

LETÍCIA DE FARIA SILVA

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

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*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2018

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

T

S586e  
2018  
Silva, Leticia de Faria, 1992-  
Estudo de associação genômica ampla (GWAS) em *Coffea  
canephora* : m / Leticia de Faria Silva. – Viçosa, MG, 2018.  
viii, 32 f. : il. ; 29 cm.

Orientador: Ney Sussumu Sakiyama.  
Dissertação (mestrado) - Universidade Federal de Viçosa.  
Referências bibliográficas: f. 30-32.

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

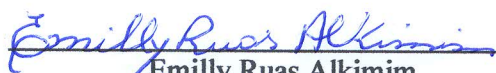
CDD 22. ed. 576.53

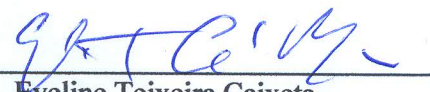
LETÍCIA DE FARIA SILVA

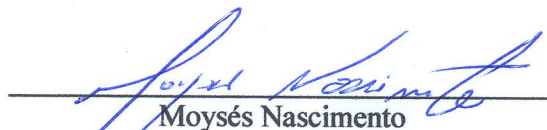
**ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)  
EM *Coffea canephora***


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: 02 de agosto de 2018.

  
Emilly Ruas Alkimim  
(Coorientadora)

  
Eveline Teixeira Caixeta  
(Coorientadora)

  
Moyes Nascimento

  
Ney Sussumu Sakiyama  
(Orientador)

*“A diferença entre ganhar e perder é muitas vezes não desistir”*

Walt Disney

À minha mãe Esméria,  
E a toda minha família,  
Dedico.

## AGRADECIMENTOS

A Deus pela sabedoria, calma e sobretudo sua benção concedida a mim nessa trajetória  
À minha mãe Esméria, por ser minha maior incentivadora e expectadora, acreditando mais em mim do que eu mesma.

Ao meu pai Naelson, pelo apoio e conselhos

À minha tia Enilda por ser meu maior exemplo de força e perseverança. As minhas avós Rita, Luzia e minha madrinha Lídia por todas as orações e palavras de conforto. Aos meus tios e primos por toda a torcida.

Ao meu orientador Dr. Ney Sussumu Sakiyama pela orientação e ensinamentos transmitidos.

À Dra. Eveline Teixeira Caixeta, por ser um exemplo de superação e positividade. Por todos esses anos de orientação não somente profissional, mas também para a vida, pela paciência, oportunidade e ensinamentos transmitidos.

Ao professor, Moysés Nascimento por aceitar participar da minha banca e por em momento algum poupar esforços para me ajudar e tranquilizar. Sem ele a realização desse trabalho não seria possível,

À Dra. Emilly Alkimim, que além de minha coorientadora é minha grande amiga me auxiliando e me apoiando em todos os momentos dessa caminhada.

Aos meus amigos do Biocafé, por toda amizade e companheirismo. Em especial ao Tiago e a Francielle por sempre me ajudarem

Aos meus amigos de infância Gabriel e Yara os irmãos que nunca tive por todos esses anos de companheirismo e amizade.

Aos amigos que eu conquistei durante o mestrado: Julia, Isa, Lívia, Fumes, Laura, Pedro Ricardo, Juan, Geleta.

## SUMÁRIO

<b>RESUMO</b> .....	vi
<b>ABSTRACT</b> .....	viii
<b>Introdução</b> .....	1
<b>Material e Métodos</b> .....	3
<i>Extração do DNA genômico</i> .....	3
<i>Análises dos marcadores SNP</i> .....	3
<i>Avaliações fenotípicas</i> .....	4
<i>Análises dos dados fenotípicos</i> .....	4
<i>Estudo de associação genômica ampla</i> .....	7
<i>Identificação de genes e anotação funcional</i> .....	8
<b>Resultados e discussão</b> .....	8
<b>Conclusão</b> .....	12

## RESUMO

SILVA, Letícia de Faria, M.Sc., Universidade Federal de Viçosa, agosto de 2018. **Estudo de associação genômica ampla (GWAS) em *Coffea canephora***. Orientador: Ney Sussumu Sakiyama. Coorientadoras: Eveline Teixeira Caixeta e Emily Ruas Alkimim.

Estudos de associação genômica ampla (GWAS) tem sido tradicionalmente utilizados para a identificação de locos de interesse responsáveis pela variação fenotípica. A utilização desse tipo de metodologia explora as ligações mais próximas existentes entre os marcadores moleculares e regiões cromossômicas de interesse para um maior entendimento das características fenotípicas da espécie em estudo, propiciando a identificação de marcas a serem utilizadas em seleções assistidas por marcadores moleculares (SAM) nos programas de melhoramento. Nesse contexto, a utilização dessa metodologia visando um maior entendimento de caracteres morfoagronômicos em um menor período de tempo se mostra uma solução viável para programas de melhoramento de espécies com longo ciclo de vida (plantas perenes) como o café. Assim, o objetivo do trabalho foi identificar regiões cromossômicas com associações significativas em populações de *C. canephora* em melhoramento para as principais características fenotípicas de importância para essa espécie, através da metodologia de Associação Genômica Ampla (GWAS). A população em estudo foi composta por 165 genótipos contendo os dois grupos varietais Conilon e Robusta além dos Híbridos provenientes desse cruzamento. Foram avaliadas oito características morfoagronômicas sendo elas altura da planta, vigor vegetativo, incidência a cercosporiose, incidência a ferrugem, diâmetro da copa, produtividade, época de maturação dos frutos e tamanho dos frutos. Os cafeeiros foram genotipados com marcadores SNP, distribuídos em todo o genoma. Foram realizadas análises de qualidade ( $MAF \geq 0,01$  e  $CR \geq 90\%$ ) que resultaram em 17.885 SNP. Após a aplicação da metodologia de GWAS, foram identificados marcadores SNP com associações significativas para cinco características (altura da planta, diâmetro da copa, vigor vegetativo, incidência a ferrugem e a cercosporiose). As análises também permitiram a identificação de genes candidatos, nos quais, os marcadores SNP estavam inseridos. Foi possível também relacionar a função dos genes candidatos a característica nos quais esses se encontravam inseridos. Foi possível a identificação da presença dos genes 08-g11560, 10-g16500, 11-g17430, 00-g17460 e 02-g38180 nas duas principais doenças do cafeeiro. Os marcadores SNP com associações



significativas se mostraram distribuídos em todo o genoma da espécie e estavam presentes em todos os cromossomos. Com destaque para os cromossomos 0,2,6,9 e 11. Por fim, a metodologia se mostrou eficiente em populações *C.canephora*, os genes candidatos encontrados inseridos nos marcadores SNP com associações significativas poderão ser avaliados com maior detalhamento e posteriormente poderão ser usados em seleções assistidas por marcadores moleculares (SAM) auxiliando programas de melhoramento da espécie.

## ABSTRACT

SILVA, Leticia de Faria, M.Sc., Universidade Federal de Viçosa, August, 2018. **Genome-wide association Study (GWAS) in *Coffea canephora***. Adviser: Ney Sussumu Sakiyama. Co-advisers: Eveline Teixeira Caixeta and Emilly Ruas Alkimim

Genome-wide Association Study (GWAS) have traditionally been used to identify loci of interest responsible for the phenotypic variation. The use of this type of methodology explores the closer links between molecular markers and chromosomal regions of interest for a better understanding of the phenotypic characteristics of the species under study, favoring the identification of markers to be used in selections assisted by molecular markers (SAM) in breeding programs. In this context, the use of this methodology aiming at a greater understanding of morphoagronomic characters in a shorter period of time is a viable solution for breeding programs of species with a long life cycle (perennial plants) such as coffee. Thus, the objective of this work was to identify chromosomal regions with significant associations in *C. canephora* populations in breeding for the main phenotypic characteristics of importance for this species, through the GWAS methodology. The study population consisted of 165 genotypes containing the two Conilon and Robusta varietal groups in addition to the hybrids from that cross. Eight morphoagronomic characteristics were evaluated: plant height, vegetative vigor, incidence of cercosporiae, incidence of rust, canopy diameter, productivity, fruit maturation time and fruit size. The coffee trees were genotyped with SNP markers, distributed throughout the genome. Quality analyzes (MAF  $\geq 0.01$  and CR  $\geq 90\%$ ) were performed, resulting in 17,885 SNPs. After application of the GWAS methodology, SNP markers were identified with significant associations for five characteristics (plant height, crown diameter, vegetative vigor, incidence of rust and cercoporiosis). The analyzes also allowed the identification of candidate genes in which the SNP markers were inserted. It was also possible to relate the function of the candidate genes to the characteristic in which they were inserted. It was possible to identify the presence of genes 08-g11560, 10-g16500, 11-g17430, 00-g17460 and 02-g38180 in the two major diseases of coffee. The SNP markers with significant associations were shown to be distributed throughout the genome of the species and were present on all chromosomes. Finally, the methodology was efficient in populations *C. canephora*, the candidate genes found inserted in the SNP markers with significant associations can be evaluated in greater detail and later can be used in molecular marker assisted selections (SAM) to support species breeding programs.

## Introdução

Os programas de melhoramento genético buscam, estratégias de seleção cada vez mais eficientes para a identificação de genótipos que apresentem desempenho superior. Os programas de melhoramento que visam melhorar plantas com longo ciclo de vida (plantas perenes) apresentam dificuldades adicionais, como períodos mais longos para a obtenção de flores para a realização dos cruzamentos, custo elevado nas avaliações de campo devido a necessidade de áreas e períodos de tempos maiores, necessidade da avaliação da precocidade e da longevidade produtiva (Ferrão, 2004; Ferrão et al., 2007). Entre as culturas perenes, a cultura do café se destaca principalmente por sua grande importância econômica para o desenvolvimento mundial (Resende et al., 2001).

As duas espécies de cafeeiro mais cultivadas no mundo são *Coffea arabica* e *Coffea canephora* (Fazuoli, 1986). O café arábica é responsável pelo aroma, sabor adocicado e, conseqüentemente, proporciona qualidade superior de bebida. O café canephora se destaca por apresentar maior quantidade de cafeína e sólidos solúveis, proporcionando corpo a bebida (Babova et al., 2016). Durante muitos anos, *C. canephora* foi considerada uma espécie inferior para a utilização em bebidas. No entanto, seu uso passou a ser explorado comercialmente nos chamados *blends* com a espécie *C. arabica*. Esse foi um dos fatores responsáveis pelo impulso mundial da produção de *C. canephora*. Além da importância na indústria, essa espécie destaca-se pela alta produtividade de grãos, menor suscetibilidade a estresses abióticos e bióticos (Ferrão et al., 2007).

A espécie *C. canephora* é dividida em dois grupos varietais, denominados Conilon e Robusta. O grupo varietal Conilon compreende os genótipos de melhor adaptação no Brasil, apresenta folhas menores, menor vigor e porte, frutos pequenos, tolerância à seca e são, em geral, suscetíveis à ferrugem. Os cafeeiros do grupo varietal Robusta são representados por plantas de porte alto, vigorosas, de folhas e frutos maiores, resistentes à ferrugem e com maior sensibilidade à seca (Musoli, 2009).

Em razão da importância econômica de *C. canephora*, programas de melhoramento dessa espécie vem sendo desenvolvidos, visando, principalmente, o aumento da produtividade, resistência a doenças e o desenvolvimento sócio-econômico das regiões de cultivo (Maia et al., 2017). Esses programas de melhoramento têm usado avaliações de características morfoagronômicas para discriminação e seleção de genótipos. No entanto, essas características fenotípicas apresentam limitações como grande influência ambiental e baixo ganho com a

seleção de características de interesse (Ferrão et al., 2007; Souza, 2011). Assim, com intuito de minimizar essas limitações, tem se destacado a implementação de marcadores moleculares, os quais detectam polimorfismos a nível de DNA (Sousa et al., 2017; Alkimim et al., 2017).

Dentre os diversos marcadores moleculares, os marcadores SNP representam o tipo de polimorfismo mais abundante no genoma, são, em geral, bialélicos e codominantes. Com o desenvolvimento das tecnologias de NGS (*Next Generation Sequencing*), estes marcadores são facilmente automatizados, o que tem proporcionado uma economia significativa de tempo e recursos para seu uso nos programas de melhoramento (Resende et al., 2008; Liao e Lee, 2010). Dessa forma, vários estudos utilizando marcadores SNP têm sido conduzidos com sucesso. Dentre estes, podem-se citar estudos de diversidade e estrutura genética de populações, estudos de seleção genômica (GWS- *Genome Wide Selection Study*) e estudos de Associação Genômica Ampla (GWAS – *Genome Wide Association Study*). (Alkimim et al., 2017; Sousa et al., 2017; Meuwissen et al., 2016).

Os estudos de GWAS permitem associar o fenótipo de interesse com o genótipo, permitindo identificar possíveis regiões do genoma que apresentam maior efeito sobre a característica. Por essa estratégia é possível, também, achar os genes contendo SNP significativos, e posteriormente identificar genes envolvidos no controle da característica e sua respectiva função biológica. A caracterização desses genes permite aumentar a compreensão da influência do genótipo sobre o fenótipo (Yang et al., 2013).

A abordagem GWAS foi pioneira em estudos relacionados a doenças humanas, posteriormente os estudos GWAS foram adaptados para uma série de organismos modelos como *Arabidopsis thaliana* e para culturas de importância comercial (Huang et al., 2012).

Apesar dos estudos de GWAS apresentarem vantagens e amplas aplicações nos programas de melhoramento, até o momento, nenhum estudo com aplicações da GWAS para a espécie *C. canephora* foi publicado. Estudos realizados por Alkimim (2017) permitiram a identificação e validação de marcadores SNP para *C. canephora*, foram utilizadas 10.000 sondas distribuídas ao longo de todo o genoma, a partir de sequências de referência provenientes do Projeto Brasileiro do Genoma Café e do banco de dados do genoma da espécie *C. canephora*. De posse dessas sondas, os genótipos foram sequenciados via plataforma *Illumina* e os marcadores identificados por abordagem de bioinformática. Sendo obtidos um total de 18.111 marcadores SNP.

Diante do exposto, objetivou-se a identificação de regiões cromossômicas com associações significativas em populações de *C. canephora* em melhoramento para as principais

características fenotípicas de importância para essa espécie, através da metodologia de Associação Genômica Ampla (GWAS).

## **Material e Métodos**

A população em estudo, alocada no município de Oratórios, Minas Gerais, faz parte do programa de melhoramento da Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig) em parceria com a Universidade Federal de Viçosa (UFV) e a Empresa Brasileira de Pesquisa Agropecuária - Café (Embrapa Café). O plantio dos cafeeiros foi realizado em espaçamento de 3,0 m entre as fileiras e 1,5 m entre as plantas, em delineamento em blocos casualizados (DBC).

A população é composta por clones dos grupos varietais Conilon e Robusta e por Híbridos interpopulacionais (Conilon x Robusta). O grupo varietal Conilon (Tabela 1) é constituído por 51 genótipos, obtidos do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER). O grupo varietal Robusta (Tabela 1) é constituído de 32 genótipos obtidos do Centro Agronômico Tropical de Investigación y Enseñanza (CATIE). Além desses, 82 Híbridos interpopulacionais foram obtidos por meio de cruzamentos artificiais obtidos a partir de um dialelo parcial formado por cinco genitores masculinos Conilon e cinco genitores femininos Robusta (Tabela 2).

### ***Extração do DNA genômico***

Seguindo a metodologia proposta por Diniz et al (2005), folhas jovens dos 165 cafeeiros tiveram seu DNA genômico extraídos no Laboratório de Biotecnologia do Cafeeiro – Biocafé/UFV. Após a extração, a qualidade do DNA foi verificada em gel de agarose 1% e sua concentração verificada em *NanoDrop 2000*. As concentrações das amostras foram padronizadas e, posteriormente, enviadas para a empresa *RAPiD Genomics* localizada na Florida/EUA, para a construção de sondas, sequenciamento e identificação de marcadores moleculares SNP.

### ***Análises dos marcadores SNP***

Os marcadores SNP foram analisados quanto à qualidade, eliminando-se os alelo menos frequente (MAF) e os marcadores que apresentam grandes quantidades de dados perdidos (CR-*call rate*). Os valores usados para eliminação foram MAF inferior a 0,01 e CR inferiores a 90%.

### *Avaliações fenotípicas*

Avaliações fenotípicas de oito características foram conduzidas durante três anos consecutivos (2014 a 2016). A característica vigor vegetativo (Vig) foi avaliada levando-se em consideração o aspecto geral da planta, examinando se o enfolhamento, coloração das folhas, estado nutricional e vitalidade (Tabela 3). As características incidência a ferrugem (Fer) e a cercosporiose (Cer) foram avaliadas com notas de 1 a 5 de acordo com os sintomas apresentados em decorrência do patógeno. A época de maturação dos frutos (Mat) do cafeeiro foi classificada em precoce, média e tardia, recebendo notas de acordo com a aparência dos frutos, e o tamanho dos frutos (Tf) foi classificado como pequeno, médio e graúdo (Tabela 3).

Para as características altura da planta (Ap) e diâmetro da copa (Dc) foram realizadas aferições em centímetros (cm), estipuladas pela medição da ramificação ortotrópica mais desenvolvida, iniciadas do nível do solo até o último ponto apical do cafeeiro. A característica produtividade (Prod) por planta foi mensurada colhendo-se todos os frutos presentes por genótipo e mensurando-se o volume em litros com o auxílio de uma proveta (Tabela 3).

### *Análises dos dados fenotípicos*

Os dados fenotípicos foram corrigidos para efeitos ambientais de anos e blocos, utilizando a metodologia de modelos mistos, por meio do software Selegen REML (Método da Máxima Verossimilhança Restrita)/ BLUP (Melhor predição linear imparcial) (Resende 2016). Os dados foram analisados de acordo com o seguinte modelo:

$$y = Xu + Tc + Wf + Zm + Qs + Sb + e$$

Em que:

*y*: Vetor das observações fenotípicas;

*u*: Vetor dos efeitos das médias (considerados fixos) somados a média geral;

*c*: Vetor de efeito da capacidade específica de combinação entre os genitores Conilon e Robusta (considerados aleatórios);

*f*: Vetor dos efeitos da capacidade geral de combinação do parental Robusta (considerado aleatório);

*m*: Vetor dos efeitos da capacidade geral de combinação do parental Conilon (considerado aleatório);

*s*: Vetor de efeito ambiental permanente de indivíduos (considerado aleatório);

$b$ : Vetor de efeito permanente dos blocos (considerado aleatório);  
 $e$ : Vetor de resíduos (assumidos como aleatório).

Após o ajuste do modelo, os valores fenotípicos ajustados ( $Y_{ijk}^*$ ) foram obtidos como a soma dos efeitos aleatórios de genótipos e do erro experimental.

As estruturas de médias e variâncias associadas ao modelo são:

$$y|r, V \sim N(Xu, V); \quad m|\sigma_m^2 \sim N(0, I\sigma_m^2); \quad f|\sigma_f^2 \sim N(0, I\sigma_f^2); \quad c|\sigma_c^2 \sim N(0, I\sigma_c^2);$$

$$s|\sigma_s^2 \sim N(0, I\sigma_s^2); \quad b|\sigma_b^2 \sim N(0, I\sigma_b^2); \quad e|e \sim N(0, I\sigma_e^2).$$

As covariâncias entre os efeitos são iguais à zero.

Em que:

$$V = \text{Variância fenotípica e equivale a: } ZZ'\sigma_m^2 + WW'\sigma_f^2 + TT'\sigma_c^2 + QQ'\sigma_s^2 + SS'\sigma_b^2 + I\sigma_e^2;$$

$I$  = Matriz Identidade;

$\sigma_m^2, \sigma_f^2, \sigma_c^2, \sigma_s^2, \sigma_b^2$ , e  $\sigma_e^2$  são as variâncias entre os genitores do grupo Conilon, entre os genitores do grupo Robusta, da capacidade específica de combinação entre os genitores dos grupos Conilon e Robusta, de ambiente permanente de indivíduo, de ambiente permanente de blocos, e residual, respectivamente.

As equações de modelos mistos para o BLUP (*Best Linear Unbiased Predictor*) equivalem a:

$$\begin{bmatrix} \hat{u} \\ \hat{m} \\ \hat{f} \\ \hat{c} \\ \hat{s} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W & X'T & X'Q & X'S \\ Z'X & Z'Z + I\lambda_1 & Z'W & Z'T & Z'Q & Z'S \\ W'X & W'Z & W'W + I\lambda_2 & W'T & W'Q & W'S \\ T'X & T'Z & T'W & T'T + I\lambda_3 & T'Q & T'S \\ Q'X & Q'Z & Q'W & Q'T & Q'Q + I\lambda_4 & Q'S \\ S'X & S'Z & S'W & S'T & S'Q & S'S + I\lambda_5 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \\ Q'y \\ S'y \end{bmatrix}$$

Em que:

$$\lambda_1 = \sigma_e^2 / \sigma_m^2 = \frac{1 - m^2 - f^2 - c^2 - s^2 - b^2}{m^2};$$

$$\lambda_2 = \sigma_e^2 / \sigma_f^2 = \frac{1-m^2-f^2-c^2-s^2-b^2}{f^2};$$

$$\lambda_3 = \sigma_e^2 / \sigma_c^2 = \frac{1-m^2-f^2-c^2-s^2-b^2}{c^2};$$

$$\lambda_4 = \sigma_e^2 / \sigma_s^2 = \frac{1-m^2-f^2-c^2-s^2-b^2}{s^2};$$

$$\lambda_5 = \sigma_e^2 / \sigma_b^2 = \frac{1-m^2-f^2-c^2-s^2-b^2}{b^2};$$

$m^2 = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$ : coeficiente de determinação da capacidade geral de combinação do parental Conilon;

$f^2 = \frac{\sigma_f^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$ : coeficiente de determinação da capacidade geral de combinação do parental robusta;

$c^2 = \frac{\sigma_g^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$ : coeficiente de determinação da capacidade específica de combinação entre os genitores Conilon e robusta;

$s^2 = \frac{\sigma_p^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$ : coeficiente de determinação de ambiente permanente de indivíduos;

$b^2 = \frac{\sigma_b^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$ : coeficiente de determinação de ambiente permanente de blocos.

Os componentes de variância foram estimados por REML (Patterson and Thompson, 1971), a partir de iterações nas equações de modelos mistos (Resende, 2007; Resende and Duarte, 2007), cujos estimadores são:

$$\sigma_m^2 = [\hat{m}'\hat{m} + \sigma_e^2 \text{tr}(C^{22})]/q;$$

$$\sigma_f^2 = [\hat{f}'\hat{f} + \sigma_e^2 \text{tr}(C^{33})]/n;$$

$$\sigma_c^2 = [\hat{c}'\hat{c} + \sigma_e^2 \text{tr}(C^{44})]/m;$$

$$\sigma_s^2 = [\hat{s}'\hat{s} + \sigma_e^2 \text{tr}(C^{55})]/N;$$

$$\sigma_b^2 = [\hat{p}'\hat{p} + \sigma_e^2 \text{tr}(C^{66})]/l;$$



$$\sigma_e^2 = [y'y - \hat{u}'X'y - \hat{m}'Z'y - \hat{f}'W'y - \hat{c}'T'y - \hat{s}'Q'y - \hat{b}'S'y]/[N - r(X)].$$

Em que:

C: Matriz inversa generalizada da matriz dos coeficientes das equações de modelos mistos;

tr: operador traço matricial;

r(X): posto da matriz X (número de colunas linearmente independentes);

N-r(X): número de graus de liberdade do erro;

N, q, n, m e l: número total de dados, de genitores do grupo Conilon, genitores do grupo Robusta, número de combinações entre os grupos Conilon e Robusta; e de blocos, respectivamente.

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

Previamente a realização das análises de associação genômica ampla (GWAS), a estrutura genética das populações utilizadas nesse estudo foram avaliadas por análise de componentes principais (PCA), utilizando-se a função *princomp* do *software* R. O PCA foi aplicado aos dados genotípicos para obter os componentes principais (CP) que capturam a maior parte da variação genética considerando um percentual mínimo de explicação da variação total igual à 80%. Após esse procedimento, o número de CP que atendiam ao critério anterior foram selecionados e utilizados como covariáveis no modelo de GWAS para detectar SNPs associados as características fenotípicas. O modelo geral do GWAS é dado por:

$$Y_{ijk}^* = \mu + SNP_i + \sum_{k=1}^5 PC_{jk} + \varepsilon_{ijk}$$

Em que:

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

$\mu$ : Equivale a média da população;

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

$PC_{kj}$ : Equivale ao efeito fixo do  $k^{th}$  componente principal individual K;

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

O vetor  $\theta = [\mu, \text{SNP}_i, PC_{jk}]^T$  representa parâmetros desconhecidos. O limiar (Threshold  $\log_{10} P$ ) adotado foi de 5.55 calculado com base na correção do método de Bonferroni para testes múltiplos para identificar SNPs significativos a 5 %.

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

Os SNPs significativos foram localizados nos 11 cromossomos do genoma de referência de *C. canephora* (*Coffee Genome Hub*). As regiões contendo os SNPs foram analisados para a identificação de genes, utilizando o *software* Phyton v 3.6. Genes contendo os marcadores SNP significativos foram analisados.

### **Resultados e discussão**

A análise de qualidade ( $\text{MAF} \geq 0,01$  e  $\text{CR} \geq 90\%$ ) resultou em 17.885 marcadores SNP do total de 18.111 marcadores SNP identificados por Alkimim (2017). Essa alta densidade de marcadores permite aumentar a chance de estudos de GWAS capturar genes de efeitos menores e, dessa forma, aumentar a capacidade da explicação genética do caráter em estudo (Resende et al., 2008; Resende et al., 2016).

Por meio da análise de GWAS, cinco características fenotípicas (Ap, Cer, Dc, Fer e Vig), avaliadas nesse estudo, apresentaram marcadores SNP com associações significativas (Tabelas 4 a 8).

A metodologia de GWAS apresenta maiores dificuldades na detecção de marcadores SNP com associações significativas para características que possuem heranças mais complexas. Isso se deve a dois possíveis fatores, a característica pode ser controlada por muitas variantes raras, cada uma delas tendo grande efeito sobre o fenótipo, ou em contraste, essa característica pode apresentar muitos variantes comuns de pequenos efeitos (Gibson, 2010). Nesse contexto, pode se citar as características produtividade (Prod), tamanho de fruto (Tf) e época de maturação dos frutos (Mat) que nesse trabalho não apresentaram marcadores SNP com associações significativas.

Ao se analisar a característica produtividade (Prod) (Figura 1), é possível perceber seu caráter poligênico e de baixa herdabilidade, ou seja, a mesma é controlada por muitos genes de pequeno efeito (Gardner, 1977). Assim, ao se realizar um estudo de associação torna-se difícil

a detecção desses genes. Ao se referir a detecção deve-se considerar que o beta ( $\beta$ ) do coeficiente de regressão é significativo (difere de zero). Dessa forma, pequenos valores de  $\beta$  dificultam a rejeição da hipótese de nulidade, propiciando a não detecção de SNP significativos.

O mesmo pode ser observado para as características Tf e Mat, estudos de GWS realizados por Alkimim 2017 utilizando-se a mesma população em melhoramento de *C.canephora*, demonstraram que essas características apresentam herdabilidade genômica baixa sendo esse valor de 0,21. Assim, apesar de se tratarem de valores consideráveis, a análise GWS não conseguiu obter boas estimativas de capacidade preditiva, que equivale a correlação entre os valores genéticos genômicos e os valores fenotípicos observados. Desse modo, em ambos os trabalhos as metodologias utilizadas não apresentaram resultados favoráveis para características que apresentaram baixa herdabilidade.

Para a característica incidência à ferrugem (Fer), foram identificados 22 SNPs significativos (Figura 2 A). Destes, nove SNPs se encontram inseridos em genes, sendo que cinco estão marcando diferentes locos que pertencem a mesma característica. Dessa forma, foram encontrados quatro genes potencialmente associados à resistência a ferrugem, localizado nos cromossomos 2, 8, 10 e 11. Um gene foi localizado no cromossomo zero dessa espécie. Estudos realizados na espécie permitiram o sequenciamento de seus cromossomos, no entanto algumas informações ainda não se encontram bem elucidadas. Desse modo, o conjunto de *scaffolds* com sequências não ordenadas foi denominada de cromossomo zero (Denoëud et al., 2014).

Ao se analisar os quatro genes candidatos encontrados para a característica, podem se destacar o gene Flavone-3-Hidroxilase (Cc08\_g11560). Os flavonoides são metabólitos secundários que atuam como moléculas sinalizadoras em interações planta- micróbio (Hahlbrock e Sheel, 1989; Dooner et al., 1991; Koes et al., 1993). Outro gene candidato é o gene Piruvato desidrogenase componente de subunidade E1 mitocondrial (Cc11\_g17430), o complexo piruvato desidrogenase mitocondrial (CPD) catalisa reações cruciais para a formação de acetil-Coa e redução de NAD<sup>+</sup>. O fluxo desse complexo é rigidamente regulado para atender as necessidades metabólicas dos diferentes tecidos presentes no organismo, e sua função está intimamente relacionada a manutenção da homeostase e em alguns organismos pode estar relacionado a sinalização de doenças (Araujo et al., 2011; Fuentes et al., 2011; Gleason et al., 2011).

Na espécie *C. canephora*, poucos estudos envolvendo a identificação de QTL associados à doença foram relatados (Ferrão et al., 2007)). A principal explicação consiste no

baixo investimento realizado na espécie em detrimento da espécie *C. arabica*. No entanto, devido aos danos causados a essa cultura em virtude da suscetibilidade, apresentada por algumas variedades, é de suma importância estudos para melhor compreender os mecanismos de resistência. Estes estudos possibilitam auxiliar os programas de melhoramento no lançamento de novas variedades resistentes (Ferrão et al., 2007).

Outra importante doença do cafeeiro é a cercosporiose (Cer), causada pelo fungo *Cercospora coffeicola*. Esse patógeno também é responsável por grandes perdas na produção, devido principalmente a desfolha intensa ocasionada (Matiello, 2006). No presente estudo, foram identificados 157 SNPs significativos associados a essa característica (Figura 2 B). Desses, 80 estão inseridos em genes, sendo que 12 estão marcando diferentes locos para a mesma característica. Dessa forma, identificou-se 65 genes associados ao fenótipo distribuídos ao longo dos onze cromossomos da espécie *C. canephora*.

Ao se examinar os genes candidatos para a característica, é possível observar que a maior parte deles se encontra relacionados a mecanismos de defesa da planta para proteção contra o patógeno. A entrada dos agentes causadores de doenças em plantas, geralmente ocorre através de uma ferida ou abertura natural, assim uma forma de aumentar a resistência a essas doenças seria fortalecer os tecidos mais propícios a entrada desses patógenos (Mauch Mani and Slusarenko, 1996). Nesse contexto, pode se citar a presença da cutícula que se encontra intimamente relacionado a proteção da planta contra estresses abióticos e bióticos. Entre os genes candidatos encontrados para essa característica foi encontrado o gene fator de transcrição responsivo ao etileno WIN 1 (ERF) (Cc04\_g04250) cuja a função está relacionado ao acúmulo dessa cera cuticular (Aharoni et al., 2004; Brown et al., 2004).

Outra forma de defesas a patógenos em plantas, compreende em uma cascata de sinalização formando uma defesa em camadas com uma resposta dinâmica ao ataque de patógenos e pragas (Zhang et al., 2011; Takahashi et al., 2007; Coll et al., 2011). O gene candidato sintaxina-71 (Cc06\_g03780) pertencente a família de proteínas integradas a membrana plasmática possui sua função relacionada a essa cascata de sinalização relacionada ao mecanismo de defesa da planta (Brown and London, 1998.; Rajedran and Simons, 2005). Essa rede de cascatas possui sua atividade regulada por proteínas transcricionais, nesse estudo foi encontrado o gene candidato fator de *splicing* rico em serina e arginina RS 2Z32(Cc06\_g03240) cuja função se baseia na regulação de *Splicing* alternativo (Long and Cáceres, 2009). Por muito tempo estudos envolvendo *splicing* alternativos e seus reguladores

em plantas era muito limitado, no entanto é sabido que sua função se encontra intimamente relacionado a resposta de estresses bióticos e abióticos em plantas (Duque, 2011).

As principais doenças do cafeeiro, apresentam herdabilidade alta, sendo Fer 0,48 e cer 0,54 e são controladas por um grande número de QTLs confirmando a natureza quantitativa dessas características. No entanto, estudos realizados no organismo modelo *Arabidopsis thaliana*, características fenotípicas avaliadas sob condições de seleção biótica apresentaram resultados expressivos e melhores quando comparados com a seleção de características abióticas. Esses resultados podem ser explicados, em parte, pelo fato de essas características apresentarem um número de locos de efeito maior (Louthan,2011).

Foi possível observar que todos os genes potencialmente associados ao controle da característica incidência à ferrugem (08-g11560, 10-g16500, 11-g17430, 00-g17460 e 02-g38180) também estão presentes no controle da característica incidência à cercosporiose. Melhoristas de culturas perenes, como o cafeeiro, buscam reduzir os ciclos seletivos em seus programas de melhoramento para garantir maior competitividade no lançamento de cultivares (Resende et al., 2005). Desse modo, estratégias de melhoramento que visam a introdução de genes que controlam duas ou mais doenças de importância podem acelerar etapas dos programas auxiliando o melhorista no lançamento de novas cultivares.

Os SNP com associações significativas responsáveis pela arquitetura da planta como altura (Ap) e diâmetro da copa (Dc) estão representados na Figura 3. Para a característica altura de planta (Ap) (Figura 3 A), foram identificados 27 SNP significativos, localizados acima do *threshold* estabelecido. Desses, 13 se encontram inseridos em genes. No entanto, como os SNP V 14979 e V14981; V 807 e V811; V8568 e V8569 estão marcando o mesmo loco, foram identificados dez genes potencialmente ligados a característica altura de planta. Para a característica diâmetro da copa (Dc), foram identificados 168 SNP com associações significativas (Figura 3 B), onde 93 se encontram inseridos em 76 genes.

Na figura 3C é possível verificar 30 SNP significativos para a característica vigor vegetativo (Vig). Desse total, 22 se encontram inseridos em 18 genes.

Nos programas de melhoramento de *C. canephora*, busca-se a seleção de cafeeiros mais vigorosos e com maior diâmetro da copa associado a uma menor altura da planta (Alkimim, 2017). No presente trabalho, foi possível observar a presença de marcadores SNP com associações significativas, nos genes C 06-g02290; C11-g11880 e C00-g17460, para essas três características. Esse resultado é de grande utilidade para os melhoristas, os quais podem

conduzir seus programas de modo a introduzir esses genes para as três características de forma conjunta.

Contudo, mesmo com as limitações apresentadas pela metodologia GWAS, a população de *C. canephora* em estudo obteve resultados promissores. Ao analisar a quantidade de SNP significativos, inseridos em genes, por cromossomos (Figura 4) foi possível observar que os SNP cobriram bem o genoma. Estes, estão presentes em todos os onze cromossomos da espécie, com destaque para os cromossomos 0, 2, 6, 9 e 11 que apresentaram 23, 29, 28, 22 e 26 marcadores SNP, respectivamente.

Dos 404 SPN com associações significativas encontrados nesse estudo 217 estavam inseridos em genes. Analisando esses 217 marcadores SNP com o auxílio do *software* Phytion, foi possível anotar 96,4% do total desses genes.

Ao se fazer uma análise geral do estudo de GWAS na população de *C.canephora*, pode-se concluir que a metodologia foi altamente eficiente, identificando um grande número de marcadores SNP com associações significativas. Parte desses bons resultados, podem ser associados ao tamanho efetivo e a estrutura genética da população. Como a população está dividida em dois grupos varietais divergentes geneticamente, é possível observar uma maximização da variância genética dentro da amostra por razões que incluem adaptações diferentes. Em virtude dessa heterogeneidade genética, a correlação entre o fenótipo e uma variante rara será diminuída, propiciando a melhor detecção de associações genômicas (Li et al., 2010).

## **Conclusão**

A metodologia de GWAS foi capaz de detectar marcadores SNP com associações significativas para cinco características fenotípicas (Ap, Dc, Vig, Fer e Cer) de grande importância no melhoramento genético do cafeeiro. O *software* Phytion permitiu identificar os genes candidatos onde esses marcadores SNP estavam inseridos, desses genes foi possível anotar funções que se encontravam diretamente relacionados a característica de interesse. Essas informações serão de grande valia para o programa de melhoramento da espécie, uma vez que essas regiões cromossômicas poderão ser utilizadas em estudos de associação genômica ampla (SAM) auxiliando o melhorista na obtenção de variedades melhoradas.

**Tabela 1.** Genótipos de *C. canephora* dos grupos varietais Conilon e Robusta analisados

<b>Genótipos Conilon</b>			
UFV 513	UFV 3628-23	UFV 3628-45	UFV 3629-31
UFV 3627-20	UFV 3628-24	UFV 3628-46	UFV 3629-34
UFV 3627-24	UFV 3628-26	UFV 3628-5	UFV 3629-4
UFV 3627-27	UFV 3628-27	UFV 3629-10	UFV 3630-10
UFV 3627-29	UFV 3628-28	UFV 3629-11	UFV 3630-2
UFV 3627-30	UFV 3628-29	UFV 3629-17	UFV 3630-5
UFV 3627-31	UFV 3628-3	UFV 3629-24	UFV 3630-6
UFV 3628-1	UFV 3628-32	UFV 3629-25	UFV 3630-7
UFV 3628-10	UFV 3628-35	UFV 3629-26	UFV 3631-10
UFV 3628-16	UFV 3628-4	UFV 3629-27	UFV 3631-11
UFV 3628-17	UFV 3628-40	UFV 3629-28	UFV 3631-13

<b>Genótipos Robusta</b>			
UFV 3356-71	UFV 3362-118	UFV 3367-101	UFV 3373-43
UFV 3356-76	UFV 3363-125	UFV 3367-105	UFV 3374-29
UFV 3357-93	UFV 3365-144	UFV 3368-52	UFV 3375-65
UFV 3358-88	UFV 3366-134	UFV 3368-58	UFV 3375-66
UFV 3360-169	UFV 3366-138	UFV 3370-50	UFV 3376-8
UFV 3360-171	UFV 3366-139	UFV 3371-19	UFV 3376-9
UFV 3361-148	UFV 3367-96	UFV 3371-20	UFV 3377-12
UFV 3361-151	UFV 3367-97	UFV 3373-36	UFV 514
UFV 3356-71	UFV 3362-118	UFV 3367-101	UFV 3373-43
UFV 3356-76	UFV 3363-125	UFV 3367-105	UFV 3374-29
UFV 3357-93	UFV 3365-144	UFV 3368-52	UFV 3375-65
UFV 3358-88	UFV 3366-134	UFV 3368-58	UFV 3375-66
UFV 3360-169	UFV 3366-138	UFV 3370-50	UFV 3376-8

**Tabela 2.** Híbridos entre cafeeiros de *C. canephora* dos grupos varietais Conilon e Robusta analisados

<b>Genótipos Híbridos</b>	<b>Cruzamentos</b>
H092-2, H092-8, H092-10, H092-20, H092-22, H092-35	UFV 3366-139 x UFV 513
H093-1, H093-10, H093-12, H093-20	UFV 3366-139 x UFV 3629-11
H094-2, H094-5, H094-12, H094-13, H094-18, H094-30	UFV 3367-98 x UFV 513
H095-9, H095-11, H095-14, H095-25, H095-28	UFV 3366-139 x UFV 3627-31
H097-8, H097-13, H097-16, H097-17, H097-18, H097-19	UFV 3374-28 x UFV 3627-31
H098-1, H098-2, H098-3, H098-5, H098-13	UFV 3365-144 x UFV 3629-11
H099-1, H099-2, H099-7, H099-8, H099-10	UFV 3365-144 x UFV 513
H0910-4, H0910-5, H0910-6, H0910-7, H0910-8	UFV 3367-98 x UFV 3629-11
H0911-2, H0911-3, H0911-4, H0911-5, H0911-6, H0911-7	UFV 3367-98 x UFV 3628-2
H0912-1, H0912-3, H0912-4, H0912-6, H0912-7, H0912-8	UFV 3366-139 x UFV 3628-2
H0913-1, H0913-3, H0913-4, H0913-5, H0913-7	UFV 3365-144 x UFV 3628-2
H0914-2, H0914-3, H0914-4, H0914-5, H0914-6	UFV 3374-28 x UFV 513
H0915-1, H0915-2, H0915-3, H0915-4, H0915-5	UFV 3373-36 x UFV 3628-2
H0916-1, H0916-2, H0916-4	UFV 3367-98 x UFV 3629-25
H0917-1, H0917-2, H0917-3	UFV 3365-144 x UFV 3629-25
H0918-1, H0918-2, H0918-3	UFV 3373-36 x UFV 513
H0919-2	UFV 3373-36 x UFV 3629-11
H0920-1	UFV 3373-36 x UFV 3629-25
H0921-1	UFV 3374-28 x UFV 3629-11
H0922-1	UFV 3373-36 x UFV 3627-31



**Tabela 3.** Avaliações fenotípicas realizadas para oito características morfoagronômicas na espécie *C.canephora*

Característica	Forma de avaliação
Vigor vegetativo (Vig)	Notas de 1 (totalmente depauperada) a 10 (altamente vigorosa).
Incidência à ferrugem do cafeeiro no campo (Fer)	Notas de 1(sem nenhum sintoma) a 5 (altamente suscetível)
Incidência de cercosporiose (Cer)	Notas de 1(sem nenhum sintoma) a 5(altamente suscetível)
Época de maturação dos frutos (Mat)	Notas de 1 a 3 para precoce, média e tardia, respectivamente.
Tamanho de fruto (Tf)	Notas de 1 a 3 para pequeno, médio e graúdo, respectivamente.
Altura da planta (Ap)	Determinada em centímetros (cm)
Diâmetro da projeção da copa (Dc)	Determinada em centímetros (cm)
Produção em litros por planta (Prod)	Volume total em litros por planta (l)

**Tabela 4:** SNP marcando regiões significativas para a característica altura da planta

SNP	P-Valor	R <sup>2</sup>
133	8,82 E-08	0,01057
807	6,79 E-03	0,092912
811	1,05 E- 03	0,090432
1044	1,16 E-09	0,0095
1781	3,92 E-07	0,089132
1807	1,15 E-08	0,07123
2151	2,86 E-06	0,090255
3293	8,08 E-08	0,07576
5675	2,00 E-07	0,09374
5681	1,05 E-08	0,081793
5768	5,90 E-08	0,15368
7728	5,68 E-06	0,051338
7730	7,41E-08	0,0512
8568	2,24 E-08	0,12066
8569	2,24 E-08	0,12066
8987	7,11 E-08	0,1242
9610	1,49 E-09	0,0556
10807	1,15 E-09	0,0599
10809	5,58 E-08	0,06661
12701	1,41 E-09	0,0805
13059	2,81 E-06	0,142584
13800	9,43 E-08	0,085621
13801	1,42 E-08	0,096601
14201	6,01 E-08	0,06465
14979	1,07 E-08	0,071123
14981	2,32 E-08	0,06677
17001	3,25 E-08	0,07194

**Tabela 5:** SNP marcando regiões significativas para a característica cercosporiose

SNP	P-Valor	R <sup>2</sup>
1	6,43E-07	0,0384
131	6,01E-09	0,01765
133	6,88E-08	0,01944
135	2,74E-07	0,01485
743	3,40E-09	0,11347
755	2,42E-09	0,08887
807	9,89E-01	0,08425
811	2,24E-02	0,08277
861	5,97E-07	0,10091
958	1,59E-08	0,03405
1044	1,12E-09	0,01716
1429	3,33E-09	0,07538
1532	1,55E-09	0,10259
1738	9,24E-08	0,03511
1746	8,03E-09	0,01681
1781	2,39E-09	0,05255
1803	2,09E-08	0,15086
1807	1,27E-07	0,13572
1858	1,75E-08	0,02507
2093	6,37E-08	0,01979
2151	4,55E-04	0,08385
2225	1,73E-09	0,0434
2227	1,74E-08	0,12801
2401	1,42E-09	0,03136
2405	2,63E-09	0,04593
2406	2,64E-08	0,06042
2407	9,24E-08	0,05363
2512	8,33E-09	0,01164
2513	5,86E-09	0,01245
2514	3,21E-09	0,01348
2544	1,86E-09	0,03485
2723	2,84E-08	0,02416
3031	9,09E-08	0,04617
3293	2,40E-09	0,05349
3294	7,52E-09	0,04657
3765	2,57E-09	0,0261
3839	1,75E-09	0,01969
3962	6,87E-08	0,00568
4195	2,23E-09	0,0566
4253	4,77E-07	0,04902
4726	3,84E-09	0,04638
4787	5,07E-09	0,05567
4790	2,35E-09	0,04902
4887	6,33E-08	0,01271
4888	6,33E-08	0,01271
4930	2,97E-09	0,01975
5215	8,42E-09	0,00959

Continuação...

<b>SNP</b>	<b>P-Valor</b>	<b>R<sup>2</sup></b>
5244	3,44E-09	0,02088
5295	2,65E-05	0,01945
5296	4,31E-07	0,01642
5298	4,31E-07	0,01642
5299	2,54E-08	0,01132
5302	2,69E-08	0,01226
5310	1,50E-09	0,07796
5681	2,90E-08	0,0522
5705	8,01E-09	0,00879
5717	1,96E-09	0,00234
5730	7,24E-09	0,05194
5768	1,01E-07	0,23035
5769	5,05E-07	0,19163
6011	1,76E-08	0,02882
6012	7,22E-09	0,02815
6442	1,65E-09	0,02913
6443	1,65E-09	0,02913
6626	4,69E-09	3,94E-05
7220	4,01E-09	0,01866
7503	3,04E-09	0,02907
7583	1,18E-09	0,01325
7728	1,38E-08	0,03136
7730	2,09E-07	0,03688
7734	1,30E-09	0,01834
7815	6,94E-07	0,05388
7988	2,96E-09	0,03506
8246	3,26E-09	0,0291
8265	3,26E-09	0,0291
8568	4,52E-08	0,16336
8569	4,52E-08	0,16336
8627	2,72E-08	0,07917
8654	1,66E-09	0,03634
8663	7,33E-07	0,04121
9258	3,83E-09	0,03953
9581	1,16E-09	0,06022
9595	3,83E-08	0,00915
9601	1,98E-09	0,0062
9607	1,64E-09	0,01153
9610	2,10E-09	0,04376
9687	4,30E-09	0,04338
9689	2,47E-09	0,04953
10109	4,12E-09	0,12276
10135	4,20E-08	0,08598
10225	8,34E-08	0,11756
10274	1,11E-09	0,09534

Continuação...

SNP	P-Valor	R <sup>2</sup>
10605	2,09E-09	0,02774
10660	1,16E-09	0,02682
11110	8,01E-09	0,01264
11133	3,65E-09	0,1274
11148	3,82E-08	0,02906
11230	2,37E-09	0,06048
11232	2,66E-09	0,06103
11236	2,68E-08	0,10541
11343	9,02E-08	0,05741
11586	4,27E-09	0,02323
11589	6,35E-09	0,07293
11629	2,26E-09	0,01956
11630	6,06E-09	0,04825
12182	3,26E-09	0,0291
12383	8,14E-07	0,03441
12626	1,11E-09	0,03507
12675	1,11E-09	0,03507
12680	1,11E-09	0,03507
12701	4,30E-06	0,1607
12702	2,77E-08	0,14726
12703	4,64E-08	0,13878
12742	1,11E-09	0,03507
12765	1,11E-09	0,03507
12783	1,11E-09	0,03507
12797	1,11E-09	0,03507
12808	1,11E-09	0,03507
12844	2,43E-09	0,05029
12989	2,38E-08	0,0474
13008	3,24E-09	0,03171
13100	1,11E-09	0,03507
13643	2,13E-08	0,10874
13800	1,55E-07	0,15363
13801	6,09E-05	0,17303
13803	4,88E-08	0,1029
13813	5,03E-09	0,09236
14703	2,89E-06	0,05101
10605	2,09E-09	0,02774
10660	1,16E-09	0,02682
11110	8,01E-09	0,01264
11133	3,65E-09	0,1274
11148	3,82E-08	0,02906
11230	2,37E-09	0,06048
11232	2,66E-09	0,06103
11236	2,68E-08	0,10541

Continuação...

<b>SNP</b>	<b>P-Valor</b>	<b>R<sup>2</sup></b>
11343	9,02E-08	0,05741
13800	1,55E-07	0,15363
13801	6,09E-05	0,17303
13803	4,88E-08	0,1029
13813	5,03E-09	0,09236
14703	2,89E-06	0,05101
14704	7,37E-06	0,04739
14720	2,17E-09	0,01629
14979	8,11E-08	0,04426
14981	2,85E-09	0,03771
15204	2,08E-09	0,13871
15410	6,12E-09	0,0482
15989	2,08E-09	0,01963
16014	6,79E-09	0,01474
16283	1,43E-09	0,07204
16429	4,64E-09	0,00323
16467	3,21E-07	0,11961
16633	9,31E-08	0,04035
16634	9,97E-06	0,1735
16726	3,47E-09	0,05951
16879	4,06E-09	0,0199
16998	6,54E-08	0,04552
17000	1,38E-09	0,05862
17001	2,73E-09	0,0452
17016	9,24E-08	0,03511
17063	2,03E-09	0,07862
17087	9,24E-08	0,03511
17098	9,24E-08	0,03511
17102	9,24E-08	0,03511
17106	9,24E-08	0,03511
17205	3,26E-09	0,0291
17229	3,26E-09	0,0291
17589	2,52E-09	0,00868
17807	1,56E-09	0,04097
17822	6,69E-08	0,01015

**Tabela 6:** SNP marcando regiões significativas para a característica diâmetro da copa

<b>SNP</b>	<b>P-Valor</b>	<b>R<sup>2</sup></b>
1	3,21E-06	0,04259
133	3,03E-09	0,0198
135	6,79E-09	0,01627
255	6,77E-09	0,08141
341	3,22E-09	0,12718
358	4,93E-09	0,06845
365	4,81E-09	0,06174
743	2,20E-09	0,109833
755	5,55E-08	0,08062
807	1,70E+00	0,08068
808	2,54E-09	0,046247
811	9,56E-03	0,08162
861	5,30E-08	0,08928
958	2,81E-08	0,03182
1044	3,88E-08	0,01811
1229	8,95E-08	0,008262
1230	6,30E-08	0,00973
1388	7,57E-09	0,11697
1429	6,30E-08	0,075108
1441	4,15E-09	0,04513
1532	2,70E-09	0,101862
1656	5,66E-09	0,0321
1738	4,80E-09	0,03515
1746	3,74E-09	0,01777
1781	2,06E-08	0,05941
1803	5,32E-07	0,15761
1807	2,36E-05	0,15072
1858	5,50E-09	0,02026
2093	3,01E-09	0,02025
2096	1,05E-09	0,027426
2151	2,64E-03	0,08862
2225	2,14E-09	0,043609
2227	5,26E-09	0,11101
2262	5,42E-09	0,13647
2406	3,18E-09	0,04611
2514	6,41E-09	0,01264
2575	4,22E-09	0,07035
2723	4,85E-07	0,02666
3031	4,48E-08	0,05361
3093	2,24E-09	0,017816
3293	1,35E-08	0,06136

Continuação...

---

3294	3,84E-08	0,05439
3765	7,52E-09	0,02879
3962	4,73E-09	0,00777
4253	1,17E-09	0,042946
4726	9,31E-07	0,06425
4790	4,18E-09	0,04488
4930	1,31E-09	0,032673
5244	7,09E-09	0,02602
5295	3,37E-09	0,01771
5296	1,23E-09	0,013717
5298	1,23E-09	0,013717
5302	7,11E-09	0,0093
5310	2,17E-08	0,08834
5681	3,54E-08	0,05975
5705	7,14E-09	0,01264
5708	4,31E-08	0,07074
5717	4,62E-08	0,00644
5730	3,22E-08	0,08472
5768	2,14E-07	0,21436
5769	4,34E-08	0,18201
6011	1,48E-08	0,03029
6012	2,12E-09	0,034543
6442	4,60E-09	0,0323
6443	4,60E-09	0,0323
6602	1,43E-09	0,10035
7220	4,10E-09	0,01806
7221	6,97E-09	0,01189
7236	7,21E-09	0,05087
7601	8,85E-08	0,132521
7728	4,58E-08	0,02562
7730	1,37E-08	0,03445
7734	2,53E-09	0,016091
7815	7,46E-06	0,07852
7988	7,33E-09	0,03664
8568	4,33E-09	0,13342
8569	4,33E-09	0,13342
8627	5,81E-08	0,072612
8654	5,49E-08	0,03544
8663	5,45E-08	0,041845
9393	6,68E-09	0,03283
9546	1,17E-09	0,066314
9581	2,30E-08	0,08575
9595	2,65E-08	0,01345
9607	2,54E-08	0,01939
9610	4,87E-07	0,0649
9689	6,70E-09	0,04424
10172	3,17E-09	0,05253

---



Continuação...

---

10184	1,77E-09	0,044444
10225	2,13E-08	0,11912
10274	7,14E-09	0,08148
11133	7,10E-09	0,12279
11148	6,77E-09	0,033
11236	4,07E-09	0,09069
11332	1,82E-09	0,053758
11343	1,13E-08	0,08415
11448	1,21E-08	0,03052
11586	3,56E-08	0,0341
11589	1,02E-09	0,098202
11629	1,70E-09	0,02354
12262	1,91E-09	0,154629
12266	9,83E-08	0,154703
12383	1,41E-09	0,030801
12626	2,87E-09	0,038509
12675	2,87E-09	0,038509
12680	2,87E-09	0,038509
12680	2,87E-09	0,03851
12701	5,71E-06	0,1506
12702	2,14E-09	0,123509
12703	2,42E-09	0,119785
12742	2,87E-09	0,03851
12765	2,87E-09	0,03851
12783	2,87E-09	0,03851
12797	2,87E-09	0,03851
12808	2,87E-09	0,03851
12844	2,66E-09	0,04625
12849	6,70E-09	0,06799
12989	6,10E-08	0,048228
13008	5,10E-09	0,03706
13077	3,60E-09	0,14707
13100	2,87E-09	0,03851
13800	3,61E-06	0,14148
13801	1,47E-06	0,16143
13803	3,69E-09	0,0878
13813	3,79E-09	0,0864
14358	6,45E-09	0,08143
14411	6,25E-09	2,16E-02
14703	8,77E-05	0,05506
14704	1,51E-06	0,0524
14720	4,34E-08	0,02059
14743	7,36E-09	0,0156
14979	1,03E-08	0,04722
14981	7,50E-08	0,039337
15540	3,97E-09	0,12047
16227	2,53E-09	0,06647

---

Continuação...

---

16283	4,76E-07	0,09163
16286	5,30E-09	0,06974
16633	3,77E-07	0,05311
16634	3,24E-05	0,1585
16726	9,88E-08	0,069124
16879	3,09E-09	0,0251
16998	2,33E-09	0,046346
17000	8,34E-08	0,056622
17001	6,67E-08	0,047656
17016	4,80E-09	0,03515
17063	8,39E-08	0,082731
17087	4,80E-09	0,03515
17098	4,80E-09	0,03515
17102	4,80E-09	0,03515
17106	4,80E-09	0,03515
17588	6,80E-09	0,00075
17589	1,21E-09	0,01294
17701	3,02E-09	0,00018
17744	4,43E-09	0,06285
17807	9,90E-08	0,057315
17816	7,20E-09	0,08661
17822	3,50E-09	0,01909
16283	4,76E-07	0,09163
16286	5,30E-09	0,06974
16633	3,77E-07	0,05311
16634	3,24E-05	0,1585
16726	9,88E-08	0,069124
16879	3,09E-09	0,0251
16998	2,33E-09	0,046346
17000	8,34E-08	0,056622
17001	6,67E-08	0,047656
17016	4,80E-09	0,03515
17063	8,39E-08	0,082731
17087	4,80E-09	0,03515
17098	4,80E-09	0,03515
17102	4,80E-09	0,03515
17106	4,80E-09	0,03515
17588	6,80E-09	0,00075
17589	1,21E-09	0,01294
17701	3,02E-09	0,00018
17744	4,43E-09	0,06285
17807	9,90E-08	0,057315
17816	7,20E-09	0,08661
17822	3,50E-09	0,01909

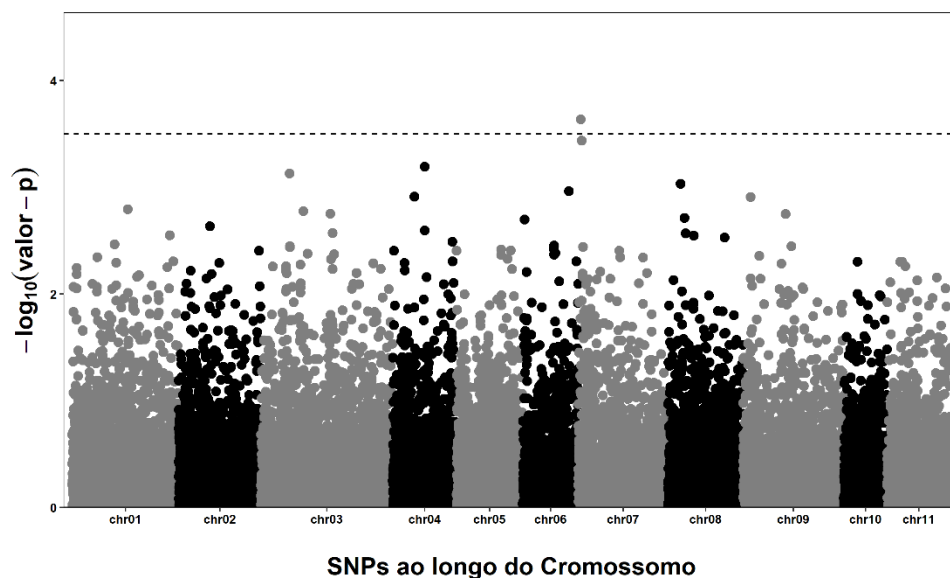
---

**Tabela 7:** SNP marcando regiões significativas para a característica incidência a ferrugem

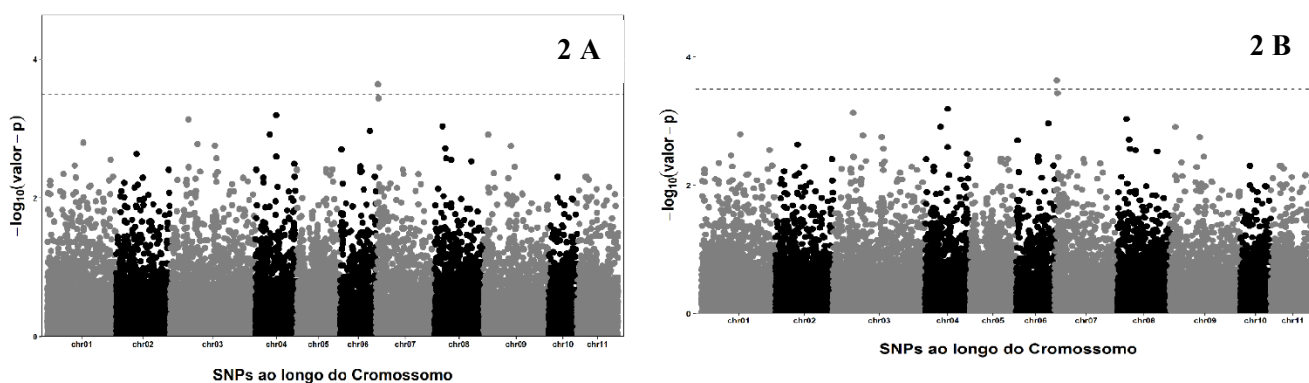
SNP	P-Valor	R <sup>2</sup>
378	3,81E-08	0,034197
743	8,91E-08	0,02873
807	6,16E-02	0,215054
808	6,29E-07	0,16663
811	2,30E-03	0,198617
1094	5,73E-08	0,046674
1426	3,53E-08	0,034466
1656	1,21E-09	0,06511
1807	3,02E-08	0,011154
2151	7,10E-08	0,14466
2152	8,58E-07	0,047741
2406	2,32E-07	0,176328
2407	7,26E-08	0,164573
4726	2,54E-08	0,155716
6011	7,27E-06	0,159832
6012	1,30E-07	0,141061
7728	2,54E-08	0,144566
7730	5,37E-05	0,178737
7815	1,23E-08	0,02862
8568	8,49E-08	0,041101
8569	8,49E-08	0,041101
16633	5,04E-08	0,1398

**Tabela 8:** SNP marcando regiões significativas para a característica vigor vegetativo.

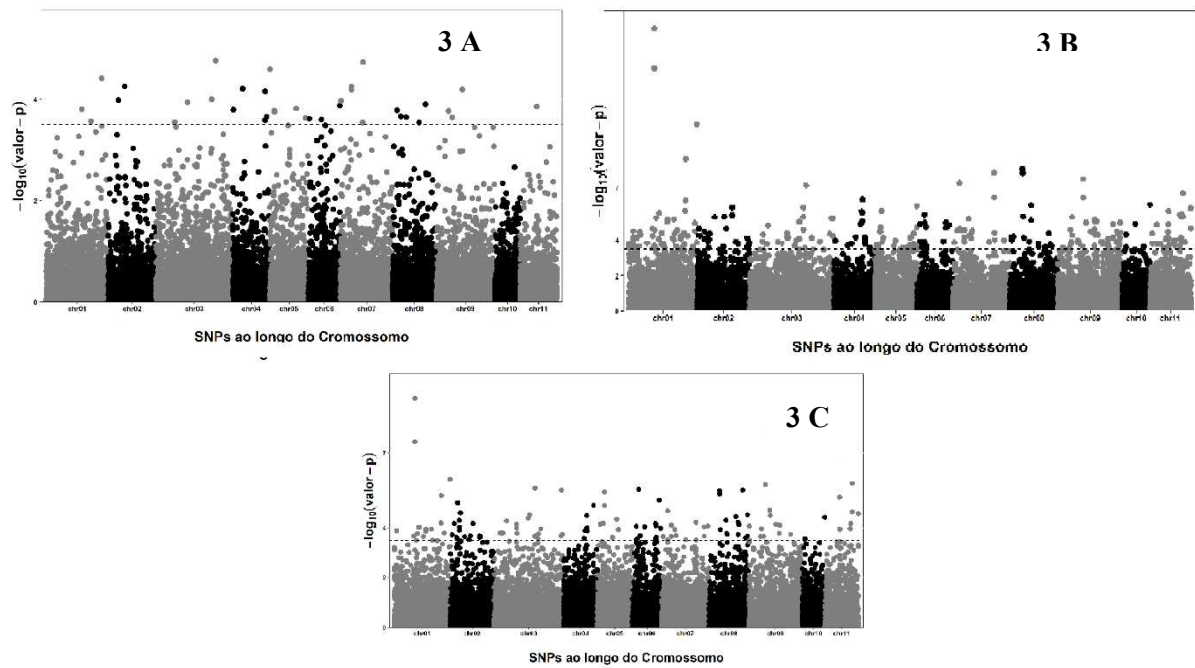
SNP	P-Valor	R <sup>2</sup>
1	1,92E-08	0,01343
807	3,55E-06	0,02209
811	6,23E-03	0,02388
1807	5,07E-08	0,13265
2151	1,12E-08	0,03157
2456	9,95E-08	0,14692
2553	2,53E-09	0,08026
5244	9,31E-06	0,06743
5295	5,90E-08	0,03292
5768	1,63E-08	0,19424
5769	2,27E-09	0,16205
6011	2,70E-09	0,00894
7399	3,00E-09	0,06896
7601	2,61E-07	0,16082
8627	3,18E-08	0,08622
9595	3,25E-09	0,00309
9856	1,28E-09	0,11979
10274	3,71E-08	0,1232
10276	1,31E-09	0,1474
11589	2,93E-08	0,09811
12383	7,75E-08	0,04369
12701	2,10E-09	0,12632
14703	3,30E-08	0,01628
14704	4,45E-08	0,015
15597	3,11E-08	0,05328
15777	2,98E-09	0,10027
16467	1,85E-08	0,15335
16633	1,98E-09	0,02032
16634	1,66E-06	0,19205
16685	3,25E-09	0,03456



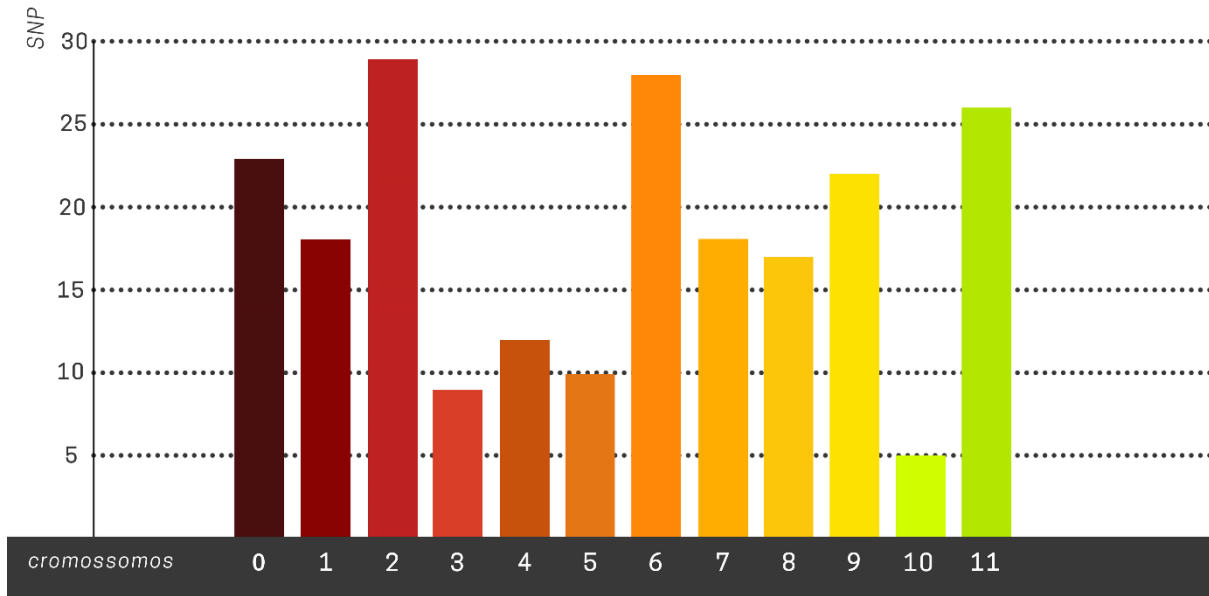
**Figura 1:** Gráfico de *Manhattan Plot* para produtividade. As linhas tracejadas mostram os limiars ajustados de Bonferroni de 5,55 para alfa igual a 5%, Eixo X indicando cromossomo e eixo Y valores de P-valor.



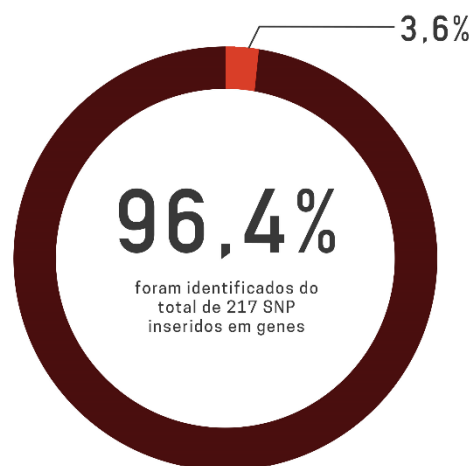
**Figura 2:** Gráficos de *Manhattan Plot* para Ferrugem (A) e Cercosporiose (B). As linhas tracejadas mostram os limiars ajustados de Bonferroni de 5,55 para alfa igual a 5%, Eixo X indicando cromossomo e eixo Y valores de P-valor



**Figura 3.** Gráficos de *Manhattan Plot* para altura (A), diâmetro da copa (B) e vigor vegetativo (C), As linhas tracejadas mostram os limiars ajustados de Bonferroni de 5,55 para alfa igual a 5%, Eixo X indicando cromossomo e eixo Y valores de P-valor



**Figura 4.** Representação SNP significativos inseridos em genes, distribuídos ao longo dos cromossomos da espécie *C. canephora*.



**Figura 5.** Anotação funcional realizada 217 SNP inseridos em genes na espécie *C.canephora* por meio do *software* Phyton.

## Referências

- Aharoni A et al (2004) . **The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in *Arabidopsis***. *Plant Cell*, Vol 16, p 2463–2480.
- Alkimim, E,R; Caixeta, E,T, Sousa, T,V, et al (2017), **Marker assisted selection provides arábica coffee with genes from other *Coffea* species targeting on multiple resistance to rust and coffee berry disease**, *Molecular Breed*, Vol 37, p 6,
- Anderson, C,A; Pettersson, F,H,; Clarke, G,M,; et al, ( 2010) **Data quality control in genetic case-control association studies**, *Nature Protocols*, v,5, n 9, p 1564-1573
- Araújo, W. L., Nunes-Nesi, A., Osorio, S., Usadel, B., Fuentes, D., Nagy, R.,Martinoia, E. (2011). **Antisense inhibition of the iron-sulphur subunit of succinate dehydrogenase enhances photosynthesis and growth in tomato via an organic acid–mediated effect on stomatal aperture**, *The Plant Cell*, p 110.
- Babova, O; Occhipinti, A; Maffei, M,E (2016), **Chemical partitioning and antioxidante capacity of green coffee (*Coffea arabica* and *Coffea canephora*) of diferente geographical origin phytochemistry**, Vol 123, p 33-39.
- Broun P, Poindexter P, Osborne E, Jiang CZ, Riechmann JL (2004). **WIN1, a transcriptional activator of epidermal wax accumulation in *Arabidopsis***. *Proc Natl Acad Sci USA*, Vol 101, p 4706–4711
- Brown D.A, London E. (1998). **Structure and origin of ordered lipid domains in biological membranes**. *J. Membr. Biol.*, Vol. 164, p 103-114.
- Cantor, R, M,; Lange, K,; Sinsheimer, J,S( 2010), **Prioritizing GWAS results: A review of statistical methods and recommendations for their application**, *American Journal of Human Genetics*, Vol, 86,p 6-22.
- Coll N.S, Epple P, Dangl J.L (2011). **Programmed cell death in the plant immune system**. *Cell Death Differ*, Vol18, p 1247–1256.
- Diniz L,E,C; Sakiyama N,S; Lashermes P, et al (2005) **Analysis of AFLP marker associated to the Mex-1 resistance locus in Icatu progênies**, *Crop Breed Appl Biotechnol* vol 5 p 387-393,
- Duque P (2011) A role for SR proteins in plant stress responses. *Plant Signal Behav* 6:49–54.
- Fazuoli, L,C, **Genética e melhoramento do cafeeiro** (1986), In: Rena, A,B; Malavolta, E,; Rocha, M,; Yamada, T, *Cultura do cafeeiro: fatores que afetam a produtividade*, Associação Brasileira para Pesquisa da Potassa e do fosfato p 86-113.
- Ferrão, R, G,, Ferrão, M, G, A,, Fonseca, A, F ,A,, and Pacova, B, E, V, (2007), **Melhoramento genético de *Coffea canephora***, Vol,1, p, 123-173, Incaper, Vitória, Brazil,
- Gardner, E,J, (1977), **Genética**, Interamericana: Rio de Janeiro, 5 ed, P, 327-347
- Gibson, G,(2011) **Rare and common variants: Twenty arguments**, *Nat Rev Genetics*, Vol 13, P, 327-347,



Li, H.; Durbin, R,C (2010), **The interactions of selection and linkage II**, *Optimus models, Genetics*, Vol 26, p 589-595.

Li, Y., Huang, Y., Bergelson, J., Nordborg, M., Borevitz, J,O (2010): **Association mapping of local climate sensitive quantitative trait loci in *Arabidopsis thaliana***, *Proc Natl Acad Sci*, Vol 49, P 21199-21204,

Long J.C, Caceres J.F (2009) **The SR protein family of splicing factors: master regulators of gene expression**. *Biochem*, Vol 417, p 15–27

Louthan, A,M,; Kay, K,M, (2011), **Comparing the adaptative landscape across trait types: Larger QTL effect size in traits under biotic selection**, *BMC Evol Biol*, Vol 11, p, 60,  
Maia, T., Badel, J,T, Fernandes, M,B, et al (2017), **Variation in Aggressiveness Components in the *Hemileia Vastatrix* Population in Brazil**, *J Phytopathol* Vol165, p 174-188.

Matiello, J,B,; Garcia, A,W,R,; Almeida, S,R, (2006), **Adulbos, corretivos e defensivos para a lavoura cafeeira**, *Varginha-MG*, P, 54-55.

Mauch-Mani, B., and Slusarenko, A.J. (1996). **Production of salicylic acid precursors is a major function of phenylalanine ammonia-lyase in the resistance of *Arabidopsis* to *Pemnospora parasitica***. *Plant Cell* , p 203-2012

Meuwissen, T.; Hayes, B.; Goddard, M (2016), **Genomic selection: A paradigma shift in animal breeding**, *Animal Frontiers*, Vol 6, p 6.

Musoli, P,; Cubry P,; Billot, C,; Dufour, M,; Bellis, F, de; Pot, D,; Bieysse, A,; Charrier, T,L (2009), **Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda**, *Genome*, Vol 7, p 634-646.

Pearson, T,A,; Manolio, T,A (2008), **How to interpret a genome wide association study**, *Journal of the american medical association*, Vol 299, p, 1335-1344,

Rajendran L, Simons K (2005). **Lipid rafts and membrane dynamics**, *J. Cell Sci*, Vol. 118, p 1099-1102.

**Resende M,D,V de (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding**, *Crop Breed Appl Biotechnol*, V,16, p 330-339,

Resende, M,D,V de; Duarte, J,B (2007) **Precisão e controle de qualidade em experimentos de avaliação de cultivares**, *Pesquisa Agropecuária Tropical* Vol 37, p 182-194.

Resende, M,D,V, et al, (2008) **Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético**, *Pesquisa Florestal Brasileira*, Vol 56,p 63-77.

Resende, M,D,V,de; Barbosa, M,H,P,; Rezende, G,D,S,P,; Aguiar, A,M,; Dias, L,A, dos S,; Sturion, J,A, (2005), **Métodos e estratégias de espécies perenes: Estado da arte e perspectivas**, In: *Congresso Brasileiro de Melhoramento de Plantas*, Gramado, Anais.

Slusarenko, A.J. (1996). **The role of lipoxygenase in, resistance of plants to infection. In *Lipoxygenase and Lipoxygenase Pathway Enzymes***, Champaign, IL: AOCS Press, p. 176-197.

Sousa, T,V; Caixeta E,T; Alkimim, E,R, et al (2017), **Molecular markers useful to discriminate Coffea arábica cultivars with high genetic similarity**, Euphytica,

Souza, F,F,D (2011), **Estudos sobre a diversidade, estrutura populacional, desequilíbrio de ligação e mapeamento associativo em Coffea canephora Pierre ex Froehner**, Universidade Federal de Viçosa, Viçosa, MG,

Takahashi Y, Nasir KHB, Ito A, Kanzaki H, Matsumura H, Saitoh H et al (2007) **A high-throughput screen of cell-death-inducing factors in Nicotiana benthamiana identifies a novel MAPKK that mediates INF1-induced cell death signaling and non-host resistance to Pseudomonas cichorii**. Plant J, Vol 49, p 1030–1040.

Taylor, J,F, et al (1998), **Candidate gene analysis of GH1 for effects on growth and carcass composition of cattle**, Animal genetics, Vol, 29, p194-201

Wang, W,Y,S,et al (2005), **Genome wide association studies: Theoretical and practical concerns**, Nature Reviews Genetics, Vol 6, p,109-118

Xu. S, Zhang. Z, Jing. B, Gannon. P, Ding. J et al (2011) **Transportin-SR is required for proper splicing of resistance genes and plant immunity**. PLoS Genet, Vol 7, p 6.

Yang, J, et al (2013), **Genome wide complex trait analysis (GCTA): Methods, Data analyses, and interpretations**, P 215-236,