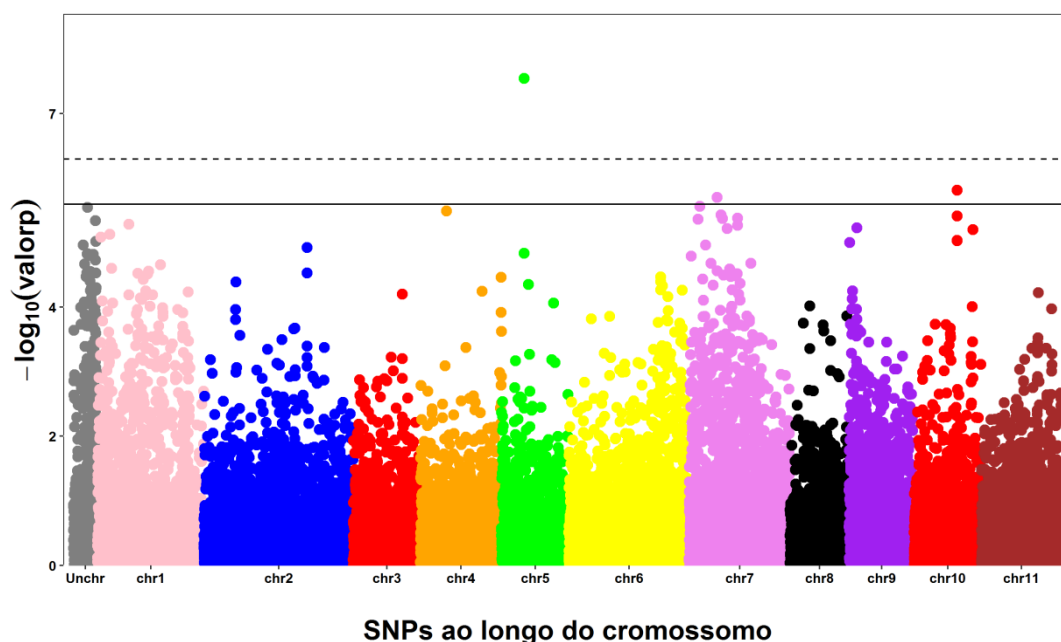
**Figura 7.** Gráfico de Manhattan para número de nós vegetativos. Os SNP acima da linha tracejada horizontal mostram a significância com um limiar moderadamente restrito de  $-\log_{10}$  (0,01/19.623). Gráficos acima da linha contínua horizontal mostram a significância com um limiar de  $-\log_{10}$  (0,05/19.623). As diferentes cores indicam gráficos para diferentes cromossomos, que seguem a ordem: Unchr (*Un characterized chromossome*)- chr11 (Cromossomo 11).

#### ***Estudo de associação genômica ampla para incidência de cercosporiose***

Por meio do GWAS, detectou-se três SNP com associações significativas ( $p < 0,05$ ) para o fenótipo corrigido de incidência de cercosporiose (Tabela 7), distribuídos nos cromossomos 5(chr5), 7(chr7) e 10(chr10) (Figura 8). Os efeitos dos SNP associados à redução da cercosporiose estão na Tabela 6. A resistência à *Cercospora coffeicola*, agente etiológico da cercosporiose do cafeeiro, tem sido buscada nos programas de melhoramento genético dessa espécie, visto que, com as mudanças climáticas tem aumentado sua incidência e, conseqüentemente, os prejuízos nos últimos anos, tendo grande impacto na produção de mudas em viveiros (Paiva et al., 2013; Carvalho et al., 2017).

**Tabela 7.** SNP com associações significativas ( $p < 0,05$ ), posição, alelos favoráveis e efeitos associados a redução da incidência de cercosporiose (Cer).

SNP	Chr	Posição	$p$ -value	$R^2$	Alelo	Alelo Favorável	Efeito(%)
V13732	chr5_e	31756982	2,87E-08	0,13177	A/G	G	3,34379
V5751	chr10_e	37767950	1,55E-06	0,11192	T/C	T/C	9,8906
V17709	chr7_e	4909405	2,01E-06	0,10439	T/C	T	13,1672



**Figura 8.** Gráfico de Manhattan para incidência de cercosporiose. Os SNP acima da linha tracejada horizontal mostram a significância com um limiar moderadamente restrito de  $-\log_{10}$  (0,01/19.623). Gráficos acima da linha contínua horizontal mostram a significância com um limiar de  $-\log_{10}$  (0,05/19.623). As diferentes cores indicam gráficos para diferentes cromossomos, que seguem a ordem: Unchr (*Un characterized chromossome*)- chr11 (Cromossomo 11).

Todos os SNP associados à redução da incidência a cercosporiose estão inseridos em genes. O SNP V5751 está localizado no cromossomo 10 ( $C^E$ ), inserido no gene LOC113712673, descrito como Proteína F-box / kelch de repetição At1g57790. O SNP V17709 está localizado no cromossomo 7 ( $C^E$ ), inserido no gene LOC113701115, descrito como enzima conjugadora de ubiquitina E2 tipo 28. De acordo com Akter et al. (2014), esses genes estão envolvidos com mecanismos de defesa descritos com ação antibacteriana, contudo, outros genes podem estar correlacionados com fatores que reduzem a incidência também de doenças fúngicas, como a Cercosporiose. Porém, são necessários mais estudos que

darão sustentação para confirmar a função e o modo de ação desses genes no controle da cercosporiose em cafeeiro.

O SNP V13732 está localizado no cromossomo 5 (C<sup>E</sup>), inserido no gene LOC113687823. Esse gene corresponde a um fator de transcrição MYB4. Fatores de transcrição atuam como reguladores da expressão gênica, levando a ativação dos genes. Além disso, a família dos fatores de transcrição MYB está envolvida na regulação da tolerância aos estresses bióticos e abióticos em plantas (Roy, 2016). Portanto, esse é um SNP associado à redução de incidência de cercosporiose, que deve ser validado, podendo resultar em um marcador funcional, contribuindo para o avanço dos programas de melhoramento dessa espécie. Sobretudo porque ainda não existem disponíveis na literatura, marcadores moleculares validados e ligados à resistência a essa doença.

### ***Estudo de associação genômica ampla para incidência de ferrugem***

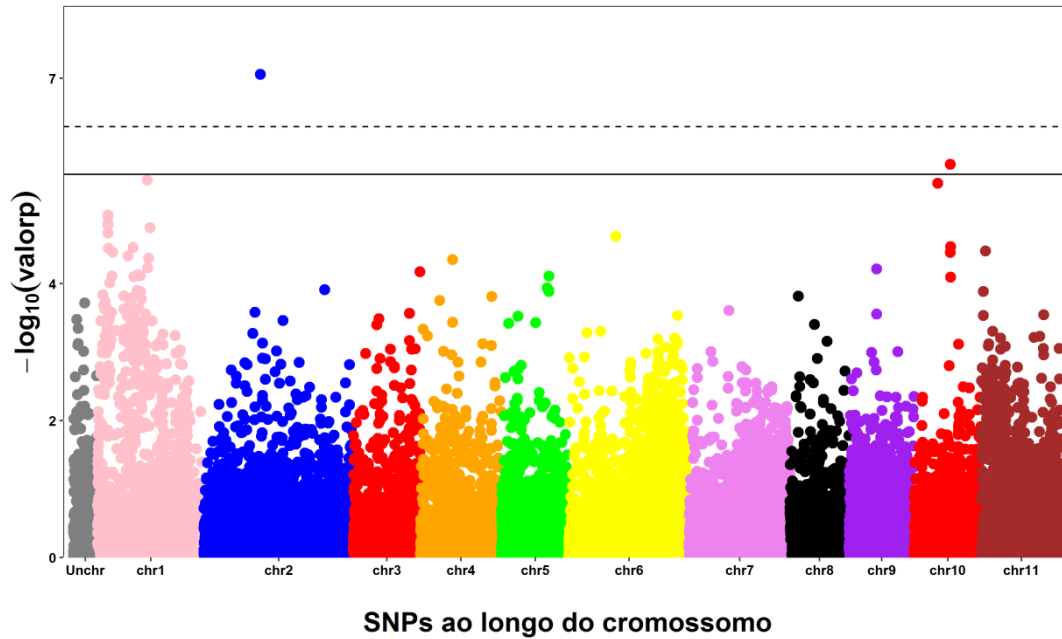
Para incidência de ferrugem foram identificados 2 SNP com associações significativas ( $p < 0,05$ ) (Tabela 8), e estão localizados nos cromossomos 2(chr2) e 10(chr10) (Figura 9). Os SNP V2369 e V2821 apresentaram efeitos de 7,09% e 7,23% associados à redução de incidência a ferrugem. A ferrugem, causada pelo fungo biotrófico *Hemileia vastatrix* é a doença de maior importância econômica que atinge a cafeicultura mundial (Zambolim, 2016). Assim, a busca por fontes resistência a essa doença é um dos principais objetivos dos programas de melhoramento do cafeeiro.

**Tabela 8.** SNP com associações significativas ( $p < 0,05$ ), posição, alelos favoráveis e efeitos associados a redução da incidência de ferrugem (Fer).

SNP	Chr	Posição	<i>p</i> -value	R <sup>2</sup>	Alelo	Alelo Favorável	Efeito(%)
V2369	chr2_e	16156335	8,89E-08	0,15572	C/G	C	7,09258
V2821	chr10_e	30568555	1,81E-06	0,12285	G/A	G/A	7,22672

Apenas o SNP V2821 foi identificado inserido em genes. Localizado no cromossomo 10 (C<sup>E</sup>), o SNP V2821 está inserido no gene LOC113711847. No entanto, de acordo com NCBI (2019), a anotação funcional desse gene não foi caracterizada.





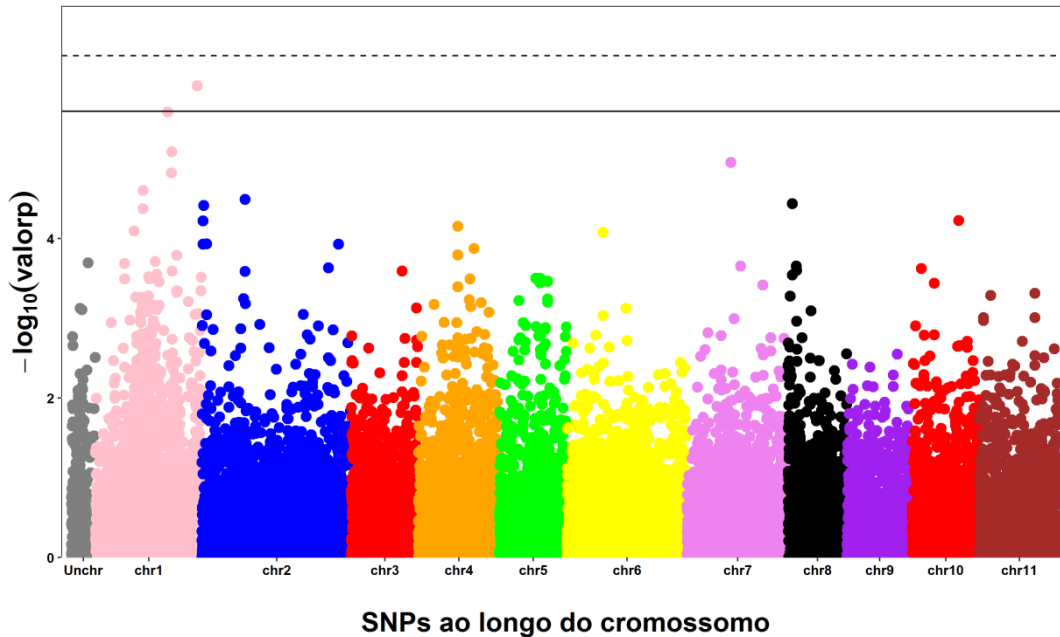
**Figura 9.** Gráfico de Manhattan para incidência de ferrugem (Fer). Os SNP acima da linha tracejada horizontal mostram a significância com um limiar moderadamente restrito de  $-\log_{10}(0,01/19.623)$ . Gráficos acima da linha contínua horizontal mostram a significância com um limiar de  $-\log_{10}(0,05/19.623)$ . As diferentes cores indicam gráficos para diferentes cromossomos, que seguem a ordem: Unchr (*Un caracterizad chromossome*)- chr11 (Cromossomo 11).

#### ***Estudo de associação genômica ampla para tamanho do fruto***

A característica tamanho de fruto (Tfr) apresentou um SNP com associação significativa ( $p < 0,05$ ), localizado no cromossomo 1 (Tabela 9). O efeito associado ao aumento de Tfr foi de 16,93%. Frutos maiores de café proporcionam maiores peneiras, e aliados a outros fatores de boa qualidade, podem resultar em maior valor agregado (Laviola et al., 2006). Além disso, o tamanho dos grãos é um dos fatores que determina o potencial produtivo da cultivares, e a sua homogeneidade garante uma torração mais uniforme (Laviola et al., 2006; Soares et al., 2019).

**Tabela 9.** SNP com associações significativas ( $p < 0,05$ ), posição, alelos favoráveis e efeitos associados ao aumento do tamanho do fruto (Tfr).

SNP	Chr	Posição	$p$ -value	$R^2$	Alelo	Alelo Favorável	Efeito(%)
V4722	chr1_c	49860517	1,21E+08	0.10983	G/A	G	16,92648



**Figura 10.** Gráfico de Manhattan para tamanho de fruto. O SNP acima da linha tracejada horizontal mostram a significância com um limiar moderadamente restrito de  $-\log_{10}(0,01/19.623)$ . Gráficos acima da linha contínua horizontal mostram a significância com um limiar de  $-\log_{10}(0,05/19.623)$ . As diferentes cores indicam gráficos para diferentes cromossomos, que seguem a ordem: Unchr (*Un caracterizad chromossome*)- chr11 (Cromossomo 11).

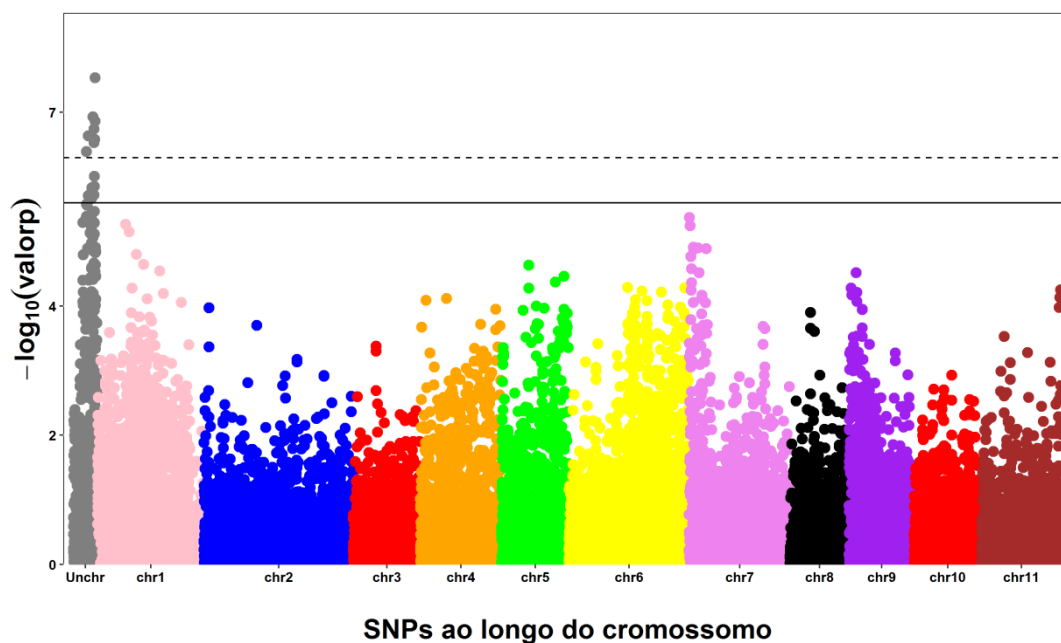
***Estudo de associação genômica ampla para diâmetro da copa***

Diâmetro da copa (Dco) apresentou 14 SNP com associações significativas ( $p < 0,05$ ) (Tabela 10). No entanto todos os SNP com associações significativas estão localizados no *Unchr* (Figura 11).

**Tabela 10.** SNP com associações significativas ( $p < 0,05$ ), posição, alelos favoráveis e efeitos associados ao aumento do diâmetro da copa (Dco).

SNP	Chr	Posição	$p$ -value	$R^2$	Alelo	Alelo Favorável	Efeito
V16933	Unchr*	1139782	2,94E-08	0,15507	G/T	G/T	7,57154
V16828	Unchr*	884969	1,20E-07	0,14606	C/G	C	19,94760
V16934	Unchr*	1139783	1,39E-07	0,14262	A/G	A/G	6,89321
V16813	Unchr*	991590	1,85E-07	0,14289	A/G	A/G	6,04344
V1310	Unchr*	453100	2,36E-07	0,14320	G/A	G/A	5,86421
V16796	Unchr*	1056119	2,67E-07	0,12349	C/T	C/T	7,66902
V16814	Unchr*	991599	2,97E-07	0,13604	T/C	T/C	6,27291
V16900	Unchr*	218187	4,12E-07	0,13074	C/T	C	15,49531
V16804	Unchr*	1045984	9,91E-07	0,13814	T/C	T	10,26386
V16805	Unchr*	1045989	1,42E-06	0,13580	A/T	A	8,96176
V16847	Unchr*	806813	1,49E-06	0,13007	C/G	C	16,94895
V16798	Unchr*	1045911	1,94E-06	0,12965	C/T	C/T	5,37956
V16881	Unchr*	497973	1,99E-06	0,12465	T/C	T/C	5,41446
V16855	Unchr*	788453	2,18E-06	0,12358	G/T	G/T	5,50469

\*SNP localizados no Unchr.



**Figura 11.** Gráfico de Manhattan para a característica diâmetro da copa. Os SNP acima da linha tracejada horizontal mostram a significância com um limiar moderadamente restrito de  $-\log_{10}(0,01/19.623)$ . Gráficos acima da linha contínua horizontal mostram a significância com um limiar de  $-\log_{10}(0,05/19.623)$ . As diferentes cores indicam gráficos para diferentes cromossomos, que seguem a ordem: Unchr (*Un characterized chromosome*)- chr11 (Cromossomo 11).

## **Conclusões**

Pela primeira vez, o GWAS foi aplicado para as principais características agronômicas utilizadas nos programas de melhoramento genético na espécie *C. arabica*. Os resultados foram promissores e possibilitaram identificar SNP com associações significativas para sete características agronômicas de importância para o cafeeiro. SNP inseridos em genes foram identificados, sendo essa informação valiosa para a elucidação de processos biológicos importantes. Assim, esses marcadores identificados apresentam potencial para serem utilizados em seleções assistidas por marcadores moleculares, garantindo maior eficiência e redução de tempo, auxiliando os programas de melhoramento da espécie. Além disso, os benefícios do GWAS são maximizados pelo fato da espécie em estudo ser perene, e, portanto, de custo elevado de manutenção de populações em campo.

## Referências Bibliográficas

- Akter A, Islam MM, Mondal SI, et al (2014). Computational identification of miRNA and targets from expressed sequence tags of coffee (*Coffea arabica*). Saudi journal of biological sciences, 21, 3-12. doi: 10.1016/j.sjbs.2013.04.007
- Alvarenga SM, Caixeta ET, Hufnagel B, et al (2011). Marcadores moleculares derivados de sequências expressas do genoma café potencialmente envolvidas na resistência à ferrugem. Pesquisa Agropecuária Brasileira, 8, 890-898. doi: 10.1590/S0100-204X2011000800015.
- Andersen JR, Lubberstedt T, (2003). Functional markers in plants. Trends in plant science, 11, 554-560. doi:10.1016/j.tplants.2003.09.010.
- Anthony F, Combes MC, Astorga C, et al (2002). The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theoretical and Applied Genetics, 104, 894-900. doi:10.1007/s00122-001-0798-8.
- Azizi P, Rafii MY, Maziah M, et al (2015). Understanding the shoot apical meristem regulation: A study of the phytohormones, auxin and cytokinin, in rice. Mechanisms of development, 135, 1-15. doi:10.1016/j.mod.2014.11.001.
- Baroni DF, Reis LSD, Baitelle DC, et al (2017). Ácido naftalenoacético associado à adubação potássica na brotação de café Conilon. Coffee Science, 12, 282 – 289. doi:10.25186/cs.v12i2.
- Bhering LL (2017). Rbio: A tool for biometric and statistical analysis using the R platform. Crop Breeding and Applied Biotechnology 17,187–190. doi: 10.1590/1984-70332017v17n2s29.
- Carvalho AM, Carvalho D, de Carvalho VL, et al (2017). Behavior of coffee cultivars under the incidence of diseases of rust and gray leaf spot in two cultivation environments. Coffee Science, 12, 100-107. doi:10.25186/cs.v12i1.1248.
- Carvalho AM, Mendes AN, Carvalho GR, Botelho CE, et al (2011). Correlação entre crescimento e produtividade de cultivares de café em diferentes regiões de Minas Gerais, Brasil. Pesquisa agropecuária brasileira, 45, 269-275
- Clarindo WR, Carvalho CR, (2008). First *Coffea arabica* karyogram showing that this species is a true allotetraploid. Plant Systematics and Evolution 274, 237–241. doi:10.1007/s00606-008-0050-y.
- Combes MC, Andrzejewski S, Anthony F, et al (2000). Characterization of microsatellite loci in *Coffea arabica* and related coffee species. Molecular ecology, 9, 1178–1180. doi: 10.1046/j.1365-294x.2000.00954-5.x.
- Denoeud F, Carretero-Paulet L, Dereeper A, et al (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 80, 345,1181–1184. doi: 10.1126/science.1255274.

Diniz LEC, Sakiyama NS, Lashermes P, et al (2005). Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. *Cropps Breeding and Applied Biotechnology* 5, 387–393. doi: 10.12702/1984-7033.v05n04a03.

Falconer DS, Mackay TFC (1996). *Introduction to quantitative genetics*, 4ed. Longmans Green, Harlow, Essex, UK, 464p.

Ferrão LFV, Caixeta ET, Pena G, et al (2015). New EST–SSR markers of *Coffea arabica*: transferability and application to studies of molecular characterization and genetic mapping. *Molecular breeding* 35, 31. doi:10.1007/s11032-015-0247-z.

Gibson G, (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics*, 13, 135-145. doi:135. 10.1038/nrg3118.

He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z, (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in plant science*, 5, 484. doi 10.3389/fpls.2014.00484.

Hongyu K, Sandanielo LM, Oliveira Junior GJ, (2016). Análise de componentes principais: resumo teórico, aplicação e interpretação. *E&S Engineering and Science*, 5, 83-90. doi: 10.18607/ES20164053.

Huang X, Han B, (2014). Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology*, 65, 531-551. doi: 10.1146/annurev-arplant-050213-035715.

Korte A, Farlow A, (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, 9, 29. doi:10.1186 / 1746-4811-9-29.

Lashermes P, Combes MC, Robert J, et al (1999). Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics MGG* 261, 259–266. doi:10.1007/s004380050965.

Laviola BG, Mauri AL, Martinez HEP, et al (2006). Influência da adubação na formação de grãos mocas e no tamanho de grãos de café (*Coffea arabica* L.). *Coffee Science*, 1, 36-42. doi: <http://dx.doi.org/10.25186/cs.v1i1>.

Mahé L, Combes MC, Várzea VM, et al (2008). Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.). *Molecular breeding*, 21, 105-113. doi: doi.org/10.1007/s11032-007-9112-z.

Miller RG, (1981) *Normal Univariate Techniques*. In: *Simultaneous Statistical Inference*. Springer Series in Statistics. Springer, New York, NY. 60-67. doi: doi.org/10.1007/978-1-4613-8122-8\_2.

Ministério da Agricultura, Pecuária e Abastecimento (MAPA), *Café no Brasil*, Disponível em: <<http://www.agricultura.gov.br/assuntos/politica-agricola/cafe/cafeicultura-brasileira>> Acesso em: 06 de Dez, de 2018.

Mishra MK, Slater A, (2012). Recent advances in the genetic transformation of coffee. *Biotechnology research international*, 1-17 doi: 10.1155/2012/580857.

National Center for Biotechnology Information  
<<https://www.ncbi.nlm.nih.gov/genome/?term=Coffea+arabica+genome>> Acesso em: 25 de Abril, de 2019.

Observatory of Economic Complexity (Oec), Coffee trade – Exporters, Disponível em: <<https://atlas.media.mit.edu/en/profile/hs92/0901/>> Acesso em: 06 de Dez, de 2018.

Oliveira ACBD, Pereira AA, Silva FLD, Rezende, J. C. D et al (2011). Prediction of genetic gains from selection in Arabica coffee progenies. *Crop Breeding and Applied Biotechnology* 11, 106-113. doi: 10.1590/S1984-70332011000200002.

Orvar BL, Sangwan V, Omann F, et al (2000). Early steps in cold sensing by plant cells: The role of actin cytoskeleton and membrane fluidity. *The Plant Journal* 23, 785–794. doi:10.1046/j.1365-3113x.2000.00845.x.

Paiva BRTL, De Souza PE, Scalco MS, et al (2013). Progresso da cercosporiose do cafeeiro sob diferentes manejos de irrigação e densidades de plantio. *Coffee Science*, 8, 166-175. doi: 10.25186/cs.v8i2.391.

Peremarti A, Twyman RM, Gómez-Galera S, et al (2010). Promoter diversity in multigene transformation. *Plant molecular biology*, 73, 363-378. doi:10.1007/s11103-010-9628-1.

Pers TH, Karjalainen JM, Chan Y, et al (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications*, 6, 5890. doi: 10.1038/ncomms6890.

Python Core Team (2018). Python: A dynamic, open source programming language. Python Software Foundation. URL <https://www.python.org/>.

R Core Team (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria. URL <http://www.R-project.org/>.

Resende M, Caixeta ET, Alkimim ER, et al (2016a). High-Throughput Targeted Genotyping of *Coffea arabica* and *Coffea canephora* Using Next Generation Sequencing. San Diego, CA, p 1.

Resende MDV de (2016b). Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breeding and Applied Biotechnology* 16, 330–339. doi: 10.1590/1984-70332016v16n4a49.

Resende RT, Resende MDV, Silva FF, et al (2017). Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. *New Phytologist*, 213, 1287-1300. doi:10.1111/nph.14266.

Roy S, (2016). Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. *Plant signaling & behavior*, 11, e1117723. doi:10.1080/15592324.2015.1117723.

- Sant'Ana GC, Pereira LFP, Pot D, et al (2018). Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica*. Scientific reports, 8, 465. doi: 10.1038/s41598-017-18800-1.
- Setotaw TA, Caixeta ET, Pereira AA, et al (2013). Coefficient of Parentage in *Coffea arabica* L. Cultivars Grown in Brazil. Crop Science 53, 1237–1247. doi: 10.2135/cropsci2012.09.054.
- Soares WL, Silva TM, Junior KSF et al (2019). Qualidade do Café Arábica por Diferentes Granulometrias. Revista Ciência Agrícola, 17, 31-35. doi:10.28998/rca.v17i1.
- Sousa TV, Caixeta ET, Alkimim ER et al (2017a). Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. Euphytica, 213,75. doi: 10.1007/s10681-017-1865-9.
- Sousa TV, Caixeta ET, Alkimim ER, et al (2017b). Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker. Tree Genetics & Genomes, 13, 124. doi:10.1007/s11295-017-1208-y.
- Sousa TV, Caixeta ET, Alkimim ER, Oliveira ACB, et al (2019). Early Selection Enabled by the Implementation of Genomic Selection in *Coffea arabica* Breeding. Frontiers in plant science, 9. doi: 10.3389/fpls.2018.01934.
- Su C, Chen K, Ding Q, et al (2018). Proteomic Analysis of the Function of a Novel Cold-Regulated Multispanning Transmembrane Protein COR413-PM1 in Arabidopsis. International journal of molecular sciences, 19, 2572. doi: 10.3390 / ijms19092572
- Su J, Zhang F, Chong X, Song A, et al (2019). Genome-wide association study identifies favorable SNP alleles and candidate genes for waterlogging tolerance in chrysanthemums. Horticulture research, 6, 21. doi: 10.1038/s41438-018-0101-7
- Teixeira AL, Gonçalves FMA, Rezende JCD, et al (2012). Seleção precoce para produção de grãos em café arábica pela avaliação de caracteres morfológicos. Pesquisa Agropecuária Brasileira 47, 1110 – 1117. doi:10.1590/S0100-204X2012000800011.
- Turner S, Armstrong LL, Bradford Y, et al (2011). Quality control procedures for genome-wide association studies. Current protocols in human genetics, 68, 1-19. doi: 10.1002 / 0471142905.hg0119s68.
- Vidal RO, Mondego JMC, Pot D, et al (2010). A High-Throughput Data Mining of Single Nucleotide Polymorphisms in *Coffea* Species Expressed Sequence Tags Suggests Differential Homeologous Gene Expression in the Allotetraploid *Coffea arabica*. Plant Physiology, 154,1053–1066. doi: 10.1104/pp.110.16243.
- Vieira LGE, Andrade AC, Colombo CA et al (2006). Brazilian coffee genome project: an EST-based genomic resource. Brazilian Journal of Plant Physiology, 18, 95-108. doi: 10.1590/S1677-04202006000100008.
- Zambolim L, (2016) Current status and management of coffee leaf rust in Brazil. Tropical Plant Pathology, 41,1-8. doi: 10.1007 / s40858-016-0065-9.