



UNIVERSIDADE
ESTADUAL DE LONDRINA

LÍVIA MARIA NOGUEIRA BRITO

**ESTUDOS DE ASSOCIAÇÃO GENÔMICA AMPLA EM UMA
POPULAÇÃO F₂ DE *COFFEA ARABICA L.***

Londrina
2019



UNIVERSIDADE ESTADUAL DE LONDRINA
CENTRO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL



Programa de
Pós-graduação em
Genética e Biologia Molecular

LÍVIA MARIA NOGUEIRA BRITO

**ESTUDOS DE ASSOCIAÇÃO GENÔMICA AMPLA EM UMA
POPULAÇÃO F₂ DE *COFFEA ARABICA L.***

Orientador: Dr. Luiz Filipe P. Pereira

Londrina
2019

LÍVIA MARIA NOGUEIRA BRITO

**ESTUDOS DE ASSOCIAÇÃO GENÔMICA AMPLA EM UMA
POPULAÇÃO F₂ DE *COFFEA ARABICA L.***

Tese apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito para a obtenção do título de Doutor.

Orientador: Dr. Luiz Filipe Protasio
Pereira

Londrina
2019

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Nogueira Brito, Lívia Maria.

Estudos de associação genômica ampla em uma população F2 de Coffea arabica L. / Lívia Maria Nogueira Brito. - Londrina, 2019. 95 f.

Orientador: Luiz Filipe Protasio Pereira.

Coorientador: Juarez Pires Tomaz.

Tese (Doutorado em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Centro de Ciências Biológicas, , 2019.

Inclui bibliografia.

1. Estudos de associação genômica ampla - GWAS - Tese. 2. Genotipagem por sequenciamento - GBS - Tese. 3. Estrutura Populacional - Tese. 4. Café - Tese. I. Pereira, Luiz Filipe Protasio. II. Tomaz, Juarez Pires. III. Universidade Estadual de Londrina. Centro de Ciências Biológicas. IV. Título.

LÍVIA MARIA NOGUEIRA BRITO

**ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA EM UMA
POPULAÇÃO F₂ DE *COFFEA ARABICA L.***

Tese apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito para a obtenção do título de Doutor.

BANCA EXAMINADORA

Orientador: Dr. Luiz Filipe Protasio Pereira
Empresa Brasileira de Pesquisa Agropecuária –
Embrapa

Dr^a. Bruna Silvestre Rodrigues da Silva
Universidade Estadual de Londrina – UEL

Dr. Gustavo Cesar Sant'Ana
Tropical Melhoramento e Genética – TMG

Dr. Gustavo Hiroshi Sera
Instituto Agrônomo do Paraná – IAPAR

Dr^a. Mirian Perez Maluf
Empresa Brasileira de Pesquisa Agropecuária –
EMBRAPA

Londrina, 31 de janeiro de 2019.

“ Comece o dia na luz da oração.

O amor de Deus nunca falha.

Aceite qualquer dificuldade sem discutir.

Hoje é o tempo de fazer o melhor”.

Psicografia de Francisco Cândido Xavier

Dedico a Deus, que iluminou meu caminho durante essa caminhada.
Ao meu esposo Salvador pelo amor e dedicação,
Aos meus queridos pais Sônia e Salviano pelo carinho e por acreditarem em mim
E ao meu filho tão querido Miguel

AGRADECIMENTOS

Agradeço a Deus, pela força e coragem que me proporcionou durante toda essa longa caminhada e pelo filho maravilhoso que nos enviou. A toda a minha família, que com muito carinho e apoio, não mediram esforços para que eu chegasse até essa etapa de minha vida.

Ao meu pai, por todo investimento e por acreditar que era capaz.

A minha mãe, pela disponibilidade de tempo, pelo seu cuidado e carinho, não só por mim, mas pelo meu filho, por cuidar de nós. Sua presença e dedicação significou segurança para poder seguir em frente.

Ao meu marido amado, pessoa que compartilho a vida. Obrigada pelo carinho, pela paciência e por sua capacidade de me trazer a tranquilidade, a paz necessária para seguir a diante com confiança. E por cuidar tão bem do nosso filho amado, no momento em que mais precisava, por ser pai tão presente para o nosso pequeno.

Ao meu irmão pelo seu carinho e disponibilidade de tempo me auxiliando nas correções.

Ao meu orientador Dr. Luiz Filipe Protasio Pereira, pela orientação, pela paciência, ajuda, pelos conhecimentos transmitidos e por depositar confiança em meu trabalho desde o início. E por ser tão humano e compreensivo em entender como é a dinâmica de quem teve um filho.

A Dra. Brígida pela parceria e auxílio na realização da fenotipagem nos grãos do café.

Ao Dr. Gustavo Sant'ana pelo seu conhecimento, idéias e discussões sobre o trabalho.

A minha amiga Rafaelle por me ajudar em etapas práticas na condução do trabalho e por todo carinho e amizade.

A minha amiga Fernanda, por me auxiliar em etapas do trabalho e por sua amizade. A minha orientada Karla por todo auxílio nas etapas de coleta e extrações do trabalho. A todos os amigos do Laboratório de Biotecnologia Vegetal pelo incentivo e amizade.

A todos pela eterna amizade, pelos conselhos e por compartilhar comigo tristezas, ansiedades e acima de tudo alegrias.

Ao IAPAR pela oportunidade de estágio e a Universidade Estadual de Londrina por minha formação acadêmica.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES), Fundação Araucária, Consórcio de Pesquisa Café e CIRAD

Muito obrigada.

LISTA DE FIGURAS

CAPÍTULO 1

- Figure 1 -** Population structure among 189 genotypes of F₂ population of *Coffea arabica*. Evolution of Δk values (y-axis) according to the number of genetic groups (x-axis). Barplot of the estimated membership coefficient (Q) based on the 1.966 SNPs for k=2 and k=441
- Figure 2 -** Bar plot of the estimated membership coefficient (Q) of the 189 genotypes in a F₂ population based on 1966 SNPs for the k = 2 were used to inferred population structure of *C. Arabica*42
- Figure 3 -** Bar plot of the estimated membership coefficient (Q) of the 189 genotypes in a F₂ population based on 1966 SNPs for the k = 4 were used to inferred population structure of *C. Arabica*43
- Figure 4 -** Principal coordinate analysis (PcoA). The PcoA individuals are coloured according to the STRUCTURE groups using k = 4. M group are coloured in grey. (A) On Figure A 189 genotypes are represented and the M group is the admixture group. (B) Figure B represents genotypes with Q value above 0.6 and according to Bayesian inference of population structure, they were separated in groups, therefore the absence of M group46

CAPÍTULO 2

- Figura 1 -** Análise de componente principal dos compostos bioquímicos em 79 genótipos de uma população F₂ de *C. arabica*. AT: Açúcares totais; CF: compostos fenólicos; AR: açúcares redutores; ACTG: ácidos clorogênicos68
- Figura 2 -** Dendrograma com base na análise de agrupamento hierárquico para os componentes químicos de 79 genótipos provenientes de uma população F₂ de *C. arábica*69
- Figura 3 -** Distribuição do teor de ácidos clorogênicos nos 79 genótipos para os SNPs associados75
- Figura 4 -** Distribuição do teor de cafeestol nos 79 genótipos para os SNPs associados.....77
- Figura 5 -** Distribuição relacionado a largura da copa dos cafeeiros nos 79 genótipos para os SNPs associados..79
- Figura 6 -** Distribuição do teor de açúcares redutores nos 79 genótipos para os SNPs associados.....81

LISTA DE TABELAS

CAPÍTULO 1

Table 1 -	Genetic similarity of F ₂ genotypes with parents L ₁ C ₁ and L ₃ C ₃ calculated by DarWin Software. The number attributed to the similarity are expressed in percentage	47
Table 2 -	Differences in sample size (N), percentage of polymorphic loci (% P), allele number average (Na), Shannon's information index (I), expected heterozygosity (He), observed heterozygosity (Ho), unbiased expected heterozygosity (uHe) and number of private alleles among groups identified by the software STRUCTURE considering k=4	50
Table 3 -	Analysis of molecular variance (AMOVA) using 1.966 SNPs of the genetic differentiation among and within groups in a F ₂ population of <i>C. arabica</i>	50
Table 4 -	Pairwise <i>Fst</i> matrix obtained using molecular data	51
Table S1 -	Pairwise <i>Fst</i> matrix obtained using molecular data	51

CAPÍTULO 2

Tabela 1 -	Valores mínimo, máximo, média, desvio padrão e coeficiente de variação para os nove compostos químicos e para as análises morfológicas altura, largura e diâmetro do caule.....	65
Tabela 2 -	Correlação entre os componentes bioquímicos ácidos clorogênicos (ACG), açúcares totais (AT), cafeína, lipídeos, proteína, sacarose, compostos fenólicos (CF), açúcares redutores (AR), cafestol, cafeol, tamanho de grãos (pequeno, médio e grande) e caracteres morfológicos da planta nos genótipos de café da população F ₂ de <i>C. arabica</i>	66
Tabela 3 -	Classificação de agrupamento hierárquico dos grupos a partir de 79 genótipos de uma população F ₂ de <i>C. arabica</i>	70
Tabela 4 -	Valores médios para nove compostos do café divididos em três grupos pela análise de agrupamento hierárquico.	70
Tabela 5 -	Marcadores SNPs de <i>C. arabica</i> mapeados nos genomas de <i>C. canephora</i> (S_C) e <i>C. eugenioides</i> (S_E) associados aos compostos relacionados à qualidade da bebida de café usando diferentes métodos de associação.....	73
Tabela 1 -	Granulometria - tamanho de grão determinado pelo diâmetro das peneiras	91

Tabela 2 -	Conteúdo de ácidos nucleicos (ACGT), açúcares totais (AT), cafeína, lipídeos, proteína, sacarose, Composto fenólico, glucose ou açúcares redutores (AR), cafestol e caveol encontrados em 79 genótipos de uma população F ₂ de <i>C. arabica</i>	93
-------------------	---	----

BRITO, Livia Maria Nogueira. **Análise fenotípica e estudos genéticos em uma população F₂ de *Coffea arabica* L.** 2019. 95 f. Tese (Doutorado em Genética e Biologia Molecular) – Universidade Estadual de Londrina, Londrina, 2019.

RESUMO

O café é uma das principais bebidas no mundo e um importante *commodity* para países tropicais. *Coffea arabica* é uma das espécies mais importantes do gênero, conhecida pela produção de cafés com bebida de qualidade superior, aromas e sabores marcantes. Entretanto apresenta uma baixa variabilidade genética o que torna relevante a busca por novas fontes de variabilidade em populações de melhoramento através da incorporação de linhagens provenientes de híbridos interespecíficos e/ou uso de genótipos selvagens do centro de origem. Nesse trabalho realizamos análise genética e fenotípica de uma população F₂ de *Coffea arabica*, composta por 189 indivíduos, e obtida a partir de um cruzamento entre plantas IAPAR 78001-L1C1 [[E335] x Catuai Vermelho IAC 46)] e IAPAR 88480-8-L₃C₃ (Sarchimor IAC 1669-33). No capítulo 1, descrevemos a genotipagem por sequenciamento (GBS) e analisamos a diversidade e estrutura genética presente em uma população F₂ de *C. arabica*. Foram identificados 1.966 SNPs após aplicação de filtros para callrate (0,8) e MAF (>0,05), que se mostraram eficientes na análise da estrutura e diversidade genética presente na população. Pelo estudo de estrutura populacional e análise de coordenada principal (PCA), foram identificados de 2 a 4 grupos (k=2 e k=4) sendo possível verificar a variabilidade genética existente entre os indivíduos da população, além de um grupo Misto, denominado grupo M, que através das estimativas, demonstrou maior variabilidade genética presente no grupo, em comparação com os outros grupos formados. Índices de Shannon (0.52), heterozigosidade esperada (0.35), heterozigosidade observada (0.46) e polimorfismo (100%) confirmaram uma maior diversidade genética presente no grupo M. A análise molecular de variância (AMOVA) demonstrou maior variabilidade presente dentro dos grupos (89%) do que entre eles (11%). No capítulo 2, o objetivo do trabalho foi identificar SNPs associados a características bioquímicas relacionadas à qualidade da bebida na população. Foram utilizados 79 indivíduos da população F₂ para um estudo de associação genômica ampla (GWAS). A análise fenotípica foi realizada por meio de espectrometria de infra-vermelho próximo (NIRs) e os genótipos caracterizados para ácidos clorogênicos, açúcares totais, cafeína, compostos fenólicos, cafestol, caveol, proteínas, lipídeos, açúcares redutores e sacarose, além de características agrônomicas como altura da planta, largura de copa e tamanho de frutos. Foram observadas correlações positivas significativas entre cafeína e proteína e entre sacarose e açúcares totais. Análises de agrupamento hierárquico classificaram os genótipos em três grupos distintos, sendo o grupo número 3, evidenciado pelo baixo teor de cafestol e alto teor de caveol. Um total de 11 SNPs foram identificados no estudo de associação pelos métodos pLARmEB, ISIS EM BLASSO e MrMLM, sendo associados aos compostos: ácidos clorogênicos (3), açúcares redutores (1), cafestol (4) e largura da copa (3). Assim, através da fenotipagem e genotipagem da população, foi possível obter SNPs tanto para discriminação da população e estimar a variabilidade genética, bem como realizar estudos de associação relacionados a bioquímica dos grãos, permitindo a seleção de genótipos promissores para serem utilizados em programas de melhoramento de café.

Palavras chaves: Diversidade. Estrutura genética. Genotipagem por sequenciamento. GWAS. Café.

Suporte financeiro: Consórcio Pesquisa Café, CAPES, Fundação Araucária, Embrapa Café e CIRAD

BRITO, Livia Maria Nogueira. **Phenotypic analysis and genetic studies in a F₂ population of *Coffea arabica* L.** 2019. 95 p. Thesis (PhD in Genetics and Molecular Biology) – Universidade Estadual de Londrina, Londrina, 2019.

ABSTRACT

Coffee is one of the main beverages in the world and a major commodity for tropical countries. *Coffea arabica* is one of the most important species of the genus, known for production of superior quality coffees, with aroma and remarkable flavor. However, has a low genetic variability, making relevant the search to increase the variability of breeding populations through the incorporation of lines from interspecific hybrids and/or the use of wild genotypes from the origin center. In this work we performed genetic and phenotypic analysis in a F₂ population of *Coffea arabica*, composed by 189 individuals, obtained from a cross between IAPAR 78001-LiC₁ [[E335] x Catuai Vermelho IAC 46)] and IAPAR 88480-8-L₃C₃ (Sarchimor IAC 1669-33) plants. In Chapter 1, we describe genotyping by sequencing (GBS) and analysis of the diversity and genetic structure of the population. A total of 1,966 SNPs were identified after filters application of Callrate (0.8) and MAF (> 0.05), which were efficient in the analysis of structure and genetic diversity present in the population. Based on the population structure and principal component analysis (PCA), we identified 2 to 4 groups (k = 2 and k = 4) been possible verify the genetic variability among the individuals of the population, as well as a mixed group called group M, which, through estimates, showed a greater genetic variability present in the group compared to the other ones. Expected heterozygosity (0.35), observed heterozygosity (0.46), and polymorphism (100%) confirmed a greater genetic diversity present in group M. The analysis of molecular variance (AMOVA) showed greater variability within the groups (89%) than among them (11%). In Chapter 2, the objective of the study was to identify SNPs associated with biochemical characteristics related to beverage quality of the population. It was used 79 individuals from the population for the genome wide association study (GWAS). The phenotypic analyse was performed using near infrared spectroscopy (NIRs) and the compounds characterized for chlorogenic acids, total sugars, caffeine, phenolic compounds, cafestol, kawool, proteins, lipids, reducing sugars and sucrose, besides agronomic characteristics as height and width plants and fruit size. Significant positive correlations were observed between caffeine and protein and sucrose and total sugars. Hierarchical grouping analyzes classified the genotypes into three distinct groups, being group number 3, evidenced by the low content of cafestol and high content of caveol. A total of 11 SNPs was identified in the association study in the pLARM EB, ISIS BLASSO and MrMLM methods, being associated to the compounds: chlorogenic acids (3), reducing sugars (1), cafestol (4) and width of treetops (3). Thus, through the phenotyping and genotyping of the population, it was possible to obtain SNPs for population discrimination and to estimate the genetic variability, as well as perform association between genotypes and phenotype related to biochemistry of grain, allowing the selection of promising genotypes to be used in coffee breeding programs.

Keywords: Diversity. Genotyping by sequencing. GWAS. Coffee.

Financial support: Consórcio Pesquisa Café, CAPES, Fundação Araucária, Embrapa Café e CIRAD.

SUMÁRIO

1	INTRODUÇÃO	15
2	OBJETIVOS	16
3	REVISÃO BIBLIOGRÁFICA	17
3.1 A	Origem do café	17
3.2	Aspectos Econômicos.....	18
3.3	Compostos bioquímicos e morfológicos	19
3.4	Genotipagem por sequenciamento (GBS).....	22
3.5	Estudo de Associação Genômica Ampla (GWAS)	23
3.6	Referências	26
4	CAPÍTULO I: Population structure and genetic diversity of an F₂ population of <i>Coffea arabica</i>	35
4.1	ABSTRACT	35
4.2	INTRODUCTION.....	36
4.3	MATERIAL AND METHODS	38
4.3.1	Plant material and DNA extraction	38
4.3.2	GBS and SNP identification.....	38
4.3.3	Genetic analysis of population structure	39
4.4	RESULTS AND DISCUSSION	39
4.4.1	Identification and analyses of SNP markers.....	39
4.4.2	Population structure analysis.....	40
4.4.3	Analyses of genetic diversity	44
4.5	CONCLUSION	52
4.6	REFERENCES.....	53
4.7	SUPPLEMENTARY MATERIAL	57
5	CAPÍTULO II: Análise fenotípica e estudos de associação genômica ampla (GWAS) em uma população F₂ de <i>Coffea arabica</i> L.	59
5.1	RESUMO	59

5.2	INTRODUÇÃO.....	60
5.3	MATERIAL E MÉTODOS.....	61
5.3.1	Material Vegetal.....	61
5.3.2	Análise de caracteres morfológicos e agronômicos de uma população F ₂ de <i>Coffea arabica</i>	62
5.3.3	Análises Estatísticas dos dados bioquímicos e morfológicos	62
5.3.4	Genotipagem por sequenciamento	63
5.3.5	Estudos de Associação Genômica Ampla (GWAS) de uma população F ₂ de <i>Coffea arabica</i>	63
5.4	RESULTADOS E DISCUSSÃO	64
5.4.1	Fenotipagem para o tamanho dos grãos	64
5.4.2	Fenotipagem para o diâmetro de caule, altura e largura das plantas.....	65
5.4.3	Correlação entre as variáveis	66
5.4.4	Análise de Componente Principal (ACP) e Análise de Agrupamento Hierárquico (AAH).....	68
5.4.5	Genotipagem por sequenciamento (GBS) e detecção de SNPs	71
5.4.6	Estudos de Associação Genômica Ampla (GWAS).....	72
5.4.7	Distribuição dos teores fenotípicos para os SNPs associados.....	75
5.5	CONCLUSÃO.....	82
5.6	REFERÊNCIAS	83
5.7	MATERIAL SUPLEMENTAR	91
6	CONCLUSÃO GERAL	95

1 INTRODUÇÃO

Coffea arabica é uma das espécies mais importantes do gênero, conhecida pela produção de cafés com bebida de qualidade superior, aromas e sabores marcantes. Nos últimos anos, muitos estudos foram realizados com o objetivo de identificar materiais promissores com tolerância a estresses bióticos e abióticos. Entretanto o trabalho de melhoramento é dificultado devido a base genética estreita de *C. arabica*. Essa baixa diversidade genética em materiais melhorados ocorreu em razão de sua recente origem, sua biologia de reprodução e pela dispersão de um número limitado de plantas do seu centro de origem.

Marcadores moleculares possibilitam análises de diversidade genética, mapeamento genético, seleção assistida e também estudos de associação. Em *C. arabica*, com a utilização de polimorfismo de base única (SNPs – *Single Nucleotide Polymorphism*), essas análises vêm sendo facilitadas, devido a abundância e facilidade de genotipagem por SNPs em relação a outros marcadores. Os SNPs caracterizam-se pela sua ampla distribuição ao longo do genoma e facilidade de detecção como por exemplo tecnologias que utilizas sequenciadores de nova geração (NGS – *Next Generation Sequencing*), tais como a genotipagem por sequenciamento (GBS – *Genotyping by Sequencing*).

A disponibilidade de um maior número de marcadores permite uma melhor precisão na análise da diversidade genética existente dentro ou entre populações, possibilitando assim, uma maior precisão nas estimativas das diferenças genéticas entre os indivíduos da população e compreensão dos padrões de distribuição da diversidade e da evolução da espécie, bem como o reconhecimento de populações como fonte de variabilidade genética para fins de melhoramento de cafeeiro.

Técnicas capazes de identificar e quantificar compostos bioquímicos presentes nos grãos de café, também se tornaram cada vez mais precisas, como por exemplo, a espectrometria de infra-vermelho próximo (NIRs – *Near Infrared Spectroscopy*). Assim, através dos dados moleculares e uma fenotipagem precisa é possível realizar estudos de associação genômica ampla (GWAS) e obter informações do quanto um marcador molecular esta associado àquela característica fenotípica. O estudo se baseia no desequilíbrio de ligação visando a identificação de marcas relacionados a fenótipos relevantes e também a identificação de genes ou vias metabólicas envolvidas na manifestação fenotípica de interesse.

Nesse contexto, esse estudo visa uma maior compreensão da diversidade genética de uma população F₂ de *C. arabica* através dos SNPs gerados a partir da GBS e estudos de associação para identificar marcas associadas a características fenotípicas dessa população.

2 OBJETIVOS

Estimar a diversidade e estrutura de uma população F₂ de *C. arabica* e identificar SNPs associados a caracteres morfológicos associados a características agronômicas de interesse e bioquímicos associados a qualidade da bebida do café.

3 REVISÃO BIBLIOGRÁFICA

3.1 A origem do café

O café é uma das principais bebidas consumidas no mundo. Pertence a família *Rubiaceae* (BIKILA *et al.*, 2017), gênero *Coffea* compreende 124 espécies (DAVIS *et al.*, 2011; HAMON *et al.*, 2017). A produção mundial de café envolve predominantemente duas espécies: *Coffea arabica* (60%) e *Coffea canephora* (40%) (ICO, 2018). *C. canephora* é uma espécie diploide ($2n = 22$ cromossomos) e autoincompatível (BERTHAUD e CHARRIER, 1988), enquanto que *C. arabica* é a única espécie tetraploide ($2n = 44$ cromossomos) e predominantemente autógama, surgindo após a hibridação natural entre duas espécies diploides (*C. eugenoides* e *C. canephora*) (LASHERMES *et al.*, 1999).

O principal centro primário de diversidade genética de *C. arabica* se encontra no Sudoeste da Etiópia e no Planalto de Boma no Sudão, entretanto também foram relatadas no Monte Imatong (Sudão), Monte Masarbit (Quênia) (BERTHAUD e CHARRIER, 1988). Espalhou-se pela Europa, chegando ao Norte do Brasil através da Guiana Francesa, em Belém, no Pará, em 1727 e devido às condições climáticas favoráveis, se difundiu (OLIVEIRA *et al.*, 2012), tornando o país o maior produtor e exportador mundial de café (CONAB, 2017).

O café produzido por *C. arabica* é conhecido como de qualidade superior e seu aroma e sabor apreciados no mundo inteiro (OLIVEIRA *et al.*, 2012), além de acessos silvestres da Etiópia possuírem resistência à ferrugem para os fatores Sh1, Sh2, Sh4 e Sh5 (BITTENCOURT *et al.*, 1981) e resistência a mancha aureolada, causada por *Pseudomonas syringae pv garcae* (MOHAN *et al.*, 1978). *C. canephora* é uma planta alógama, com maior diversidade genética (LEROY *et al.*, 2014). É conhecida como fonte de genes de resistência incluindo ferrugem do cafeeiro (*Hemileia vastatrix*), antracnose dos frutos, causada por *Colletotrichum kahawae* e resistência a nematóide do gênero *Meloidogyne sp.* (LASHERMES *et al.*, 2000).

A espécie *C. arabica* é conhecida por apresentar base genética estreita, resultante da recém origem da espécie, por derivar de poucas plantas e pelo seu sistema de reprodução autógama (SETOTAW *et al.*, 2013). A baixa diversidade genética foi confirmada por estudos utilizando marcadores moleculares (LASHERMES *et al.*, 2011; FERRÃO *et al.*, 2015).

A fim de superar a baixa diversidade, programas de melhoramento de *C. arabica* trabalham com a introgressão de características agrônômicas de interesse de outras espécies do gênero (LASHERMES *et al.*, 2000). Híbridos naturais e artificiais originados do

cruzamento entre *C. arabica* e *C. canephora* vêm sendo utilizados, e um exemplo disso são Híbridos de Timor. Esses híbridos têm sido explorados por possuir genes de resistência à ferrugem derivados de *C. canephora*, fatores esses que não estão presentes em *C. arabica* (FAZUOLI *et al.*, 1996). A espécie *C. arabica* possui resistência a ferrugem apenas para o fator Sh 1 (MORAIS *et al.*, 1974; PETEK *et al.*, 2006; SHIGUEOKA *et al.*, 2014).

Institutos de pesquisa como o Instituto Agrônômico do Paraná (IAPAR) e o Instituto Agrônômico de Campinas (IAC) tem desenvolvido populações derivados de Híbrido de Timor como por exemplo o Tupi IAC 1669-33 e também de cultivares tradicionais, como os genótipos pertencentes ao grupo Catuaí (FERNANDES *et al.*, 2012; REZENDE *et al.*, 2014).

Tupi IAC 1669-33 é uma cultivar derivada do híbrido H361/4, que se originou de um cruzamento entre Villa Sarchí e Híbrido de Timor CIF 832/2 (HdT) (híbrido natural entre *C. canephora* e *C. arabica*) (FERNANDES *et al.*, 2012). O Híbrido de Timor é conhecido por ser fonte de resistência a *Hemileia vastatrix* (SERA *et al.*, 2007; DEL GROSSI *et al.*, 2013), resistência a *Meloidogyne exigua* e resistência intermediária a *M. paranaensis* (MUNIZ *et al.*, 2009) e *M. Incognita* (GONÇALVES e FERRAZ, 1987).

Já Catuaí se originou em 1949, a partir do cruzamento entre Caturra e Mundo Novo e destaca-se pelo seu rendimento e arquitetura das plantas (CARVALHO e MÔNACO, 1972). É conhecida por produzir plantas com alto vigor vegetativo, porte baixo e ampla adaptação ao meio ambiente (GUERREIRO FILHO *et al.*, 2018). Novas cultivares foram geradas, como é o caso do Catuaí Vermelho, plantas de frutos vermelhos, produtivas e homozigotas para os alelos Caturra CtCt. Ela está no mercado a mais de 40 anos e continua sendo usada como parental de novas linhas, uma vez que possui porte baixo e ramificações secundárias e terciárias muito fortes, sendo produzida em larga escala (CARVALHO e MÔNACO, 1972).

3.2 Aspectos econômicos

O café é um dos principais produtos da agricultura brasileira, onde no país encontrou condições climáticas favoráveis para seu desenvolvimento, desempenhando um papel fundamental no cenário econômico e social do país (CONSORCIO PESQUISA CAFÉ, 2019). Se tornou uma importante fonte de receita para centenas de municípios, além de ser responsável por gerar mais de 8 milhões de empregos na agropecuária nacional e proporcionar renda, acesso à saúde e à educação para os trabalhadores e suas famílias, conferindo ao produtor, sustentabilidade econômica (MAPA, 2018).

A produção mundial de café atingiu 160 milhões de sacas de 60kg no ano de 2017-2018 com um consumo de 159 milhões de sacas e crescimento de 1,2% em relação a safra

anterior (ICO, 2019). Dados do Departamento de Agricultura dos Estados Unidos (USDA, 2018) estimam uma produção mundial de café na safra de 2018-2019 de aproximadamente 174 milhões de sacas, sendo considerado volume de safra recorde, caso seja confirmado. Países que mais se destacaram no ano-safra de 2017-2018 pelo volume de produção foram: Brasil com 51 milhões de sacas, seguido pelo Vietnã com 29,5 milhões e Colômbia com 14 milhões de sacas (ICO, 2019).

De acordo com a CONAB (Companhia Nacional de Abastecimento) a safra brasileira de 2018 foi de 59,9 milhões de sacas sendo 45,94 de *C. arabica*, e 13,97 de *C. canephora*. Tal resultado gerou um crescimento de aproximadamente 35% em comparação ao ano de 2017. Esse crescimento foi decorrente da bienalidade da cultura, as condições favoráveis do clima e maiores investimentos nas lavouras e sobretudo a tecnologia disponível (CONAB, 2018). Os Estados que mais se destacaram foram Minas Gerais, que representou 53% da produção, Espírito Santo com 22%, seguidos de São Paulo 10%, Bahia com 8%, Rondônia com 4% e Paraná com 2% e 1 milhão de sacas (CONAB, 2018).

3.3 Compostos bioquímicos e morfológicos

O gênero *Coffea* abrange mais de 100 espécies e entre elas foram detectadas muitas variações em relação a composição química dos grãos e qualidade de bebida (CLIFFORD, 1985). As espécies *C. arabica* e *C. canephora* se destacam por possuir uma qualidade superior de bebida (NEBESNY e BUDRYN, 2006).

Diferenças em relação a composição química dos grãos tem sido reportado em diversos estudos e são associados a qualidade da bebida. Alguns fatores são determinantes na formação da composição química, como origem genética e fatores ambientais (BANDEIRA *et al.*, 2009; JOËT *et al.*, 2010, SCHOLZ *et al.*, 2011), bem como tipos e níveis de adubação e demanda hídrica (SILVA *et al.*, 2005). Os principais compostos relacionados à qualidade da bebida do café são carboidratos, proteínas, lipídeos, ácidos clorogênicos, cafeína e trigonelina (BANDEIRA *et al.*, 2009; SCHOLZ *et al.*, 2011).

Entretanto, poucos são os trabalhos disponíveis que buscam identificar marcadores relacionados com a produção desses compostos que estariam ligados na determinação da qualidade da bebida do café (LEROY *et al.*, 2011; AMIN *et al.*, 2012; HASMANN *et al.*, 2014; SANT'ANA *et al.*, 2018).

Os ácidos clorogênicos (ACG) são metabólitos secundários que após processo de torrefação dos grãos e degradação se transformam, originando compostos fenólicos que são responsáveis pelo amargor (CAMPA *et al.*, 2005) e a coloração dos grãos (FARAH e

DONANGELO, 2006). Sua concentração está associada a formação e na maturação dos grãos e o seu teor varia muito de acordo com a espécie (FARAH *et al.*, 2006). Dependendo da concentração, se muito alta, atua negativamente na qualidade da bebida (FARAH e DONANGELO, 2006).

Seu consumo tem sido associado à redução no risco de doenças cardiovasculares e diabetes do tipo II (BHANDARKAR *et al.*, 2019). Foram observados teores de ACG em genótipos de *C. arabica*, variando de 6 a 12g 100g⁻¹ (FARAH *et al.*, 2005). Genótipos derivados de Híbrido de Timor, tiveram valores mais elevados de ácidos clorogênicos, quando comparados com materiais que não tiveram introgressão de *C. canephora* (BERTRAND *et al.*, 2003).

Já teores de proteínas são variáveis e dependem das técnicas de manejo, tipos de fertilizantes que serão aplicados, assim como a concentração adotada (RODRIGUES *et al.*, 2010). De acordo com a literatura, para cafés tradicionais plantados no Brasil, tais como Bourbon e Catuaí, os teores de proteína variam de 14 a 17g 100g⁻¹ (SCHOLZ *et al.*, 2011) e em cultivares recentes, tais como IAPAR 59 e IPRs, variam de 16 a 18g 100g⁻¹ (KITZBERGER *et al.*, 2013).

Com relação aos açúcares, os principais observados no café são sacarose, frutose e glicose. São compostos precursores de aroma e sabor gerados durante a torrefação. Seus teores são associados a estágios de maturação dos grãos (VAAST, 2006), ou seja, ao longo do desenvolvimento do grão, os teores de sacarose aumentam, enquanto que os teores de glicose e frutose diminuem (VASCONCELOS *et al.*, 2007; KITZBERGER *et al.*, 2013). Podemos encontrar em *C. arabica*, teores de sacarose que giram em média de 9% do peso seco (KY *et al.*, 2001).

A cafeína foi o primeiro composto estudado em cafés e sua produção varia muito, pois sofre muita influência das variações ambientais e de manejo (JOËT *et al.*, 2010), além de variações decorrentes da espécie. Naturalmente, alguns materiais já possuem um teor de cafeína baixo, como é o caso dos acessos de *C. arabica* da Etiópia, quando comparados com teores produzidos pela espécie *C. canephora* (SILVAROLLA *et al.*, 2000; SCHOLZ *et al.*, 2016).

Genótipos de café brasileiro e café etiópes citados na literatura variam seu teor de cafeína de 1.000 a 1.500mg 100g⁻¹ (SILVAROLLA *et al.*, 2000; SILVAROLLA *et al.*, 2004; TESSEMA *et al.*, 2011). Estudos recentes em *C. arabica* obtiveram uma variação entre 1.285 a 1.521mg 100g⁻¹ em cultivares tradicionais e novas, enquanto que em híbridos e arábica puros seus teores variam de 1.05 a 1.530mg 100g⁻¹ (KITZBERGER *et al.*, 2013). Resultados

anteriores confirmam maiores concentrações de cafeína em cafés híbridos. Cruzamentos utilizando Híbrido de Timor e Villa Sarchi tiveram concentrações que variaram de 700 a 1.700 mg 100g⁻¹ (MAZZAFERA *et al.*, 1998).

Apesar dos benefícios da cafeína à saúde, por atuar como estimulante do sistema nervoso central e reduzir a fadiga (MAZZAFERA *et al.*, 1998; ALI *et al.*, 2012), a identificação de genótipos com baixo teor de cafeína (SILVAROLLA *et al.*, 2000) é interessante para programas de melhoramento, devido ao mercado específico de cafés descafeinados.

O café verde de *C. arabica* contém de 7-17% de frações lipídicas (SCHOLZ *et al.*, 2016). Essa fração de óleo é composta por esteróis, tocoferóis, ceramidas, entre outros compostos bioativos (ALVES *et al.*, 2009). Os lipídios são precursores de aroma e sabor e são associados a presença de espuma na bebida, devido a formação de compostos secundários, tais como aldeídos voláteis, cetonas e alcoois (BARBOSA *et al.*, 2016). Em diferentes cultivares de *C. canephora* foram encontrados valores que variaram entre 10.69 e 16.75g 100g⁻¹ (AGUIAR *et al.*, 2005), já em *C. arabica* os teores variam de 13,00 a 16,24g 100g⁻¹ em café verde (BARBOSA *et al.*, 2016).

Caveol e cafestol são diterpenos pentacíclicos que fazem parte da fração lipídica do café e ambos diferem somente por uma ligação dupla entre os carbonos 1 e 2, presentes no caveol, e podem ser diferenciados através de um pico de absorção no espectro de comprimento de onda (KITZBERGER *et al.*, 2013). Suas produções variam muito de acordo com a espécie e cultivar (KITZBERGER *et al.*, 2013) e vários são os efeitos fisiológicos atribuídos a essa fração insaponificável que ocorre naturalmente nas frações lipídicas do café (FERRARI *et al.*, 2010).

O efeito desses dois compostos na saúde humana tem sido objeto de ampla investigação (caveol e cafestol), e podem atuar como anti-inflamatórios, antioxidantes, anti-carcinogênicos e antidiabéticos (KIM *et al.*, 2004; ALVES *et al.*, 2009; CHU *et al.*, 2011, FUMIMOTO *et al.*, 2012). Entretanto, altos teores, especialmente de cafestol, estão relacionados ao aumento de colesterol no sangue quando a bebida não é filtrada, por conter um alto teor do composto, quando comparado com a bebida filtrada, em razão de grande parte ficar retida no filtro (RENDÓN *et al.*, 2017).

3.4 Genotipagem por sequenciamento (GBS)

Os avanços na tecnologia de sequenciamento, denominadas sequenciamento de nova geração (NGS – *Next Generation Sequencing*) promoveram o sequenciamento de DNA em plataformas capazes de sequenciar milhões de bases nucleotídicas em uma única análise (CARVALHO *et al.*, 2010, KIM *et al.*, 2016; BARABASCHI *et al.*, 2016).

Genotipagem via sequenciamento (GBS – *Genotyping by Sequencing*) é uma aplicação do NGS para descobrir e genotipar milhares de SNPs em genótipos de coleções ou populações (SPINDEL *et al.*, 2013). Ela é baseada no sequenciamento de regiões de representação reduzida do genoma de forma simultânea, permitindo uma genotipagem em larga escala (HAMON *et al.*, 2017). Fragmentos do genoma de diversos indivíduos são sequenciados de forma representativa e alinhados na busca por SNPs (ELSHIRE *et al.*, 2011; CAIXETA *et al.*, 2013; BHAT *et al.*, 2016).

Através do volume de informações geradas com a técnica, e uma genotipagem de forma representativa, a um custo reduzido por amostra, tornou-se possível a genotipagem de genomas complexos (ELSHIRE *et al.*, 2011), mesmo considerando sua extensão, (SPINDEL *et al.*, 2013) variações cromossômicas e níveis de ploidia (LU *et al.*, 2013). Essa técnica tem sido utilizada em diversas espécies (HAMON *et al.*, 2017; ELTAHER *et al.*, 2018), inclusive aquelas sem genoma de referência (CHARMETANT *et al.*, 2016; HAMON *et al.*, 2017).

A GBS inclui digestão do DNA genômico com enzimas de restrição sensíveis a regiões metiladas. Essas enzimas são usadas para reduzir a complexidade do genoma e evitar a clivagem de regiões de DNA (*Deoxynucleic Acid*) repetitivo. As amostras são identificadas através da ligação de adaptadores com códigos de barras (barcodes), ocorrendo na sequência, amplificação do DNA via PCR (*Polymerase Chain Reaction*) e sequenciamento do pool de amostras (HE *et al.*, 2014). O sistema multiplex utilizado nas amostras, permite a unificação de mais de uma amostra para a realização de PCR (SPINDEL *et al.*, 2013). Portanto, a técnica GBS é considerada de alto rendimento na qual uma de suas finalidades é explorar a diversidade genética em populações (ELSHIRE *et al.*, 2011).

Como toda técnica, existem limitantes. É fundamental a identificação e descarte de sequências repetidas, erros de alinhamento e montagem, segmentos não informativos do genoma (SPINDEL *et al.*, 2013), que comprometem a análise dos dados, podendo gerar falsos SNP's. Com a disponibilidade cada vez maior de dados gerados também são necessários “pipelines” de bioinformática capazes de analisar e interpretar conjuntos de dados GBS (HE *et al.*, 2014). Avanços no campo da bioinformática também foram surgindo, a fim de suprir essa necessidade. O programa TASSEL é um exemplo de ferramenta de bioinformática

robusta e eficiente, com o objetivo de processar dados de GBS, utilizando as sequências de todas as amostras na busca por SNPs (GLAUBITZ *et al.*, 2014).

A vantagem desse método de genotipagem além do baixo custo, é o fácil manuseio de amostras e menos etapas de purificação (DAVIS *et al.*, 2011). Tem sido utilizado com sucesso na implementação de estudos de associação genômica ampla (GWAS – *Genome Wide Association Studies*)), estudo de diversidade genética, análise de ligação genética, descoberta de marcadores moleculares e seleção genômica em programas de melhoramento genético. GBS tornou-se ao longo dos anos uma alternativa competitiva em comparação com outras plataformas de genotipagem genômica completa e vem sendo extensivamente aplicada a GWAS, Seleção Assistida por Marcadores (SAM) e seleção genômica (GS – *Genomic Selection*) (HE *et al.*, 2014).

3.5 Estudo de Associação Genômica Ampla (GWAS)

A fim de melhor compreender a natureza da relação entre regiões genômicas e fenótipos e com o objetivo de entender a arquitetura genética de um genótipo, surgiram os estudos de GWAS. Esse estudo tem proporcionado a identificação de marcas, e verifica se um conjunto de variantes genéticas, como por exemplo SNPs (*Single Nucleotide Polymorphism*), em um grupo de indivíduos estão associados a características fenotípicas (CHARMETANT *et al.*, 2016; KIM *et al.*, 2016).

A identificação de regiões cromossômicas associadas as variações fenotípicas (RESENDE *et al.*, 2013; CUBRY *et al.*, 2013) baseia-se nas técnicas de genotipagem em larga escala, fenotipagem precisa e desequilíbrio de ligação de populações (FERREIRA e RANGEL, 2011). Com a disponibilidade de milhares de SNPs, hoje gerados através de plataformas de NGS e a GBS, está mais rotineiro na identificação de genes e marcadores associados a característica de interesse (NORDBORG e WEIGEL 2008).

Essa relação entre genótipo e fenótipo sempre foi de grande interesse para estudiosos e foi em ervilhas que ocorreu o primeiro estudo relacionando características quantitativas (QTL – *Quantitative Trait Loci*) usando marcadores (SAX, 1923). Mapas de ligação convencional ou mapas de características quantitativas (QTL) são eficazes na identificação de genes associados a características de interesse, entretanto, os genes identificados normalmente são restritos àquelas linhagens endogâmica recombinantes (RIL – *Recombinant Inbred Lines*). Em comparação com o GWAS, a resolução é geralmente baixa devido ao número limitado de eventos de recombinação, que ocorrem durante a criação dos RILs (WANG *et al.*, 2015).

Mapas de QTL identificam blocos gerais para características que tem aplicações diretas dentro de uma cultura (ZHANG *et al.*, 2019) e sua aplicação é de forma generalizada, uma visão macro. Já o GWAS, identificam blocos de haplótipos associados a uma característica específica dentro de uma população e possui uma resolução mais refinada. Para identificar a arquitetura dos genes, o GWAS é o mais indicado (ZHANG *et al.*, 2019).

Estudos de GWAS em populações recombinantes (F_2) e acessos tem sido realizados (GUO *et al.*, 2017; LI *et al.*, 2019; ZHANG *et al.*, 2019), demonstrando a eficiência do estudo em explorar a máxima variabilidade genética, permitindo a detecção de SNPs altamente favoráveis para características de interesse, possibilitando a identificação de loci genômicos que poderiam ser usados na SAM (KORTE e FARLOW 2013).

GWAS possui uma abordagem eficiente para análise de características complexas (KORTE e FARLOW 2013) em comparação com outros métodos. É realizado com populações e a genotipagem é feita de forma representativa, sendo estatisticamente mais poderosa que outros mapas. Sendo assim, sua principal vantagem é explorar eventos de recombinação que ocorreram ao longo de toda a história evolutiva presente no material em estudo (NORDBORG e WEIGEL 2008). Os SNPs associados podem ser utilizados na SAM (GUPTA *et al.*, 2005) como estratégia para melhorar a qualidade bioquímica dos grãos (SANT'ANA *et al.*, 2018).

Para um estudo de associação, alguns aspectos devem ser estabelecidos previamente. Com a planta modelo, *Arabidopsis thaliana*, Korte e Farlow (2013), foi possível confirmar a necessidade de uma quantidade mínima de indivíduos para o estudo, que a heterogeneidade também deve ser considerada, para análise dos dados, programas estatísticos devem ser definidos para interpretação do conjunto de dados gerados, bem como cálculos de Desequilíbrio de Ligação (DL).

O desequilíbrio de ligação é necessário para a aplicação de GWAS, e pode ser observado em nível populacional, entre alelos dos marcadores e alelos dos genes que controlam o caráter na população em estudo (SVED e HILL, 2018). A identificação dessas variações na sequência de DNA é avaliada e associadas com variações nos fenótipos do caráter (RESENDE *et al.*, 2013).

O estudo possui algumas limitações, como a necessidade de genotipagem de um grande número de indivíduos, o que pode ser caro para grandes populações, apesar da redução de custos (TRAN *et al.*, 2018). Outro aspecto relevante é que a técnica muitas vezes gera falsos positivos devido a estrutura populacional (MARCHINI *et al.*, 2004).

A fim de controlar a estrutura populacional e reduzir a presença de SNP's falso positivos, um estudo de estrutura populacional se faz necessário, podendo ser realizada no STRUCTURE, como uma alternativa (PRITCHARD *et al.*, 2000). Dados da estrutura populacional baseados em marcadores selecionados aleatoriamente ou SNPs, são utilizados e incorporados na análise de associação (WANG *et al.*, 2015). E para capturar ainda mais a relação entre os indivíduos, é utilizado uma matriz kinship criada no TASSEL (WANG *et al.*, 2016).

GWAS já foi implementado com sucesso em diversas plantas como *Arabidopsis* (ATWELL *et al.*, 2010), milho (ZHU *et al.*, 2018), arroz (GONG *et al.*, 2017), trigo (ALOMARI *et al.*, 2018), soja (ZHANG *et al.*, 2018) e café (SANT'ANA *et al.*, 2018; TRAN *et al.*, 2018).

Estudos de associação genômica em *C. arabica* foram realizados recentemente por nosso grupo, que identificaram regiões genômicas associadas aos teores de lipídeos e diterpenos (SANT'ANA *et al.*, 2018). Esse foi o primeiro estudo em café correlacionando a variabilidade desses compostos com a diversidade de nucleotídeos, sendo encontrados 9 genes candidatos relacionados a via metabólica de compostos envolvidos na qualidade da bebida.

Análise de associação também foi realizada por Tran e colaboradores (2018) na busca por loci envolvidos na produção de cafeína e trigonelina em *C. arabica*. Utilizando os genomas de referência de *C. canephora* e *C. eugenioides*, eles identificaram 40 SNPs associados a cafeína, sendo 9 fortemente associados a genes que codificam enzimas que participam nas vias biossintéticas da cafeína e 24 associados a trigonelina, ligados a genes que codificam enzimas que participam das vias de biossíntese de trigonelina.

3.6 REFERÊNCIAS

- AGUIAR, A.T.E.; FAZUOLI, L.C.; SALVA, T.J.G.; FAVARIN, J.L. Variação no teor de lipídios em grãos de variedades de *Coffea canephora*. **Pesqui. Agropecu. Bras.**, v.40(12): 1251–1254, 2005.
- ALI, M. M.; EISA, M., TAHA, M.I., ZAKARIA, B.A. AND ELBASHIR, A.A. Determination of caffeine in some Sudanese beverages by High Performance Liquid Chromatography. **Pakistan Journal of Nutrition**, v. 11, n. 4, 336-342, 2012.
- ALOMARI, D.Z.; EGGERT, K.; WIRÉN, N.; POLLEY, A.; PLIESKE, J.; GANAL, M.W.; LIU, F.; PILLEN, K.; RÖDER, M.S. Whole-genome association mapping and genomic prediction for iron concentration in wheat grains. **Internation Journal of Molecular Sciences**, 2018.
- ALVES, R.; CASAL, S.; ALVES, M.; OLIVEIRA, M. Discrimination between arabica and robusta coffee species on the basis of their tocopherol profiles. **Food Chemistry**, v.114(1): 295–299, 2009.
- AMIN, N.; BYRNE, E.; JOHNSON, J.; CHENEVIX-TRENCH, G.; WALTER, S.; NOLTE, I. M.; KEERS, J.C. Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM. **Molecular psychiatry**, v.17(11): 1116, 2012.
- ATWELL, S.; HUANG, Y.S.; VILHJALMSSON, B.J.; WILLEMS, G.; HORTON, M.; LI, Y.; MENG, D.; PLATT, A.; TARONE, A.M.; HU, T.T. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. **Nature**, 465(7298):627 – 631, 2010.
- BANDEIRA, R.D.C.C.; TOCI, A.T.; TRUGO, L.C.; FARAH, A. Composição volátil dos defeitos intrínsecos do café por CG, EM-headspace. **Química Nova**, v.32(2): 309-314, 2009.
- BARABASCHI, D.; TONDELLI, A.; DESIDERIO, F.; VOLANTE, A.; VACCINO, P.; VALÈ, G.; CATTIVELLI, L. Next generation breeding. **Plant Science**, v.243:3-13, 2016.
- BARBOSA, M.S.G.; SCHOLZ, M.B.S.; KITZBERGER, C.S.G.; BENASSI, M.T. Teores de proteínas e lipídeos em cafés *Coffea arabica* procedentes de concurso de qualidade. 6º Congresso de Ciências Farmacêuticas do Mercosul, Nov.2016.
- BERTRAND, J.; CHARRIER, A. Genetic resources of *Coffea*. In: CLARKE, R.J.; MACRAE, R.; (Eds.). *Coffee Agronomy*, London: **Elshire Applied Science**, v.4:1-42, 1988.
- BERTRAND, B.; GUYOT, B.; ANTHONY, F.; LASHERMES, P. Impact of the *Coffea canephora* gene introgression on beverage quality of *C. arabica*. *Theoretical and Applied Genetics*, v.107, 387-394, 2003.
- BETTENCOURT, A. J. Melhoramento genético do cafeeiro: transferência de factores de resistência à Hemileia vastatrix Berk. and Br. para as principais cultivares de *Coffea arabica* L. Lisboa: Junta de investigações científicas do ULTRAMAR - **Centro de Investigação das Ferrugens do cafeeiro**, Oeiras, 93, 1981.

BHANDARKAR, N. S.; BROWN, L.; PANCHAL, S. K. Chlorogenic acid attenuates high-carbohydrate, high-fat diet-induced cardiovascular, liver, and metabolic changes in rats. **Nutrition Research**, v.62: 78-88, 2019.

BHAT, J.A.; ALI, S.; SALGOTRA, R.K.; MIR, Z.A.; DUTTA, S.; JADON, V.; TYAGI, A.; MUSHTAQ, M.; JAIN, N. ; SINGH, P.K.; SINGH, G.P.; PRABHU, K.V. Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding., **Frontiers in Genetics**. v.7, 221p., 2016

BIKILA, B.A.; SAKIYAMA, N.S.; CAIXETA, E.T.; SNPs based molecular diversity of *Coffea canephora*. **Journal of Microbiology and experimentation**, v.5:1, 2017.

CAIXETA, E.T.; FERRÃO, L.F.V.; ZAMBOLIM, E.M. Marcadores Moleculares. 2: 31-68 In: BORÉM, A.; FRITSCH NETO, R. **Biotecnologia aplicada ao melhoramento de plantas**. Viçosa: Universidade Federal de Viçosa, p.335, 2013.

CAMPA C, DOULBEAU S, DUSSERT S, HAMON S, NOIROT M Qualitative relationship between caffeine and chlorogenic acid contents among wild *Coffea* species. **Food chemistry**, v.93(1):135-139, 2005.

CARVALHO, A.; MÔNACO, L.C. Transferência do fator caturra para o cultivar Mundo Novo de *Coffea arabica*. **Bregantia**, v.31:379-399, 1972.

CARVALHO, M.C.C.G; SILVA, D.C.G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, Santa Maria, v.40(3): 735-744, 2010.

CHARMETANT, P.; LEROY, T. Genome-wide association studies in *Coffea arabica* using GBS. In: **The 26th International Conference on Coffee Science – ASIC**. Kunming, China, 2016.

CHU, Y.F.; CHEN, Y.; BLACK, R.M.; BROWN, P.H.; LYLE, B.J.; LIU, R.H.; OU, B. Type 2 diabetes-related bioactivities of coffee: assessment of antioxidant activity, NF- κ B inhibition, and stimulation of glucose uptake. **Food chemistry**, v.124(3): 914-920, 2011.

CLIFFORD, M.N. CHEMICAL AND PHYSICAL ASPECTS OF GREEN COFFEE AND COFFEE PRODUCTS. In: Clifford MN, Willson KC (eds), *Coffee: botany, biochemistry and production of beans and beverage*. Avi Publishing Company, Westport, Connecticut, USA, v.40(4): 305-374, 1985.

CONAB - Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira-Café**. v.5, safra 2018, n.4 – Quarto levantamento, 2018.

CONAB – Companhia Nacional de Abastecimento. **Compêndia de Estudos Conab. A cultura do café: Análise dos custos de produção e da rentabilidade nos anos 2008 a 2017**. v.12, 2017.

CONSORCIO PESQUISA CAFÉ. Disponível em:
<<http://www.consorciopesquisacafe.com.br/index.php/consorcio/separador2/historico>>
Acesso em: 09 Jan. 2019.

CUBRY, P.; BELLIS, F.; AVIA, K.; BOUCHET, S.; POT, D.; DUFOUR, M.; LEGNATE, H.; LEROY, T. An initial assessment of linkage disequilibrium (LD) in coffee trees: LD patterns in groups of *Coffea canephora* Pierre using microsatellite analysis. **BMC Genomics**, 16;14:10, 2013.

DAVIS, A.P.; TOSH, J.; RUCH, N.; FAY, M.F. Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. **Botanical Journal of the Linnean Society**, v.167(4): 357-377, 2011.

DEL GROSSI, L.; SERA, T.; SERA, G. H.; FONSECA, I.C.D.B.; ITO, D.S.; SHIGUEOKA, L. H.; CARVALHO, F.G. Rust resistance in Arabic Coffee cultivars in northern Paraná. **Brazilian Archives of Biology and Technology**, v.56(1): 27-33, 2013.

ELSHIRE, R. J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K.; BUCKER, E.S.; MITCHELL, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS One** 6, 2011.

ELTAHER, S.; SALLAN, A., BELAMKAR, V.; EMARA, H.A.; NOWER, A.A.; SALEM, K.F.M.; POLAND, J.; BAENZIGER, P.S. Genetic diversity and population structure of F3:6 Nebraska winter wheat genotypes using genotyping by sequencing. **Frontiers in Genetics**, v.9 (76), 2018.

FARAH, A.; DONANGELO, C.M. Phenolic compounds in coffee. **Brazilian Journal of Plant Physiology**, v.18, 23-26, 2006.

FARAH, A.; MONTEIRO, M.C.; CALADO, V.; FRANCA, A.S.; TRUGO, L.C. Correlation between cup quality and chemical attributes of Brazilian coffee. *Food Chemistry*, 98, 373-380, 2006.

FARAH, A.F.; DE PAULIS, T.; TRUGO, L.C.; MARTIN, P.R. Effect of roasting on the formation and degradation of chlorogenic acid lactones in roasted coffee. **Journal of Agricultural and Food Chemistry**, v.53(5): 1505-1513, 2005.

FAZUOLI, L.C.; MEDINA, H.P.; GERREIRO FILHO, O.; LIMA, M.M.A.; SILVAROLLA, M.B.; GALLO, P.B.; COSTA, W.M. Obatã (IAC 1669-20) e Tupi (IAC 1669-33), cultivares de café de porte baixo e resistentes a ferrugem. **22º Congresso Brasileiro de Pesquisas cafeiras**, Águas de Lindóia, SP, 149-150, 1996.

FERNANDES, A.L.T.; PARTELLI, F.L.; BONOMO, R.; GOLYNSKI, A. A moderna cafeicultura dos cerrados brasileiros. **Pesquisa Agropecuária Tropical**, v.42, 231-240, 2012.
FERRÃO, L.F.V.; CAIXETA, E.T.; PENA, G.; ZAMBOLIM, E.M.; CRUZ, C.D.; ZAMBOLIM, L.; FERRAO, M.A.G.; SAKIYAMA, N.S. New EST-SSR markers of *Coffea arabica*: transferability and application to studies of molecular characterization and genetic mapping. **Mol Breed** 35(1):1-5, 2015.

FERRARI, M.; RAVERA, F.; DE ANGELIS, E.; LIVERANI, F. S.; NAVARINI, L. Interfacial properties of coffee oil. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, v.365(1–3): 79–82, 2010.

FERREIRA, M.E.; RANGEL, P.H.N. Aporte biotecnológico ao pré-melhoramento vegetal. 3: 61-84 In: LOPES, M.A.; FAVERO, A.P.; FERREIRA, M.A.J. F.; FALEIRO, F.G.; FOLLE, S.M.; GUIMARÃES, E.P. **Pré-melhoramento de plantas: estado da arte e experiências de sucesso**. Brasília, DF: Embrapa Informação tecnológica, p.614, 2011.

FUMIMOTO, R.; SAKAI, E.; YAMAGUCHI, Y.; SAKAMOTO, H.; FUKUMA, Y.; NISHISHITA, K.; TSUKUBA, T. The coffee Diterpene Kahweol prevents Osteoclastogenesis via impairment of NFATc1 expression and blocking of Erk phosphorylation. **Journal of Pharmacological Sciences**, v.118(4): 479–486, 2012.

GLAUBITZ, J.C.; CASSTEVENS, T.M.; LU, F.; HARRIMAN, J.; ELSHIRE, R.J.; SUN, Q.; BUCKLER, E.S. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. **Plos one**, v.9, 2014.

GONÇALVEZ, W.; FERRAZ, L.C.C.B. Resistência do cafeeiro a nematoídes. II Testes de progênies e híbridos para *Meloidogyne incognita*/raça 3. **Nematologia Brasileira (Brasil)**, v. 11:125-142, 1987.

GONG, J.; MIAO, J.; ZHAO, Y.; ZHAO, Q.; FENG, Q.; ZHAN, Q.; CHENG, B.; XIA, J.; HUANG, X.; YANG, S.; HAN, B. Dissecting genetic basis of grain shape and chalkiness traits in hybrid rice using multiple collaborative populations. *Mol. Plant*, v.10 (10): 1353–1356, 2017.

GUERREIRO FILHO, O.; RAMALHO, M.A.P.; ANDRADE, V.T. Alcides Carvalho and the selection of Catuaí cultivar: interpreting the past and drawing lessons for the future. **Crop Breeding and Applied Biotechnology**, v.18(4): 460-466, 2018.

GUO, Y.; HUANG, Y.; HOU, L.; MA, J.; CHEN, C.; AI, H.; HUANG, L.; REN, J. Genome-wide detection of genetic markers associated with growth and fatness in four pig populations using four approaches. **Genetics Selection Evolution**, v.49(1): 21, 2017.

GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. Linkage disequilibrium and association studies in higher plants: present status and future prospects. **Plant. Mol. Biol**, v.57: 461–485 , 2005.

HAMON, P.; GROVER, C.E.; DAVIS, A.P.; RAKOTOMALALA, J.J.; RAHARIMALALA, N.E.; ALBERT, V.A.; SREENATH, H.L.; STOFFELEN, P.; MITCHELL, S.E.; COUTURON, E.; HAMON, S.; KOCHIKO, A.; CROUZILLAT, D.; RIGOREAU, M.; SUMIRAT, U.; AKAFFOU, S.; GUYOT, R. Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species: GBS coffee phylogeny and the evolution of caffeine content. *Mol. Phylogenet. Evol.*, April, 109:351-361, 2017.

HASSMANN, U.; HAUPT, L. M.; SMITH, R. A.; WINKLER, S.; BYTOF, G.; LANTZ, I.; MARKO, D. Potential antioxidant response to coffee—A matter of genotype?. **Meta gene 2**, 525-539, 2014.

HE, J.; ZHAO, X.; LAROCHE, A.; LU, Z.X.; LIU, H.; LI, Z. Genotyping by sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. **Front. Plant Sci.**, v.5, 484, 2014.

ICO - INTERNATIONAL COFFEE ORGANIZATION. Disponível em: <<http://www.ico.org/prices/po-production.pdf>> Acesso em: 26 Jan. 2018.

ICO - INTERNATIONAL COFFEE ORGANIZATION. Disponível em: <<http://www.ico.org/prices/po-production.pdf>> Acesso em: 09 Jan. 2019.

JOËT, T.; LAFFARGUE, A.; DESCROIX, F.; DOULBEAU, S.; BERTRAND, B.; DUSSERT, S. Influence of environmental factors, wet processing and their interactions on the biochemical composition of green Arabica coffee beans. **Food chemistry**, 118(3): 693-701, 2010.

KIM, C.; GUO, H.; KONG, W.; CHANDNANI, R.; SHUANG, L.S.; PATERSON, A.H. Application of genotyping by sequencing technology to a variety of crop breeding programs. **Plant Science**, 242, 14-22, 2016.

KIM, J.Y.; JUNG, K. S.; LEE, K. J.; NA, H. K.; CHUN, H.K.; KHO, Y. H.; & JEONG, H.G. The coffee diterpene kahweol suppress the inducible nitric oxide synthase expression in macrophages. **Cancer Letters**, 213(2): 147–154, 2004.

KIM, N.; JEONG, Y.M.; JEONG, S.; KIM, G.B.; BAEK, S.; KWON, Y.E. Identification of candidate domestication regions in the radish genome based on high-depth resequencing analysis of 17 genotypes. **Theor Appl Genet.**129: 1797–1814, 2016.

KITZBERGER, C.S.G.; SCHOLZ, M.B. dos S.; PEREIRA, L.F.P.; BENASSI, M.T. Composição química de cafés arábica de cultivares tradicionais e modernas. **Pesq. agropec. bras.**, Brasília, v.48, n.11, 1498-1506, 2013.

KORTE, A.; FARLOW, A. The advantages and limitations of trait analysis with GWAS: a review. **Plant Methods**, v.9 (1):29, 2013.

KY, C.L.; LOUARN, J.; DUSSERT, S.; GUYOT, B.; HAMON, S.; NOIROT, M. Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. **Food chemistry**, v.75(2): 223-230, 2001.

LASHERMES, P.; COMBES, M. C.; ANSALDI, C.; GICHURU, E.; NOIR, S. Analysis of alien introgression in coffee tree (*Coffea arabica* L.). **Molecular breeding**, 27(2), 223-232, 2011.

LASHERMES, P.; COMBES, M.C.; ROBERT, J.; TROUSLOT, P.; D'HONT, A.; ANTHONY, F.; CHARRIER, A. Molecular characterization and origin of the *Coffea arabica* L. **Genome. Mol.Genet.**, v.261, 259-266, 1999.

LASHERMES, P.; ANDRZEJEWSKI, S.; BERTRAND, B.; COMBES, M.C.; DUSSERT, S.; GRAZIOSI, G.; TROUSLOT, P.; ANTHONY, F. Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). **Theoretical and applied genetics**, v.100(1):139-146, 2000.

- LEROY, T.; BELLIS, F.D.; LEGNATE, H. ; KANANURA, E. ; GONZALES, G.; PEREIRA, L.F.; ANDRADE, A.C.; CHARMETANT, P.; MONTAGNON, C.; CUBRY, P.; MARRACCINI, P.; POT, D.; KOCHKO, A. Improving the quality of African Robustas: QTLs for yield and quality related traits in *coffea canephora*. **Tree genetics and genomes**, v.7:781-798, 2011.
- LEROY, T.; DE BELLIS, F.; LEGNATE, H.; MUSOLI, P.; KALONJI, A.; SOLÓRZANO, R.G.L.; CUBRY, P. Developing core collections to optimize the management and the exploitation of diversity of the coffee *Coffea canephora*. **Genetica**, v.142(3): 185-199, 2014.
- LI, D.; ZHAO, X.; HAN, Y.; LI, W.; XIE, F. Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. **Genomics**, v.111(1): 90-95, 2019.
- LU, F.; LIPKA, A.E.; GLAUBITZ1, J.; ELSHIRE, R.; CHERNEY, J.H.; CASLER, M.D.; BUCKLER, E.S.; COSTICH, D.E. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery. **Plos one**, 9(1): p.1-14, 2013.
- MAPA– Ministério da Agricultura, Pecuária e Abastecimento, Versão 09.2018. <<http://www.agricultura.gov.br/assuntos/politica-agricola/cafe/cafeicultura-brasileira>> Acesso em: 30, janeiro, 2019.
- MARCHINI, J.; GARDON, L.; PHILLIPS, M.; DONNELLY, P. The effects of human population structure on large genetic association studies. **Nature Genetics**, v.36: 512-517, 2004.
- MAZZAFERA, P.; SOAVE, D.; ZULLO, M.A.T.; GUERREIRO FILHO, O. Oil content of green coffee beans from some coffee species. **Bragantia**, v.57, 45-48, 1998.
- MOHAN, S.K.; CARDOSO, R.L.; PAVAN, M.A. Resistência em germoplasma de *Coffea* ao crestamento bacteriano incitado por *Pseudomonas garcae* Amaral et al. **Pesquisa Agropecuária Brasileira**, v.13: 53-64, 1978.
- MORAES, S.A.; SUGIMORE, M.H.; TOMAZELLO-FILHO, M.; CARVALHO, P. C. T. Resistência de cafeeiros à *Pseudomonas garcae* In: CONGRESSO BRASILEIRO DE PESQUISAS CAFEEIRAS, Poços de Caldas (MG). **Resumos**, Rio de Janeiro, p.183, 1974.
- MUNIZ, M. D. F. S.; CAMPOS, V. P.; MOITA, A. W.; GONÇALVES, W.; ALMEIDA, M. R. A.; SOUSA, F. R. D.; CARNEIRO, R. M. D. Reaction of coffee genotypes to different populations of *Meloidogyne* spp.: detection of a naturally virulent *M. exigua* population. **Tropical Plant Pathology**, v.34(6): 370-378, 2009.
- NEBESNY, E.; BUDRYN, G. Evaluation of sensory attributes of coffee brews from robusta coffee roasted under different conditions. **European Food Research Technology**, v. 224(2):159-165, 2006.
- NORDBORG, M.; WEIGEL, D. Next-generation genetics in plants. **Nature**, 456, 720-723, 2008.

OLIVEIRA, I.P.; OLIVEIRA, L.C.; MOURA, C.S.F.T. Cultura do café: histórico, classificação botânica e fases de crescimento. **Revista Faculdade Montes Belos**, v.5, n.4, 2012.

PEREIRA, T. B.; SETOTAW, T. A.; SANTOS, D. N.,; MENDES, A. N. G.; SALGADO, S. M. L.; CARVALHO, G. R.; REZENDE, R. M. Identification of microsatellite markers in coffee associated with resistance to *Meloidogyne exigua*. **Genetics and Molecular Research**, 15(3): 2-3, 2016.

PETEK, M.R.; SERA, T.; SERA, G.H.; FONSECA, I.C.D.B.; ITO, D.S. Seleção de progênies de *Coffea arabica* com resistência simultânea à mancha aureolada e à ferrugem alaranjada, 2006. *Bragantia*, v.65(1), 2006.

PRITCHARD, J.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, 155(2): 945-959, 2000.

RENDÓN, M. Y.; SANTOS SCHOLZ, M. B.; BRAGAGNOLO, N. Is cafestol retained on the paper filter in the preparation of filter coffee? **Food Research International**, v.100, 798–803, 2017.

RESENDE, M.D.V.; SILVA, F.F.; RESENDE JUNIOR, M.F.R. Genética de Associação (GWS). 4: 119-150. In: BORÉM, A.; FRITSCH NETO, R. **Biotecnologia aplicada ao melhoramento de plantas**, Viçosa: Universidade Federal de Viçosa, 335p, 2013.

REZENDE, R.M.; CARVALHO, G.R.; REZENDE, J.C.; BOTELHO, C.E.; PASQUALOTTO, A.T.; VELOSSO, J.A.; CARVALHO, A.M. Agronomic traits of coffee tree progenies from Timor Hybrid x Catuai crossing. **Ciência Rural**, Santa Maria, v.44, n.5, 775-780, 2014.

RODRIGUES, C.I.; MAIA, R.; MÁGUAS, C. Comparing total nitrogen and crude protein content of green coffee beans (*Coffea* spp.) from different geographical origins. **Coffee Sci.**, v.5(3): 197–205, 2010.

SANT'ANA, G.C.; PEREIRA, L.P.P; POT, D.; IVAMOTO, S.T.; DOMINGUES, D.S.; FERREIRA, R.V.; PAGIATTO, N.F.; SILVA, B.S.R.; NOGUEIRA, L.M.; KITZBERGER, C.S.G; SCHOLZ, M.B.S.; OLIVEIRA, F.F.; SERA, G.H.; PADILHA, L.; LABOUISSSE, J.P.; GUYOT, R.; CHARMETANT, P. LEROY, T. Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. **Scientific Reports**, v.8(1):465, 2018.

SAX, K. The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. **Genetics**, 8, 552–560, 1923.

SCHOLZ, M.B.S.; KITZBERGER, C.S.G.; PAGIATTO, N.F.; PEREIRA, L.F.P.; DAVRIEUX, F.; POT, D.; CHARMETANT, P.; LEROY, T. Chemical composition in wild ethiopian Arabica coffee accessions. **Euphytica**, v. 209(2):429-438, 2016.

SCHOLZ, M.B.S.; FIGUEIREDO, V.R.G.D., SILVA, J.V.N.D. AND KITZBERGER, C.S.G. Características físico-químicas de grãos verdes e torrados de cultivares de café (*Coffea arabica* L.) do IAPAR. *Coffee Science*, v.6(3): 245-255, 2011.

SERA, G. H.; SERA, T.; ITO, D. S.; AZEVEDO, J.A.D.; MATA, J.S.D.; RIBEIRO FILHO, C. Selection for durable resistance to leaf rust using test-crosses on IAPAR-59 and Tupi IAC 1669-33 cultivars of *Coffea arabica*. **Brazilian Archives of Biology and Technology**, v.50(4): 565-570, 2007.

SETOTAW, T.A.; CAIXETA, E.T.; PEREIRA, A.A.; OLIVEIRA, A.C.B.; CRUZ, C.D.; ZAMBOLIM, E.M.; ZAMBOLIM, L.; SAKIYAMA, N.S. Coefficient of parentage in *Coffea arabica* L. cultivars grown in Brazil. **Crop Sci**, v.53:1237– 1247, 2013.

SHIGUEOKA, L.H.; SERA, G.H.; SERA, T. ; FONSECA, I.C.D.B.; MARIUCCI JUNIOR, V.; ANDREAZI, E.; CARVALHO, F.G.; GARDIANO, C.G.; CARDUCCI, F.C. Selection of Arabic coffee progenies with rust resistance. **Crop Breeding and Applied Biotechnology**, v.14(2): 88-93, 2014.

SILVA, E.A. da; MAZZAFERA, P.; BRUNINI, O.; SAKAI, E.; ARRUDA, F.B.; M ATTOSO, L.H.C.; CARVALHO, C.R.L.; PIRES, R.C.M. The influence of water management and environmental conditions on the chemical composition and beverage quality of coffee beans. **Brazilian Journal of Plant Physiology**, v.17, 229-238, 2005.

SILVAROLLA, M.B.; MAZZAFERA, P.; FAZUOLI, L.C. Plant biochemistry: a naturally decaffeinated Arabica coffee. **Nature**, 429-826, 2004.

SILVAROLLA, M.B.; MAZZAFERA, P.; LIMA, M.M.A. Caffeine content of Ethiopian coffee arabica beans. **Genet Mol Biol.**, v.23:213–215, 2000.

SPINDEL, J.; WRIGHT, M.; CHEN, C.; COBB, J.; GAGE, J.; HARRINGTON, S.; LORIEUX, M.; AHMADI, N.; MCCOUCH, S. Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. **Theor Appl Genet**, v.126 (11):2699-2716, 2013.

SVED, J.A.; HILL, W.G. One hundred years of linkage disequilibrium. **Genetics**, v.209:629-636, 2018.

TESSEMA, A.; ALAMEREW, S.; KUFA, T.; GAREDEW, W. Variability and association of quality biochemical attributes in some promising *Coffea arabica* germplasm collections in Southwestern Ethiopia. **Int J Plant Breed Genet.**, v.5:302–316, 2011.

TRAN, H.T.M.; FURTADO, A.; VARGAS, C.A.C.; SMYTH, H.; LEE, L.S.; HENRY, R. SNP in *Coffea arabica* genome associated with coffee quality. **Tree Genetics and Genomes**, 14:72, 2018.

USDA – UNITED STATES DEPARTMENT OF AGRICULTURE. **Coffee: world markets and trade**. World production, markets and trade reports, December 14, 2018.

VAAST, P.; BERTRAND, B.; PERRIOT, J.J.; GUYOT, B.; GÉNARD, M. Fruit thinning and shade improve bean characteristics and beverage quality of coffee (*Coffea arabica* L.) under optimal conditions. **Journal of the Science of Food and Agriculture**, v.86, 197-204, 2006.

VASCONCELOS, A.L.; FRANÇA, A.S.; GLÓRIA, M.B.A.; MENDONÇA, J.E.F. A comparative study of chemical attributes and levels of amines in defective green and roasted coffee bean. **S. Food Chem.**, v.100(1): 26–32, 2007.

WANG, L.; WARE, D.; LUSHBOUGH, C.; MERCHANT, N.; STEIN, L. A genome-wide association study platform built on iPlant cyber-infrastructure. **Concurrency Computat.: Pract. Exper.**, v.27:420-432, 2015.

WANG, S.; FENG, J.Y.; REN, W.L.; NI, J.L.; WEN, Y.J.; HUANG, B.; ZHOU, L.; WEN, Y.J.; ZHANG, J.; DUMWELL, J.M.; XU, S.; ZHANG, Y.M. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. **Scientific Reports**, 6, 2016.

ZHANG, C.; LI, L.; LIU, Q.; GU, L.; HUANG, J.; WEI, H.; WANG, H.; YU, S. Identification of loci and candidate genes responsible for fiber length in upland cotton (*Gossypium hirsutum* L.) via association mapping and linkage analyses. **Frontiers in Plant Science**, v.10: 53, 2019.

ZHANG, J.; WANG, X.; LU, Y.; BHUSAL, S.J.; SONG, Q.; CREGAN, P.B.; YEN, Y.; BROWN, M.; JIANG, G.L. Genome wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. **Molecular plant**, v.11(3):460-472, 2018.

ZHU, X.M.; SHAO, X.Y.; PEI, Y.H.; GUO, X.M.; LI, J.; SONG, X.Y.; ZHAO, M.A. Genetic diversity and genome wide association study of major ear quantitative traits using high density SNPs in maize. **Front Plant Sci**, v.9:966, 2018.

4 CAPÍTULO I: Population structure and genetic diversity of F₂ population of *Coffea arabica* L.

4.1 ABSTRACT

Arabica coffee (*Coffea arabica* L.) is one of the most important commodities in the world. It is an allotetraploid with low genetic diversity, the main concern for coffee breeding programs. The present study was carried out to investigate the genetic diversity of a F₂ population of *C. arabica*. The F₂ population was obtained from self-pollinated F₁ plants from an initial cross between IAPAR 78001-L₁C₁ [(E335) x Catuai Vermelho IAC 46)] and IAPAR 88480-8-L₃C₃ (Sarchimor IAC 1669-33) plants. The genetic diversity was studied in 189 genotypes (F₂ population plus two parents) using SNP markers originated by GBS (genotyping by sequencing) technique. After removing SNPs due to missing data, low minor allele and heterozygosity (MAF > 0.05, call rate > 0.8 and Ho < 0.9), 1.966 SNP were selected from the coffee population and used to analyze the population structure and genetic diversity. The analysis of population structure showed the presence of two to four groups (K = 2 and K = 4). The K = 2 is related to parents composition and K = 4 was observed from the genetic variability between individuals. The principal coordinate analysis (PcoA) was similar to STRUCTURE, suggesting four groups and the M group (representing a mixed group), which indicate a higher genetic diversity when compared to the other groups. Analysis of variance identified higher genetic variability within groups 89% than among groups 11%. The diversity indices also showed the high genetic diversity for the M group with Shanno's index (I) = 0.52, observed heterozygosity (Ho) = 0.46 and unbiased expected heterozygosity (uHe) = 0.35. The number of SNP markers was efficient to discriminate the population. It was possible to observe the effect of parents in the population and identify genotypes who has up to 80% of the parental genome, allowing a better selection of promising genotypes for use in coffee breeding programs.

Keywords: *Coffea arabica* L., F₂ population, genetic diversity

4.2 INTRODUCTION

Arabica coffee (*Coffea arabica* L.) is one of the most important commodities for tropical developing countries (ICO, 2018). It is a perennial shrub or small tree from the *Rubiaceae* family, with a recent evolution history (LASHERMES *et al.*, 2008). *C. arabica* has been reported with a reduction of genetic variability in wild collections and cultivars used in breeding programs (LASHERMES *et al.*, 2011; SETOTAW *et al.*, 2013; FERRÃO *et al.*, 2015).

This low genetic variability pattern may be a combined effect of its recent origin, its mainly autogamous reproduction and the fact that only a few genotypes were dispersed from the primary genetic diversity center and used for breeding (LASHERMES, *et al.*, 1999; ANTHONY *et al.*, 2001; TERESSA *et al.*, 2010; SETOTAW *et al.*, 2013). It has been a challenge for coffee breeding programs the identification of genotypes with high productivity, abiotic stress tolerance, resistance against pathogens, as well as biochemical traits related to coffee quality (TRAN *et al.*, 2016).

In recent decades, an increase of the genetic diversity among *C. arabica* cultivars was obtained, due to the introduction of parental lines with more diverse genetic base (SETOTAW *et al.*, 2013). Despite the genetic diversity available, the coffee breeding process is still long and difficult (LASHERMES *et al.*, 2008). Thus, genomic approach offers strategies to introduce genetic variability and the introgression of desirable agronomic characteristics, populations with diverse background have been produced in Agronomic Institute of Campinas (IAC) and Agronomic Institute of Paraná (IAPAR) using Sarchimor IAC 1669-33. This genotype is derived from hybrid H361/4 originated from the crossing between Villa Sarchí and Híbrido de Timor CIFIC 971/10 (a natural hybrid between *C. canephora* and *C. arabica*).

Also, since 1928, several materials have been transferred from *C. arabica* origin center to other Tropical Countries, in order to preserve and explore the *C. arabica* genetic variability (ANTHONY, 2001). The IAC, through a partnership with the Tropical Agronomic Center for Research and Teaching (CATIE - Costa Rica), obtained genotypes from Ethiopia FAO expedition (FAO, 1968) and part of these materials were sent to the IAPAR, where they have been kept for more than 40 years, representing an important source of variability available to be used in crossings.

One effective way to detect genetic diversity is through the use of Single Nucleotide Polymorphism (SNP), due to its higher resolution, accurate discrimination, abundant distribution along the genome and low cost (CAIXETA *et al.*, 2013). SNPs have been used to explore the genetic variability of collections, as well as the populations for genotypic

discrimination and genetic relationships between cultivated and wild populations of *C. arabica* and *C. canephora* (TERESSA *et al.*, 2010; FERRÃO *et al.*, 2015, SOUSA *et al.*, 2017, BIKILA *et al.*, 2017; SILVA *et al.*, 2018).

With the advance of high throughput sequencing, Genotyping by Sequencing (GBS) has been used as a methodology to discover and genotyping single nucleotide polymorphisms (SNP). The procedure for constructing GBS libraries is made based on reducing genome complexity with restriction enzymes followed by the ligation of barcode adapter and PCR (Polymerase Chain Reaction) amplification. GBS technique allows a genotyping of multiple DNA samples generating after the PCR and the GBS libraries are generated through high-throughput sequencing technologies (NGS – Next Generation Sequencing)) (ELSHIRE *et al.*, 2011). The GBS has been widely applied to genomic diversity studies, genetic, linkage analysis, molecular marker discovery and Genome-Wide Association Studies (GWAS) in plant breeding programs (HE *et al.*, 2014).

In this work we analyze the genetic diversity and structure from a F₂ population of *C. arabica*, using SNPs from GBS data. The plant material used in this study was created by IAPAR Coffee breeding program. The F₂ population include the genetic Ethiopian background and the access E335, that has resistance to *Pseudomonas syringae* pv. *garcae* (MOHAN *et al.*, 1978) and of Catuai Vermelho IAC46, which has high vegetative vigor, low plant size and wide environment adaptation (PEREIRA *et al.*, 2010), generating the parent IAPAR 78001-L₁C₁, which was crossed with IAPAR 88480-8-L₃C₃.

The parent L₃C₃ is derived from Sarchimor IAC 1669-33 and stands out for having small plant size, high resistance to *Hemileia vastatrix* (SERA *et al.*, 2007; DEL GROSSI *et al.*, 2013), high resistance to population Mexi1 of *Meloidogyne exigua* (MUNIZ *et al.*, 2009), besides intermediate resistance to *M. paranaensis* (MUNIZ *et al.*, 2009; SHIGUEOKA *et al.*, 2016) and *M. incognita* (GONÇALVES and FERRAZ, 1987). However, it is susceptible to *Pseudomonas syringae* pv. *garcae* (ANDREAZI *et al.*, 2015).

4.3 MATERIAL AND METHODS

4.3.1 Plant material and DNA extraction

An F₂ population of *C. arabica* of IAPAR was selected for this study. The population consisted of 187 genotypes plus the parents (IAPAR 78001-L₁C₁ and 88480-8-L₃C₃), totalizing 189 individuals. The F₂ population was generated from a self-pollination of F₁ plant (IAPAR H9904-3) obtained from a cross between L₁C₁ [(E335) x Catuai Vermelho IAC 46] and L₃C₃ (Sarchimor IAC 1669-33). This F₁ plant has the high resistance to *P. s. pv. garcae* and to *Hemileia vastatrix* originated from L₁C₁ and L₃C₃, respectively (SERA *et al.*, 2007; ANDREAZI *et al.*, 2015). The self-pollination was carried out in 2007 and planted in the field in 2009 at IAPAR (Londrina, PR, Brazil).

Leaves from 189 genotypes (F₂ population plus two parents) were collected, and the genomic DNA (Deoxynucleic Acid) was extracted using the CTAB method with modification (HEALEY *et al.*, 2014). The DNA quality was verified in 1% agarose gel, and the quantification was measured using Qubit fluorometer (Invitrogen®) and diluted in a final concentration of 100ng/ul.

4.3.2 GBS and SNP identification

The GBS was performed using the protocol described by Elshire *et al.* (2011), and it was conducted by Genomic Diversity Facility LIMS, at Cornell University (Ithaca, NY – USA) to discover single nucleotide polymorphisms (SNP). The GBS library was sequenced using the Illumina HiSeq2000. The *Pst*I restriction enzyme was used for library preparation to reduce genome complexity and genotype multiple DNA samples with 189 samples in two 96 well multiplex plates.

The sequences were trimmed for quality control to remove SNPs with low frequency and genotyping errors. The reference genomes used to map the tags were *C. canephora* (DENOEUDE *et al.*, 2014) and *C. eugenoides*, kindly provided by Arabica Coffee Genome Consortium (ACGC).

TASSEL-GBS pipeline (TASSEL 5.0 version 3.0.166) was used for SNPs identification. Only tags aligned to a unique position on the reference genomes were selected. The quality control parameters were call rate (>80%), minor allele frequency (MAF>5%) and heterozygosity (Ho<90%). The indels and the multiallelic SNPs were removed. It was also performed an imputation called LD k-nearest neighbor algorithm (LD KNNi imputation) provided by TASSEL 5.0 software (BRADBURY *et al.*, 2007).

4.3.3 Genetic analysis of population structure

Population structure was estimated using a Bayesian clustering method from the STRUCTURE software version 2.3.4 (PRITCHARD *et al.*, 2000). A set of 1.966 SNPs in 189 individuals from the F₂ population and parental L₁C₁ and L₃C₃ were evaluated and the allelic frequency estimated in each cluster and the estimating the number of subpopulation (k) made based on Evanno's criterion (EVANNO *et al.*, 2005). The K value ranged from 1 to 10, using 100.000 burn-in periods and 100.000 interaction MCMC (Markov Chain Monte Carlo) and tested ten times for each K value. Population number was defined and estimated based on ΔK values on STRUCTURE HARVESTER and after that the upper level (EARL and VON HOLDT, 2012). The level considered to assign the genotypes to the different groups was 0.6.

The genetic similarity matrix was obtained by DARwin software version 6.0.017 (JACQUEMOUD-COLLET, 2018) using Simple Matching Coefficient (SMC). The Principal Coordinate Analysis (PcoA) was performed using GenAlEx 6 software (PEAKALL and SMOUSE, 2012). It was estimated for each genetic group the mean Number of Alleles (Na), the percentage of polymorphic loci (P), expected heterozygosity (He), observed heterozygosity and unbiased expected heterozygosity (uHe) and Shannon's information index (I) using GenAlEx 6 software (PEAKALL and SMOUSE, 2012).

Analysis of molecular variance (AMOVA) (EXCOFFIER *et al.*, 1992) and *Fst* calculation also was performed using the GenAlEx 6 software (PEAKALL and SMOUSE, 2012), according to results generated by STRUCTURE analyses.

4.4 RESULTS AND DISCUSSION

4.4.1 Identification and analyses of SNP markers

From the raw GBS data, we obtained 417.938.880 reads filtrated by quality, which produced 26.462.867 tags. Mapping of the GBS tags results in 52% tags unaligned, 20% aligned to multiple positions and 29% aligned to a unique position. A total of 21.495 SNP markers were identified throughout the coffee genome using *C. canephora* and *C. eugenioides* as the reference genome.

We use those genomes since they are the two diploid ancestral species that originate the allotetraploid *C. arabica* (LASHERMES *et al.*, 1999). A public *C. arabica* genome was only released recently (https://www.ncbi.nlm.nih.gov/assembly/GCF_003713225.1), which might improve the number of SNPs detected.

A similar study using *C. arabica* GBS data and only *C. canephora* as a reference genome mapped 22% of tags in GBS data (SANT'ANA *et al.*, 2018). We identified 29% of tags, considering both genomes (*C. canephora* and *C. eugenioides*). This proportion could be explained because we did not use the *C. arabica* as a reference genome and due to sequencing errors and difficulties in genome assembly. Previous studies (FERRÃO *et al.*, 2017), had similar numbers of *C. canephora* GBS tags mapped in the *C. canephora* genome (32%).

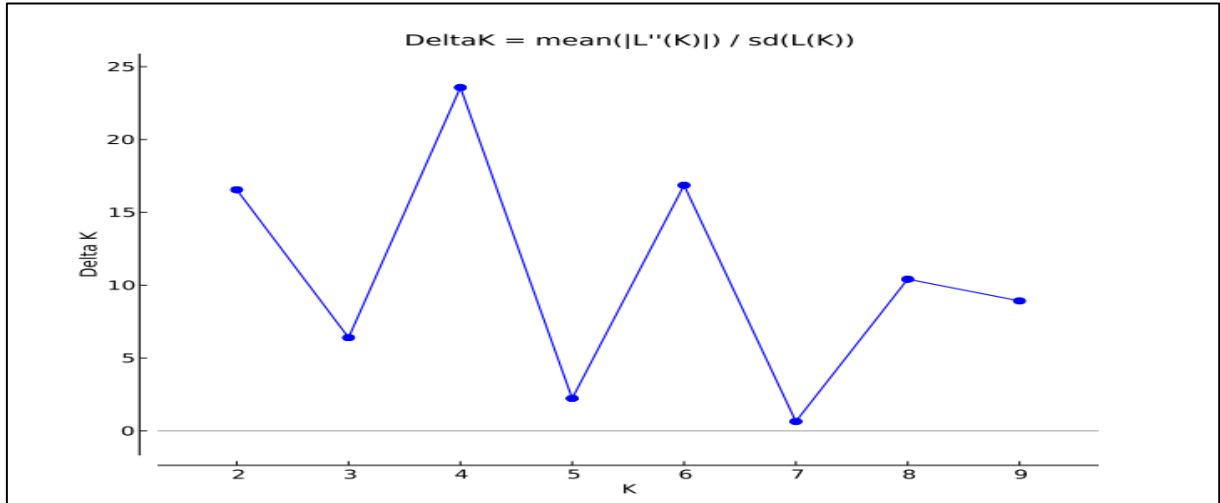
The SNPs were reduced to 4,833 SNPs after filtering for MAF (Minor Allele Frequency) and call rate, and 2,577 when indels and multiallelic SNPs were removed. We applied the imputation using LD Knni from TASSEL, which did not change the number of SNPs. Finally, we applied the heterozygosity (<0.9) filter resulting in a set of 1,966 SNP markers.

A genetic map was constructed by Moncada *et al.*, (2016) initially using 848 (SSR/SNPs) markers. However, later they reported a total set of 1008 markers, being included 159 of the randomly imputed markers, originating the first highest resolution genetic map in coffee. This mapping population of *C. arabica*, is provenient from (Caturra x CCC1046) and they also used a less restrictive filtering (MAF>2.5%). Thus, considering the work within a F₂ population and the low genetic diversity of *C. arabica*, we can consider we had an expressive number of SNPs. Recently, Merot – L'Anthiène *et al.*, (2018) showed that from a *Coffea* SNP array of 8,5 K only 10% were useful to discriminate *C. arabica* genotypes.

4.4.2 Population structure analysis

Bayesian clustering conducted using the STRUCTURE software revealed distinct subpopulations, with the highest and most reliable probabilities between a K of 2, 4 and 6 (Figure 1) with a set of 1,966 SNPs in 189 genotypes. However, the bar plot of K = 6 was not analyzed in this study, due to the amount of mixture observed between the genotypes.

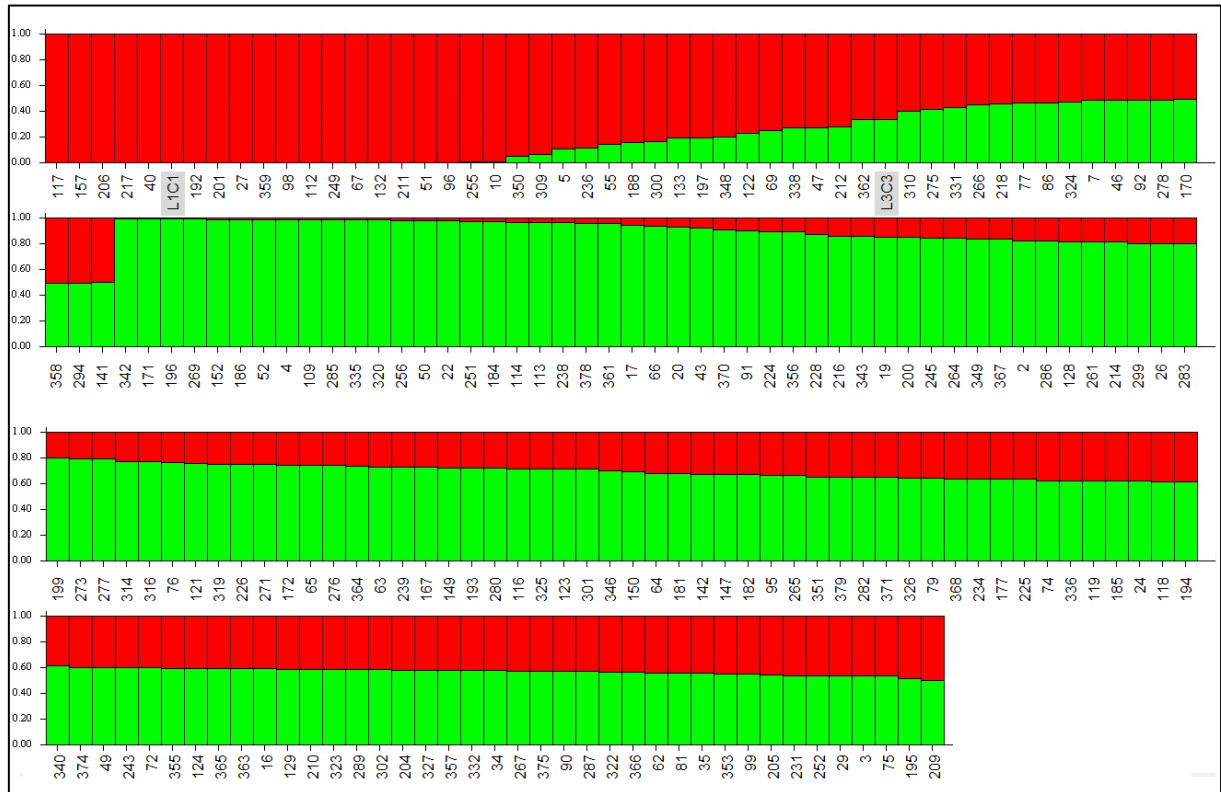
Figure 1. Population structure among 189 genotypes of F₂ population of *Coffea arabica*. Evolution of Δk values (y-axis) according to the number of genetic groups (x-axis). Barplot of the estimated membership coefficient (Q) based on the 1.966 SNPs for k=2 and k=4.



The structure (k=2) showed the difference between parents and their progeny (Figure 2). L₁C₁ comes from a cross between accession E335 and Catuaí, whereas L₃C₃ derives from an introgression by *C. canephora*. Comparing both parents, one portion of the genetic composition of L₃C₃, which is in red, is similar to L₁C₁. However, the L₁C₁ does not have any green portion in its composition. Thus, this result indicates that the parents had substantial amounts of genetic difference.

Nineteen genotypes have a genetic formation similar to the L₁C₁, with low or no contribution from L₃C₃. On the other hand, there is a variation in the genetic distribution in many genotypes, when compared with L₃C₃, a predominant part is coming from this parent.

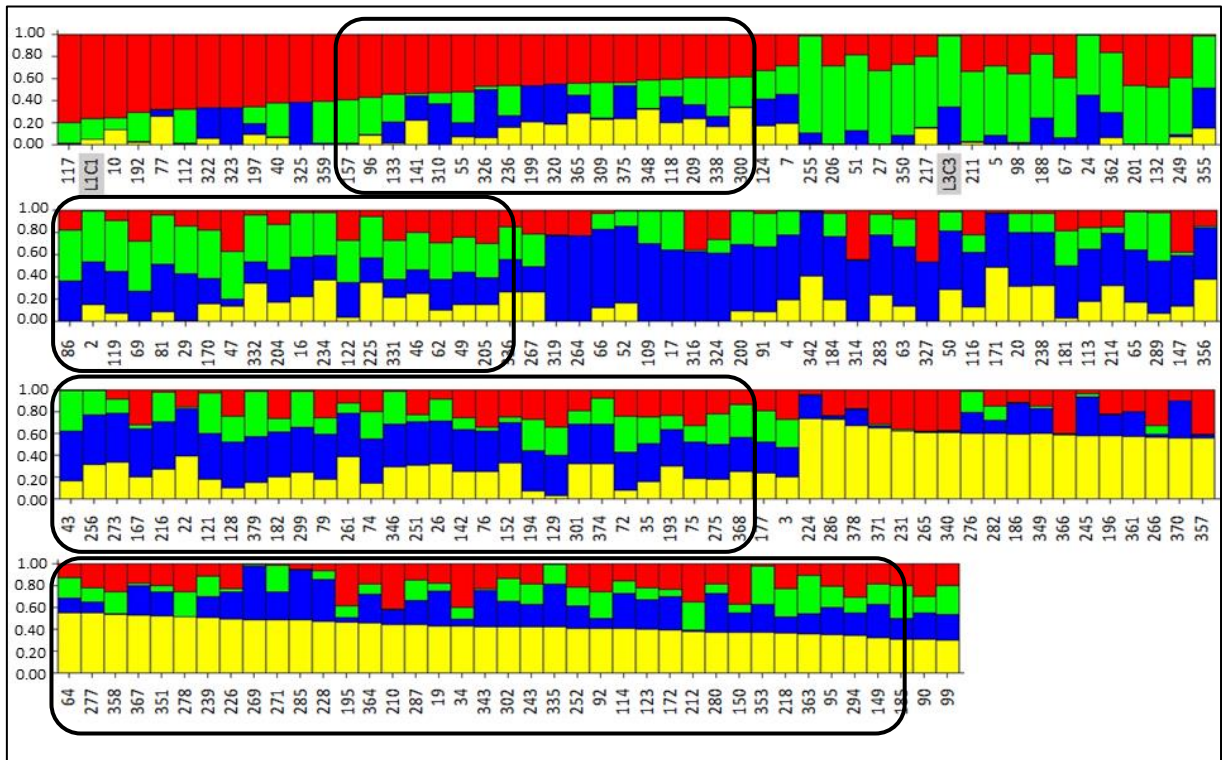
Figure 2. Bar plot of the estimated membership coefficient (Q) of the 189 genotypes in a F₂ population based on 1966 SNPs for the k = 2 were used to inferred population structure of *C. arabica*.



The structure using K=4 showed the distinct background between L₁C₁ and L₃C₃ population parents (Figure 3). Some F₂ individuals are very close to the L₁C₁ composition (accession E335 x Catuai). For example, the genotypes 10 and 117 with inferred ancestry of individuals, Q₁ value of 0.755 and 0.794, respectively. Other genotypes were close to the L₃C₃ (Tupi IAC) such as 206 and 255 with Q₂ value of 0.708 and 0.882, respectively (Supplementary Table S1).

With k=4, L₃C₃ has a large portion represented by a green color which is also in L₁C₁ formation but in a small proportion. We also identified on the L₃C₃ formation a small contribution with blue color, but unlike the parental, we could recognize in part of the population an extensive contribution of the blue color (genotypes 264 and 319, with the Q₃ value of 0.772 and 0.774, respectively). This contribution may be related to the genetic background obtained through L₃C₃ parent that has introgression of *C. canephora*, from HdT 832/2.

Figure 3. Bar plot of the estimated membership coefficient (Q) of the 189 genotypes in a F₂ population based on 1966 SNPs for the k = 4 were used to inferred population structure of *C. arabica*.



Studies have shown a great genetic distance between genotypes belonging to F₂ progenies obtained with SNP analysis in crosses between Catuaí and HdT (SOUSA *et al.*, 2017). According to the degree of introgression of *C. canephora*, we precised the division of the individuals into subgroups. The same occurred in L₁C₁ parent with a small portion represented by the yellow color. However, we have genotypes with a higher contribution of the yellow color (genotype 224, 286 and 378 with Q₄ value of 0.740, 0.733 and 0.674), respectively. Thus, these results indicate that this contribution could come from accession E335 used in the crossing to generate the L₁C₁ parent.

We also observed individuals who has a very different compositions and proportion, but with background color for both parents, called M group, representing the admixture group (Figure 3). The M group are all the individuals, who have a value of inferred clusters below 0.6. (Q value).

Those are promising genotypes for future crossings or self-pollinations due to genetic variability in the population or even select based on the parental profile, considering the possible heritability for coffee rust and nematode resistance originated from Sarchimor or *Pseudomonas* resistance originated from E335. It is worth emphasizing the importance of this

group, once we are studying a segregating population trying to recognize genetic variation among the individuals.

Studies have shown that the identification of rust resistant cultivars are fundamental for coffee breeding programs (CAIXETA *et al.*, 2009) and it has been the main measure against this disease (VÁRZEA and MARQUES, 2005). Some races have been described, but the race II is the most prevalent in Brazil (SILVA *et al.*, 2018) being Híbrido de Timor (HdT) the most important source of resistance to leaf rust used in breeding programs.

A study testing 152 HdT genotypes that were phenotyped with *Hemileia vastratrix* races II and XXXIII and molecular analysis carried out with 29 SSR markers. They identified that all genotypes were resistant to race II and 141 were resistant to race XXXIII to *H. vastratrix* and based on the presence of markers, a fingerprint was determined for each individual, using 52 alleles from 22 SSR markers (SILVA *et al.*, 2018).

Considering we have an F₂ segregating population, that used L₃C₃ (coming from HdT) as one of the parents, we have a valuable source of individuals with potential agronomic traits (such as resistance to rust) to be explored by breeding programs, such as genotypes 255, 51 and 99, for example. The structure of the population can help to identify these genotypes and principally, through the genetic similarity analysis, it is possible to identify individuals with different ranges of parental genome recovery providing helpful information for this selection.

The STRUCTURE analysis showed that most individuals in each group share the same allelic pool evidencing a high genetic variability within groups. Despite being a single population, according to molecular data, the genotypes were divided into subgroups (called group 1, 2, 3 and 4). Individuals who had Q value below 0.6 were considered the admixture group, called M and these individuals are dispersed among all groups.

4.4.3 Analyses of genetic diversity

In a principal coordinate analysis (PcoA), most of the variation is explained by the first coordinate, with 39.68%. The coordinates 1 and 2 explained 73.47% of the total genetic variation of the panel. Similar to the analysis using STRUCTURE, there was a division of the whole panel in 5 groups, it was formed 4 groups and the M group represent individuals with a mixture (Figure 4 A and B).

Group one (red) is composed of 12 individuals including the parental L₁C₁. Based on the PcoA, the genotypes 10 and 117 are very close to the parental L₁C₁. Group two (green, second group) contains 10 individuals including the parental L₃C₃, with genotypes 255 and 51

being close parental L₃C₃. Group three (blue, third group) has seven individuals, and population four (yellow, fourth group), contains 11 genotypes. Among the 4 groups formed, we could recognize individual with a mixed genetic structure that we arbitrarily referred to as 'M group', composed by 149 genotypes, which are distributed throughout as coordinates. This distribution of PcoA supports the results generated by STRUCTURE with a total correspondence between the groups.

According to the analysis of genetic similarity (Table 1), we recognize individuals whose similarity is closer to the parents. Genotype 255 is 81% similar to the parental L₃C₃ and the genotypes 51, 67, 99 and 362, and 80% similar to the same parental. In contrast, the genotypes 10, 40, 117 and 359 are 80% similar to parental L₁C₁.

Despite the parental plants originated from different crosses, the analysis among parents showed 73% of similarity between them (Table 1), which is an indication of a narrow genetic basis of *C. arabica*. Thus, we can consider 27% of variation between them. We were able to identified F₂ genotypes with admixture that has proportions above 73% such as in genotypes 69, 275, 365 and 366 in which the similarity was 61/61%; 59/60%; 63/63% and 64/62%, to the parental genotypes L₁C₁ and L₃C₃, respectively.

They presented the lowest frequency of common similarity among parents and it is possible to verify several intermediate levels in these genotypes composition. This fact could be due to presence of parents that came from more than one cross (acession E335, Catuaí, Villa Sarchi and Híbrido de Timor). This could indicate that maybe some characteristics is controlled by certain number of genes called genes of primary effects, located in different loci that jointly affect the expression of the characteristic or even a transgressive segregation for various characteristics.

Probably, part of the diversity present in the parents (L₁C₁ and L₃C₃), it was hidden and could evidenced only later, in the genotypes of the population. Thus, this genotypes could represent a source of variability present in F₂ population, which can be explored and being relevant to breeding programs.

Figure 4. Principal coordinate analysis (PcoA). The PcoA individuals are coloured according to the STRUCTURE groups using $k = 4$. M group are coloured in grey. (A) On Figure A 189 genotypes are represented and the M group is the admixture group. (B) Figure B represents genotypes with Q value above 0.6 and according to Bayesian inference of population structure, they were separated in groups, therefore the absence of M group.

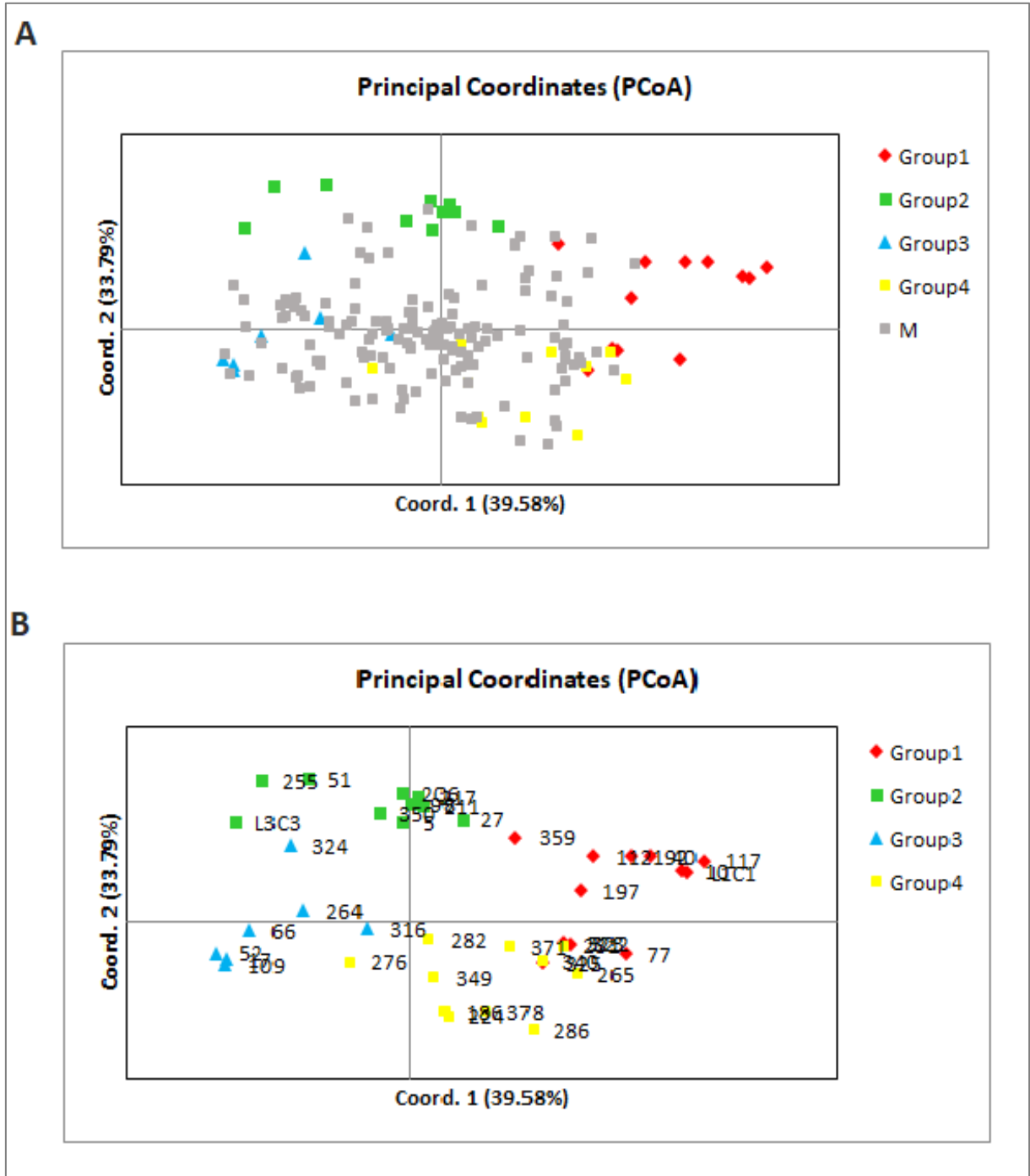


Table 1. Genetic similarity of F₂ genotypes with parents L₁C₁ and L₃C₃ calculated by DARwin Software. The number attributed to the similarity are expressed in percentage.

Sample	L ₁ C ₁	L ₃ C ₃	L ₁ C ₁ xL ₃ C ₃	Sample	L ₁ C ₁	L ₃ C ₃	L ₁ C ₁ xL ₃ C ₃	Sample	L ₁ C ₁	L ₃ C ₃	L ₁ C ₁ xL ₃ C ₃
L ₁ C ₁	1,00		0	186	0,74	0,75	0,555	27	0,75	0,76	0,57
L ₃ C ₃	0,73	1,00	0,73	195	0,75	0,74	0,555	35	0,75	0,76	0,57
275	0,59	0,60	0,354	245	0,75	0,74	0,555	116	0,75	0,76	0,57
325	0,60	0,59	0,354	261	0,74	0,75	0,555	128	0,76	0,75	0,57
69	0,61	0,61	0,3721	285	0,75	0,74	0,555	141	0,76	0,75	0,57
366	0,64	0,62	0,3968	349	0,74	0,75	0,555	170	0,75	0,76	0,57
365	0,63	0,63	0,3969	370	0,75	0,74	0,555	193	0,75	0,76	0,57
351	0,66	0,66	0,4356	342	0,72	0,78	0,5616	194	0,75	0,76	0,57
326	0,66	0,67	0,4422	119	0,73	0,77	0,5621	205	0,75	0,76	0,57
66	0,65	0,69	0,4485	184	0,73	0,77	0,5621	209	0,76	0,75	0,57
243	0,68	0,69	0,4692	273	0,73	0,77	0,5621	210	0,76	0,75	0,57
2	0,69	0,75	0,5175	19	0,74	0,76	0,5624	214	0,75	0,76	0,57
3	0,73	0,73	0,5329	22	0,74	0,76	0,5624	228	0,75	0,76	0,57
286	0,76	0,71	0,5396	50	0,74	0,76	0,5624	238	0,76	0,75	0,57
196	0,75	0,72	0,54	92	0,74	0,76	0,5624	264	0,75	0,76	0,57
17	0,72	0,76	0,5472	98	0,74	0,76	0,5624	278	0,75	0,76	0,57
26	0,72	0,76	0,5472	124	0,76	0,74	0,5624	302	0,75	0,76	0,57
276	0,72	0,76	0,5472	142	0,74	0,76	0,5624	327	0,76	0,75	0,57
355	0,72	0,76	0,5472	182	0,74	0,76	0,5624	336	0,75	0,76	0,57
114	0,74	0,74	0,5476	226	0,76	0,74	0,5624	340	0,76	0,75	0,57
224	0,74	0,74	0,5476	234	0,74	0,76	0,5624	356	0,76	0,75	0,57
43	0,71	0,78	0,5538	251	0,76	0,74	0,5624	371	0,75	0,76	0,57
4	0,72	0,77	0,5544	294	0,76	0,74	0,5624	378	0,76	0,75	0,57
16	0,72	0,77	0,5544	343	0,76	0,74	0,5624	117	0,80	0,72	0,576
197	0,77	0,72	0,5544	46	0,75	0,75	0,5625	24	0,73	0,79	0,5767
200	0,72	0,77	0,5544	113	0,75	0,75	0,5625	52	0,73	0,79	0,5767
277	0,77	0,72	0,5544	157	0,75	0,75	0,5625	65	0,73	0,79	0,5767
20	0,73	0,76	0,5548	172	0,75	0,75	0,5625	271	0,73	0,79	0,5767
34	0,76	0,73	0,5548	181	0,75	0,75	0,5625	346	0,73	0,79	0,5767
64	0,73	0,76	0,5548	252	0,75	0,75	0,5625	109	0,74	0,78	0,5772
91	0,73	0,76	0,5548	361	0,75	0,75	0,5625	217	0,74	0,78	0,5772
150	0,76	0,73	0,5548	353	0,73	0,78	0,5694	265	0,78	0,74	0,5772
152	0,76	0,73	0,5548	121	0,74	0,77	0,5698	320	0,78	0,74	0,5772
171	0,73	0,76	0,5548	256	0,74	0,77	0,5698	47	0,77	0,75	0,5775
199	0,76	0,73	0,5548	269	0,74	0,77	0,5698	63	0,75	0,77	0,5775
123	0,74	0,75	0,555	283	0,74	0,77	0,5698	72	0,75	0,77	0,5775
149	0,74	0,75	0,555	335	0,74	0,77	0,5698	76	0,75	0,77	0,5775
167	0,75	0,74	0,555	358	0,77	0,74	0,5698	95	0,75	0,77	0,5775

Continued - **Table 1.** Genetic similarity of F₂ genotypes with parents L₁C₁ and L₃C₃ calculated by DARwin Software. The number attributed to the similarity are expressed in percentage.

Sample	L ₁ C ₁	L ₃ C ₃	L ₁ C ₁ xL ₃ C ₃	Sample	L ₁ C ₁	L ₃ C ₃	L ₁ C ₁ xL ₃ C ₃
118	0,77	0,75	0,5775	300	0,77	0,76	0,5852
129	0,75	0,77	0,5775	367	0,76	0,77	0,5852
185	0,75	0,77	0,5775	10	0,80	0,74	0,592
204	0,75	0,77	0,5775	40	0,80	0,74	0,592
212	0,77	0,75	0,5775	99	0,74	0,80	0,592
267	0,75	0,77	0,5775	266	0,77	0,76	0,5852
280	0,75	0,77	0,5775	77	0,79	0,75	0,5925
301	0,75	0,77	0,5775	133	0,79	0,75	0,5925
363	0,75	0,77	0,5775	188	0,75	0,79	0,5925
364	0,75	0,77	0,5775	324	0,75	0,79	0,5925
375	0,77	0,75	0,5775	5	0,76	0,78	0,5928
379	0,75	0,77	0,5775	75	0,76	0,78	0,5928
7	0,76	0,76	0,5776	122	0,76	0,78	0,5928
147	0,76	0,76	0,5776	206	0,76	0,78	0,5928
216	0,76	0,76	0,5776	249	0,76	0,78	0,5928
231	0,76	0,76	0,5776	289	0,76	0,78	0,5928
236	0,76	0,76	0,5776	338	0,78	0,76	0,5928
287	0,76	0,76	0,5776	348	0,78	0,76	0,5928
299	0,76	0,76	0,5776	55	0,77	0,77	0,5929
309	0,76	0,76	0,5776	112	0,77	0,77	0,5929
81	0,74	0,79	0,5846	132	0,77	0,77	0,5929
192	0,79	0,74	0,5846	314	0,77	0,77	0,5929
218	0,74	0,79	0,5846	255	0,74	0,81	0,5994
74	0,75	0,78	0,585	86	0,76	0,79	0,6004
177	0,75	0,78	0,585	316	0,76	0,79	0,6004
225	0,75	0,78	0,585	323	0,79	0,76	0,6004
282	0,75	0,78	0,585	331	0,76	0,79	0,6004
319	0,75	0,78	0,585	90	0,77	0,78	0,6006
332	0,75	0,78	0,585	211	0,77	0,78	0,6006
357	0,78	0,75	0,585	310	0,78	0,77	0,6006
374	0,75	0,78	0,585	368	0,77	0,78	0,6006
29	0,76	0,77	0,5852	51	0,76	0,8	0,608
49	0,76	0,77	0,5852	67	0,76	0,8	0,608
62	0,76	0,77	0,5852	359	0,8	0,76	0,608
79	0,76	0,77	0,5852	362	0,76	0,8	0,608
96	0,77	0,76	0,5852	350	0,77	0,79	0,6083
201	0,77	0,76	0,5852	322	0,79	0,78	0,6162
239	0,77	0,76	0,5852				

Genotypes as 275, 365 and 366 has a yellow color contribution, specially the 366, who has the biggest composition from yellow color. This contribution could come from accession E335, perhaps as a result of a transgressive segregation for a specific characteristic. This type of segregation possibly is explained owing to an overdominance effect, which could justify the identification of individuals with a different phenotype from the progenitors for one or more characters.

Both Bayesian and PcoA analyses evidenced individuals 69, 275, 365 and 366 in the M group, which indicate the higher genetic diversity of this group when compared to other groups. According to STRUCTURE analysis, the formation of individual 99 is very heterogeneous, which explains why this individual was incorporated into the M group. However, the genetic similarity analysis showed that this individual is 80% similar to the parent L₃C₃.

In accordance to STRUCTURE and the PcoA results, the individual 359 is part of the group one, and it is very close to the parental L₁C₁. The same result was found here, considering the genetic similarity data (Table 1), in which the individual 359 has 80% of similarity with the parental L₁C₁. We also have individuals whose combined contribution of both parents had the highest values of similarity as 0,608% for genotypes 51, 67, 359, 362, and 0.6083 and 0.6162% for genotypes 350 and 322, respectively.

The M group also had the highest genetic diversity according to the allele number average (Na), the percentage of polymorphic loci (% P), Shannon's information index (I), and expected heterozygosity (He) with means of 2.01, 100%, 0.52, 0.35 and 0.46, respectively. The unbiased expected heterozygosity (uHe) values, which measure the genetic diversity weighted by the sample size in each group, had a higher value on groups two, four and M. Due to a high percentage of polymorphic loci and observed heterozygosity, we expect that the M group is a valuable source of favorable alleles (Table 2).

Table 2. Differences in sample size (N), percentage of polymorphic loci (% P), allele number average (Na), Shannon's information index (I), expected heterozygosity (He), observed heterozygosity (Ho), unbiased expected heterozygosity (uHe) and number of private alleles among groups identified by the software STRUCTURE considering k=4.

Group	N	% P	Na	I	He	Ho	uHe
1	12	91,61%	1,92	0,45	0,29	0,36	0,31
2	10	91,25%	1,92	0,49	0,33	0,47	0,35
3	7	80,01%	1,80	0,43	0,29	0,40	0,31
4	11	92,32%	1,93	0,49	0,33	0,45	0,35
M	149	100,00%	2,00	0,52	0,35	0,46	0,35
Total	189	91,04%	1,92	0,47	0,32	0,43	0,33

The AMOVA revealed 89% of the variation within groups, while the rest of the variation (11%) was among groups (Table 3). This result can be explained by the fact that, we are working with F₂ segregating population, so we expect a higher genetic variability between individuals than in the groups.

Similar results were found in studies of genetic diversity and population structure in a F_{3:6} population of wheat. They identified 25.566 SNPs by GBS technique and the analysis of molecular variance had shown 8% of the variance among subpopulation and 92% within subpopulation, indicating a high level of genetic diversity within-subpopulations. They explained that the high variation within groups was due to selection for specific agronomic traits in the wheat breeding program of Nebraska and the low diversity among subpopulation was due to a high value of gene flow (ELTAHER *et al.*, 2018). We do not expect in our population the gene flow, once we have only one population. However, the variability between individuals is expected due to contrasting parents of the population.

Table 3. Analysis of molecular variance (AMOVA) using 1.966 SNPs of the genetic differentiation among and within groups in a F₂ population of *C. arabica*.

Source of Variation	d.f.	Sum of squares	Variance Component	Variation percentage
Among groups	4	6231	61,2	11%
Within groups	184	91290	496,1	89%

The *Fst* value was $Fst < 0.1$ and all different from zero (Table 4). *Fst* value measures the magnitude of genetic differentiation between populations, and the results showed low differentiation among groups. Value of 0.0 is considered we have identical allelic composition and a maximum value is 1.0, corresponding to the highest value representing the magnitude of genetic differentiation among genotypes. In this sense, the results of AMOVA and *Fst* value

indicated that most of the variation is within groups. Although the structure has indicated the formation of four groups, there is a low variation between them according to the *Fst* results.

Table 4. Pairwise *Fst* matrix obtained using molecular data.

Pairwise Population Fst Values					
	Group1	Group2	Group3	Group4	M group
Group1	0,000				
Group2	0,057	0,000			
Group3	0,066	0,065	0,000		
Group4	0,048	0,063	0,065	0,000	
M group	0,032	0,033	0,031	0,021	0,000

Our results indicated that there is a genetic difference within groups inside the population, and this genetic structure could be mostly related to the expected segregation from the F₂ population. Similar results have been described in coffee progenies derived from Catuaí and Híbrido de Timor (HdT). In this case, they identified mixtures within the progenies and the contribution of HdT (introgression of *C. canephora*) (SOUSA *et al.*, 2017).

4.5 CONCLUSION

Through the GBS technique, it was possible to genotype an F₂ population allowed through the presence of 1.966 SNPs. The availability of information on the genetic diversity and population structure of this coffee population made it possible to identify recombinant genotypes as 69, 275, 365 and 366 for selection, with new recombination of alleles and also according to the distributions of genotypes in the cluster groups, the identification of individuals who have recovered up to 80% of the parental genome, as 10, 40, 51, 67, 117, 99, 255, 359 and 362, showing the effect of parental line contribution in the population. The population showed to have a huge genetic potential to be explored and can help breeders to choose genotypes to carry on their program, reducing efforts on analysis of their genetic resources.

4.6 REFERENCES

- ANDREAZI, E.; SERA, G.H.; FARIA, R.T.; SERA, T.; SHIGUEOKA, L.H.; CARVALHO, F.G.; CARDUCCI, F.C.; CHAMLET, D. Desempenho de híbridos F₁ de café arábica com resistência simultânea a ferrugem, mancha aureolada e bicho mineiro. **Coffee Science**, v.10: 375-382, 2015.
- ANTHONY, F.; BERTRAND, B.; QUIROS, O.; WILCHES, A.; LASHERMES, P.; BERTHAUD, J.; CHARRIER, A. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. **Euphytica** 118, 53–65, 2001.
- BIKILA, B.A.; SAKIYAMA, N.S.; CAIXETA, E.T.; SNPs based molecular diversity of *Coffea canephora*. **Journal of Microbiology and experimentation**, v.5:1, 2017.
- BRADBURY, P.J., ZHANG, Z., KROON, D.E., CASSTEVENS, T.M., RAMDOSS, Y. AND BUCKLER, E.S. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 23, 2633-2635, 2007.
- CAIXETA, E.T.; OLIVEIRA, A.C.B.; BRITO, G.G.; SAKIYAMA, N.S. Tipos de marcadores moleculares. In: BORÉM, A.; CAIXETA, E.T. (eds) **Marcadores Moleculares**. UFV, Viçosa, 11–93p, 2009.
- CAIXETA, E.T.; FERRÃO, L.F.V.; ZAMBOLIM, E.M. Marcadores Moleculares. 2: 31-68 In: BORÉM, A.; FRITSCH NETO, R. **Biotecnologia aplicada ao melhoramento de plantas**. Viçosa: Universidade Federal de Viçosa, p.335, 2013.
- DEL GROSSI, L.; SERA, T.; SERA, G. H.; FONSECA, I.C.D.B.; ITO, D.S.; SHIGUEOKA, L. H.; CARVALHO, F.G. Rust resistance in Arabic Coffee cultivars in northern Paraná. **Brazilian Archives of Biology and Technology**, v.56(1): 27-33, 2013.
- DENOEUDE, F.; CARRETERO-PAULET, L.; DEREPPER, A.; DROC, G.; GUYOT, R.; PIETRELLA, M.; ZHENG C.; ALBERTI, A.; ANTHONY, F.; APREA, G.; AURY, J.M.; BENTO, P.; BERNARD, M. et. al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v.345:6201, 1181-4, 2014.
- EARL, D. A. & VON HOLDT, B. M. Structure harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. **Conserv. Genet. Resour.** v.4, 359–361, 2012.
- ELTAHER, S.; SALLAM, A.; BELAMKAR, V.; EMARA, H.A.; NOWER, A.A.; SALEM, K.F.M.; POLAND, J.; BAENZIGER, P.S. Genetic diversity and population structure of F3:6 Nebraska winter wheat genotypes using genotyping by sequencing. **Front. Genet.**, v.9, article 76, 2018.
- EVANNO, G., REGNAUT, S. & GOUDET, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Mol. Ecol.** 14, 2611–2620, 2005.

EXCOFFIER, L.; SMOUSE, P.E.; QUATTRO, J.M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics**, 131(2):479–491, 1992. ELSHIRE, R. J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K.; BUCKER, E.S.; MITCHELL, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS One** 6, 2011.

FERRÃO, L.F.V.; CAIXETA, E.T.; PENA, G.; ZAMBOLIM, E.M.; CRUZ, C.D.; ZAMBOLIM, L.; FERRAO, M.A.G.; SAKIYAMA, N.S. New EST–SSR markers of *Coffea arabica*: transferability and application to studies of molecular characterization and genetic mapping. **Mol Breed** 35(1):1–5, 2015.

FERRÃO, L.F.V.; FERRÃO, R.G.; FERRAO, M.A.G.; FRANCISCO, A.; GARCIA, A.A.F. A mixed model to multiple model harvest location trials applied to phenomic prediction in Coffee canephora . **Tree Genet. Genom.** v.12: 95, 2017.

GONÇALVES, W.; FERRAZ, L.C.C.B. Resistência do cafeeiro a nematoídes. II Testes de progênies e híbridos para Meloidogyne incognita/raça 3. **Nematologia Brasileira (Brasil)**, v. 11:125-142, 1987.

HEALEY, A.; FURTADO, A.; COOPER, T.; HENRY, R.J. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. **Plant Methods**,10:21, 2014.

HE, J.; ZHAO, X.; LAROCHE, A.; LU, Z.X.; LIU, H.; LI, Z. Genotyping by sequencing (GBS) , an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. **Front. Plant Sci.**, v.5, 484, 2014.

ICO – International Coffee organization. Trade statistics tables, Version 11.2018. <http://www.ico.org/trade_statistics.asp> Acesso em: 22, janeiro, 2019.

JACQUEMOUD-COLLET, J.P. DARwin software, version 6.0.17, 2018. <<http://darwin.cirad.fr/darwin>>

LASHERMES, P.; COMBES, M.C.; ROBERT, J.; TROUSLOT, P.; D’HONT, A.; ANTHONY, F.; CHARRIER, A. Molecular characterization and origin of the *Coffea arabica* L. **Genome. Mol.Genet.**, v.261, 259-266, 1999.

LASHERMES, P.; ANDRADE, A.C.; ETIENNE, H. Genomics of Coffee One of the World’s Largest Traded Commodities. In: Genomics of Tropical Crop Plants. Plant Genetics and Genomics: Crops and Models, vol 1. Springer, New York, 2008.

LASHERMES, P.; COMBES, M.C.; ANSALDI, C.; GICHURU, E.; NOIR, S. Analysis of alien introgression in coffee tree (*Coffea arabica* L.). **Molecular breeding**, v.27(2): 223-232, 2011.

MEROT-L'ANTHOENE, V.; TOURNEBIZE, R.; DARRACQ, O.; RATTINA, V.; LEPELLEY, M.; BELLANGER, L.; TRANCHANT-DUBREUIL, C.; COULEE, M.; PEGARD, M.; METAIRON, S.; FOURNIER, C. Development and evaluation of a genome-wide Coffee 8.5 K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. **Plant biotechnology journal**, 2018.

MOHAN, S.K.; CARDOSO, R.L.; PAVAN, M.A. Resistência em germoplasma de *Coffea* ao cretamento bacteriano incitado por *Pseudomonas garcae* Amaral et al. **Pesquisa Agropecuária Brasileira**, v.13: 53-64, 1978.

MONCADA, M. P.; TOVAR, E.; MONTOYA, J.C.; GONZALEZ, A.; SPINDEL, J.; MCCOUCH, S. A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size. **Tree Genet. Genom.**, 12, 5, 2016.

MUNIZ, M. D. F. S., CAMPOS, V. P., MOITA, A. W., GONÇALVES, W., ALMEIDA, M. R. A., SOUSA, F. R. D., & CARNEIRO, R. M. D. Reaction of coffee genotypes to different populations of *Meloidogyne* spp.: detection of a naturally virulent *M. exigua* population. **Tropical Plant Pathology**, v.34(6): 370-378, 2009.

PEAKALL, R. & SMOUSE, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research an update. **Bioinformatics**, 28, 2537–2539, 2012.

PEREIRA, M.C.; CHALFOUN, S.M.; CARVALHO, G.R.; SAVIAN, T.V. Multivariate analysis of sensory characteristics of coffee grains (*Coffea arabica* L.) in the region of upper Paranaíba. *Acta Scientiarum Agronomy*, v.32, n.4, 2010.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, 155, 945–959, 2000.

SANT'ANA, G.C.; PEREIRA, L.P.P; POT, D.; IVAMOTO, S.T.; DOMINGUES, D.S.; FERREIRA, R.V.; PAGIATTO, N.F.; SILVA, B.S.R.; NOGUEIRA, L.M.;KITZBERGER, C.S.G; SCHOLZ, M.B.S.; OLIVEIRA, F.F.; SERA, G.H.; PADILHA, L.; LABOUISSSE, J.P.; GUYOT, R.; CHARMETANT, P. LEROY, T. Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. **Scientific Reports**, v.8(1):465, 2018.

SERA, G. H.; SERA, T.; ITO, D. S.; AZEVEDO, J.A.D.; MATA, J.S.D.; RIBEIRO FILHO, C. Selection for durable resistance to leaf rust using test-crosses on IAPAR-59 and Tupi IAC 1669-33 cultivars of *Coffea arabica*. **Brazilian Archives of Biology and Technology**, v.50(4): 565-570, 2007.

SETOTAW, T.A.; CAIXETA, E.T.; PEREIRA, A.A.; OLIVEIRA, A.C.B.; CRUZ, C.D.; ZAMBOLIM, E.M.; ZAMBOLIM, L.; SAKIYAMA, N.S. Coefficient of parentage in *Coffea arabica* L. cultivars grown in Brazil. **Crop Sci**, v.53:1237– 1247, 2013.

SILVA, R.A.; ZAMBOLIM, L.; CASTRO, I.S.L.; RODRIGUES, H.S.; CRUZ, C.D.; CAIXETA. The Híbrido de Timor germplasm: identification of molecular diversity and resistance sources to coffee berry disease and leaf rust. **Euphytica**, v.214: 153, 2018.

SOUSA, T.V.; CAIXETA, E.T.; ALKIMIM, E.R.; OLIVEIRA, A.C.B.; PEREIRA, A.A.; SAKIYAMA, N.S.; RESENDE-JUNIOR, M.F.R.; ZAMBOLIM, L. 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, 2017.

SHIGUEOKA, L. H., SERA, G. H., SERA, T., SILVA, S. A., FONSECA, I. D. B., & MACHADO, A. C. Z. Host reaction of arabica coffee genotypes derived from “Sarchimor” to *Meloidogyne paranaensis*. **Nematoda**, v.3: 10-16, 2016.

TRAN, H.; SLADE LEE, L.; FURTADO, A.; SMYTH, H.; HENRY, R. Advances in genomics for the improvement of quality in Coffee. **Journal of the Science of Food and Agriculture**, 96, p. 3310–3312, 2016.

TERESSA, A.; CROUZILLAT, D.; PETIARD, V.; BROUHAN, P. Genetic diversity of Arabica coffee (*Coffea arabica* L.) collections. **EJAST**, v.1: 63-79, 2010.

VÁRZEA, V.M.P.; MARQUES, D.V. Population variability of *Hemileia vastatrix* vs. Coffee durable resistance. In: ZAMBOLIN, L.; ZAMBOLIN, E.M.; VÁRZEA, V.M.P. (eds.) **Durable resistance to coffee leaf rust**. Viçosa, UFV, 53-74p, 2005

4.7 SUPPLEMENTARY MATERIAL

Supplementary Table S1. Bayesian inference of population structure for ancestry.

genotype	Q1	Q2	Q3	Q4	genotype	Q1	Q2	Q3	Q4	genotype	Q1	Q2	Q3	Q4
2	0,00	0,46	0,39	0,15	79	0,25	0,16	0,41	0,19	181	0,19	0,31	0,47	0,03
3	0,27	0,25	0,27	0,21	81	0,04	0,45	0,43	0,09	182	0,26	0,12	0,42	0,20
4	0,00	0,21	0,59	0,20	86	0,17	0,46	0,36	0,01	184	0,03	0,21	0,57	0,20
5	0,28	0,63	0,08	0,01	90	0,29	0,15	0,25	0,31	185	0,19	0,30	0,19	0,31
7	0,28	0,26	0,26	0,20	91	0,02	0,30	0,59	0,09	186	0,11	0,00	0,29	0,60
10	0,76	0,11	0,00	0,14	92	0,25	0,24	0,10	0,41	188	0,18	0,58	0,24	0,00
16	0,02	0,40	0,36	0,22	95	0,20	0,19	0,25	0,36	192	0,71	0,26	0,00	0,03
17	0,00	0,35	0,64	0,01	96	0,57	0,34	0,00	0,09	193	0,23	0,13	0,34	0,30
19	0,18	0,07	0,32	0,43	98	0,35	0,62	0,02	0,01	194	0,27	0,29	0,37	0,08
20	0,03	0,17	0,49	0,32	99	0,20	0,26	0,23	0,31	195	0,38	0,10	0,04	0,47
22	0,15	0,02	0,44	0,39	109	0,01	0,29	0,70	0,01	196	0,22	0,00	0,20	0,58
24	0,01	0,54	0,45	0,01	112	0,67	0,31	0,00	0,01	197	0,65	0,15	0,10	0,10
26	0,08	0,20	0,39	0,32	113	0,15	0,20	0,47	0,18	199	0,46	0,00	0,33	0,21
27	0,32	0,67	0,00	0,01	114	0,15	0,12	0,32	0,41	200	0,01	0,30	0,59	0,10
29	0,14	0,43	0,42	0,01	116	0,22	0,16	0,50	0,13	201	0,46	0,53	0,01	0,01
34	0,40	0,11	0,07	0,43	117	0,79	0,19	0,00	0,01	204	0,13	0,41	0,29	0,18
35	0,25	0,25	0,35	0,16	118	0,40	0,16	0,24	0,20	205	0,29	0,31	0,24	0,15
40	0,62	0,31	0,00	0,07	119	0,09	0,45	0,38	0,07	206	0,28	0,71	0,01	0,00
43	0,00	0,37	0,46	0,17	121	0,03	0,37	0,43	0,18	209	0,39	0,24	0,13	0,24
46	0,20	0,33	0,21	0,26	122	0,27	0,38	0,32	0,04	210	0,41	0,01	0,13	0,45
47	0,37	0,43	0,06	0,14	123	0,22	0,11	0,27	0,41	211	0,33	0,64	0,01	0,02
49	0,24	0,31	0,29	0,16	124	0,33	0,25	0,25	0,17	212	0,34	0,26	0,01	0,39
50	0,01	0,17	0,53	0,29	128	0,24	0,24	0,42	0,11	214	0,15	0,06	0,47	0,33
51	0,18	0,69	0,13	0,00	129	0,34	0,26	0,37	0,03	216	0,02	0,27	0,44	0,27
52	0,01	0,13	0,70	0,17	132	0,47	0,52	0,01	0,00	217	0,20	0,65	0,00	0,16
55	0,52	0,28	0,13	0,07	133	0,54	0,25	0,20	0,02	218	0,22	0,26	0,15	0,37
62	0,29	0,32	0,28	0,10	141	0,53	0,02	0,22	0,23	224	0,04	0,01	0,22	0,74
63	0,08	0,25	0,54	0,14	142	0,25	0,11	0,39	0,25	225	0,06	0,37	0,22	0,36
64	0,12	0,19	0,14	0,55	147	0,37	0,03	0,46	0,14	226	0,23	0,03	0,26	0,49
65	0,01	0,35	0,47	0,18	149	0,18	0,19	0,31	0,33	228	0,06	0,07	0,39	0,48
66	0,03	0,14	0,71	0,13	150	0,37	0,08	0,18	0,38	231	0,36	0,01	0,00	0,63
67	0,39	0,55	0,06	0,01	152	0,25	0,05	0,37	0,33	234	0,02	0,39	0,22	0,38
69	0,27	0,45	0,27	0,01	157	0,59	0,40	0,00	0,01	236	0,46	0,27	0,11	0,17
72	0,24	0,33	0,35	0,08	167	0,32	0,04	0,44	0,21	238	0,03	0,17	0,48	0,33
74	0,20	0,25	0,40	0,15	170	0,18	0,43	0,23	0,16	239	0,11	0,18	0,20	0,51
75	0,32	0,15	0,34	0,19	171	0,02	0,01	0,49	0,49	243	0,18	0,19	0,21	0,42
76	0,34	0,04	0,38	0,25	172	0,23	0,06	0,31	0,40	245	0,03	0,03	0,36	0,58
77	0,68	0,01	0,05	0,26	177	0,19	0,28	0,29	0,24	249	0,39	0,51	0,02	0,08

SUPPLEMENTARY TABLE S1. (continued) - Bayesian inference of population structure for ancestry.

genotype	Q1	Q2	Q3	Q4
251	0,22	0,07	0,40	0,31
252	0,21	0,17	0,20	0,41
255	0,01	0,88	0,10	0,01
256	0,01	0,22	0,45	0,32
261	0,12	0,09	0,40	0,39
264	0,22	0,00	0,77	0,00
265	0,38	0,00	0,01	0,61
266	0,32	0,08	0,03	0,57
267	0,21	0,29	0,23	0,27
269	0,01	0,01	0,49	0,49
271	0,01	0,25	0,25	0,49
273	0,08	0,13	0,45	0,34
275	0,22	0,28	0,32	0,18
276	0,01	0,20	0,19	0,61
277	0,22	0,13	0,10	0,55
278	0,26	0,22	0,01	0,52
280	0,18	0,09	0,35	0,38
282	0,15	0,13	0,13	0,60
283	0,03	0,18	0,54	0,24
285	0,05	0,01	0,46	0,49
286	0,24	0,00	0,03	0,73
287	0,14	0,19	0,22	0,45
289	0,02	0,44	0,47	0,08
294	0,31	0,14	0,21	0,35
299	0,01	0,32	0,42	0,25
300	0,39	0,27	0,01	0,34
301	0,19	0,12	0,36	0,33
302	0,13	0,21	0,24	0,43
309	0,43	0,32	0,02	0,23
310	0,52	0,10	0,37	0,01
314	0,44	0,00	0,55	0,00
316	0,36	0,01	0,63	0,01
319	0,22	0,00	0,77	0,00
320	0,45	0,00	0,36	0,19
322	0,66	0,00	0,28	0,06
323	0,66	0,00	0,33	0,01
324	0,26	0,12	0,61	0,00
325	0,61	0,00	0,39	0,00

Genotype	Q1	Q2	Q3	Q4
326	0,47	0,03	0,44	0,07
327	0,46	0,00	0,54	0,00
331	0,27	0,35	0,16	0,22
332	0,04	0,42	0,19	0,35
335	0,01	0,18	0,39	0,42
336	0,14	0,30	0,29	0,27
338	0,39	0,35	0,09	0,17
340	0,37	0,01	0,01	0,61
342	0,00	0,01	0,57	0,41
343	0,23	0,01	0,34	0,43
346	0,01	0,30	0,40	0,29
348	0,41	0,26	0,01	0,32
349	0,15	0,02	0,23	0,60
350	0,27	0,65	0,08	0,01
351	0,20	0,06	0,23	0,52
353	0,02	0,35	0,26	0,37
355	0,01	0,47	0,36	0,16
356	0,14	0,02	0,46	0,38
357	0,40	0,01	0,03	0,56
358	0,25	0,20	0,01	0,54
359	0,60	0,38	0,01	0,01
361	0,20	0,00	0,23	0,57
362	0,16	0,54	0,23	0,07
363	0,11	0,35	0,19	0,36
364	0,18	0,09	0,26	0,46
365	0,44	0,10	0,17	0,29
366	0,39	0,01	0,01	0,59
367	0,18	0,02	0,27	0,53
368	0,13	0,30	0,31	0,26
370	0,10	0,01	0,34	0,56
371	0,31	0,01	0,02	0,66
374	0,07	0,24	0,36	0,33
375	0,43	0,03	0,30	0,24
378	0,17	0,01	0,15	0,67
379	0,01	0,41	0,42	0,16
L1C1	0,76	0,19	0,00	0,05
L3C3	0,01	0,64	0,34	0,01

5 CAPITULO II: Análise fenotípica e estudos de associação genômica ampla (GWAS) em uma população F₂ de *Coffea arabica* L.

5.1 RESUMO

Um estudo de associação genômica ampla permitiu identificarmos SNPs associados a características agronômicas relacionadas à qualidade da bebida em uma população F₂ de *Coffea arabica*. Utilizando a espectrometria de infra-vermelho próximo (NIRs) caracterizamos 79 genótipos da F₃ para os compostos: ácidos clorogênicos, açúcares totais, cafeína, compostos fenólicos, cafestol, caveol, proteínas, lipídeos, açúcares redutores e sacarose. Correlações positivas significativas entre sacarose e açúcares totais, cafeína e proteína foram observadas, assim como correlações negativas significativas entre ácidos clorogênicos e cafeína. Na análise de componentes principais os dois primeiros componentes explicaram 53,58% da variação total. De um lado formou-se um grupo com genótipos contendo maiores teores de ácidos clorogênicos, caveol, açúcares redutores, cafestol e lipídios, enquanto que um segundo grupo de genótipos continha maiores teores de açúcares totais, sacarose, proteína, cafeína e compostos fenólicos. Análises de agrupamento hierárquico classificaram os genótipos em três grupos distintos, sendo o grupo número 3, composto por 19 genótipos que se destacaram pelo baixo teor de cafestol e alto teor de caveol, tendo uma combinação bioquímica interessante, por trazer benefícios a saúde. Através da genotipagem via sequenciamento de 189 indivíduos nós identificamos um total de 21.495 SNPs que após a filtragem de qualidade (MAF>0,05, call rate >0,8 e Ho<0,9) permitiu a identificação de 1.966 SNPs. Um total de 11 SNPs foram identificados no estudo de associação pelos métodos (pLARM EB, ISIS EM BLASSO e MrMLM) e potencialmente associados a ácidos clorogênicos (3), açúcares redutores (1), cafestol (4) e largura da copa (3). É possível que estes SNPs sejam relacionados a variação do fenótipo permitindo a aplicação de seleção assistida para aceleração do programa de melhoramento genético do cafeeiro. Também se abre possibilidade de identificar genes candidatos relacionados as características de interesse.

Palavras chave: Estudo de associação genômica ampla, *Coffea arabica*, composição bioquímica.

5.2 INTRODUÇÃO

O café (*Coffea arabica* L.) é uma das principais commodities agroindustriais mundial e o Brasil se destaca como o maior produtor e exportador do grão (CONAB, 2017). O gênero *Coffea* compreende 124 espécies, no entanto somente 10 são cultivadas em todo o mundo. São produzidas em mais de 60 países (DAVIS *et al.*, 2011) e as espécies economicamente importantes são *Coffea arabica* e *Coffea canephora*, que respondem por cerca de 60 e 40% da produção mundial respectivamente (ICO, 2018). *C. arabica* é uma espécie alotetraplóide ($2n = 4 \times = 44$), derivada de uma hibridação espontânea entre duas espécies diplóides ($2n = 2 \times = 22$), *C. eugenioides* e *C. canephora* (LASHERMES *et al.*, 1999).

Coffea arabica possui uma base genética estreita e devido a sua importância e seu valor econômico, programas de melhoramento tem desempenhado um papel fundamental na seleção e desenvolvimento de populações (SETOTAW *et al.*, 2013) a fim de proporcionar um aumento da variabilidade genética da espécie. Pesquisas estão sendo realizadas com o objetivo de identificar materiais genéticos que possuam características agrônômicas de interesse, tais como resistência a ferrugem do cafeeiro, além de superioridade na qualidade da bebida (MENDONÇA *et al.*, 2007; TRAN *et al.*, 2016).

Nos programas de melhoramento, o uso de marcadores moleculares tem sido explorado para estimar a variabilidade genética dos indivíduos, a discriminação genotípica, bem como a identificação de marcas associadas a características de interesse (FERRÃO *et al.*, 2015; SOUSA *et al.*, 2017a). Dentre as marcas moleculares presentes no genoma, o polimorfismo de base única (SNP – *Single Nucleotide Polymorphism*) se caracterizam pela abundante distribuição no genoma, facilidade de genotipagem e baixo custo (CAIXETA *et al.*, 2013).

A genotipagem por sequenciamento (GBS – *Genotyping by Sequencing*) tem sido amplamente aplicada na identificação de SNPs (ELSHIRE *et al.*, 2011). Baseada no sequenciamento de representação gênômica e de forma simultânea, a técnica permite uma genotipagem em larga escala (HAMON *et al.*, 2017). Fragmentos localizados em regiões similares do genoma de diversos indivíduos são sequenciados de forma representativa visando detectar SNPs entre os genótipos da população (ELSHIRE *et al.*, 2011; CAIXETA *et al.*, 2013; BHAT *et al.*, 2016). A GBS tem sido aplicada em diversas culturas como cevada, trigo (RIMBERT *et al.*, 2018), milho (SU *et al.*, 2017), ervilha (ANNICCHIARICO *et al.*, 2017) e gramíneas, na qual o desafio é ainda maior quando observado variações cromossômicas e níveis de ploidia (LU *et al.*, 2013). Essa técnica também pode ser utilizada em espécies cujo

genoma de referência não está disponível, como é o caso do *C. arabica* (PEREIRA *et al.*, 2016; HAMON *et al.*, 2017).

Sendo assim, no presente estudo, foi realizada a caracterização fenotípica e genotípica de uma população F₂ de *Coffea arabica*, visando identificar marcadores e regiões cromossômicas potencialmente associadas a características agrônomicas de interesse em cafeeiros.

5.3 MATERIAL E MÉTODOS

5.3.1 Material Vegetal

Foram utilizados 79 genótipos de uma população F₂ de *Coffea arabica*, originada de uma autofecundação de uma planta F₁ obtida a partir dos parentais IAPAR 78001-L₁C₁ [(E335) x Catuai Vermelho IAC 46)] e 88480-8-L₃C₃ (Sarchimor IAC 1669-33). Este último foi originado através de um cruzamento entre Villa Sarchí e Híbrido de Timor CIF 832/2 (HdT) (FERNANDES *et al.*, 2012), sendo HdT conhecido por ser fonte de resistência a *Hemileia vastatrix* (SERA *et al.*, 2007; DEL GROSSI *et al.*, 2013) e a *Meloidogyne exigua* (MUNIZ *et al.*, 2009).

Os genótipos parentais foram selecionados por possuírem características desejáveis como alta produtividade (IAC, 2018), resistência a *Pseudomonas* (MOHAN *et al.*, 1978) para o genótipo E335 (MOHAN *et al.*, 1978), resistência a ferrugem do cafeeiro e porte baixo para o genótipo Tupi (CARVALHO *et al.*, 2008). O genótipo E335 é proveniente da coleção da Etiópia (FAO, 1968). As plantas F₁ foram autofecundadas em 2007 e plantadas em 2009, no campo experimental do Instituto Agrônomico do Paraná (IAPAR) – Londrina-PR.

A coleta do tecido foliar da progênie F₂ foi realizada no ano de 2015 e o material foi armazenado em freezer -80 °C, para posterior extração do DNA. Os frutos de 79 genótipos da F₃ foram colhidos no mesmo ano, de abril a junho durante o período em que a maioria dos frutos estavam maduros. Os frutos foram secos pelo processamento natural, em pleno sol, em peneiras com fundo de malha de 0,5 mm até atingir umidade de aproximadamente 12%. Os grãos foram moídos em moinho de disco PERTEN 3600 e após moagem, foram armazenados em frascos de plástico a -20°C até a análise.

5.3.2 Análises de caracteres morfológicos e agronômicos de uma população F₂ de *Coffea arabica*

Foram realizadas análises morfológicas de largura e altura de planta, além de diâmetro do caule. Também foi avaliado o tamanho e peso de frutos e grãos. Os dados do diâmetro do caule foram obtidos utilizando paquímetro que foram acoplados ao caule das plantas, 5 cm acima da superfície do solo.

Para verificar o tamanho dos grãos (foram utilizadas peneiras com diferentes tamanhos de malha com tela 14 (0,56 cm), 15 (0,59 cm), 16 (0,63 cm), 17 (0,67 cm) e 18 (0,71 cm). Os grãos foram classificados em chato graúdo (G) para peneiras (19/18/17); chato médio (M) para peneiras (16/15); e chato miúdo (P) para peneiras (14 ou menor) (BRASIL, 2016). Cada genótipo passou por 6 peneiras e as sementes que ficaram retidas em cada uma delas, foram pesadas. Os pesos das peneiras 19, 18 e 17 foram somadas, a das peneiras 16 e 15 também, e todo o restante desde sementes inteiras quebradas e moca nas peneiras 14 ou menor. Para o estudo de associação, as peneiras (14 ou menor) por ficarem retidos os menores grãos, foi atribuído o número 1; para as peneiras (16/15) com tamanhos de grãos médios, foi atribuído o número 2 e para as peneiras (19/18/17) por ficarem retidos os maiores grãos, foi atribuído o número 3.

Para as análises bioquímicas, todos os resultados foram expressos em base seca em 100g de grãos coletados em cada indivíduo da população F₂ de 79 genótipos, porém os grãos representaram a F₃. Através da espectrometria de infra-vermelho próximo (NIRs) (COZZOLINO, 2014) foram quantificados compostos relacionados para a qualidade da bebida, como diterpenos caveol e cafestol, ácidos clorogênicos, açúcares totais, sacarose, cafeína e proteínas, lipídeos, açúcares redutores e compostos fenólicos (SCHOLZ *et al.*, 2014; SCHOLZ *et al.*, 2016).

5.3.3 Análises Estatísticas dos dados bioquímicos e morfológicos

Foram realizadas análise de componente principal (ACP) para analisar a estrutura da população em relação à sua composição química, análise de agrupamento hierárquico (AAH) para os componentes químicos e análises estatísticas de correlação de Pearson. Foram realizadas utilizando o algoritmo de Ward com distância Euclidiana, realizadas através do programa XLSTAT (ADDINSOFT, 2010), utilizando os dados fenotípicos dos componentes químicos.

5.3.4 Genotipagem por sequenciamento

A população F₂ de *C. arabica* foi genotipada através da técnica de GBS realizada pela Genomic Diversity Facility LIMS, na Universidade de Cornell (Ithaca, NY – EUA). A enzima de restrição utilizada foi a *Pst*I para a preparação da biblioteca de GBS seguindo o protocolo de Elshire *et al.*, (2011). Antes do envio das amostras, ainda para testar a qualidade do DNA, cerca de 10% das amostras extraídas foram digeridas através da enzima de restrição *Hind*III.

A biblioteca de GBS foi sequenciada utilizando a plataforma Illumina HiSeq2000 com 189 amostras em duas placas de 96. Os genomas de referência utilizados foram os de *C. canephora* (DENOEUDE *et al.*, 2014) e *C. eugenioides*, gentilmente cedido pelo Consórcio Genoma Café Arábica (ACGC).

Para identificar os SNPs foi utilizado o TASSEL-GBS pipeline utilizando o software TASSEL5.0 (<http://www.maizegenetics.net/idex.php?option=comcontent&task=view&id=89&Itemid=119>). Os SNPs identificados foram filtrados para remoção de SNPs de baixa qualidade, como os com frequências de alelos menores que 5% MAF (>0,05), call rate (>0,80) e heterozigosidade (<0,90). Também foi realizado a imputação dos SNPs utilizando o software TASSEL 5.0 e o modo de imputação LD k-nearest neighbor algorithm (LD KNNi imputation) (BRADBURY *et al.*, 2007).

5.3.5 Estudos de Associação Genômica Ampla (GWAS) de uma população F₂ de *Coffea arabica*

Para identificar SNPs e genes candidatos associados a produção dos compostos bioquímicos citados acima, foi utilizado o programa R versão 3.4.1 e o pacote denominado mrMLM usando quatro métodos: mrMLM (WANG *et al.*, 2016), FASTmrEMMA (WEN *et al.*, 2017), ISIS-EM-BLASSO (TAMBA *et al.*, 2017) e pLARmEB (ZHANG *et al.*, 2017).

O método mrMLM é um método de modelo linear misto que inclui marcadores selecionados, que trata o efeito de um SNP como sendo aleatório e permite uma correção de Bonferroni modificada para calcular o p-value para testes de significância, com critérios menos rigorosos (WANG *et al.*, 2016). O método FASTmrEMMA também é um modelo linear misto, porém quando comparado com o mrMLM é mais poderoso na detecção de características quantitativas e menos viés na estimativa dos efeitos dessas características e não utiliza o Bonferroni, sendo mais rigorosa com ajuste de modelo e robustez, porém requer bem menos tempo para executá-lo (WEN *et al.*, 2017).

Tanto o modelo ISIS EM-BLASSO quanto o PLARmEB são modelos Bayesianos, porém enquanto o ISIS utiliza uma abordagem de triagem independente e o BLASSO é usado para estimar todos os efeitos dos SNP's selecionados (TAMBA *et al.*, 2017). O PLARmEB é um modelo em que marcadores em um cromossomo são incluídos simultaneamente em modelo multi-locus (ZHANG *et al.*, 2017).

Foi gerado no programa TASSEL uma matriz de Kinship e no programa STRUCTURE uma matriz Q de estrutura populacional ($k = 4$) a fim de controlar a estrutura populacional e reduzir a presença dos SNP's falsos positivos.

Foram construídos gráficos box plots utilizando o programa XLStat versão 2018.1 para comparar valores fenotípicos dos indivíduos em diferentes genótipos para os SNPs significativos (homozigotos ou heterozigotos para os diferentes alelos (ADDINSOFT, 2018).

5.4 RESULTADOS E DISCUSSÃO

5.4.1 Fenotipagem para o tamanho dos grãos

Para grãos considerados chato graúdo, o genótipo que apresentou maior porcentagem de grãos desse tamanho foi o genótipo 121 da F₃, com 78% dos grãos retidos nas peneiras 17,18 e 19, considerado tamanho G e número 3 para a tabela de GWAS – *Genome Wide Association Studies* (Tabela Suplementar 1). Em contrapartida, o genótipo 331 e o parental L₃C₃ apresentaram somente 3% de sementes com esse tamanho.

Para tamanho de grão médio, o genótipo 4 foi o que apresentou maior porcentagem, com 73% dos grãos retidos na peneira 15 e 16, e por isso foi atribuído o tamanho M e foi classificado como número 2 para o GWAS, considerado tamanho médio de semente.

Para grãos considerados chato miúdo, o que apresentou maior porcentagem de grãos com esse tamanho foi o parental L₃C₃ e o genótipo 331, com 59% e 50% respectivamente e por essa razão, apresentaram as menores porcentagens de sementes graúdas. A última peneira (14 ou menor) inclui uma porcentagem grande grãos quebrados e moca, que para qualidade da bebida não é muito indicado, uma vez que, a semente moca no processo de torrefação, torra mais do que as outras sementes, comprometendo a qualidade da bebida, causando amargor e adstringência (BRASIL, 2016).

O tamanho de grão considerado ideal para comercialização são os chatos graúdos (peneira 19, 18 e 17). A qualidade do café está diretamente relacionada aos diversos constituintes físicoquímicos, que são responsáveis pela aparência do grão, pelo

sabor e aroma característicos das bebidas, podendo nessa primeira etapa, selecionar alguns genótipos de acordo com o tamanho do grão.

5.4.2 Fenotipagem para o diâmetro de caule, altura e largura das plantas

Foram realizadas análises fenotípicas do diâmetro do caule, altura e largura das plantas, realizado para avaliar o desenvolvimento das plantas no campo. A Tabela 1 mostra os valores mínimo, máximo, a média, o desvio padrão e o coeficiente de variação para as características analisadas.

O desvio padrão obtidos para altura da planta (31,1), largura da copa (22,88) e diâmetro do caule (13,53) (Tabela 1) corroboram com as observações visuais a campo no IAPAR. Apesar das plantas terem sido plantadas no mesmo ano e período, observou-se uma variabilidade genética entre os indivíduos da população, para essas três características.

Tabela 1. Valores mínimo, máximo, média, desvio padrão e coeficiente de variação para os nove compostos químicos e para as análises morfológicas altura, largura e diâmetro do caule.

Variável	Mínimo	Máximo	Média	Desvio padrão	CV %
Altura (cm)	88	250	151,99	31,1	20,46
Largura (cm)	60	196	117,77	22,88	19,43
Diâmetro do caule (mm)	37,42	116,15	60,71	13,53	22,28
ACGT ^a	3,62	7,6	5,55	1	18,00
Cafeína ^a	0,91	1,74	1,26	0,14	11,29
Lipídeos ^a	12,38	15,4	14,1	0,58	4,09
Proteína ^a	13,03	16,2	14,67	0,68	4,63
Sacarose ^a	4,8	7,94	6,36	0,66	10,37
CF ^a	4,68	6,62	5,58	0,38	6,80%
AT ^a	5,4	8,4	6,23	0,64	10,28
AR ^a	0,13	0,64	0,38	0,12	30,72
Cafestol ^a	204,3	958,9	547,85	145,41	26,54
Caveol ^a	365,69	954,79	679,06	107,77	15,87
Chato graúdo	3%	78%	39%	0,21	54,23
Chato médio	15%	75%	42%	0,16	38,78
Chato miúdo	1%	59%	14%	0,10	75,66

ACGT ácidos clorogênicos, CF composto fenólico, AR açúcares redutores, Cafeína, Lipídeos, Proteína, Sacarose

CV. Coeficiente de variancia, AT açúcares totais

^a Expresso em g 100⁻¹

5.4.3 Correlação entre as variáveis

Através da espectrometria de infra-vermelho próximo (NIRs – *Near Infrared Spectroscopy*) foi possível quantificar compostos relacionados com qualidade da bebida como diterpenos, ácidos clorogênicos, açúcares totais, sacarose, cafeína, proteínas, lipídeos, açúcares redutores e morfológicos. Esses nove componentes bioquímicos foram analisados e quantificados em 79 genótipos de uma população F₃.

Foi observada uma grande variabilidade de composição bioquímica na população e os compostos com maior desvio padrão foram cafestol com (145,41) e caveol com (107,7) (Tabela 1). O menor valor observado para teor de cafestol foi de 204,3, apresentado pelo parental L₃C₃ e valor máximo de 958,9 apresentado pelo parental L₁C₁. Para teores de caveol o parental L₁C₁ produziu 647,34, enquanto que o parental L₃C₃ produziu 954,79. Estudos conduzidos por Kitzberger e colaboradores (2013a e 2013b) mostraram diferenças na produção de diterpenos em cafés arábica puros e cafés arábica com introgressões de *C. canephora*.

Em Iapar 59, considerado uma cultivar moderna, com introgressão de *C. canephora* através do Híbrido de Timor, foi observado um teor de 326g 100⁻¹ para cafestol e 898g 100⁻¹ para caveol. Entretanto, para cultivares tradicionais como Catuaí foram encontrados valores para cafestol de 604g 100⁻¹ e 371g 100⁻¹ para caveol, indicando que cafés com a introgressão de genes de *C. canephora*, como é o caso do Iapar 59 e Tupi, tem valores mais altos de caveol e menores de cafestol. É possível observar esse tipo de composição em alguns materiais da população, nos quais são encontrados altos níveis de caveol e baixo de cafestol (Suplementar 2), evidenciando que possivelmente esses materiais possuem uma grande contribuição genética do parental L₃C₃ que também possui introgressão de *C. canephora*. Essa composição entre os dois diterpenos é recomendada devido ao aumento negativo do colesterol sérico relacionado ao cafestol (NAIDOO *et al.*, 2011), possibilitando uma seleção de genótipos para produção de cafés com maiores benefícios para a saúde.

O teor de ácidos clorogênicos da população F₂ para valores mínimo e máximo, variou de 3,62 a 7,6 g 100⁻¹. Valores de ácidos clorogênicos considerados baixos variam entre 3,4 a 4,8% (UPADHYAY; MOHAN RAO, 2013) em *C. arabica*. Ferreira (2017) observou no acesso da Etiópia E335, um dos parentais da população, uma concentração de 3,57 g 100⁻¹ para esse composto. Já trabalho realizado por Link e colaboradores (2014) observaram altas concentrações de ácidos clorogênicos de 11,14g 100⁻¹ em Tupi e 8,14g 100⁻¹ em Catuaí. Em *C. canephora* foram observadas concentrações que variaram de 7,9 a 14,5 g 100⁻¹ (KY *et al.*, 2001).

Acredita-se que as variações encontradas no estudo dos valores mínimo e máximo de ácidos clorogênicos encontrados na população F₂ de *C. arabica* vêm da contribuição dos pais para a progênie e dados do acesso da Etiópia 335 no estudo de Ferreira (2017) sugerem que esse acesso estaria contribuindo na redução dos teores de ácidos clorogênicos. Portanto na população, é possível identificar genótipos que derivaram do acesso E335, sendo possível selecionar indivíduos com valores reduzidos para a produção desse composto.

Os ácidos clorogênicos são compostos importantes por estarem relacionados a formação e maturação dos grãos (JESZKA-SKOWRON *et al.*, 2016). Tem efeito positivo ou negativo sobre o sabor final da bebida, dependendo de suas concentrações. Eles conferem ao café sabor, adstringência e amargor, porém em altas concentrações se torna muito amargo e ácido comprometendo a qualidade da bebida (CAMPA *et al.*, 2005), sendo importante portanto a identificação de cafés contendo baixas concentrações de ácidos clorogênicos.

Os açúcares são importantes precursores dos compostos aromáticos gerados durante a torrefação (VASCONCELOS *et al.*, 2007) e genótipos com alto teor de açúcares são preferidos, pois agregam intenso sabor a bebida. Em nossos dados, o teor de sacarose variou entre 4,8 a 7,94, demonstrando grande variabilidade quanto a esse composto. Alguns trabalhos relatam um teor de sacarose que variam de 6,4 a 9,06 g 100⁻¹ (TESSEMA *et al.*, 2011).

Já o teor de açúcares redutores variou de 0,13 a 0,64 g 100⁻¹. Trabalho realizado por Kitzberger e colaboradores (2013b) analisaram a composição química de cafés arábicas de cultivares tradicionais e modernas, e para teor de sacarose obtiveram em Catuaí 7,89g 100⁻¹ e 6,91g 100⁻¹ em Iapar 59, que é bem próximo de Híbrido de Timor. Para o teor de açúcares redutores, obtiveram 0,20 e 0,348g 100⁻¹ respectivamente. Essa variação na produção dos açúcares (sacarose e açúcares redutores) pode ser atribuída às diferenças de respostas das cultivares ao ambiente à adaptação desses materiais ao local e a condição de produção ou simplesmente devido à variabilidade genética. Sabe-se que durante o desenvolvimento do grão ocorre uma variação na produção dos açúcares e que enquanto sacarose tende a aumentar, a produção de açúcares redutores tende a diminuir (KITZBERGER *et al.*, 2013b).

O teor de cafeína variou de 0,91 a 1,74g 100⁻¹ com um desvio padrão de 0,14. Em trabalho realizado por Scholz *et al.*, (2016) foi encontrado um teor de cafeína entre 0,72 e 1,23 em acessos originários da Etiópia e em trabalho realizado por Ferreira (2017) o teor de cafeína encontrado no acesso E335 foi de 1,05 g 100⁻¹. Link e colaboradores (2014) também identificaram teor de cafeína de 1,31 g 100⁻¹ em Tupi IAC 1669-33 e 1,01 g 100⁻¹ em Catuaí,

Para esse composto, como nenhum dos parentais possuía um teor muito baixo de cafeína apresentou pouca variação entre os indivíduos da população.

Por sua vez, os compostos químicos que apresentaram menor desvio padrão foram açúcares redutores com 0,12 e cafeína com 0,14 e conseqüentemente, com pouca variação entre

os valores mínimo e máximo dentro da população.

Entre os compostos, obtivemos 32 correlações positivas significativas e 25 correlações negativas (Tabela 2). Os valores da correlação positiva significativa (r) encontrada entre sacarose e açúcares totais foram de 0,984 e entre cafeína e proteína o valor obtido foi 0,764 e entre altura e largura de planta foi de 0,619. Cafeína e proteína são compostos que dependem dos compostos nitrogenados que se acumulam durante a formação dos frutos (SCHOLZ *et al.*, 2016) e fornecem aroma e sabor à bebida.

No presente estudo, a maior correlação negativa e significativa obtida, foi entre os tamanho de grão graúdo e médio ($r = -0,775$) e entre tamanho de grão pequeno e graúdo ($r = -0,713$). Esses resultados vão de acordo com dados gerados com esse coeficiente, uma vez que correlações negativas significa que enquanto uma variável aumenta a outra, obrigatoriamente, sempre diminui. Considerando a característica de tamanho de grão, quanto maior a quantidade de grãos de um tamanho encontrado no genótipo, menor é porcentagem obtida para os outros tamanhos.

Também foi obtida uma correlação negativa e significativa entre tamanho de grão graúdo e ácidos clorogênicos (ACG) ($r = -0,461$). Essa correlação negativa demonstra que o tamanho de grão não está diretamente relacionado a produção do composto. Esse resultado é interessante, em razão de uma busca pela produção de grãos de café com tamanho graúdo, pelo valor econômico agregado. Já teores de ácidos clorogênicos são benéficos tendo em vista sua atividade antioxidante, entretanto teores muito elevados podem promover um impacto sensorial negativo (CLIFFORD, 1983).

Entre ácidos clorogênicos (ACG) e cafeína ($r = -0,430$), também foi observada de forma significativa ($r = -0,47$) em trabalho realizado por SCHOLZ e colaboradores (2016). Outros trabalhos encontraram uma correlação positiva e significativa para os compostos (ANTHONY, *et al.*, 1993), inclusive trabalhos realizados por nosso grupo de pesquisa (FERREIRA, 2017; SILVA, 2018), que identificaram uma correlação positiva de ($r = 0,396$ no ano de 2012 e $r = 0,171$ no ano de 2015) e ($r = 0,17$ no ano de 2016), respectivamente. Nosso estudo, por se tratar de uma população F₂ de *C. arabica* específica, diferentemente dos

trabalhos anteriores, não segue o comportamento das demais populações já previamente estabelecidas, por isso os valores negativos de correlação.

Tabela 2. Correlação entre os componentes bioquímicos ácidos clorogênicos (ACG), açúcares totais (AT), cafeína, lipídeos, proteína, sacarose, compostos fenólicos (CF), açúcares redutores (AR), cafestol, caveol, tamanho de grãos (pequeno, médio e graúdo e caracteres morfológicos da planta nos genótipos de café da população F₂ de *C. arabica*.

Variáveis	Cafeína	Caveol	Cafestol	Proteína	Lipídeos	Sacarose	ACG	CP	AR	AT	P	G	M	altura	largura	diâmetro
Cafeína	1															
Caveol	-0,087	1														
Cafestol	-0,429	-0,136	1													
Proteína	0,764	0,000	-0,325	1												
Lipídeos	-0,156	0,386	0,293	-0,185	1											
Sacarose	0,135	0,060	-0,367	0,209	-0,257	1										
ACG	-0,430	0,280	0,495	-0,199	0,150	-0,041	1									
CP	0,373	-0,102	-0,058	0,142	0,229	-0,150	-0,375	1								
AR	-0,407	0,357	0,285	-0,232	0,339	-0,249	0,506	-0,332	1							
AT	0,065	0,128	-0,326	0,173	-0,202	0,984	0,050	-0,216	-0,074	1						
P	-0,139	0,220	0,093	0,030	-0,006	-0,157	0,383	-0,277	0,192	-0,127	1					
G	0,206	-0,139	-0,349	-0,012	0,035	0,288	-0,461	0,139	-0,303	0,241	-0,713	1				
M	-0,236	0,084	0,438	-0,028	0,071	-0,289	0,365	0,053	0,345	-0,234	0,246	-0,775	1			
altura	-0,296	0,242	0,335	-0,160	0,295	-0,360	0,265	-0,002	0,446	-0,289	0,084	-0,280	0,432	1		
largura	-0,166	0,193	0,102	-0,057	0,152	-0,115	0,138	0,106	0,278	-0,068	-0,020	-0,057	0,217	0,619	1	
diâmetro	-0,129	0,022	0,203	0,024	0,164	-0,226	0,038	0,032	0,269	-0,183	0,104	-0,116	0,249	0,510	0,325	1

Os valores em negrito são diferentes de 0 com um nível de significância $\alpha=0,05$

Trabalhos realizados por Campa e colaboradores (2005) revelam que em populações selvagens a produção de ACG pode ser acumulada sem estar ligado à cafeína. Eles observaram que os ACG sempre foram mais abundantes que a cafeína, indicando que toda a produção de ACG não está vinculado com o clorogenato de cafeína na semente. Ky e colaboradores (2013) confirmaram que de fato cada processo pode ser caracterizado por dois traços correlatos altamente independentes entre cafeína e ACG e que conforme o teor de cafeína aumenta o de ACG aumenta proporcionalmente, no entanto, essa relação não é absoluta. Outra situação é quando há menos cafeína do que ácidos clorogênicos, indicando que apenas uma parte dos ACG parece estar complexado.

Esses últimos resultados corroboram com os encontrados no nosso trabalho, sendo possível observar genótipos com altos teores de ácidos clorogênicos e baixo teores de cafeína, tais como: 267 e 289 que produziram 7,11 e 7,6g 100⁻¹ de ácidos clorogênicos e 1,25 e 1,04g 100⁻¹ de cafeína. Apesar dos parentais terem altos teores de ácidos clorogênicos como 6,67 g 100⁻¹ para L₁C₁ e 7,57 g 100⁻¹ para L₃C₃, possuem baixos teores de cafeína, 1,07 e 1,35g 100⁻¹), respectivamente. Baixos teores de ácidos clorogênicos também foram encontrados em indivíduos na população. Como utilizamos o acesso E335 proveniente da Etiópia, que produziu um baixo teor de ácidos clorogênicos de 3,57 g 100⁻¹ (FERREIRA, 2017), esse acesso (E335) pode ter influenciado na ocorrência de uma correlação negativa entre os dois compostos.

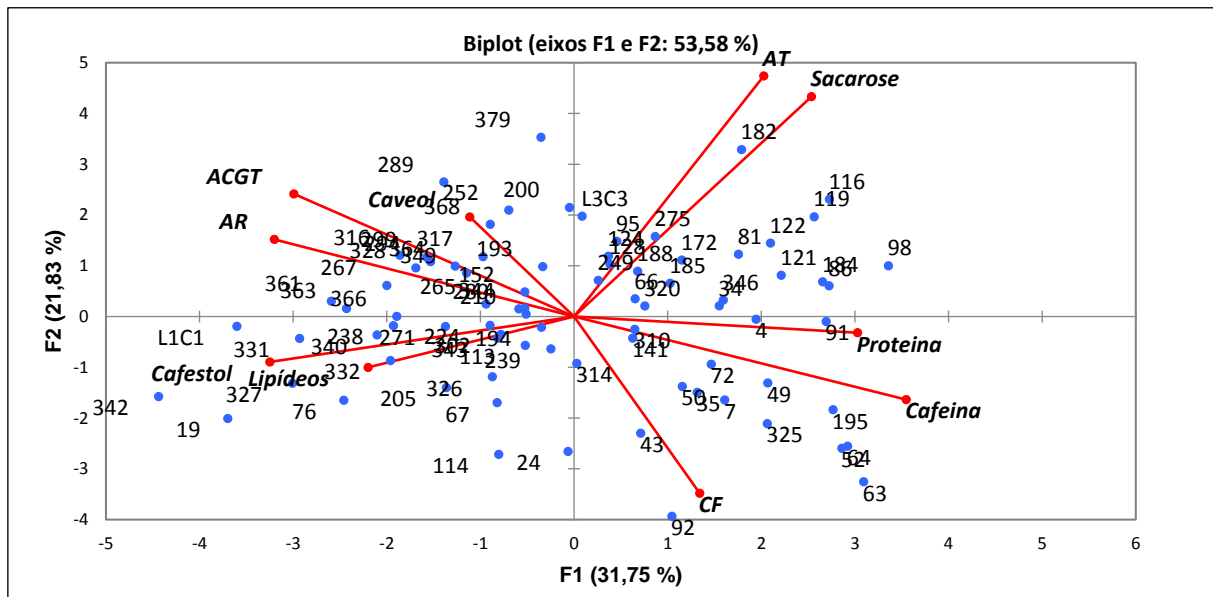
Como referência para o parental L₃C₃ temos os valores de Iapar 59, em trabalhos realizados por Scholz e colaboradores (2011), cultivar bem próxima de Tupi IAC, em que valores de referência são de 6,59% para teores de ácidos clorogênicos em plantas em condições de pleno sol e 1,26 e 1,23%, em plantas cultivadas em Itaguajé e Paranavaí.

Nós observamos também uma correlação negativa entre as concentrações de caveol e cafestol ($r = -0,136$). Trabalho realizado por Sant'Ana e colaboradores (2018) também identificaram correlação negativa entre caveol e cafestol ($r = -0,30$). Na etapa final de maturação do fruto existe uma inversão no acúmulo desses compostos e estudos sugerem que a síntese de um pode ser inversamente dependente do outro em sua via metabólica (IVAMOTO *et al.*, 2016).

5.4.4 Análise de Componente Principal (ACP) e Análise de Agrupamento Hierárquico (AAH)

Pela análise de componente principal (ACP) foi possível verificar simultaneamente a influência dos compostos na discriminação dos genótipos. Houve discriminação dos genótipos em razão da composição química de cada um. É possível observar na Figura 1 os genótipos e as variáveis, onde os dois primeiros eixos explicam 53,58% da variação total (31,75% para PC1 e 21,83% para a PC2). Os genótipos localizados a esquerda apresentam altos níveis de ácidos clorogênicos, caveol, açúcares redutores, cafestol e lipídeos. Enquanto que os genótipos a direita possuem maiores níveis de açúcares totais, sacarose, proteína, cafeína e compostos fenólicos.

Figura 1. Análise de componente principal dos compostos bioquímicos em 79 genótipos de uma população F₂ de *C. arabica*. AT: Açúcares totais; CF: compostos fenólicos; AR: açúcares redutores; ACTG: ácidos clorogênicos.



Alguns genótipos são aproximados de acordo com a alta produção de alguns compostos como é o caso do indivíduo 182 com alto nível de sacarose e açúcares totais, o indivíduo 4 e 91 para produção de proteínas e cafeína, uma vez que encontram-se bem próximos. É possível observar um agrupamento de genótipos que demonstram materiais com alto teor de ácidos clorogênicos e açúcares redutores e na mesma condição genótipos com alto teor de cafestol e lipídeos (Figura 1).

Observando-se a possível existência de grupos de acordo com a produção dos compostos químicos, foi feita uma análise de agrupamento hierárquico (AAH) e a construção

de um dendrograma com base em uma matriz de dissimilaridade (Figura 2). Foram formados três grupos com base na classificação de agrupamento hierárquico. Os genótipos que compuseram cada grupo são demonstrados na Tabela 3.

Figura 2. Dendrograma com base na análise de agrupamento hierárquico para os componentes químicos de 79 genótipos provenientes de uma população F₂ de *C. arabica*.

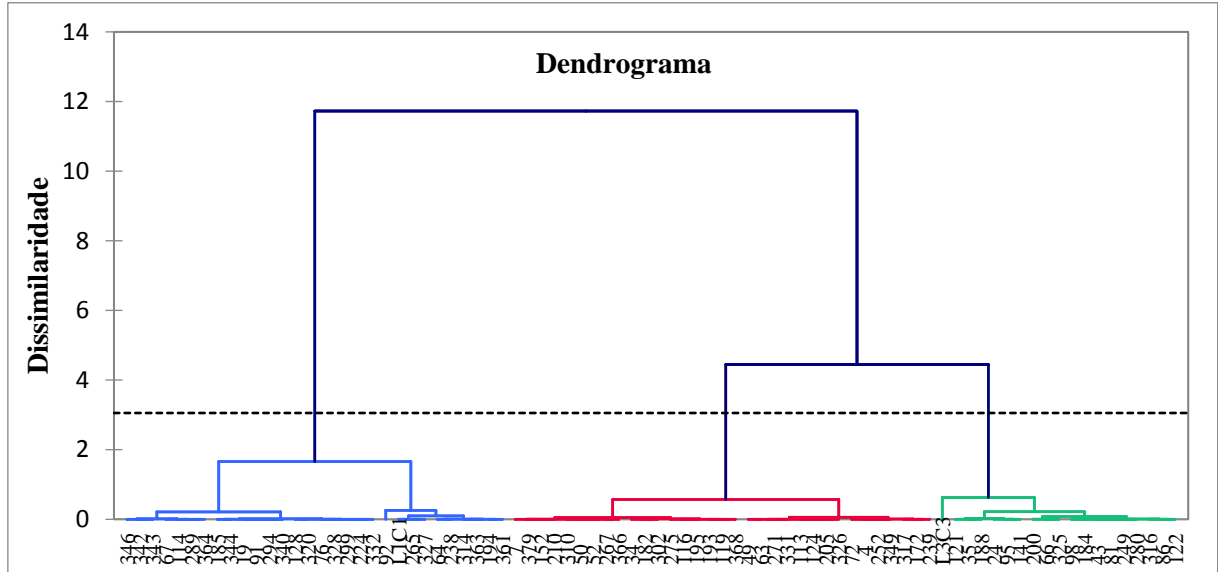


Tabela 3. Classificação de agrupamento hierárquico dos grupos a partir de 79 genótipos de uma população F₂ de *C. arabica*.

Classe	1	2	3
Objetos	32	28	19
	4	19	24
	7	64	35
	34	67	43
	49	76	66
	50	91	81
	52	92	86
	63	114	95
	72	128	98
	113	185	121
	116	194	122
	119	224	141
	124	238	184
	152	265	188
	172	289	200
	182	294	249
	193	299	280
	195	314	316
	205	320	325
	210	327	L3C3
	239	332	
	252	340	
	267	342	
	271	343	
	275	346	
	302	361	
	310	363	
	326	364	
	331	L1C1	
	349		
	366		
	368		
	379		

Para uma matriz de dissimilaridade, um agrupamento hierárquico foi realizado com base nos dados dos compostos de cada indivíduo da população (Tabela 4), formando assim os grupos com base na média. O primeiro grupo foi composto por 32 indivíduos, o segundo e o terceiro por 28 e 19 indivíduos, respectivamente. Os valores em negrito indicam o melhor grupo dentro de cada composto em relação a produção do composto.

Tabela 4. Valores médios para nove compostos do café divididos em três grupos pela análise de agrupamento hierárquico.

Classe	ACGT ^a	Cafeína ^a	Lipídeos ^a	Proteína ^a	Sacarose ^a	CF ^a	AR ^a	Cafestol ^b	Caveol ^b
1 (32)	5,35	1,29	14,02	14,77	6,46	5,54	0,38	529,44	690,23
2 (28)	5,98	1,21	14,19	14,46	6,13	5,56	0,40	676,78	616,67
3 (19)	5,22	1,31	14,08	14,82	6,56	5,66	0,35	376,26	758,19

^a Expresso em g 100⁻¹ ^b Expresso em mg 100⁻¹

Para teor de cafeína, a menor média foi observada no grupo 2. Se verificarmos os resultados da Tabela 1 observamos que a mínima foi de 0,91 e a máxima de 1,74, confirmando que não houve muita variação na produção da cafeína e que o teor do composto esta equivalente a trabalhos realizados anteriormente quando utilizado Catuai e Tupi (LINK *et al.*, 2014). A identificação de materiais com baixo teor de cafeína é interessante para produção de cafés descafeínados.

O grupo 3 se destaca por possuir 19 genótipos com baixo teor de ácidos clorogênicos que estão relacionados ao ciclo de maturação dos grãos, além do fator genético desses materiais que contribuem no perfil de baixa produção do composto. Quanto menor o teor desse composto, mais rapidamente os frutos atingem a maturação (JESZKA-SKOWRON *et al.*, 2016). Outro aspecto interessante desse grupo é a alta concentração de sacarose, diretamente envolvida na qualidade de bebida.

Também no grupo 3, observou-se genótipos com baixo teor de cafeol, sendo significativamente diferentes. Altos teores de cafeol foram observados nesse grupo em relação ao cafeol, e corroboram com os dados obtidos por Kitzberger *et al.*, (2013a) e dados fornecidos na Tabela 1. Podemos observar que em decorrência do cruzamento entre o parental L₁C₁ e L₃C₃ foi possível identificar genótipos que apresentam maiores teores de cafeol e menores teores de cafeol.

A partir das análises realizadas, tornou-se possível a seleção de genótipos de café promissores de acordo com a composição química observada. Eles deverão ser avaliados em análise sensorial. Portanto, para um perfil contendo baixo teor de ácidos clorogênicos, alto teor de sacarose e açúcares totais, baixo teor de cafeol e alto de cafeol, foi possível a seleção de 15 genótipos de acordo com dados fornecidos na Tabela Suplementar 2, foram eles: 81, 86, 95, 98, 116, 119, 121, 122, 182, 188, 193, 200, 252, 275 e 379, sendo a grande maioria presente no grupo 3. Essa seleção pode viabilizar a utilização desses genótipos para avanço de geração de autofecundação em programas de melhoramento para desenvolver cultivares com esses caracteres.

5.4.5 Genotipagem por sequenciamento (GBS) e detecção de SNPs

Através da genotipagem por sequenciamento de 189 indivíduos da população F₂, foram obtidos 417.938.880 *reads* filtrados para qualidade, que produziram 26.462.867 de *tags*. O alinhamento dos *tags* foi realizado utilizando os dois genomas de referência, de *C. canephora* e *C. eugenioides*. Cerca de aproximadamente 29% estavam alinhados em uma única posição e 21.495 SNPs foram identificados.

Os SNPs foram filtrados usando os seguintes parâmetros: a frequência do menor alelo ($MAF > 0.05$), call rate (> 0.80) reduzindo a quantidade para 4.833 SNPs nos 189 genótipos e 4.722 SNPs nos 79 genótipos. SNPs com $H_o > 0.9$ também foram descartados. Após a heterozigosidade, um total de 1.966 SNPs em 189 genótipos foram identificados, porém para os 79 genótipos que foram utilizados para o estudo de associação foram obtidos 1.917 SNPs.

5.4.6 Estudos de Associação Genômica Ampla (GWAS)

Para verificar a associação entre o fenótipo e os SNPs obtidos, foi utilizado o pacote mrMLM no programa R e aplicados quatro métodos diferentes (mrMLM, ISIS EM-BLASSO, pLARmEB e FASTmrEMMA) com 79 indivíduos da população F_2 para características morfológicas da planta e bioquímicas do grão, a fim de detectar regiões genômicas associadas com variação natural para as características de cafeol, cafeol, açúcares redutores, açúcares totais, cafeína, lipídeos, proteínas, ácidos clorogênicos, compostos fenólicos, diterpenos e sacarose, além das características morfológicas para tamanho dos grãos, altura, largura e diâmetro de caule das plantas.

Entre os quatro métodos testados, os que tiveram maior número de SNPs associados foram os métodos de ISIS EM_BLASSO e pLARmEB, com 22 e 29 marcas, respectivamente (Tabela 5). Para o métodos de mrMLM foram observadas 16 marcas, enquanto que nenhum SNP foi significativo pelo método de FASTmrEMMA.

Tabela 5. Marcadores SNPs de *C. arabica* mapeados nos genomas de *C. canephora* (S_C) e *C. eugenioides* (S_E) associados aos compostos relacionados a qualidade da bebida de café usando diferentes métodos de associação.

Características	SNP	MrMLm (p-value)	ISIS EM-BLASSO (LOD value)	pLARmEB (p-value)
ácidos clorogênicos	S8E_16969233	4.9979		
	S2E_49388794	4.0862		
	S10C_5135307		3.8517	0.0024
	S11C_12948382		4.2844	
	S1E_31925645		4.3785	2.00E-04
	S8E_11353682		3.9856	9.46E-08
	S1C_31015752			3.00E-04
	S1C_32642377			4.62E-06
	S4C_1885406			4.00E-04
	S6C_17723108			1.99E-05
	S11C_6427759			3.79E-06
S7E_9975489			9.00E-04	
açúcares redutores	S3C_8522081	5.777	3.8416	
	S11E_22478529	3.0967		
	S6C_17723108	3.2032		
	S2E_4741240		2.2925	
	S7E_3902786		2.5925	
cafestol	S1C_32686917	3.3726		7.08E-06
	S2C_4296485	4.3801		8.11E-05
	S4C_4715896	3.6347		4.29E-05
	S3E_8046550	3.3618		8.33E-05
	S2C_7520913		4.0817	
	S7C_15139626		2.794	
	S16E_3737965		4.8299	
Diâmetro do caule	S4C_16451686	3.0378		
	S9E_1950726	4.3334		
Largura da copa	S1C_25414931	3.5474	3.1939	1.00E-04
	S1C_27866832	3.1939	3.2095	1.00E-04
	S7C_4621935	3.2095	3.5474	5.30E-05
Proteína	S3E_8219473	3.7451		
	S4C_2042402	4.1269		
	S1C_27071416		8.7209	
	S1C_32106594		3.3806	
	S2C_896257		5.6759	
	S2C_24361040		4.5862	
	S5C_25610115		3.3926	
S9C_11285870		4.1111		
Tamanho de grão	S3C_23850757		2.5235	
	S2E_40017002		3.3578	
	S7E_4350651		2.9437	

Continuação Tabela 5. Marcadores SNPs de *C. arabica* mapeados nos genomas de *C. canephora* (S_C) e *C. eugenioides* (S_E) associados as compostos relacionados a qualidade da bebida de café usando diferentes métodos de associação.

Características	SNP	MrMLM (p- value)	ISIS EM-BLASSO (LOD value)	PLARmEB (p-value)
Cafestol	S1C_36080726			3.10E-05
	S2C_4296485			1.00E-04
	S4E_3737965			3.20E-05
Sacarose	S11C_14597913			0.0014
	S11C_27943863			6.00E-04
	S2E_45675190			3.00E-05
Açúcares totais	S2C_25020614			0.0011
	S2C_7142877			2.57E-07
	S5C_32341546			2.00E-04
	S7C_7833006			0.0013
	S11C_27943863			1.83E-05
	S2E_51785183			4.00E-04
	S6E_16419275			4.96E-05

Foram identificados 53 SNPs associados a 10 características, sendo 18 mapeados no genoma de *C. eugenioides* e 35 no *C. canephora* (Tabela 5). Dentre eles, 8 SNPs estavam presentes em dois métodos para as características ácidos clorogênicos, açúcares redutores e cafestol e 3 presentes nos três métodos associados para largura de planta. Enquanto que o método mrMLM é um método multi-locus, sem a necessidade da correção de Bonferroni para calcular o limiar p-value, enquanto que os métodos ISIS e pLARmEB são modelos Bayesianos.

Dos 11 SNPs, 8 são derivados de *C. canephora* e 3 de *C. eugenioides*. Recentemente Sant'Ana e colaboradores (2018) identificaram um total de 6.696 SNPs e após filtragem dos dados, foram obtidos 2.587 SNPs. Com o estudo de associação foram identificados 21 SNPs associados a lipídeos (5), cafestol (4), caveol (3) e cafestol/caveol (9) utilizando o mesmo método de análise (mrMLM, pLARmEB e ISIS EM-BLASSO), porém selecionaram apenas aqueles que foram identificados em pelo menos dois métodos. Eles encontraram SNPs posicionados dentro ou perto de regiões genômicas que codificam proteínas envolvidas em vias metabólicas lipídicas e de diterpenos.

A razão de termos encontrado os 3 SNPs para largura de planta nos três métodos é um indicativo de que esse SNP esteja fortemente associado a regiões genômicas ligadas a genes de interesse. Essa característica está ligada a estágios de desenvolvimento da planta, a eficiência do crescimento e a habilidade de adaptação às condições ambientais em que estas

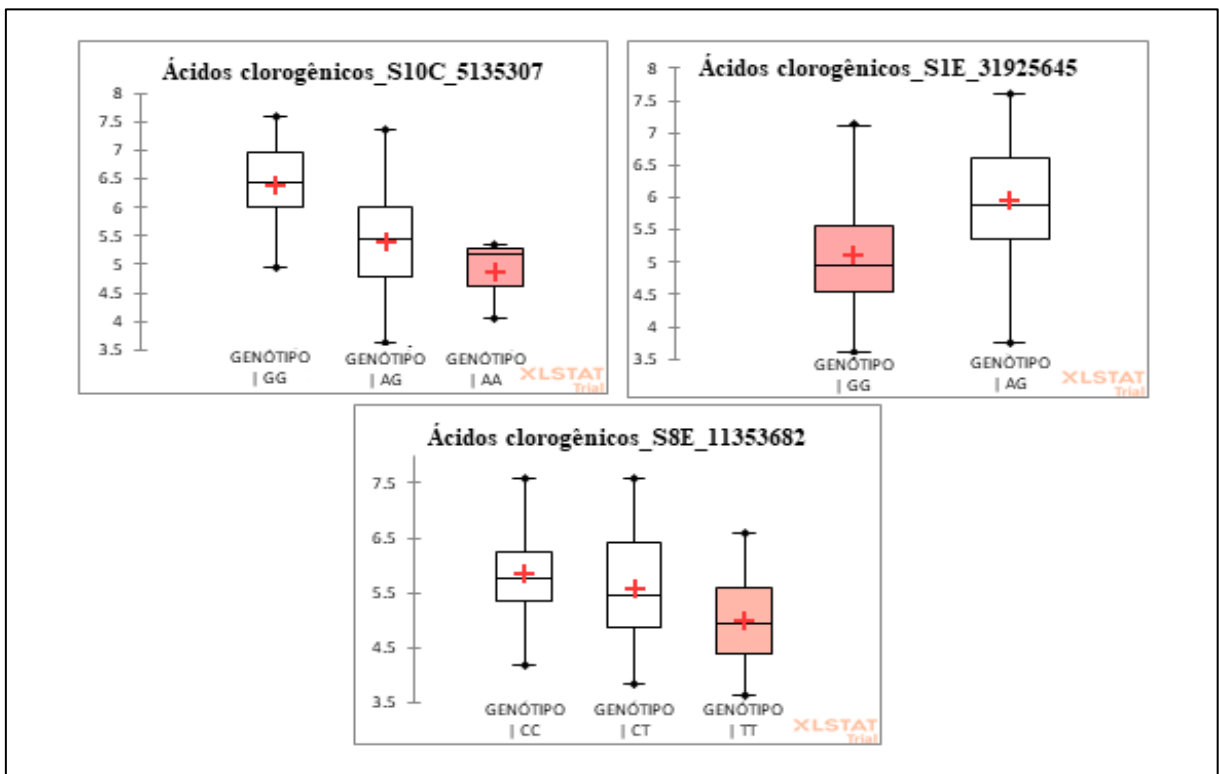
plantas crescem. Com a tendência cada vez mais de mecanização das lavouras cafeeiras, a identificação de marcadores ligados a largura da planta poderá ser importante para desenvolvimento de cafeeiros adaptados a condições de plantio recomendado para mecanização.

Para todas as outras características como caveol, açúcares totais, cafeína, lipídeos, proteínas, composto fenólico e altura, nenhum SNP foi associado.

5.4.7 Distribuição dos teores fenotípicos para os SNPs associados

Comparando os valores de ácidos clorogênicos para o SNP S10C_5135307 significativamente associado, observamos que os indivíduos com o genótipo homozigoto AA para o alelo alternativo apresentaram baixas concentrações de ácidos clorogênicos e o genótipo homozigoto GG para o alelo de referência, apresentaram altos teores de ácidos clorogênicos. Somente três indivíduos possuíam o genótipo com alelo AA, favorável a baixa produção do composto, enquanto que 66 possuíam o genótipo AG e 10 o genótipo GG (Figura 3). Esse marcador foi identificado no genoma de *C. canephora*.

Figura 3. Distribuição do teor de ácidos clorogênicos nos 79 genótipos para os SNPs associados.



Para o SNP S1E_31925645 também associado à produção de ácidos clorogênicos, o genótipo homozigoto GG está relacionado ao alelo de referência e à baixa produção do composto, enquanto que o alelo AG seria o alelo alternativo. Trinta e seis apresentaram indivíduos com o genótipo AG e 43 o genótipo GG. Para o SNP S8E_11353682, o indivíduo com o genótipo homozigoto TT para o alelo alternativo, foi o que apresentou menores concentrações de ácidos clorogênicos, enquanto que o genótipo CC foi o genótipo com o alelo de referência. Para o genótipo CC encontramos 19 indivíduos, 42 para o genótipo CT e 18 para o TT. Ambos os marcadores foram identificados no genoma de *C. eugenoides*.

Portanto, para os teores de ácidos clorogênicos, os haplótipos significativamente associados acima, sugerem que os indivíduos com os genótipos homozigotos (AA, GG e TT) apresentaram menores concentrações de ácidos clorogênicos em relação aos outros indivíduos da população. Sendo assim, a identificação desse conjunto de variantes genéticas (SNPs) associado a essa característica fenotípica, possibilita a identificação de regiões cromossômicas associadas a essa variação fenotípica, possibilitando a seleção de genótipos capazes de produzir uma bebida menos amarga, menos ácida, proporcionando ao café mais aroma e sabor.

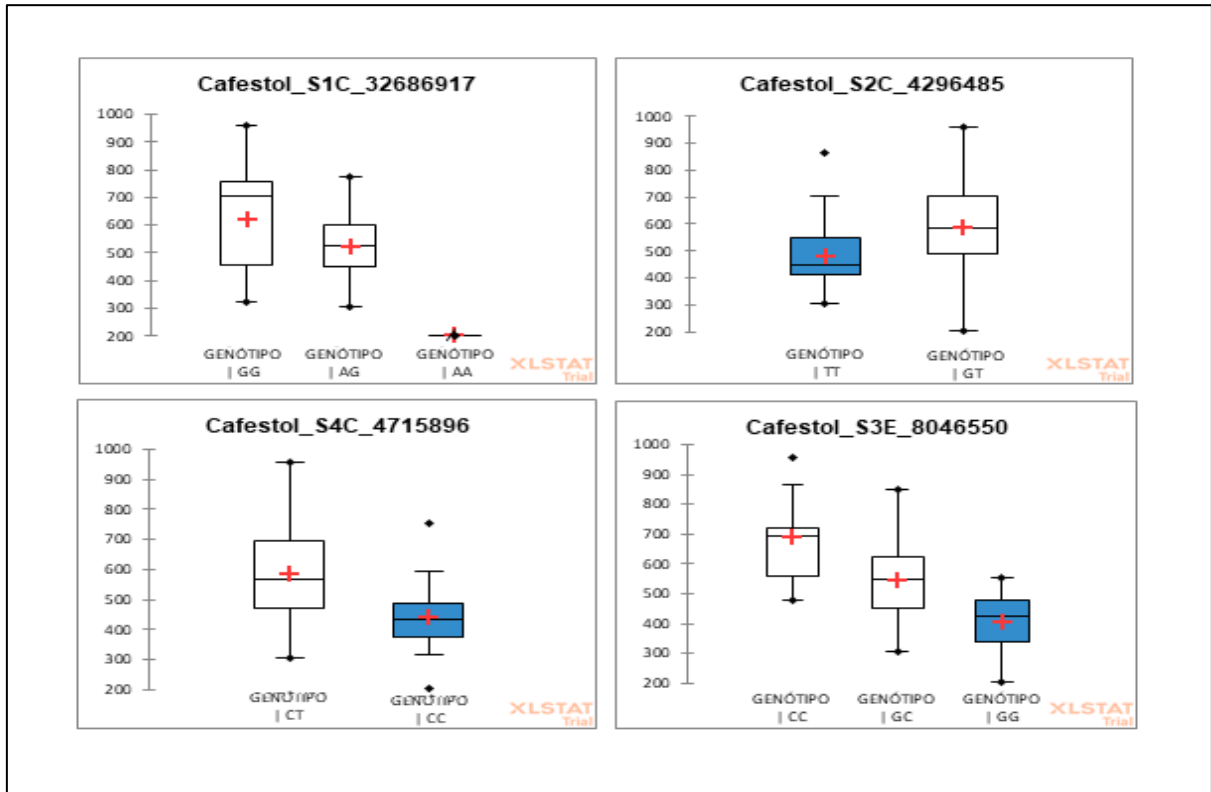
Leroy e colaboradores (2011) visando estudos sobre a qualidade da bebida identificaram QTLs (*Quantitative Trait Loci*) associados a várias características bioquímicas em *Coffea canephora*. Foram identificados QTLs relacionados a ácidos clorogênicos com efeitos variando de 22 a 35% da variação fenotípica. Esses QTLs foram co-localizados a genes candidatos (CCoAOMT), citados em estudos prévios (CAMPA *et al.*, 2003) e mais 3 genes (4CL) relacionados a biossíntese desse composto, sendo co-localizados nos cromossomos 2, 6, 7 e 8 de *C. canephora* (LEROY *et al.*, 2011), enzimas envolvidas na via de cadeia de produção do composto.

E o trabalho acima identifica QTLs específicos em um determinado estágio das plantas e de um grupo bem específico investigado, de uma descendência de retrocruzamentos entre grupos e por citar genes relacionados a biossíntese dos ácidos clorogênicos. Entretanto, vale ressaltar que, no nosso trabalho, não fizemos a busca por genes causais próximos aos SNPs identificados, porém foram encontrados associações entre SNPs e a produção de ácidos clorogênicos, que apesar de terem sido localizados em cromossomos diferentes, temos que considerar que foi realizado em uma população F₂ segregante específica utilizando *C. canephora* e *C. eugenoides* como genomas de referência.

Na Figura 4, comparando os valores de cafestol entre diferentes genótipos para o SNP S1C_32686917 significativamente associado, observamos que os indivíduos com o genótipo

homozigoto AA, com alelo alternativo, apresentou baixa concentração de cafestol e aqueles com genótipo homozigoto GG, estão associados com o alelo de referência. Apenas um indivíduo foi identificado com o genótipo AA, 57 com o genótipo AG e 21 com o genótipo GG.

Figura 4. Distribuição do teor de cafestol nos 79 genótipos para os SNPs associados.



Para o SNP S2C_4296485, observamos 34 indivíduos com o genótipo homozigoto TT, com alelo de referência, e foram os que apresentaram os menores teores de cafestol. Já o genótipo heterozigoto GT, com o alelo alternativo, apresentou 45 indivíduos com esse genótipo. Para o SNP S4_4715896 significativamente associado, os indivíduos com os genótipos heterozigotos CT, com o alelo de referência, apresentaram maiores teores de cafestol, somando 57 indivíduos. Os indivíduos homozigotos CC, 22 indivíduos e com o alelo de referência foram os que apresentaram menores teores de cafestol.

Foram identificados três SNPs associados ao genoma de *C. eugenioides*, o SNP S1E_31925645, S8E_11353682 citado acima e o SNP S3E_8046550 significativamente associado a teores de cafestol (Figura 4). Os indivíduos com o genótipo homozigoto GG, com o alelo alternativo, foram identificados com baixos teores de cafestol. O genótipo homozigoto CC, com alelo de referência com altos teores de cafestol. Observamos 9 indivíduos com o genótipo GG, 61 com o genótipo GC e 9 com o genótipo CC.

Sendo assim, os haplótipos significativamente associados sugerem que a presença dos genótipos AA, TT, CC e GG estão relacionados ao baixo teor de cafestol. Consumo de café com altos teores de cafestol pode aumentar o nível de colesterol no sangue (RENDÓN *et al.*, 2017), portanto a identificação de genótipos com baixo teor desse composto na bebida do café é altamente promissoras. Assim, a identificação de cafés com baixo teor do composto é de grande interesse dos programas de melhoramento.

Em trabalho realizado por Silva (2018), foram identificados 17 SNPs de *C. arabica* mapeados em *C. canephora*, co-localizados a 13 genes candidatos envolvendo vias metabólicas de compostos no café, porém somente 4 associados a teores de diterpenos. Um bloco de haplótipos (S2_14.156.413, S2_14.156.429, S2_14.156.434 e S2_14.156.457) foi significativamente associado a cafeol e razão caf/cav e co-localizados aos genes (Cc02_g15760 e Cc02_g15870), responsáveis pela codificação de duas enzimas do citocromo P450.

Sant'Ana e colaboradores (2018) utilizaram o genoma de referência de *C. canephora* e encontraram 4 SNPs associados a teores de cafestol, localizados nos cromossomos 3, 6 e 11 (S3_7990620, S6_7853861, S11_29778697 e S11_30776239). Destes, dois SNPs S6_7853861 e S11_29778697 foram localizados próximos a genes para monooxigenases que podem estar envolvidas na biossíntese de metabólitos secundários.

No nosso trabalho identificamos os haplótipos significativamente associados para teores de cafestol, nos cromossomos 1, 2 e 4 de *C. canephora*. Um dos SNPs (S2C_4296485) foi localizado no cromossomo 2, porém bem distante dos apresentados por Silva (2018). Não localizamos nenhum SNP nos cromossomos 6 e 11 conforme Sant'Ana e colaboradores (2018). Vale ressaltar que os trabalhos citados acima foram realizados utilizando acessos de *C. arabica* oriundos da Etiópia, enquanto que o presente estudo utilizou uma população F₂ segregante específica, sendo assim, não necessariamente deveríamos encontrar os mesmos SNPs.

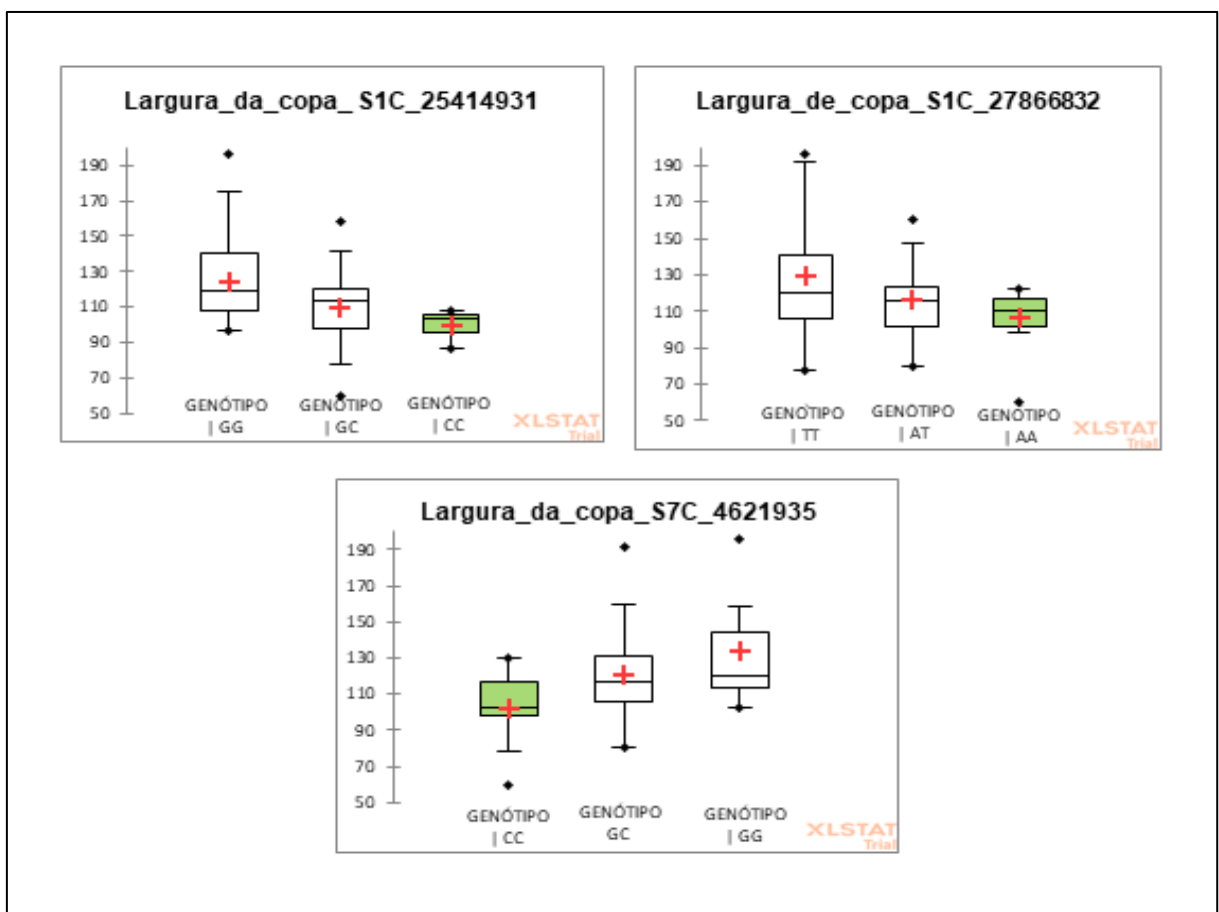
Outro SNP também foi localizado no cromossomo 3 de *C. eugenioides* (S3E_8046550). No caso do cafestol, a presença do SNP, estaria associada a uma redução significativa na produção do composto. Uma outra etapa da pesquisa será investigar essas regiões, na qual foram localizados esses SNPs, na tentativa de localizarmos genes candidatos próximos que poderiam estar influenciando na concentração do composto nesses materiais.

Para resultados gerados a partir da medição da largura da copa das plantas no campo (Figura 5), encontramos 3 SNPs associados, 2 alocados no cromossomo 1 e o outro no cromossomo 7. Na figura 3, para o SNP S1C_25414931, observamos que o genótipo

homozigoto CC, para o alelo alternativo, apresentou menor comprimento referente a largura das plantas e o genótipo homozigoto GG, para o alelo de referência, apresentou plantas com maior largura da copa. Três apresentaram o genótipo CC, 30 o genótipo GC e 46 o genótipo GG. Para o SNP S1C_27866832, o genótipo homozigoto AA para o alelo alternativo, apresentou menor comprimento de largura de copa, em contrapartida, o genótipo homozigoto TT para o alelo de referência apresentou valores superiores. Doze apresentaram genótipos AA, 48 AT e 19 TT.

Para o SNP S7C_4621935 o genótipo homozigoto CC, para o alelo de referência apresentou menor largura de copa, enquanto que o genótipo homozigoto GG, para o alelo alternativo, apresentou maior largura de copa, com 17 indivíduos com o genótipo CC, 54 com o genótipo CG e 8 para o GG (Figura 5). Sendo assim, os blocos de haplótipos significativamente associados sugerem que indivíduos com a presença dos genótipos homozigotos CC e AA possuem uma baixa largura das copas em cafeeiros.

Figura 5. Distribuição relacionado a largura da copa dos cafeeiros nos 79 genótipos para os SNPs associados.



Moncada e colaboradores (2016) realizaram um mapeamento de QLT, na qual obtiveram 2 QTLs associados a altura da plantas e tamanho de grão e 1 QTL associado a rendimento. Devido a correlação entre as características avaliadas por moncada e colaboradores (2016) podemos levar em consideração informações a respeito dos QTLs encontrados para altura de planta. Os dois QTLs altamente significativos encontrados para altura de planta explicaram 25% da variação fenotípica no grupo de ligação 1 e o outro explicou 15,6% da variação fenotípica no grupo de ligação 4.

Características como altura, largura de planta e quantidade de frutos produzidos, são muito importantes para o desenvolvimento de qualquer cultura, pois avaliam o rendimento agrícola. Em culturas perenes como o café, são necessárias medições por pelo menos cinco anos em cada geração (MONCADA *et al.*, 2016). Entretanto, nosso objetivo, ao fenotipar altura, largura de planta e diâmetro do caule, foi avaliar o desenvolvimento e densidade das plantas no campo.

O parental L₁C₁, vem de um cruzamento de Catuaí Vermelho IAC46, que por sua vez possui como uma das características, porte pequeno de plantas (PEREIRA *et al.*, 2010). Já o acesso da Etiópia E335, produz plantas de porte alto e Sarchimor IAC 1669-33, plantas de porte baixo. Com o objetivo de reduzir a altura de plantas, estudos foram reportados envolvendo a introdução do gene Ct (Caturra) em genótipos, visando redução da altura das plantas a fim de facilitar tanto a colheita manual, quanto a mecânica (CARVALHO e MÔNACO, 1972).

O pequeno porte presente em cafés como Villa Sarchi parecem estar associados a esse fator denominado fator caturra, controlados por um par de alelos dominantes, Caturra (Ct). Catuaí também são portadores do mesmo fator genético que reduz o porte das plantas (Ctct), originários dos parentais Villar Sarchi e Caturra (LARA *et al.*, 2014).

Análises genéticas realizadas por um grupo de pesquisadores do IAC, sugerem a existência de três a cinco locos gênicos envolvendo a expressão desse caráter. Esse gene Ct, por promover não só o encurtamento dos internós no eixo principal, mas sim dos ramos laterais (CARVALHO *et al.*, 1991), também está envolvido e influencia no comprimento da largura da copa.

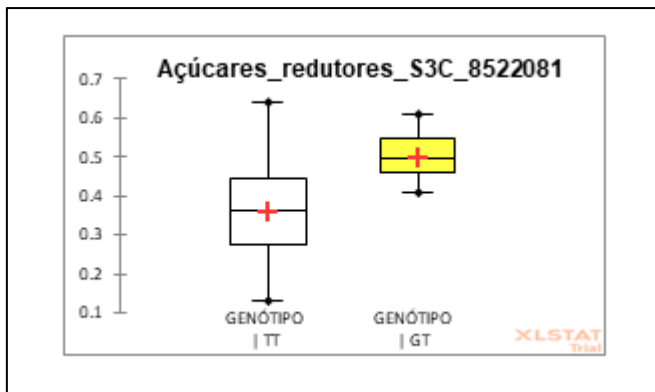
No presente estudo, foi possível identificar plantas com uma variedade de tamanhos na população F₂ segregante e inclusive genótipos com pequeno porte, sendo possível verificar vários níveis para expressão dessa característica.

Cafeeiros de pequeno porte permitem uma maior densidade de plantio no campo, preferencialmente plantas com baixa estatura e alto rendimento (PEREIRA *et al.*, 2010;

FERNANDES, *et al.*, 2012). A identificação de SNPs associados a essa característica fenotípica, valoriza ainda mais os resultados obtidos no nosso estudo.

Para o composto açúcares redutores encontramos um SNP significativamente associado (Figura 6). O SNP S3C_8522081 o genótipo heterozigoto GT, com o alelo alternativo sendo o G e o de referência o T, foi o que possuiu maiores teores de açúcares redutores, enquanto que o genótipo TT apresentou menores concentrações. Apenas 12 indivíduos apresentaram o genótipo GT, enquanto que 67 apresentaram o genótipo TT. Portanto, o genótipo heterozigoto GT estaria influenciando a alta produção de açúcares redutores nos genótipos em estudo.

Figura 6. Distribuição do teor de açúcares redutores nos 79 genótipos para os SNPs associados.



Vale lembrar que a variação nos teores de açúcares pode ocorrer devido a adaptação dos materiais ao local e a diferença de resposta das cultivares ao ambiente e que durante a produção do grão, o teor de sacarose tende a aumentar e o de açúcares redutores a diminuir. Entretanto, estudos revelaram que a diversidade genética confere variabilidade na composição química do café, permitindo a discriminação de cultivares modernas e tradicionais (KITZBERGER *et al.*, 2013b).

As condições de cultivo no presente estudo foram as mesmas e as plantas cultivadas em um mesmo local e a colheita dos frutos realizada no mesmo ponto de maturação visual, a fim de minimizar possíveis interferências nas análises bioquímicas. Sendo assim, a variação dos açúcares pode estar associada a variabilidade genética dos genótipos, propiciando a associação do SNP.

5.5 CONCLUSÃO

A análise da composição química da população F₂ de *C. arabica* revelou a variabilidade existente nesses materiais, com um grupo de genótipos com baixo teor de cafeol e alto de cafeína, baixo teor de cafeína, além de alto teor de açúcar. Com o conhecimento da composição química dos grãos, foi possível identificar grupos com características bioquímicas em comum. É possível verificar vários níveis intermediários para expressão de algumas características, o que poderia indicar características controladas por um determinado número de genes de efeitos secundários (*minor genes* em inglês), localizados em diferentes loci, que conjuntamente afetam a expressão da característica ou até mesmo uma segregação transgressiva para várias características. Dessa forma, com base na composição química e tamanho de grão, torna-se possível selecionar genótipos promissores para uma análise sensorial para verificar a qualidade da bebida.

Foram identificados 11 SNPs potencialmente associados a características agronômicas de interesse. É possível afirmar que os SNPs foram significativamente associados aos compostos ácidos clorogênicos, açúcares redutores e largura de copa. A análise da frequência alélica dos SNPs, de acordo com os fenótipos associados nos Box-pots demonstram o potencial uso desses SNPs nos programas de melhoramento. Também se abre possibilidade de identificar genes próximos a essas marcas e que possam estar relacionados as características de interesse.

Os resultados obtidos revelam importante variabilidade genética na população, evidenciando a possibilidade de seu uso em programas de melhoramento.

5.6 REFERÊNCIAS

ADDINSOFT. XLStat: Software for Statistical Analysis. Paris. Versão 2010, 2010.

ADDINSOFT. XLSTAT: your data analysis solution. Paris: Versão 2018, 2018.

ANTHONY, F.; CLIFFORD, M.N.; NOIROT, M. Biochemical diversity in the genus *Coffea* L.: chlorogenic acids, caffeine and mozambioside contents. **Genetic Resources and Crop Evolution**, 40(2):61-70, 1993.

ANNICCHIARICO, P.; NAZZICARI, N.; PECETTI, L.; ROMANI, M.; FERRARI, B.; WEI, Y.; BRUMMER, E.C. GBS-Based Genomic Selection for Pea Grain Yield under Severe Terminal Drought. **Plant Genome**, v.10(2), 2017.

BHAT, J.A.; ALI, S.; SALGOTRA, R.K.; MIR, Z.A.; DUTTA, S.; JADON, V.; TYAGI, A.; MUSHTAQ, M.; JAIN, N. ; SINGH, P.K.; SINGH, G.P.; PRABHU, K.V. Genomic Selection in the Era of Next Generation Sequencing for Complex Traits in Plant Breeding., **Frontiers in Genetics**. v.7, 221p., 2016

BRADBURY, P.J., ZHANG, Z., KROON, D.E., CASSTEVENS, T.M., RAMDOSS, Y. AND BUCKLER, E.S. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 23, 2633-2635, 2007.

BRASIL. Decreto-lei n. 4.629, de 21 de março de 2003. Estabelece critérios técnicos de identidade e de qualidade para a classificação do café beneficiado grão cru. Diário Oficial [da] República Federativa do Brasil, Brasília, DF, 2003. Seção 1, p. 4. Disponível em: <<http://www.agricultura.gov.br/legislacao/SISLEGIS>>. Acesso em: 07 mar. 2016.

CAIXETA, E.T.; FERRÃO, L.F.V.; ZAMBOLIM, E.M. Marcadores Moleculares. 2: 31-68 In: BORÉM, A.; FRITSCHÉ NETO, R. **Biotecnologia aplicada ao melhoramento de plantas**. Viçosa: Universidade Federal de Viçosa, p.335, 2013.

CAMPA, C.; NOIROT, M.; BOURGEOIS, M.; PERVENT, M.; KY, C.L.; CHRESTIN, H.; HAMON, S.; DE KOCHKO, A. Genetic mapping of a caffeoylcoenzyme A 3-O-methyltransferase gene in coffee trees. Impact on chlorogenic acid content. **Theor Appl Genet**, v.107:751–756, 2003.

CAMPA, C.; DOULBEAU, S.; DUSSERT, S.; HAMON, S.; NOIROT, M. Qualitative relationship between caffeine and chlorogenic acid contents among wild *Coffea* species. **Food Chemistry**, v.93, 135-139, 2005.

CARVALHO, A.; MÔNACO, L.C. Transferência do fator caturra para o cultivar Mundo Novo de *Coffea arabica*. **Bregantia**, v.31:379-399, 1972.

CARVALHO, A.; MEDINA FILHO, H.P.; FAZUOLI, L.C.; GUERREIRO FILHO, O.; LIMA, M.M. Aspectos genéticos do cafeeiro. **Revista Brasileira de Genética**, v.14,135-183, 1991.

CARVALHO, C.H.S et al. Cultivares de Café Arabica de Porte Baixo. 9: 157-226 In: CARVALHO, H.S.C. (Ed.). **Cultivares de café: origem, características e recomendações**. Brasília, DF: Embrapa Café, 334p, 2008.

CHARMETANT, P.; LEROY, T. Genome-wide association studies in *Coffea arabica* using GBS. In: **The 26th International Conference on Coffee Science – ASIC**. Kunming, China, 2016.

CHEN, W.; GAO, Y.; XIE, W.; GONG, L.; LU, K.; WANG, W.; LI, Y.; LIU, X.; ZHANG, H.; DONG, H.; ZHANG, W.; ZHANG, L.; YU, S.; WANG, G.; LIAN, X.; LUO, J. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. **Nature genetic**, 46: 714-721, 2014.

CHENG, B; FURTADO, A; SMYTH, H.E; HENRY, R.J. Influence of genotype and environment on coffee quality. **Trends Food Sci Technol**, 57:20–30, 2016.

CLARKE, W.E.; HIGGINS, E.E.; PLIESKE, J.; WIESEKE, R.; SIDEBOTTOM, C.; KHEDIKAR, Y.; BATTLE, J.; EDWARDS, D.; MENG, J.; LI, R.; LAWLEY, C.T.; PAUQUET, J.; LAGA, B.; CHEUNG, W. A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. **Theor. Appl. Genet.**, v.129, 1887-1899, 2016.

CLIFFORD, M.N.; OHIOKPEHAI, O. Coffee astringency. **Anal. Proc.** v. 20, 83-86. 1983.

CONAB – Companhia Nacional de Abastecimento. Compêndio de Estudos Conab. **A cultura do café: Análise dos custos de produção e da rentabilidade nos anos 2008 a 2017**. v.12, 2017.

COZZOLINO, D. Use of Infrared Spectroscopy for In-Field Measurement and Phenotyping of Plant Properties: Instrumentation, Data Analysis, and Examples. **Applied Spectroscopy Reviews**, v.49, 7, 2014.

CUBRY, P.; BELLIS, F.; AVIA, K.; BOUCHET, S.; POT, D.; DUFOUR, M.; LEGNATE, H.; LEROY, T. An initial assessment of linkage disequilibrium (LD) in coffee trees: LD patterns in groups of *Coffea canephora* Pierre using microsatellite analysis. **BMC Genomics**, 16;14:10, 2013.

D'AMELIO, N. FONTANIVE, L.; UGGERI, F.; SUGGI-LIVERANI, F.; NAVARINI, L. NMR reinvestigation of the caffeine–chlorogenate complex in aqueous solution and in coffee brews. **Food Biophys**, 4:321–330, 2009.

DAVIS, A.P.; TOSH, J.; RUCH, N.; FAY, M.F. Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Botanical Journal of the Linnean Society* 167(4): 357-377, 2011.

DEBIBAKAS, S.; ROCHER, S.; GARSMEUR, O.; TOUBI, L.; ROQUES, D.; D'HONT, A.; HOARAU J.Y.; DAUGROIS, J.H. Prospecting sugarcane resistance to Sugarcane yellow leaf virus by genome-wide association. **Theor Appl Genet**, 127:1719–1732, 2014.

DEL GROSSI, L.; SERA, T.; SERA, G. H.; FONSECA, I.C.D.B.; ITO, D.S.; SHIGUEOKA, L. H.; CARVALHO, F.G. Rust resistance in Arabic Coffee cultivars in northern Paraná. **Brazilian Archives of Biology and Technology**, v.56(1): 27-33, 2013.

DENOEUDE, F.; CARRETERO-PAULET, L.; DEREPPER, A.; DROC, G.; GUYOT, R.; PIETRELLA, M.; ZHENG C.; ALBERTI, A.; ANTHONY, F.; APREA, G.; AURY, J.M.; BENTO, P.; BERNARD, M. et. al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v.345:6201, 1181-4, 2014.

ELSHIRE, R. J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K.; BUCKER, E.S.; MITCHELL, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS One** 6, 2011.

FAO, Coffea Mission to Ethiopian, 1964-54. FAO, Rome, Italy, 1968.

FERNANDES, A.L.T.; PARTELLI, F.L.; BONOMO, R.; GOLYNSKI, A. A moderna cafeicultura dos cerrados brasileiros. **Pesquisa Agropecuária Tropical**, v.42, 231–240, 2012.

FERRÃO, L.F.V.; CAIXETA, E.T.; PENA, G.; ZAMBOLIM, E.M.; CRUZ, C.D.; ZAMBOLIM, L.; FERRAO, M.A.G.; SAKIYAMA, N.S. New EST–SSR ~ markers of *Coffea arabica*: transferability and application to studies of molecular characterization and genetic mapping. **Mol Breed** 35(1):1–5, 2015.

FERREIRA, M.E.; RANGEL, P.H.N. Aporte biotecnológico ao pré-melhoramento vegetal. 3: 61-84 In: LOPES, M.A.; FAVERO, A.P.; FERREIRA, M.A.J. F.; FALEIRO, F.G.; FOLLE, S.M.; GUIMARÃES, E.P. **Pré-melhoramento de plantas: estado da arte e experiências de sucesso**. Brasília, DF: Embrapa Informação tecnológica, p.614, 2011.

FERREIRA, R.V. **Fenotipagem e estudos de associação genômica ampla (GWAS) em uma coleção de *Coffea arabica***. [Dissertação de Doutorado] Londrina: Pós-Graduação em Genética e Biologia Molecular – Universidade Estadual de Londrina, 2017.

HAMON, P.; GROVER, C.E.; DAVIS, A.P.; RAKOTOMALALA, J.J.; RAHARIMALALA, N.E.; ALBERT, V.A.; SREENATH, H.L.; STOFFELEN, P.; MITCHELL, S.E.; COUTURON, E.; HAMON, S.; KOCHIKO, A.; CROUZILLAT, D.; RIGOREAU, M.; SUMIRAT, U.; AKAFFOU, S.; GUYOT, R. Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species: GBS coffee phylogeny and the evolution of caffeine content. **Mol. Phylogenet. Evol.**, April 109:351-361, 2017.

HEALEY, A.; FURTADO, A.; COOPER, T.; HENRY, R.J. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. **Plant Methods**, 10:21, 2014.

IAC - Instituto Agronômico de Campinas. **Centro de café**. Disponível em: <http://www.iac.sp.gov.br/areasdepesquisa/cafe/centrodecafe.php>. Acesso em: 23, fevereiro, 2018.

INTERNATIONAL COFFEE ORGANIZATION (ICO). Disponível em: <http://www.ico.org/prices/po-production.pdf> Acesso em: 26 Jan. 2018.

IVAMOTO, S.T.; POT, D, CHARMETANT P, MARRACCINI P, FERREIRA LP, DOMINGUES DS, VIEIRA, LGE, PEREIRA, LFP. Diterpenes in *Coffea arabica*: biochemical aspects and transcriptional analysis of candidate CYPs involved in cafestol and kahweol biosynthesis. **Plant Physiol Biochem**, 111, 340-347. 2016.

JESZKA-SKOWRON M, SENTKOWSKA A, PYRZYN'SKA K, DE PEÑA MP Chlorogenic acids, caffeine content and antioxidant properties of green coffee extracts: influence of green coffee bean preparation. **Eur Food Res Technol** 242:1403–1409, 2016.

KIM, N., JEONG, Y.M.; JEONG, S.; KIM, G.B.; BAEK, S.; KWON, Y.E. Identification of candidate domestication regions in the radish genome based on high-depth resequencing analysis of 17 genotypes. **Theor Appl Genet.**129: 1797–1814, 2016.

KITZBERGER, C.S.G.; SCHOLZ, M.B. dos S.; PEREIRA, L.F.P.; VIEIRA, L.G.E.; SERA, T.; SILVA, J.B.G.; BENASSI, M.T. Diterpenes in green and roasted coffee of *Coffea arabica* cultivars growing in the same edapho-climatic conditions. **Journal of Food Composition and Analysis**, v.30, p.52-57, 2013a.

KITZBERGER, C.S.G.; SCHOLZ, M.B. dos S.; PEREIRA, L.F.P.; BENASSI, M.T. Composição química de cafés arábica de cultivares tradicionais e modernas. **Pesq. agropec. bras.**, Brasília, v.48, n.11, 1498-1506, 2013b.

KY, C. L., LOUARN, J., DUSSERT, S., GUYOT, B., HAMON, S., NOIROT, M. Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. *Food Chemistry*, 75, 223–230, 2001.

KY, C.L.; BARRE, P.; NOIROT, M. Genetic investigations on the caffeine and chlorogenic acid relationship in an interspecific cross between *Coffea liberica dewevrei* and *C. pseudozanguebariae*. **Tree Genetics and Genomes**, v.9, 1043-1049, 2013.

LARA, J.M.R.A.; REZENDE, J.C.D.; CARVALHO, G.R.; PEREIRA, A.A.; CARVALHO, J.P.F.; GUERREIRO FILHO, O. Caracterização de formas botânicas diversas do Banco Ativo de Germoplasma de cafeeiros do Estado de Minas Gerais, **Bragância-Campinas**, v.73(4): 383-389, 2014.

LASHERMES, P.; COMBES, M.C.; ROBERT, J.; TROUSLOT, P.; D'HONT, A.; ANTHONY, F.; CHARRIER, A. Molecular characterization and origin of the *Coffea arabica* L. **Genome. Mol.Genet.**, v.261, 259-266, 1999.

LEROY, T.; BELLIS, F.D.; LEGNATE, H.; KANANURA, E.; GONZALES, G.; PEREIRA, L.F.; ANDRADE, A.C.; CHARMETANT, P.; MONTAGNON, C.; CUBRY, P.; MARRACCINI, P.; POT, D.; KOCHKO, A. Improving the quality of African Robustas: QTLs for yield and quality related traits in *coffea canephora*. **Tree genetics and genomes**, v.7:781-798, 2011.

LINK, J.V.; LEMES, A.L.G.; MARQUETTI, I.; SCHOLZ, M.B.S.; BONA, E. Geographical and genotypic segmentation of arabica coffee using self-organizing maps. **Food Research Internation**, v.59, 1-7, 2014.

LU, F.; LIPKA, A.E.; GLAUBITZ1, J.; ELSHIRE, R.; CHERNEY, J.H.; CASLER, M.D.; BUCKLER, E.S.; COSTICH, D.E. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery. **Plos one**, 9(1): p.1-14, 2013.

MENDONÇA, L.M.V.L.; PEREIRA, R.G.F.A.; MENDES, A.N.G.; BORÉM, F.M.; MARQUES, E.R. Composição química de grãos crus de cultivares de *Coffea arabica* L. suscetíveis e resistentes à *Hemileia vastatrix* Berg et Br. **Ciência e Agrotecnologia**, v.31, p.413-419, 2007.

MOHAN, S.K.; CARDOSO, R.M.L.; PAIVA, M.A. Resistência em germoplasma de *Coffea* ao crestamento bacteriano incitado por *Pseudomonas garcae* Amaral et al. **Pesquisa Agropecuária Brasileira**, Brasília, v.13(1): 53-64, 1978.

MONCADA, M. P.; TOVAR, E.; MONTOYA, J.C.; GONZALEZ, A.; SPINDEL, J.; MCCOUCH, S. A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size. **Tree Genet. Genom.**, 12, 5, 2016.

MUNIZ, M. D. F. S., CAMPOS, V. P., MOITA, A. W., GONÇALVES, W., ALMEIDA, M. R. A., SOUSA, F. R. D., & CARNEIRO, R. M. D. Reaction of coffee genotypes to different populations of *Meloidogyne* spp.: detection of a naturally virulent *M. exigua* population. **Tropical Plant Pathology**, v.34(6): 370-378, 2009.

NAIDOO, N; CHEN, C; REBELLO, S. A; SPEER, K; TAI, E. S; LEE, J & VAN DAM, R. M. Cholesterol-raising diterpenes in types of coffee commonly consumed in Singapore, Indonesia and India and associations with blood lipids: A survey and cross-sectional study. **Nutr J**, 10, 48, 2011.

OLIVEIRA, A. P. L. R.; CORRÊA, P. C.; REIS, E. L.; & DE OLIVEIRA, G. H. H. Comparative Study of the Physical and Chemical Characteristics of Coffee and Sensorial Analysis by Principal Components. **Food Analytical Methods**, v.8, n.5, 1303–1314, 2014.

PEREIRA, M.C.; CHALFOUN, S.M.; CARVALHO, G.R.; SAVIAN, T.V. Multivariate analysis of sensory characteristics of coffee grains (*Coffea arabica* L.) in the region of upper Paranaíba. *Acta Scientiarum Agronomy*, v.32, n.4, 2010.

PEREIRA, L.F.P.; SANT'ANA, G.C.; FERREIRA, R.; SILVA, B.; NOGUEIRA, L.; PADILHA, L.; PAGIATTO, N.; KITZBERGER, C.; SCHOLZ, M.B.; SERA, G.; DOMINGUES, D.; GUYOT, R.; POT, D.; CHARMETANT, P.; LEROY, P. Genome-wide association studies in *Coffea arabica* using GBS. In: **The 26th International Conference on Coffea Science** – ASIC. Kunming, China, 2016.

RENDÓN, M. Y.; SANTOS SCHOLZ, M. B.; BRAGAGNOLO, N. Is cafestol retained on the paper filter in the preparation of filter coffee? **Food Research International**, v.100, 798–803, 2017.

RESENDE, M.D.V.; SILVA, F.F.; RESENDE JUNIOR, M.F.R. Genética de Associação (GWS). 4: 119-150. In: BORÉM, A.; FRITSCHÉ NETO, R. **Biotecnologia aplicada ao melhoramento de plantas**, Viçosa: Universidade Federal de Viçosa, 335p., 2013.

RIMBERT, H.; DARRIER, B.; NAVARRO, J.; KITT, J.; CHOULET, F.; LEVEUGLE, M.; DUARTE, J.; RIVIÈRE, N.; EVERSOLE, K.; GOUIS, J.L. ; DAVASSI, A.; BALFOURIER, F.; PASLIER, M.C.; BERARD, A.; BRUNEL, D.; FEUILLET, C.; PONCET, C.; SOURDILLE, P.; PAUX, E. High throughput SNP discovery and genotyping in hexaploid wheat. **Plos One**, 1-19, 2018.

ROMERO, G.; VÁSQUEZ, L.M.; LASHERMES, P.; HERRERA, J.C. Identification of a major QTL for adult plant resistance to coffee leaf rust (*Hemileia vastatrix*) in the natural Timor hybrid (*Coffea arabica* x *C. canephora*). **Plant Breeding**, v.133, 121–129, 2014.

SANT'ANA, G.C.; PEREIRA, L.P.P; POT, D.; IVAMOTO, S.T.; DOMINGUES, D.S.; FERREIRA, R.V.; PAGIATTO, N.F.; SILVA, B.S.R.; NOGUEIRA, L.M.;KITZBERGER, C.S.G; SCHOLZ, M.B.S.; OLIVEIRA, F.F.; SERA, G.H.; PADILHA, L.; LABOUISSSE, J.P.; GUYOT, R.; CHARMETANT, P. LEROY, T. Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. **Scientific Reports**, v.8(1):465, 2018.

SERA, G. H.; SERA, T.; ITO, D. S.; AZEVEDO, J.A.D.; MATA, J.S.D.; RIBEIRO FILHO, C. Selection for durable resistance to leaf rust using test-crosses on IAPAR-59 and Tupi IAC 1669-33 cultivars of *Coffea arabica*. **Brazilian Archives of Biology and Technology**, v.50(4): 565-570, 2007.

SETOTAW, T.A.; CAIXETA, E.T.; PEREIRA, A.A.; OLIVEIRA, A.C.B.; CRUZ, C.D.; ZAMBOLIM, E.M.; ZAMBOLIM, L.; SAKIYAMA, N.S. Coefficient of parentage in *Coffea arabica* L. cultivars grown in Brazil. **Crop Sci**, 53:1237– 1247, 2013.

SILVA, B.S.R. **Análise da diversidade genética, fenotípica e mapeamento por associação (GWAS) em um painel de C. arabica, incluindo acessos do centro de origem.** [Dissertação de Doutorado] Londrina: Pós-Graduação em Genética e Biologia Molecular – Universidade Estadual de Londrina, 2018.

SCHOLZ, M.B.S; FIGUEIREDO, V.R.G.; SILVA, J.V.N.; KITZBERGER, C.S.G. Características físico-químicas de grãos verdes e torrados de cultivares de café (*Coffea arabica* L.) do Iapar. **Coffee Science**, v.6, n.3, 245-255, 2011.

SCHOLZ, M.B.S; KITZBERGER, C.S.G; PEREIRA, L.F.P; DAVRIEUX, F; POT, D; CHARMETANT, P; LEROY, T. Application of near infrared spectroscopy for green coffee biochemical phenotyping. **JNear Infrared Spectrosc**, 22:411–421, 2014.

SCHOLZ, M. B. S.; KITZBERGER, C.S.G; PAGIATTO, N.F.; PEREIRA, L.F.P.; DAVRIEUX, F.; POT, D.; CHARMETANT, P.; LEROY, T. Chemical composition in wild Ethiopian Arabica coffee accessions. **Euphytica** 209, 429–438, 2016.

SOUSA, T.V.; CAIXETA, E.T.; ALKIMIM, E.R.; OLIVEIRA, A. C. B. de; PEREIRA, A. A.; ZAMBOLIM, L.; SAKIYAMA, N. S. Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. **Euphytica** 213:75, 2017a.

- SOUSA, T.V.; CAIXETA, E.T.; ALKIMIM, E.R.; OLIVEIRA, A. C. B. de; PEREIRA, A. A.; ZAMBOLIM, L.; SAKIYAMA, N. S. RESENDE JUNIOR, M.F.R.; ZAMBOLIM, L. Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker. **Tree Genetics and Genomes**, v.13, 124, 2017b.
- SU, C.; WANG, W.; GONG, S.; ZUO, J.; LI, S.; XU, S. High density linkage map construction and mapping of yield trait QTLs in Maize (*Zea mays*) using the Genotyping-by-sequencer (GBS) Technology. **Front Plant Sci.**, 8:706, 2017.
- SYRÉN, P. O.; HENCHE, S.; EICHLER, A.; NESTL, B. M & HAUER, B. Squalene-hopene cyclases-evolution, dynamics and catalytic scope. *Curr. Opin. Struct. Biol.*, v.41: 73–82, 2016.
- TAMBA, C.L.; NI, Y.L.; ZHANG, Y.M. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. **Plos Computational Biology**, v. 13: 1, 2017.
- TRAN, H.; SLADE LEE, L.; FURTADO, A.; SMYTH, H.; HENRY, R. Advances in genomics for the improvement of quality in Coffee. **Journal of the Science of Food and Agriculture**, 96, p. 3310–3312, 2016.
- TESSEMA, A.; ALAMEREW, S.; KUFA, T.; GAREDEW, W. Variability and association of quality biochemical attributes in some promising *Coffea arabica* germplasm collections in Southwestern Ethiopia. **Int. J. Plant Breed Genet** 5: 302-316, 2011.
- UPADHYAY, R., MOHAN RAO, L.J. An Outlook on Chlorogenic Acids—Occurrence, Chemistry, Technology, and Biological Activities. *Critical Review in Food Science and Nutrition*, v.53, 9, 2013.
- VASCONCELOS, A.L.; FRANÇA, A.S.; GLÓRIA, M.B.A.; MENDONÇA, J.C.F. “A comparative study of chemical attributes and levels of amines in defective green and roasted coffee beans”, **Food Chem.** 100(1), 26–32, 2007.
- WANG, S.; FENG, J.Y; REN, W.L.; NI, J.L.; WEN, Y.J.; HUANG, B.; ZHOU, L.; WEN, Y.J.; ZHANG, J.; DUMWELL, J.M.; XU, S.; ZHANG, Y.M. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. **Scientific Reports**, 6, 2016.
- WEN, Y.J.; ZHANG, H.; HUANG, B.; ZHANG, J.; FENG, J.Y, WANG, S.B.; DUNWELL, J.M.; ZHANG, Y.M.; WU, R. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. **Brief Bioinform**, 2017.
- WEN, T.; WU, M.; SHEN, C.; GAO, B.; ZHU, D.; ZHANG, X.; YOU, C.; LIN, Z. Linkage and association mapping reveals the genetic basic of brown fibre (*Gossypium hirsutum*). **Plant Biotechnology J.**, 2018.
- WERNER, C.R.; QIAN. L.; VOSS-FELS, K.P.; ABBADI, A.; LECKBAND, G.; FRISCH, M.; SNOWDON, R.J. Genome-wide regression models considering general and specific combining ability predict hybrid performance in oilseed rape with similar accuracy regardless of trait architecture. **Theoretical and Applied Genetics**, v.131:2, 299-317, 2018.

ZHAO, Y.; WANG, H.; CHEN, W.; LI, Y. Genetic structure, linkage disequilibrium and association mapping of *Verticillium* wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. **Plos One**, 9(1): 1-15, 2014.

ZHANG, T.; HU, Y.; JIANG, W.; FANG, L.; GUAN, X.; CHEN, J.; ZHANG, J.; SASKI, C.A.; SCHEFFLER, B.E.; STELLY, D.M.; HULSE-KEMP, A.M.; WAN, Q.; LIU, B.; LIU, C.; WANG, S.; PAN, M.; WANG, D.; YE, W.; CHANG, L.; ZHANG, W.; SONG, Q.; KIRKBRIDE, R.C.; CHEN, X.; DENNIS, E.; LLEWELLYN, D.J.; PETERSON, D.G.; THAXTON, P.; JONES, D.C.; WANG, Q.; XU, X.; ZHANG, H.; WU, H.; ZHOU, L.; MEI, G.; CHEN, S.; TIAN, Y.; XIANG, D.; LI, X.; DING, J.; ZUO Q.; TAO, L.; LIU, Y.; LI, J.; LIN, Y.; HUI, Y.; CAO, Z.; CAI, C.; ZHU, X.; JIANG, Z.; ZHOU, B.; GUO, W.; LI, R.; CHEN, Z.J. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. **Nat. biotechnol.** 33, 531-537, 2015.

ZHANG, J.; FENG, J.Y.; NI, J.L.; WEN, Y.J.; NIU, Y.; TAMBA, C.L.; YUE, C.; SONG, Q.; ZHANG, J.M. pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. **Heredity**, v.118, 517-524, 2017.

5.7 MATERIAL SUPLEMENTAR

Tabela Suplementar 1. Granulometria - tamanho de grão determinado pelo diâmetro das peneiras.

Amostr as	chato graúdo (G) - 17,18 e 19		chato médio (M) - 15 e 16		chato miúdo (P) - 14 ou menor		Classificaç ão por tamanho	Classificaç ão para GWAS
	peso	porcentagem	peso	porcentagem	peso	porcentagem		
4	6,25	13%	36,43	73%	7,75	16%	M	2
7	29,39	59%	18,36	35%	2,69	5%	G	3
19	7,55	15%	36,99	72%	6,03	12%	M	2
24	36,41	73%	11,27	20%	3,33	7%	G	3
34	34,91	70%	12,08	24%	3,42	7%	G	3
35	26,28	53%	25,02	39%	3,25	7%	G	3
43	24,69	49%	17,57	29%	5,50	11%	G	3
49	30,40	61%	16,19	32%	2,87	6%	G	3
50	33,65	67%	14,37	22%	1,97	4%	G	3
52	15,73	31%	26,80	46%	3,44	7%	M	2
63	23,30	47%	22,88	46%	3,64	7%	G	3
64	31,16	62%	16,74	15%	2,28	5%	G	3
66	30,05	60%	17,97	36%	2,00	4%	G	3
67	17,54	35%	24,61	49%	7,98	16%	M	2
72	36,91	74%	10,81	22%	2,45	5%	G	3
76	8,76	18%	35,44	64%	6,05	12%	M	2
81	33,58	67%	11,90	22%	3,53	7%	G	3
86	25,95	52%	18,81	34%	5,57	11%	G	3
91	11,11	22%	33,79	67%	5,21	10%	M	2
92	30,80	62%	17,27	35%	2,22	4%	G	3
95	29,53	59%	16,17	33%	4,56	9%	G	3
98	15,39	31%	21,26	36%	8,41	17%	M	2
113	2,98	6%	32,55	65%	5,60	11%	M	2
114	13,76	28%	32,25	65%	4,00	8%	M	2
116	19,81	40%	24,05	26%	6,06	12%	G	3
119	21,46	43%	23,63	44%	4,85	10%	M	2
121	39,00	78%	9,84	20%	1,18	2%	G	3
122	18,25	36%	26,57	48%	5,26	11%	M	2
124	13,48	27%	27,58	52%	8,91	18%	M	2
128	10,67	21%	30,00	60%	9,21	18%	M	2
141	23,69	47%	21,99	24%	8,63	17%	G	3
152	6,57	13%	37,03	67%	6,35	13%	M	2
172	17,60	35%	23,75	25%	8,54	17%	G	3
182	21,80	44%	23,20	40%	4,95	10%	G	3
184	26,50	53%	20,41	18%	2,95	6%	G	3
185	29,85	60%	18,60	33%	1,59	3%	G	3
188	24,05	48%	21,83	39%	5,37	11%	G	3
193	15,78	32%	27,79	55%	6,46	13%	M	2
194	32,31	65%	17,64	35%	1,83	4%	G	3
195	29,16	58%	17,88	36%	2,90	6%	G	3
200	24,63	49%	19,36	20%	6,05	12%	G	3
205	13,66	27%	30,00	60%	6,28	13%	M	2
210	15,06	30%	26,42	48%	8,71	17%	M	2
224	19,65	39%	23,98	41%	6,45	13%	M	2
238	16,18	32%	27,05	50%	6,81	14%	M	2

Peso: em gramas (g) correspondente a um peso total de 50g

Continuação Tabela Suplementar 1. Granulometria - tamanho de grão determinado pelo diâmetro das peneiras,

Amostas	chato graúdo (G) - 17,18 e 19		chato médio (M) - 15 e 16		chato miúdo (P) - 14 ou menor		Classifica ção por tamanho	Classifica ção para GWAS
	peso (g)	porcentagem (%)	peso (g)	porcentagem (%)	peso (g)	porcentagem (%)		
239	11,02	22%	22,71	33%	6,50	13%	M	2
249	6,56	13%	27,95	47%	15,59	31%	M	2
252	31,82	64%	15,88	32%	2,33	5%	G	3
265	4,48	9%	34,14	68%	9,62	19%	M	2
267	27,72	55%	19,93	40%	2,18	4%	G	3
271	22,57	45%	18,51	25%	9,07	18%	G	3
275	24,85	50%	20,44	37%	4,83	10%	G	3
280	18,89	38%	27,16	54%	3,83	8%	M	2
289	16,97	34%	25,18	46%	7,89	16%	M	2
294	26,43	53%	18,98	38%	4,51	9%	G	3
299	19,16	38%	28,70	44%	3,55	7%	M	2
302	10,98	22%	31,28	63%	7,77	16%	M	2
310	38,79	78%	9,48	19%	1,68	3%	G	3
314	7,70	15%	30,96	56%	11,47	23%	M	2
316	4,72	9%	31,63	64%	13,57	27%	M	2
320	1,96	4%	31,32	60%	16,70	33%	M	2
325	10,11	20%	28,35	39%	10,31	21%	M	2
326	21,16	42%	18,10	20%	10,58	21%	G	3
327	2,66	5%	37,74	75%	9,57	19%	M	2
331	1,37	3%	23,42	38%	24,95	50%	P	1
332	20,66	41%	23,91	48%	5,53	11%	M	2
340	3,32	7%	33,50	67%	13,05	26%	M	2
342	3,25	6%	34,36	69%	12,46	25%	M	2
343	13,87	28%	25,73	52%	10,35	21%	M	2
346	6,67	13%	20,18	32%	23,29	47%	P	1
349	31,48	63%	14,69	23%	3,78	8%	G	3
361	6,67	13%	39,76	69%	3,50	7%	M	2
363	23,89	48%	22,11	44%	4,01	8%	G	3
364	10,01	20%	25,51	34%	14,20	28%	M	2
366	25,97	52%	18,80	17%	4,63	9%	G	3
368	17,27	35%	25,24	50%	7,61	15%	M	2
379	37,50	75%	12,01	24%	0,43	1%	G	3
L1C1	22,83	46%	18,18	28%	8,71	17%	G	3
L3C3	1,51	3%	18,51	37%	29,66	59%	P	1

Peso: em gramas (g) correspondente a um peso total de 50g

Tabela Suplementar 2. Conteúdo de ácidos nucleicos (ACGT), açúcares totais (AT), cafeína, lipídeos, proteína, sacarose, Composto fenólico, glucose ou açúcares redutores (AR), cafestol e caveol encontrados em 79 genótipos de uma população F₂ de *C. arabica*.

Amostras	ACGT	AT	Cafeína	Lipídeos	Proteína	Sacarose	Comp. gluc Fenol.	(AR)	Cafestol	Caveol
4	5,54	7,29	1,56	14,37	16,10	6,90	6,10	0,39	593,62	754,73
7	4,12	6,53	1,39	14,75	15,39	6,25	5,96	0,29	455,53	629,13
19	6,98	5,74	1,12	15,26	13,26	5,23	6,35	0,51	703,15	729,77
24	3,76	5,49	1,28	14,61	14,62	5,08	5,67	0,40	318,25	689,42
34	4,35	7,05	1,28	14,18	15,04	6,76	5,44	0,28	427,42	651,23
35	4,04	6,44	1,39	15,01	14,90	6,15	6,00	0,29	344,05	777,64
43	4,08	5,91	1,37	14,79	15,16	5,52	5,83	0,39	347,83	662,13
49	3,97	6,88	1,36	14,23	14,71	6,64	6,07	0,24	470,78	536,48
50	4,48	6,49	1,38	14,41	15,23	6,09	5,81	0,40	424,08	572,84
52	3,87	6,09	1,42	13,47	15,49	5,94	6,04	0,14	411,29	553,88
63	3,62	6,05	1,53	13,91	15,70	5,92	6,23	0,13	486,14	556,21
64	4,77	6,50	1,74	14,59	16,20	6,26	6,04	0,24	598,17	491,41
66	5,44	7,04	1,24	13,90	14,23	6,65	5,79	0,39	377,75	629,51
67	5,10	6,54	1,39	15,12	15,12	6,06	6,33	0,48	848,57	751,34
72	4,94	6,80	1,41	14,50	15,14	6,53	5,96	0,27	513,41	646,04
76	5,46	5,89	1,10	15,03	13,90	5,42	5,80	0,47	623,13	662,13
**81	4,78	7,64	1,32	14,19	14,73	7,28	5,72	0,36	374,42	718,52
**86	3,97	7,68	1,34	14,58	15,21	7,50	6,03	0,18	437,44	811,13
91	5,42	7,47	1,55	13,62	15,47	7,10	6,19	0,36	527,25	556,42
92	4,78	5,72	1,32	13,25	14,39	5,54	6,62	0,17	690,90	365,69
**95	4,44	7,71	0,91	14,12	13,52	7,41	6,14	0,30	375,04	769,87
**98	4,94	7,17	1,40	12,38	15,90	6,90	5,23	0,27	328,02	632,52
113	5,46	6,45	1,18	13,73	14,55	6,10	5,90	0,36	652,88	767,13
114	4,95	5,78	1,25	14,65	13,94	5,56	6,06	0,22	694,37	632,77
**116	5,97	7,99	1,40	13,39	15,90	7,67	5,35	0,32	452,80	643,01
**119	4,66	8,19	1,28	13,52	14,56	7,89	5,78	0,30	435,19	629,52
**121	4,17	7,57	1,30	14,13	14,97	7,20	5,94	0,38	303,86	728,04
**122	5,99	7,53	1,48	14,20	15,96	7,22	5,58	0,31	455,13	847,27
124	6,11	7,42	1,23	14,17	14,82	7,06	5,65	0,36	574,65	665,43
128	5,23	7,13	1,22	13,82	14,63	6,71	5,05	0,43	481,79	528,52
141	4,95	6,59	1,30	14,37	14,09	6,39	5,85	0,20	428,60	873,44
152	5,75	6,72	1,27	13,53	14,58	6,24	5,57	0,49	584,84	808,27
172	4,60	7,41	1,17	13,64	13,78	7,16	5,28	0,25	450,28	538,90
**182	6,12	8,40	1,33	13,63	15,05	7,94	5,31	0,46	425,25	640,20
184	5,60	6,98	1,46	13,31	15,74	6,74	5,17	0,24	336,04	652,34
185	4,05	7,33	1,18	13,55	14,10	7,01	5,44	0,32	553,08	562,04
**188	5,44	7,12	1,27	13,76	14,50	6,64	5,72	0,47	321,64	745,56
**193	4,82	7,09	1,32	14,28	14,25	6,59	5,26	0,50	549,91	801,70
194	6,66	6,80	1,22	14,18	14,13	6,52	5,72	0,28	623,25	491,49
195	3,84	6,41	1,52	13,41	14,85	6,24	5,92	0,17	449,46	636,79
**200	6,06	7,50	1,27	14,28	14,23	7,06	5,08	0,44	449,43	716,56
205	5,58	5,78	1,22	14,37	14,74	5,39	5,39	0,39	615,53	709,16
210	4,92	6,55	1,22	14,32	14,75	6,14	5,28	0,41	560,88	771,64
224	5,68	6,67	1,29	14,54	14,17	6,16	5,46	0,50	544,18	583,01
238	6,38	6,39	1,19	13,99	15,08	5,76	5,42	0,64	715,72	594,51
239	5,76	6,47	1,25	14,19	15,14	6,02	5,75	0,45	534,39	642,91
249	5,34	6,79	1,34	13,79	15,00	6,30	5,38	0,49	436,99	843,24
**252	6,46	7,43	1,16	14,13	14,07	7,04	5,19	0,40	564,13	715,52
265	5,16	6,76	1,14	13,97	13,93	6,40	5,03	0,36	710,63	503,01
267	7,11	6,53	1,25	14,60	14,23	6,07	5,38	0,46	610,65	833,39
271	5,37	6,30	1,07	13,47	13,41	5,84	5,30	0,46	488,57	550,79
**275	5,57	7,33	1,40	13,89	15,76	6,86	5,16	0,48	560,23	796,88
280	6,27	6,59	1,25	14,20	14,39	6,17	5,86	0,41	450,32	866,80
289	7,60	7,38	1,04	13,70	14,14	7,12	4,87	0,26	754,03	688,50

** Grupo composto por 15 genótipos, selecionados de acordo com o perfil bioquímico com baixo teor de ácidos clorogênicos, alto teor de sacarose e açúcares totais, baixo teor de cafestol e alto de caveol.

Continuação suplementar 2. Conteúdo de ácidos nucleicos (ACGT), açúcares totais (AT), cafeína, lipídeos, proteína, sacarose, composto fenólico, glucose ou açúcares redutores (AR), cafestol e caveol encontrados em 79 genótipos de uma população F₂ de *C. arabica*.

Amostras	ACGT	AT	Cafeína	Lipídeos	Proteína	Sacarose	Comp. Fenol.	gluc (AR)	Cafestol	Caveol
294	5,84	7,11	1,15	14,71	14,16	6,66	5,33	0,45	703,92	736,08
299	6,34	7,02	1,16	14,80	14,55	6,54	5,20	0,48	650,34	706,54
302	5,77	6,16	1,31	13,56	14,84	5,58	5,37	0,58	480,50	727,71
310	5,02	6,58	1,38	13,89	15,27	6,21	5,39	0,37	559,97	739,16
314	7,10	6,11	1,36	12,95	15,33	5,79	5,54	0,32	695,28	557,63
316	7,14	6,53	1,11	14,29	14,92	6,01	5,20	0,52	478,04	846,03
320	4,70	6,43	1,20	12,89	14,71	6,13	4,68	0,31	466,45	519,94
325	5,19	5,93	1,51	12,91	14,58	5,74	5,99	0,19	381,73	640,79
326	5,19	5,92	1,23	13,98	14,14	5,66	5,37	0,26	648,05	745,43
327	6,67	5,83	1,00	14,01	14,13	5,42	5,70	0,40	867,20	610,57
331	6,67	6,06	1,27	15,40	14,70	5,50	5,23	0,56	707,72	804,54
332	5,79	6,01	1,21	14,04	14,01	5,47	5,43	0,54	600,10	646,63
340	6,58	6,40	1,23	14,82	15,25	5,80	5,73	0,60	719,60	758,37
342	6,03	5,40	0,97	14,69	13,89	4,80	5,52	0,61	778,33	745,82
343	5,95	6,24	1,21	14,20	15,19	5,90	5,15	0,34	668,03	646,29
346	5,32	6,99	1,23	12,91	14,92	6,74	5,42	0,25	520,18	494,79
349	5,30	6,84	1,19	13,92	13,88	6,33	5,21	0,51	567,94	719,83
361	6,66	6,68	1,09	14,37	13,85	6,17	5,49	0,52	776,10	600,84
363	6,58	6,65	1,10	14,45	13,93	6,11	5,52	0,53	720,81	579,75
364	7,38	7,15	1,15	14,56	14,61	6,82	5,70	0,33	763,14	700,51
366	5,60	6,18	1,14	13,74	13,48	5,72	5,21	0,46	558,14	753,95
368	6,78	7,32	1,14	14,52	14,57	6,91	5,46	0,41	552,76	796,64
**379	6,32	7,97	1,11	13,78	14,09	7,52	4,86	0,45	503,36	700,06
L1C1	6,67	6,42	1,07	14,87	13,03	6,05	5,39	0,37	958,90	647,34
L3C3	7,57	7,11	1,35	14,67	14,96	6,68	5,45	0,43	204,30	954,79

** Grupo composto por 15 genótipos, selecionados de acordo com o perfil bioquímico com baixo teor de ácidos clorogênicos, alto teor de sacarose e açúcares totais, baixo teor de cafestol e alto de caveol.

6 CONCLUSÃO GERAL

Os resultados demonstram a importância da produção de uma população F₂ específica de *C. arabica* com o objetivo de desenvolver materiais com características agronômicas de interesse. Esse trabalho representa um avanço no conhecimento genético e molecular dessa população. Foram identificados genótipos de interesse para o programa de melhoramento, tanto pela análise fenotípica, quanto genotípica. As análises de diversidade genética e estrutura populacional revelaram uma importante variabilidade genética presente na população, possibilitando seu uso em programas de melhoramento. Foram identificados marcadores associados aos compostos ácidos clorogênicos, cafestol, açúcares redutores e para característica agronômica largura da copa, que após etapas de validação, poderão ser utilizados por programas de melhoramento.

Portanto, essa abordagem, também abre a possibilidade para pesquisas futuras, com o objetivo de auxiliar os programas de melhoramento genético visando a seleção de materiais com qualidade de bebida para produção de café especiais.