



UNIVERSIDADE
ESTADUAL DE LONDRINA

RAFAELLE VECCHIA FERREIRA

**FENOTIPAGEM BIOQUÍMICA E ESTUDOS DE
ASSOCIAÇÃO GENÔMICA AMPLA EM UMA COLEÇÃO
DE *Coffea arabica***

Londrina
2017



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



Programa de
Pós-graduação em
Genética e Biologia Molecular

RAFAELLE VECCHIA FERREIRA

**FENOTIPAGEM BIOQUÍMICA E ESTUDOS DE ASSOCIAÇÃO
GENÔMICA AMPLA EM UMA COLEÇÃO DE *Coffea arabica***

Londrina

2017

RAFAELLE VECCHIA FERREIRA

**FENOTIPAGEM BIOQUÍMICA E ESTUDOS DE
ASSOCIAÇÃO GENÔMICA AMPLA EM UMA COLEÇÃO
DE *Coffea arabica***

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

Orientador: Dr. Luiz Filipe Protasio Pereira

Londrina
2017

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

Ferreira, Rafaelle Vecchia.

Fenotipagem bioquímica e estudos de associação genômica ampla em uma coleção de Coffea arabica / Rafaelle Vecchia Ferreira. - Londrina, 2017.
99 f. : il.

Orientador: Luiz Filipe Protasio Pereira.

Dissertação (Mestrado em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Genética e Biologia Molecular, 2017.

Inclui bibliografia.

1. Café - Melhoramento genético - Tese. 2. Café - Genética - Tese. 3. Espectroscopia de infravermelho - Tese. 4. Bioquímica - Tese. I. Pereira, Luiz Filipe Protasio. II. Universidade Estadual de Londrina. Centro de Ciências Biológicas. Programa de Pós-Graduação em Genética e Biologia Molecular. III. Título.

RAFAELLE VECCHIA FERREIRA

**FENOTIPAGEM BIOQUÍMICA E ESTUDOS DE ASSOCIAÇÃO
GENÔMICA AMPLA EM UMA COLEÇÃO DE *Coffea arabica***

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

BANCA EXAMINADORA

Orientador: Prof. Dr. Luiz Filipe Protasio
Pereira
Instituto Agronômico do Paraná – IAPAR

Dra. Lilian Padilha
Empresa Brasileira de Pesquisa
Agropecuária – EMBRAPA

Dr. Gustavo César Sant'Ana
Universidade Federal de Viçosa – UFV

Dr. Leandro Simões Azeredo Gonçalves
Universidade Estadual de Londrina – UEL

Dra. Maria Brígida dos Santos Scholz
Instituto Agronômico do Paraná – IAPAR

Londrina, 30 de maio de 2017.

Ao meu esposo Gabriel,
Aos meus queridos Miguel e Luiza.

DEDICO

AGRADECIMENTOS

Acima de tudo agradeço a DEUS, por cuidar de mim e por ser minha real força. Pelo o que eu sou e por tudo aquilo que Ele me capacita ser.

Aos meus pais e à minha irmã, que sempre acreditaram em minha capacidade de ir mais além.

Ao meu esposo Gabriel, por todo o amor, carinho, dedicação e paciência, sempre ao meu lado.

A todos meus amigos que me acompanharam durante todos esses anos de estudo e dedicação. Obrigada por compreender minhas ausências quando foram necessárias.

Ao meu orientador Dr. Luiz Filipe, por me ajudar a construir um longo caminho, depositando sua confiança em meu trabalho desde o final da graduação com o TCC até o fim desse doutorado.

À Dra. Brígida pela imensa ajuda na idealização do Capítulo I e por tudo que eu aprendi em relação a fenotipagem do cafeeiro.

Ao Dr. Gustavo pelo incansável auxílio para que esse trabalho pudesse ser finalizado. Obrigada por todas as suas ideias, conversas, discussões e conclusões.

À minha amiga Lívia por me acompanhar na condução desse trabalho e pela grande ajuda na preparação de todo material. À Giselly, Bruna e Fernanda, pela grande ajuda nas coletas e posteriores cuidados com os frutos.

Ao Consórcio de Pesquisa Café, CAPES e CIRAD pelo apoio financeiro neste trabalho e a Capes, pela concessão da bolsa de estudo.

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

A todos que de alguma forma, contribuíram para a realização desse trabalho.

Muito obrigada.

FERREIRA, Rafaelle Vecchia. **Fenotipagem bioquímica e estudos de associação genômica ampla em uma coleção de *Coffea arabica***. 99 f. 2017. Tese (Doutorado em Genética e Biologia Molecular) – Universidade Estadual de Londrina, Londrina, 2017.

RESUMO

O Brasil ocupa a posição de maior produtor de café do mundo. Essa liderança mundial é resultado, em parte, dos esforços desenvolvidos pela pesquisa brasileira para colocar à disposição dos agricultores, cultivares com altas capacidades produtivas. O melhoramento tradicional do cafeeiro tem mostrado resultados expressivos, entretanto, sua eficiência é restringida devido às características botânicas do gênero *Coffea* e à complexidade das características alvo da seleção como qualidade da bebida. *Coffea arabica* apresenta maior importância econômica e apesar dessa espécie apresentar baixa diversidade genética dentre os genótipos cultivados, uma variabilidade genética maior é encontrada em coleções provenientes do centro de origem, a Etiópia. Neste trabalho, os genótipos de uma coleção *C. arabica* originários da Etiópia foram submetidos a uma avaliação bioquímica pela metodologia da Espectroscopia no Infravermelho Próximo (NIRS), com o objetivo de explorar a composição química da espécie. Nós avaliamos os teores de ácidos clorogênicos, cafeína, açúcares totais, sacarose, lipídios, proteínas e taninos de 66 genótipos, nos anos de 2012 e 2015. Essa avaliação confirmou a variabilidade genética existente nesse material e apresentou, nos dois anos analisados, um grupo de genótipos com baixos níveis de ácidos clorogênicos e cafeína, além de um teor maior de açúcar, podendo ser útil aos programas de melhoramento. Ainda foi verificada a importância de uma coleta mais homogênea de frutos maduros, a fim de não haver diferenças entre os teores químicos avaliados no mesmo material em diferentes anos. Visando à produção de ferramentas biotecnológicas que facilitem o melhoramento de cafeeiros, esses dados fenotípicos foram utilizados em estudos de associação com marcadores moleculares. Marcadores SNPs foram identificados a partir da genotipagem por sequenciamento, utilizando como genoma de referência um dos ancestrais diploides de *C. arabica*, o *Coffea canephora*. Um total de 6210920 tags produzidas pelos *reads single-end*, a partir das bibliotecas formadas pelo GBS, foi mapeado no genoma de referência identificando 6.696 SNPs, com profundidade média de 39X. Os SNPs foram filtrados para MAF >0.05, Call Rate >0.8 e heterozigosidade < 0.9, fornecendo 2460 marcas que foram usados nas análises de associação da coleção da Etiópia do IAPAR. A estratégia de GWAS pelo *software* TASSEL 5.2.20, em um conjunto de 66 genótipos de *C. arabica*, foi utilizada para identificar marcas associadas aos conteúdos bioquímicos dos genótipos dessa coleção. Foram encontrados 44 SNPs significativos associados a ácidos clorogênicos, cafeína, açúcares totais, sacarose, lipídios, proteínas e taninos. Também foram identificadas marcas que estariam associadas às vias de ácidos clorogênicos, cafeína, lipídios, proteínas e sacarose, responsáveis pela qualidade da bebida, dentro de uma coleção de genótipos com maior variabilidade genética.

Palavras-chave: NIR. GBS. GWAS. Marcadores SNP. Café.

FERREIRA, Rafaelle Vecchia. **Biochemical phenotyping and Genome-Wide Association Study in a *Coffea arabica* collection.** 99 p. 2017. Thesis (Doctorate in Genetics and Molecular Biology) – Universidade Estadual de Londrina, Londrina, 2017.

ABSTRACT

Brazil is the greatest coffee producer in the world. This global leadership results in part from the efforts made by Brazilian research to make cultivars with high productive capacities available to farmers. The traditional coffee breeding has shown significant results; however, its efficiency is restricted due to the botanical characteristics of the genus *Coffea* and to the complexity of the target characteristics of the selection as cup quality. *Coffea arabica* has greater economic importance and although this species presents low genetic diversity among cultivated genotypes, a greater genetic variability is found in collections from the center of origin, Ethiopia. In this work, the genotypes of a *C. arabica* collection originating in Ethiopia were submitted to a biochemical evaluation by the Near Infrared Spectroscopy (NIRS) methodology, to explore the chemical composition of the species. We evaluated the levels of chlorogenic acids, caffeine, total sugars, sucrose, lipids, proteins and tannins from 66 genotypes in the years of 2012 and 2015. This evaluation confirmed the genetic variability in this material and presented, during the two years analyzed, a group of genotypes with low levels of chlorogenic acids and caffeine, in addition to a higher sugar content, which may be useful for breeding programs. It was also verified the importance of a more homogeneous collection of mature fruits, to avoid differences between the chemical contents evaluated in the same material in different years. Aiming at the production of biotechnological tools that facilitate the improvement of coffee trees, these phenotypic data were used in association studies with molecular markers. SNP markers were identified from genotyping by sequencing, using as reference genome one of the diploid ancestors of *C. arabica*, *Coffea canephora*. A total of 6210920 tags produced by single-end read, from libraries formed by GBS, was mapped into the reference genome identifying 6,696 SNPs, with a mean depth of 39X. SNPs were identified, with an average depth of 39X. SNPs were filtered for MAF > 0.05, Call Rate > 0.8 and heterozygosity < 0.9, providing 2460 marks that were used in the association analysis of the IAPAR Ethiopian collection. The GWAS strategy for TASSEL software 5.2.20, in a set of 66 *C. arabica* genotypes, was used to identify marks associated with the biochemical contents of the genotypes of this collection. We found 44 significant SNPs associated with chlorogenic acids, caffeine, total sugars, sucrose, lipids, proteins and tannins. Also identified were marks that would be associated with the pathways of chlorogenic acids, caffeine, lipids, proteins and sucrose, compounds responsible for the cup quality, within a collection of genotypes with greater genetic variability.

Keywords: NIR. GBS. GWAS. SNP markers. Coffee.

SUMÁRIO

1	INTRODUÇÃO	9
2	REVISÃO BIBLIOGRÁFICA	11
2.1	Café	11
2.1.1	<i>Origem</i>	11
2.1.2	<i>Aspectos Econômicos</i>	12
2.1.3	<i>Melhoramento do Cafeeiro</i>	13
2.2	Fenotipagem	14
2.2.1	<i>Compostos Bioquímicos</i>	14
2.2.2	<i>NIR</i>	18
2.3	Genotipagem	19
2.3.1	<i>SNPs</i>	19
2.3.2	<i>GBS</i>	20
2.3.3	<i>GWAS</i>	22
2.4	Referências	25
3	OBJETIVOS	34
3.1	Geral	34
3.2	Específicos	34
4	CAPÍTULO I: Análise geral da composição química de genótipos de café arábica da Etiópia	35
4.1	RESUMO	35
4.2	ABSTRACT	36
4.3	INTRODUÇÃO	37
4.4	MATERIAIS E MÉTODOS	39
4.4.1	<i>Material Vegetal</i>	39
4.4.2	<i>Análise no NIRS</i>	39
4.4.3	<i>Análise Estatística</i>	39
4.5	RESULTADOS E DISCUSSÃO	41
4.5.1	<i>Correlação entre as variáveis</i>	43
4.5.2	<i>Análise de Componentes Principais (ACP) e Análise de Agrupamento Hierárquico (AAH)</i>	45
4.6	CONCLUSÃO	51
4.7	REFERÊNCIAS	52
4.8	MATERIAL SUPLEMENTAR	55
5	CAPÍTULO II: Estudo de associação genômica ampla (GWAS) em uma coleção de Coffea arabica	61
5.1	RESUMO	61
5.2	ABSTRACT	62
5.3	INTRODUÇÃO	63
5.4	MATERIAIS E MÉTODOS	66
5.4.1	<i>Material Vegetal</i>	66

5.4.2	<i>Genotipagem por Sequenciamento</i>	66
5.4.3	<i>Fenotipagem para Compostos Bioquímicos</i>	66
5.4.4	<i>Estrutura Genética da Coleção e Desequilíbrio de Ligação</i>	67
5.4.5	<i>Associação Genômica Ampla</i>	68
5.5	RESULTADOS E DISCUSSÃO	68
5.5.1	<i>Análise Fenotípica</i>	68
5.5.2	<i>GBS e Detecção dos SNPs</i>	71
5.5.3	<i>Análise da Estrutura Genética da Coleção e Desequilíbrio de Ligação</i>	71
5.5.4	<i>Estudos de associação para os componentes bioquímicos do café verde</i>	73
5.6	CONCLUSÃO	88
5.7	REFERÊNCIAS	89
5.8	MATERIAL SUPLEMENTAR	95
6	CONCLUSÃO GERAL	98

1 INTRODUÇÃO

O melhoramento tradicional de cafeeiros tem mostrado expressivos resultados em relação ao aumento da produtividade e tolerância a estresses bióticos e abióticos. Entretanto, a eficiência do melhoramento convencional é limitada devido à característica perene dessa espécie, ao longo período juvenil e a complexidade do controle genético das características de interesse que são, na maioria dos casos, poligênicas, com custos de caracterização altos. Portanto, a identificação de marcadores moleculares relacionados a características de interesse permitiria maiores ganhos genéticos por ciclo de seleção.

Coffea arabica, espécie com maior importância econômica, apresenta uma origem recente, resultado de uma hibridização natural entre os progenitores diploides *Coffea eugenioides* e *Coffea canephora*. Além disso, poucas plantas originaram os genótipos cultivados, atribuindo a essa espécie uma baixa diversidade genética. Esse fato gera restrições tanto para os programas de melhoramento, quanto para os trabalhos de busca de marcadores polimórficos para auxiliarem esses programas. No entanto, uma variabilidade genética maior é encontrada nas coleções de genótipos provenientes do centro de origem, a Etiópia, e que podem ser utilizadas como fonte de genes de interesse para o melhoramento do cafeeiro.

O Instituto Agrônomo do Paraná (IAPAR) possui uma coleção de 132 genótipos de *C. arabica* originários da Etiópia e a sua avaliação com marcadores moleculares associados a características fenotípicas constitui uma ferramenta importante para a caracterização da diversidade genética e para o mapeamento de características de interesse. A caracterização fenotípica e genotípica dessa coleção pode ser o ponto de partida para o mapeamento de genes ou marcas de interesse agrônomo.

Para isso é fundamental o investimento na busca de novos marcadores, por meio de estratégias e técnicas modernas da biologia molecular, assim como a utilização de técnicas modernas de fenotipagem que permitam avaliação acurada e rápida das características analisadas.

Neste trabalho foram identificados marcadores SNPs a partir da Genotipagem por Sequenciamento (GBS), feita em plataformas de sequenciamento genômico de DNA de alto desempenho (Illumina, Inc.). Ao mesmo tempo, os genótipos desta coleção foram submetidos a uma avaliação bioquímica pela metodologia da Espectroscopia no Infravermelho Próximo (NIRS), com o objetivo de explorar a composição química da espécie. Além disso, procurou-se associar esses dois elementos, fenótipo e genótipo, importantes para estudos de envolvimento dos marcadores no determinismo de características de interesse agrônomo.

2 REVISÃO BIBLIOGRÁFICA

2.1 Café

2.1.1 Origem

África, Ásia Tropical e Australásia correspondem aos centros de origem do gênero *Coffea*, que compreende 124 espécies (Davis et al., 2011; Hamon et al., 2017). O cultivo de café tem como base duas espécies economicamente mais importantes, *Coffea arabica* L. e *Coffea canephora* Pierre ex A. Froehner (Baruah et al., 2003), que respondem por 70% e 30% da produção mundial, respectivamente (Yu et al., 2011).

C. canephora e as demais espécies do gênero são autoincompatíveis e diplóides ($2n = 2x = 22$ cromossomos) (Berthaud & Charrier, 1988). *C. arabica*, no entanto, é classificada como alotetraploide ($2n = 4x = 44$ cromossomos) e apresenta sistema reprodutivo predominantemente autógamo (Charrier & Berthaud, 1985; Prakash et al., 2002). A hibridização natural entre ancestrais diploides *C. eugenoides* e *C. canephora* deu origem ao *C. arabica* (Lashermes et al., 1999).

O centro primário de diversidade de *C. arabica* se encontra no sudoeste da Etiópia (Anthony et al., 2001; Silvestrini et al., 2007), em regiões montanhosas, entre 1.000 a 2.500 m de altitude (Miranda, Pereira & Bergo, 1999). O início do seu cultivo ocorreu no Yemen e no começo do século XVIII uma única planta da Indonésia cultivada na Europa foi trazida por expedições para a América Latina (Chevalier & Dagrón, 1928; Carvalho, 1946). No Brasil foi introduzida no ano de 1727, já com grande valor comercial, trazida da Guiana Francesa a pedido do governador do Maranhão (Martins, 2008).

Devido a sua origem recente (Lashermes et al., 1999; Vidal et al., 2010) e a sua biologia de reprodução autógama (Prakash et al., 2002) o *C. arabica* apresenta uma baixa diversidade genética (Lashermes et al., 1996). No entanto, uma maior variabilidade é encontrada no centro de origem de *C. arabica* (Anthony et al., 2001; Silvestrini et al., 2007).

Prospecções foram organizadas para coletar genótipos originários dessa região pela *Food and Agriculture Organization* (FAO) entre os anos 1964-1965 (FAO, 1968) e as amostras coletadas foram encaminhadas para seis institutos: na Índia, na Tanzânia, na Etiópia, na Costa Rica, no Peru e em Portugal (Lashermes et al., 1996; Silvestrini et al., 2007).

O Instituto Agrônomo de Campinas (IAC) obteve parte destes genótipos do *Centro Agronómico Tropical de Investigación y Enseñanza* (CATIE – Costa Rica), e desses, 132 genótipos foram fornecidos ao Instituto Agrônomo do Paraná (IAPAR). As plantas vêm sendo cultivadas em campo e, nos últimos anos, caracterizadas fenotípica e genotipicamente (Ferreira, 2013; Pagiatto, 2013; Scholz et al. 2016).

2.1.2 Aspectos Econômicos

Responsável por pouco mais de 30% da produção mundial de café, o Brasil lidera o mercado produtor, seguido de Vietnã, Colômbia e Indonésia. Além disso, é também o maior exportador e recentemente o maior consumidor dentro desse cenário internacional (Silva, 2016). No centro-sul do país destacam-se, respectivamente, os estados: Minas Gerais, Espírito Santo, São Paulo e Paraná, que apresentam maior área em produção (em hectares), produtividade (sacas/hectare) e produção (mil sacas beneficiadas). (CONAB, 2017). Na região norte o destaque é para o estado de Rondônia e na região nordeste as maiores áreas em produção são encontradas no estado da Bahia.

Cerca de 70% da produção total é de *C. arabica* ou café arábica enquanto que *C. canephora*, também chamado de conilon ou robusta, representa aproximadamente 30%. Em 2016 as maiores áreas em produção de café arábica se concentravam nos estados de Minas Gerais, São Paulo, Espírito Santo, Bahia e Paraná, enquanto a produção do café robusta era distribuída principalmente entre Espírito Santo, Bahia e Rondônia (CONAB, 2017).

O primeiro levantamento da safra brasileira de 2017 aponta que a produção total de café (arábica e conilon) ficará entre 23 e 25 sacas/hectare, ou seja, de 46,65 a 47,51 milhões de sacas (60 Kg de café beneficiado).

Haverá assim uma redução de 15 a 7,5%, em relação à safra de 2016 de 51,37 milhões de sacas (60 Kg de café beneficiado), devido ao ano de bienalidade negativa, característica de plantas perenes como é o caso do cafeeiro (CONAB, 2017).

O café arábica tem uma estimativa de produtividade total entre 23,48 e 25,40 sacas/hectare para 2017. O estado do Paraná já foi o maior produtor nacional e hoje conta com uma área de 46.240 mil hectares, com uma produtividade média de 27,03 sc/ha. Além disso, estima-se um aumento de 14,6 a 24,2% em relação à safra de 2016, devido a uma inversão da bienalidade para o estado, que será positiva, após a geada de 2013 (CONAB, 2017).

2.1.3 Melhoria do Cafeeiro

O melhoramento tradicional do cafeeiro tem mostrado resultados expressivos, entretanto, sua eficiência é restringida devido às características botânicas do gênero *Coffea*, sendo uma planta perene e que apresenta um longo período juvenil. Ainda há a complexidade do controle genético das características de interesse que são, na maioria dos casos, poligênicas, com custos de caracterização altos.

Como a base genética de *C. arabica* é estreita (Lashermes et al., 1996), há um limite na diversidade molecular observada e conseqüentemente nas possibilidades de melhoramento. No entanto, devido sua importância e seu valor econômico (Hein & Gatzweiler, 2006), os genótipos de café arábica originários da Etiópia podem ser explorados, pois constituem uma grande fonte de alelos para o melhoramento genético. Entre as características variáveis observadas nos genótipos da Etiópia observa-se, por exemplo, diferentes níveis de cafeína (Silvarolla, Mazzafera & Fazuoli, 2004) e diterpenos (Scholz et al., 2016), resistência a doenças e pragas (Van Der Vossen & Walyaro, 1980; Van Der Graaff, 1981; Kushalappa & Eskes, 1989; Anzueto et al., 2001) e produtividade (Bertrand et al., 2005).

A caracterização dos recursos genéticos de *C. arabica*, de grande importância para o melhoramento do cafeeiro, vem sendo realizada, tanto

fenotípica (Carvalho et al., 1983; Montagnon & Bouarmont, 1996; Mazzafera et al., 1990; Silvarolla, Mazzafera & Fazuoli, 2004), como genotipicamente (Lashermes et al., 1996; Anthony et al., 2001; Anthony et al., 2002; Aga et al., 2003; Chaparro et al., 2004; Aga et al., 2005, Silvestrini et al., 2007), mostrando que existe uma diferenciação tanto molecular como fenotípica, relacionada ao centro de origem da espécie. No entanto, esses estudos não identificaram associações entre marcadores e características de interesse para o melhoramento genético.

O desenvolvimento de novas tecnologias, principalmente com os sequenciamentos de nova geração (NGS – Next Generation Sequencing), vem permitindo um avanço no melhoramento vegetal em geral (Metzker, 2010). O sequenciamento do genoma inteiro ou parte dele facilita a obtenção de polimorfismos, que podem gerar os marcadores moleculares, responsáveis pela distinção de indivíduos geneticamente diferentes (Faleiro, 2007). Microsatélites (SSR), ou mais atualmente, SNPs vêm sendo utilizados no melhoramento do cafeeiro, como em estudos prévios de análise de estrutura de populações (Aggarwal et al., 2007; Silvestrini et al., 2007; Lopes-Gartner et al., 2009), resistência a doenças (Alvarenga et al., 2011; Pereira et al., 2011) e diversidade e mapeamento genético (Cubry et al., 2008; Geleta et al., 2012; Moncada et al., 2016).

2.2 Fenotipagem

2.2.1 Compostos Bioquímicos

Por apresentar melhor qualidade na bebida, o café arábica é preferido entre as demais espécies (De Castro & Marraccini, 2006) e por isso é responsável por 70% do café consumido no mundo, liderando o mercado consumidor (CONAB, 2017).

O café possui, além da mais conhecida e estudada cafeína, muitos componentes importantes presentes desde o grão até a bebida final. A composição química dos grãos do café cru faz parte de muitos fatores que determinam a qualidade do café. A identificação desses componentes faz-se

importante para indicar as características desejáveis da bebida. (Ferrão et al., 2003).

Encontrada em uma quantidade de 1 a 4% do peso seco, a cafeína (1,3,7-trimetilxantina), principal alcaloide presente nos grãos de café, apresenta propriedades estimulantes ao sistema nervoso central (Esquivel & Jimenez 2012). De acordo com alguns estudos que relacionam o consumo de cafeína com a saúde, houve um menor risco de doença de Parkinson e diabetes tipo 2 associado a ingestão de café. Outros, porém fazem a associação do consumo exagerado da bebida com desenvolvimento de alguns tipos de câncer e doenças cardiovasculares (Cornelis, 2014).

No cafeeiro, a cafeína encontrada em folhas jovens, atua como pesticida natural (Nathanson, 1984; Hollingsworth, Armstrong & Campbell, 2003; Guerreiro Filho & Mazzafera, 2003). O conteúdo de cafeína na bebida final contribui para o amargor, estando assim relacionado à sua qualidade (Farah et al., 2006). Mesmo com o processo de torra, este composto da fração hidrossolúvel do grão, apresenta-se estável e é pouco afetado pela degradação (Moreira, Trugo & De Maria, 2000).

Em razão dos seus efeitos na saúde e em como esses teriam relação com a genética do indivíduo, podendo associá-los tanto a riscos quanto a benefícios (The Coffee and Caffeine Genetics Consortium et al., 2014), busca-se uma bebida de melhor qualidade, como o café arábica, e que possa atender a baixos índices de cafeína (Teixeira, 2011) ou mesmo que descafeinado, seja por processos de descafeinização, ou naturalmente (Silvarolla, Mazzafera & Fazuoli, 2004).

Outro grupo da fração hidrossolúvel do grão e também importante para a qualidade da bebida são os ácidos clorogênicos (ACG), que correspondem de 6 a 12% do peso seco (Farah et al., 2005). Estes compostos secundários são formados por reações de esterificação entre o ácido quínico junto a ácidos hidroxinâmicos, como o ácido caféico e assim são denominados polifenóis (Garambone & Rosa, 2007; Farah et al, 2008).

O grão de café verde apresenta maiores teores de ACG (Farah et al. 2006; Jeszka-Skowron et al. 2016) e o processo de torra destrói e transforma progressivamente esse composto no grão (Garambone & Rosa, 2007), resultando em substâncias fenólicas que contribuem para o amargor da bebida (Clifford, 1985; Campa et al. 2005). Também no grão de café os ACGs apresentam atividades antioxidantes (Jeszka-Skowron et al. 2016) o que promove propriedades benéficas para a saúde, como atividades antivirais, hipoglicêmicas, hepatoprotetoras e antiespasmódicas (Farah & Donangelo, 2006; Tfouni et al., 2014).

Especialistas em café indicam o café arábica apresentando melhor bebida do que o café robusta (De Castro & Marraccini, 2006; Jeszka-Skowron et al. 2016). Se considerarmos uma bebida de melhor qualidade como contendo um baixo teor de ACG, corrobora com o conteúdo de ACG encontrado nessas duas espécies de café: *C. canephora* (robusta) apresenta maiores teores (11.3% do peso seco) de ACG do que *C. arabica* (4.1% do peso seco) (Ky et al., 2001).

A sacarose representa mais de 90% de todos os oligossacarídeos presentes nos grãos de café verde (Esquivel & Jiménez, 2012). Com o processo de torra os carboidratos podem ser encontrados de 38 a 42% do peso seco do grão (Belitz, Grosch & Schieberle, 2009) e contribuem para a formação do aroma, reagindo com aminoácidos e determinando o perfil dos compostos voláteis (Ky et al., 2000). O café arábica apresenta teores mais elevados de açúcar (9.2% do peso seco em média) do que o café robusta (5.5% do peso seco em média) (Ky et al., 2001). Como é um importante componente do aroma do café, quanto maior for seu teor no grão verde, mais intenso será o sabor da bebida (Ky et al., 2000).

Triacilgliceróis, esteróis, tocoferóis e diterpenos compõem a fração lipídica dos grãos de café verde (Esquivel & Jiménez, 2012). Localizados no endosperma dos grãos, o teor lipídico no café arábica é em média 15% do peso seco e no café robusta contém cerca de 10% (Speer & Kölling-Speer, 2006). O óleo do café pode ser extraído e comumente é utilizado em indústrias

cosméticas, por contribuir com a umidade natural da pele e apresentar propriedades de proteção solar (Ferrari et al., 2010; Wagemaker et al., 2011).

Compreendendo em até 20% dos lipídios totais, os diterpenos – cafestol e caveol – agem contra toxinas, podendo assim ter um efeito hepatoprotetor nos consumidores do café (Speer & Kölling-Speer, 2006; Lee, Choi & Jeong, 2007). Além disso, é relatado que os diterpenos podem apresentar propriedades antioxidantes, anti-inflamatórias e anticarcinogênicas (Cavin et al., 2002; Esquivel & Jimenez, 2012). No entanto, observou-se também que um aumento no nível de colesterol estaria correlacionado aos diterpenos (Kurzrock & Speer, 2001), e que seu teor presente na bebida poderia ser maior ou menor de acordo com o preparo do café (Urgert et al., 1995; Urgert & Katan, 1997; Ricketts et al., 2007).

Apesar de contribuir para a formação do sabor da bebida, ainda há poucos estudos sobre as propriedades bioquímicas e moleculares das proteínas do café (Acuña et al., 1999; Shimizu & Mazzafera, 2000). O conteúdo médio nas sementes de café arábica é de 10 a 14% do peso seco (Clifford, 1985). As proteínas se acumulam durante o desenvolvimento da semente e após a germinação vão sendo degradadas em aminoácidos (Shimizu & Mazzafera, 2000). Durante a torra, as proteínas que ainda estão estocadas nas sementes sofrem desnaturação, e com a hidrólise das ligações peptídicas são liberadas aminas e carbonilas, que contribuem para o aroma da bebida. A composição de proteínas nos grãos verdes, como também a espécie e/ou variedade do café, além do tempo de maturação do fruto, tudo isso implica no teor de proteínas encontrado nos grãos torrados (Fernandes et al., 2001).

Os taninos são substâncias complexas classificadas como polifenóis, e segundo sua estrutura química, podem ser classificados em taninos hidrolisáveis e taninos condensados. Quando são solúveis em água, apresentam a propriedade de adstringência, o que interfere na qualidade da bebida (Ferrão et al., 2003). Os taninos condensados são responsáveis pela produção de pigmentos avermelhados (antocianidinas), sendo encontrados em maior quantidade nas variedades de café amarelo do que em café vermelho (Ferrão et al. 2003; Esquivel & Jimenez, 2012). É descrito que os taninos

podem apresentar propriedades bactericidas, fungicidas e antivirais (Scalbert, 1991; Okuda, Yoshida & Hatano, 1993), além de ação antitumoral (Okuda, Yoshida & Hatano, 1989).

Como observado, a ocorrência de muitos componentes determinam a composição química dos grãos de café cru, fazendo com que diferentes fatores impliquem na qualidade da bebida. Os estudos já citados demonstraram como cada um desses fatores pode contribuir nas características desejáveis da bebida, e, além disso, comentaram que a composição química depende tanto da espécie estudada como também das condições ambientais que proporcionam essa composição (Ferrão et al. 2003).

2.2.2 NIR

A espectroscopia do infravermelho próximo (*Near Infrared Spectroscopy* – NIR) é uma técnica que emprega um tipo de espectroscopia vibracional de fótons de energia que corresponde à faixa de comprimento de onda de 750 a 2.500 nm, ou seja, a região do espectro eletromagnético logo após a região visível (Pasquini, 2003). Quando há a incidência dessa radiação em determinada molécula, esta apresenta uma vibração de seus átomos em maior amplitude, pois há uma transferência de energia de radiação para energia mecânica dos átomos. Assim, a diferença entre as forças das ligações químicas dos compostos vai resultar diferentes espectros NIR (Pasquini, 2003; Araujo, 2007; De Beer et al., 2011).

Determinações diretas das amostras, sem a necessidade de pré-tratamento; resultados rápidos; quantificação de amostras líquidas ou sólidas, de forma simples, rápida e segura, além da realização de análises não destrutivas são algumas das vantagens em se utilizar a técnica de NIR (Pasquini, 2003; Conesa et al., 2007). No entanto há a necessidade de uma etapa anterior, para uma correlação entre os espectros obtidos e os valores atribuídos como parâmetro daquele determinado grupo de amostras, que é a chamada calibração (Araujo, 2007; Morgano et al., 2007; Pagiatto, 2013).

Em cafeeiro, a técnica de NIR tem sido recentemente empregada na verificação da diversidade dos compostos químicos do café (Aluka, et al.,

2016), na identificação de origem geográfica e genotípica de cultivares (Villegas et al., 2014; Marquetti et al., 2016), na verificação do grau de torrefação entre café arábica e robusta (Bertone et al., 2016) e também na quantificação de compostos químicos, com base em modelos de calibração (Pagiatto, 2013; Villegas et al., 2014; Marquetti et al., 2016), como a determinação de cafeína em cafés solúveis (Gorla et al., 2015).

A quantificação de compostos químicos para *C. arabica* da Etiópia da coleção do IAPAR também vem sendo analisada pela técnica de NIR (Pagiatto, 2013; Scholz et al. 2011, 2016), metodologia também utilizada neste trabalho.

2.3 Genotipagem

2.3.1 SNPs

Para muitas culturas, a aplicação de marcadores moleculares nos estudos de diversidade genética e/ou melhoramento é uma realidade. No caso de *C. arabica* existe uma limitação na identificação de marcadores polimórficos, o que compromete o desenvolvimento de ferramentas que ajudem no melhoramento genético do cafeeiro. As causas para essa dificuldade são principalmente a natureza poliploide da espécie e a estreita base genética dos bancos de germoplasma disponíveis (Anthony et al., 2002; Poncet et al., 2006; Aggarwal et al., 2007).

Os marcadores moleculares são ferramentas que permitem detectar variabilidade no DNA dos indivíduos analisados. Podem ser utilizados em várias etapas do melhoramento de plantas, por exemplo, na análise de diversidade genética, na genotipagem de indivíduos ou também no mapeamento de regiões genômicas de interesse (Ferreira & Grattapaglia, 1998). Uma classe importante de marcadores moleculares são os polimorfismos de nucleotídeo único (*Single Nucleotide Polymorphism* – SNP). Estes marcadores apresentam alta frequência no genoma na maioria dos organismos, incluindo plantas, podendo potencialmente identificar mutações. Vale ressaltar que a maioria dos polimorfismos entre indivíduos de uma mesma

espécie está associada à diferença de um único par de bases na sequência do DNA (Rafalski, 2002).

Os SNPs surgem por mutação e são herdados como variantes alélicas. Se os SNPs estiverem localizados em regiões não codantes são denominados anônimos, porém, em regiões codantes podem ser sinônimos (não alteram o codón de tradução), neutros (alteram o códon de tradução, mas não a proteína codificada) ou não sinônimos (alteram o códon e consequentemente a proteína) (Pierce, 2011). Como constituem as formas mais abundantes de variação do genoma, os SNPs podem fornecer uma fonte para saturação de mapas genéticos, aumentando a sua precisão. Também podem ser utilizados na análise de caracterização de recursos genéticos, por meio da identificação de genótipos e/ou cultivares, além da atuação como ferramentas para seleção assistida por marcadores (Rafalski, 2002; Moncada et al., 2016), estudos de associação (Wang et al., 2015), análises de diversidade genética (Tessema et al., 2011) e mapas genéticos (Moncada et al., 2016).

SNPs podem ser detectados por meio de várias técnicas, a maioria delas envolvendo um conhecimento prévio das sequências de DNA de diferentes genótipos, o que no passado encarecia o custo da implementação para a sua identificação. As novas tecnologias de sequenciamento em larga escala (Metzker, 2010), permitem a análise de várias sequências de indivíduos e oferecem a possibilidade da descoberta de polimorfismos de nucleotídeo a custos mais baixos (Wiedmann et al., 2008; Maughan et al., 2009).

2.3.2 GBS

A genotipagem por sequenciamento (*Genotyping by Sequencing* – GBS) é uma das novas metodologias de genotipagem em larga escala que utiliza o sequenciamento de última geração, apresentando-se como um sistema altamente multiplexado, mas ao mesmo tempo simples e ideal para estudos de população, como caracterização e mapeamento, contribuindo com o melhoramento de diversos organismos (Elshire et al., 2011).

É um método direto para pequenos genomas, mas também assegura o enriquecimento ou a redução da complexidade de genomas grandes, garantindo sobreposição suficiente na cobertura das sequências dessas espécies. A redução da complexidade do genoma é feita por enzimas de restrição, que é realizada de maneira fácil, rápida, extremamente específica e altamente reprodutível, podendo atingir as regiões importantes do genoma que são inacessíveis aos métodos de captura de sequência (Mamanova, et al., 2010; Elshire et al., 2011).

O método de sequenciamento multiplex é possível com a marcação dos fragmentos de DNA, cortados pela enzima de restrição, por *barcodes*, sequências curtas de DNA que se ligam a esses fragmentos, *upstream* a região de corte da enzima (Craig et al., 2008; Elshire et al., 2011). Uma placa de 96 poços pode ser sequenciada simultaneamente: a amostra de DNA, (mais o *barcode* e seus adaptadores) é digerida pela enzima de restrição, fazendo com que os adaptadores se liguem as extremidades dos fragmentos; após são adicionados *primers* específicos aos adaptadores e é realizada uma PCR. Logo depois há uma etapa de purificação e assim os fragmentos são sequenciados em seguida (Elshire et al., 2011). Os SNPs são identificados após a sobreposição dos fragmentos de uma mesma região genômica dos diferentes genótipos e o mapeamento em um genoma de referência.

O baixo custo da genotipagem por sequenciamento tem atraído mais buscas por marcadores e mapas saturados (Spindel et al., 2013). Vantajoso também para espécies que não possuem genoma de referência, o GBS proporciona a descoberta de novos SNPs e a genotipagem com os mesmos, de forma simultânea (Kim et al., 2016). Uma desvantagem no uso dessa técnica seria o grande número de dados gerados, e a dificuldade nas análises de bioinformática (Spindel et al., 2013). No entanto programas cada vez mais simples e que conseguem resolver alguns problemas gerados por esse grande número de dados vem sendo desenvolvidos, ajudando nas análises (Glaubitz et al., 2014).

Espécies com genoma de referência bem estabelecidos têm mostrado bons resultados com a técnica de GBS. Utilizando essa técnica um

grande número de SNPs foi encontrado por Lu et al. (2013), em *Panicum virgatum* L., uma gramínea perene. Para este e outros estudos a técnica de GBS também foi usada para mapeamento (Spindel et al., 2013; Velmurugan, et al., 2016; Henning et al., 2016; Celik et al., 2016), utilizando um genoma de referência.

Alguns avanços com espécies poliploides e que não apresentam genoma de referência também tem sido mostrados, sendo benéfico para muitas culturas (Kim et al., 2016). Como exemplo de espécie aloploiploide, destaca-se o algodão (*Gossypium hirsutum* L.). Esforços em busca de um genoma de referência de qualidade para essa espécie têm sido mostrados, usando diversas tecnologias, dentre elas também o GBS (Li et al., 2015; Zhang et al., 2015). Tanto esses como em outros trabalhos, destaca-se o desafio em trabalhar com os complexos genomas poliploides, apresentando algumas alternativas para compreendê-los.

O uso da tecnologia de GBS em café também tem apresentado resultados expressivos. Moncada et al. (2016) desenvolveram um mapa de ligação para *C. arabica*, a partir de marcadores microssatélites (*Simple Sequence Repeats* – SSR) associados a QTLs (*Quantitative Trait Loci* – QTL) juntamente com SNPs obtidos a partir de GBS, a fim de aumentar a robustez do mapeamento. Com isso um mapa genético com mais de 800 marcadores SNP e SSR foi desenvolvido, identificando QTLs associados com a produção de café, altura da planta e tamanho do grão, representando um avanço na implementação da seleção assistida por marcadores, em programas de melhoramento genético do cafeeiro.

Um recente estudo, trabalhando com diferentes espécies do gênero *Coffea*, demonstrou o uso do GBS para a análise filogenética desse gênero. Através dessa técnica pode-se demonstrar que a África ou a Ásia seria a localidade ancestral mais provável para a origem do gênero. Além disso, a distribuição evolutiva do conteúdo de cafeína, importante para a qualidade da bebida, também foi detalhada (Hamon et al., 2017).

2.3.3 GWAS

Estudos de associação e mapeamento genético, combinados a dados fenotípicos fornecem a base para o relacionamento marcador-característica (Kim et al., 2016). Estudos de associação genômica ampla (*Genome-Wide Association Study* – GWAS) verificam se um conjunto de variantes genéticas, como os SNPs, em diferentes indivíduos estaria associado a alguma característica de interesse destes indivíduos. O GWAS comumente necessita da genotipagem e a fenotipagem de um grupo de indivíduos, detecção de variantes e imputação, extração dos dados das marcas e mapeamento associativo (Wang et al., 2015).

Em espécies poliploides, os resultados demonstrados pelo GWAS mostraram-se uma ótima ferramenta, pois mesmo nesses genomas complexos foi possível obter associação de marcadores (Li et al., 2015; Zhang et al., 2015), a fatores ambientais (Jia et al., 2014) e na identificação de genes candidatos a características específicas (Su et al., 2016).

Estudos de associação têm avançado em café, como demonstrou um recente estudo com *C. canephora* (Andrade et al., 2017). Nele foram validados SNPs no genoma de *C. canephora*, com ampla cobertura. Outros estudos em café, utilizando a técnica de GBS para o desenvolvimento de estudos de associação, envolvem os efeitos na ingestão do café (Amin et al., 2012; The Coffee and Caffeine Genetics Consortium et al., 2015), ambos ressaltando o efeito da cafeína presente na bebida.

Estudos de associação iniciais com *C. arabica* proveniente do centro de origem demonstraram resultados satisfatórios até o momento (Pereira et al., 2016), com a identificação de 6696 SNPs obtidos através de GBS, com cobertura de 39 vezes no genoma de referência *C. canephora*. Além disso, a associação entre os SNPs e a fenotipagem da coleção de *C. arabica* demonstrou diferenças de distribuição geográfica dos genótipos e SNPs relacionados com a produção de compostos bioquímicos envolvidos com a qualidade da bebida como lipídios e diterpenos.

Neste trabalho busca-se a associação de SNPs obtidos por GBS com nova caracterização fenotípica da coleção de *C. arabica* proveniente da

Etiópia, a fim de aumentar o conhecimento genético da espécie e contribuir para o melhoramento genético do cafeeiro.

2.4 Referências

- Acuña, R.; Bassüner, R.; Beilinson, V.; Cortina, H.; Cadena-Gómez, G.; Montes, V.; Nielsen, N.C. **Coffee seeds contain 11S storage proteins**. *Physiologia Plantarum* 105(1):122-131, 1999.
- Andrade, A.C., De A. Carneiro, F.; Da Silva Jr, O.B.; Marraccini, P.; Grattapaglia, D. **Towards GWAS and genomic prediction in coffee: development and validation of a 26K SNP chip for *Coffea canephora***. In: Proceedings Plant and Animal Genome XXV Conference. San Diego, 2017.
- Aga, E.; Bryngelsson, T.; Bekele, E.; Salomon, B. **Genetic diversity of forest arabica coffee (*Coffea arabica* L.) in Ethiopia as revealed by random amplified polymorphic DNA (RAPD) analysis**. *Hereditas* 138:36-46, 2003.
- Aga, E.; Bekele, E.; Bryngelsson, T. **Inter-simple sequence repeat (ISSR) variation in forest coffee trees (*Coffea arabica* L.) populations from Ethiopia**. *Genetica* 124:213-221, 2005.
- Aggarwal, R.K.; Hendre, P.S.; Varshney, R.K.; Bhat, P.R.; Krishnakumar, V.; Singh, L. **Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species**. *Theor Appl Genet*, 114:359-372, 2007.
- Aluka, P.; Davrieux, F.; Ngugi, K.; Omaria, R.; Gerard, F.; Leroy, T.; Dufour, M.; Fabrice, P. **The Diversity of Green Bean Biochemical Compounds in Robusta Coffee (*Coffea canephora* Pierre ex A. Froehner) as Evaluated by Near Infrared Spectroscopy**. *American Journal of Experimental Agriculture*, 12(2), 2016.
- Alvarenga, S.M.; Caixeta, E.T.; Hufnagel, B.; Thiebaut, F.; Maciel-Zambolim, E.; Zambolim, L.; Sakiyama, N.S. **Marcadores moleculares derivados de sequências expressas do genoma café potencialmente envolvidas na resistência à ferrugem**. *Pesq. agropec. Bras* 46,8: 890-898, 2011.
- Amin, Najaf et al. **Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM**. *Molecular psychiatry*, v. 17, n. 11, p. 1116-1129, 2012.
- 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.
- Anthony, F.; Combes, M.C.; Astorga, C.; Bertrand, B.; Graziosi, G.; Lashermes, P. **The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers**. *Theoretical and Applied Genetics* 104:894-900, 2002.
- Anzueto, F.; Bertrand, B.; Sarah, J.L.; Eskes, A.B.; Decazy, B. **Resistance to *Meloidogyne incognita* in Ethiopian *Coffea arabica* accessions**. *Euphyta* 118: 1-8, 2001.
- Araujo, A.M. **Monitoramento de processos em tempo real via espectrofotometria no infravermelho próximo**. [Dissertação de Mestrado] Aracajú: Pós-Graduação em Engenharia de Processos – Universidade de Tiradentes, 2007.
- Baruah, A.; Naik, V.; Hendre, P.S.; Rajkumar, R.; Rajendrakumar, P.; Aggarwal, R.K. **Isolation and characterization of nine microsatellite markers from *Coffea arabica***

- L., showing wide cross-species amplifications.** *Molecular Ecology Notes* 3, 647-650, 2003.
- Belitz, H.D.; Grosch, W.; Schieberle, P. **Food chemistry** (4th ed.). Heidelberg: Springer (Chapter 21), 2009.
- Berthaud, J.; Charrier, A. **Genetic resources of *Coffea***. In: Clarke, R. J., Macrae, R. (Eds.). *Coffee Agronomy*. London: Elsevier Applied Science 4: 1-42, 1988.
- Bertone, E.; Venturello, A.; Giraud, A.; Pellegrino, G.; Geobaldo, F. **Simultaneous determination by NIR spectroscopy of the roasting degree and Arabica/Robusta ratio in roasted and ground coffee.** *Food Control*, 59, 683-689, 2016.
- Bertrand, B.; Etienne, H.; Cilas, C.; Charrier, A.; Baradat, P. ***Coffea arabica* hybrid performance for yield, fertility and bean weight.** *Euphytica*. 141(3): 255-262, 2005.
- Campa C.; Doubeau S.; Dussert S.; Hamon S.; Noirot M. **Qualitative relationship between caffeine and chlorogenic acid contents among wild *Coffea* species.** *Food chemistry* 93(1):135-139, 2005.
- Carvalho, A. **Distribuição geográfica e classificação botânica do gênero *Coffea* com referência especial à espécie Arábica.** *Boletim da Superintendência dos Serviços do café* 21: 174-180, 1946.
- Carvalho, A.; Fazuoli, L.C.; Levy, F.A.; Guerreiro-Filho, O.; Mazzafera, P. **Observações sobre características dos frutos de introduções de *Coffea arabica* da Etiópia.** In: Resumos. 10º Congresso Brasileiro de Pesquisas Cafeeiras. Poços de Caldas, MG, pp 90-92, 1983.
- Cavin, C.; Holzhaeuser, D.; Scharf, G.; Constable, A.; Huber, W.W.; Schilter, B. **Cafestol and kahweol, two coffee specific diterpenes with anticarcinogenic activity.** *Food and Chemical Toxicology*, Oxford, v. 40, n. 8, p. 1155-1163, 2002.
- Celik, I., Bodur, S., Frary, A.; Doganlar, S. **Genome-wide SNP discovery and genetic linkage map construction in sunflower (*Helianthus annuus* L.) using a genotyping by sequencing (GBS) approach.** *Molecular Breeding*, 36(9), 133, 2016.
- Chagné, D. **Chapter One-Whole Genome Sequencing of Fruit Tree Species.** *Advances in Botanical Research*, 74, 1-37, 2015.
- Chaparro, A.P.; Cristancho, M.A.; Cortina, H.A.; Gaita'n, A.L. **Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs.** *Genet Resour Crop Evol* 51:291-297, 2004.
- Charrier, A.; Berthaud, J. **Botanical classification of coffee.** In: Clifford, M.N. & Wilson, K.C. (ed.). *Coffee: botany, biochemistry and production of beans and beverage*. London: Croom Helm, p.13-47, 1985.
- Chevalier, A.; Dagrón, M. **Recherches historiques sur les débuts de la culture du caféier en Amérique.** *Communications et Actes de l'Académie des Sciences Coloniales*, Paris, 1928.
- Clifford, M.N. **Chlorogenic acids.** In R. J. Clarke & R. Macrae (Eds.), *Coffee volume 1: chemistry* (pp. 153-202). London: Elsevier Applied Science, 1985.
- Clifford, M.N. **Chemical and physical aspects of green coffee and coffee products.** In. Clifford, M.N. & Wilson, K.C. (Eds.) *Coffee: Botany, Biochemistry and Production of Beans and Beverage*, Westport, AVI Publishing Company, Inc., 305-374, 1985.

CONAB – Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira – Café**. V. 4 - safra 2017 - N. 1 - Primeiro levantamento, 2017.

Conesa, A.; Gumí, T.; Coello, J.; Palet, C. **Near infrared spectroscopy: A novel technique for classifying and characterizing polysulfone membranes**. Journal of Membrane Science, Amsterdam, v. 300, n. 1-2, p. 122-130, 2007.

Cornelis, M.C. **Gene-coffee interactions and health**. Curr Nutr Rep 3: 178-195, 2014.

Craig, D.W.; Pearson, J.V.; Szelinger, S.; Sekar, A.; Redman, M.; Comeveaux, J.J.; Pawlowski, T.L.; Laub, T.; Nunn, G.; Stephan, D.S.; Homer, N.; Huentelman, M. **Identification of genetic variants using bar-coded multiplexed sequencing**. Nature methods, 5(10), 887-893, 2008.

Cubry, P.; Musoli, P.; Legnaté, H.; Pot, D.; Bellis, F.; Poncet, V.; Anthony, F.; Dufour, M.; Leroy, T. **Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding**. Genome 51: 50-63, 2008.

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.

De Beer, T.; Burggraeve, A.; Fonteyne, M.; Saerens, L.; Remon, J. P.; Vervaet, C. **Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes**. International Journal of Pharmaceutics, Amsterdam, v. 417, n. 1-2, p. 32-47, 2011.

De Castro, R. D.; Marraccini, P. **Cytology, biochemistry and molecular changes during coffee fruit development**. Brazilian Journal of Plant Physiology, v.18, n.1, p.175-199, 2006.

Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S. Mitchell, S.E. **A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species**. PloS one, 6(5), e19379, 2011.

Esquivel, P.; Jiménez, V.M. **Functional properties of coffee and coffee by-products**. Food Research International 46(2) 488-495, 2012.

Faleiro, F.G. **Marcadores genético-moleculares aplicados a programas de conservação e uso de recursos genéticos**. Planaltina: Embrapa Cerrados, 102 p., 2007.

FAO – Food and Agriculture Organization. **Coffee mission to Ethiopia 1964–65**. FAO, Rome, 1968.

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, 53(5), 1505-1513, 2005.

Farah, A.; Donangelo, C.M. **Phenolic compounds in coffee**. Brazilian Journal of Plant Physiology, 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.; Monteiro, M.; Donangelo, C.M.; Lafay, S. **Chlorogenic acids from green coffee extract are highly bioavailable in humans.** The Journal of nutrition, 138(12), 2309-2315, 2008.

Fernandes, S.M.; Pinto, N.A.V.D.; Thé, P.M.P.; Pereira, R.G.F.A.; Carvalho, V.D. **Teores de polifenóis, ácido clorogênico, cafeína e proteína em café torrado.** Rev. Bras. de Agrociência, v.7 n 3, p.197-199, 2001.

Ferreira, R. V. **Genotipagem de uma coleção de *Coffea arabica* utilizando marcadores microssatélites.** [Dissertação de Mestrado] Londrina: Pós Graduação em Genética e Biologia Molecular – Universidade Estadual de Londrina, 2013.

Ferreira, M.E.; Grattapaglia, D. **Introdução ao uso de marcadores moleculares em análise genética.** 3ª ed., Brasília: EMBRAPA-CENARGEN, 1998.

Ferrari, M.; Ravera, F.; De Angelis, E.; Liverani, F.S.; Navarini, L. **Interfacial properties of coffee oils.** Colloids and Surfaces A: Physicochemical Engineering Aspects, 365, 79-82, 2010.

Ferrão, M.F.; Furtado, J.C.; Neumann, L.G.; Konzen, P.H.A.; Morgano, M.A.; Bragagnolo, N.; Ferreira, M.M.C. **Técnica não destrutiva de análise de tanino em café empregando espectroscopia no infravermelho e algoritmo genético.** Tecnol. Santa Cruz do Sul, v.7, n.1, p. 9-26, 2003.

Geleta, M.; Herrera, I.; Monzon, A.; Bryngelsson, T. **Genetic Diversity of Arabica Coffee (*Coffea arabica* L.) in Nicaragua as Estimated by Simple Sequence Repeat Markers.** The Scientific World Journal, V. 2012, 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, 9(2), e90346, 2014.

Gorla, F. A.; Sanches, J.G.; Felício, A.L.D.S.M.; Nixdorf, S.L. **Avaliação do Emprego da Técnica NIRS para a Determinação de Cafeína em Café Solúvel.** UNOPAR Científica Ciências Exatas e Tecnológicas, 12(1), 2015.

Guerreiro Filho, O.; Mazzafera, P. **Caffeine and resistance of coffee to the berry borer *Hypothenemus hampei* (Coleoptera: Scolytidae).** Journal of agricultural and food chemistry, v. 51, n. 24, p. 6987-6991, 2003.

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.; Kochko, 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.** Molecular Phylogenetics and Evolution, 109, 351-361, 2017.

Hein, L.; Gatzweiler, F. **The economic value of coffee (*Coffea arabica*) genetic resources.** Ecological Economics 60: 176-185, 2006.

Henning, J.A.; Gent, D.H., Twomey, M.C., Townsend, M.S., Pitra, N.J.; Matthews, P.D. **Genotyping-by-sequencing of a bi-parental mapping population segregating for downy mildew resistance in hop (*Humulus lupulus* L.).** Euphytica, 208(3), 545-559, 2016.

Hollingsworth, R. G.; Armstrong, J. W.; Campbell, E. **Caffeine as a novel toxicant for slugs and snails.** Ann. Appl. Biol. 142 ,91-97, 2003.

- Jia Y, Sun X, Sun J, Pan Z, Wang X, et al. **Association Mapping for Epistasis and Environmental Interaction of Yield Traits in 323 Cotton Cultivars under 9 Different Environments**. PLoS ONE 9(5): e95882, 2014.
- Kaul, S.; Koo, H. L.; Jenkins, J.; Rizzo, M.; Rooney, T.; Tallon, L. J., et al. (2000). **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana***. Nature, 408(6814), 796-815, 2000.
- 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.
- Kurzrock, T.; Speer, K. **Diterpenes and diterpenes esters in coffee**. Food Reviews International, New York, v. 17, n. 4, p. 433-450, 2001.
- Kushalappa, A.C.; Eskes, A.B. **Advances in coffee rust research**. Ann Rev Phytopath 27, 503-531, 1989.
- Ky, C.L.; Doubeau, B.; Guyot, B.; Akaffou, S.; Charrier, A.; Hamon, S.; Louarn, J.; Noirot, M. **Inheritance of coffee bean sucrose content in the interspecific cross *Coffea pseudozanguebariae* × *Coffea liberica* 'dewevrei'**. Plant Breeding, 119: 165-168, 2000.
- 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(2):223-230, 2001.
- Lashermes, P.; Trouslot, P.; Anthony, F.; Combes, M.C.; Charrier, A. **Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica***. Euphytica 87:59-64, 1996.
- 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 Gen Genet 261:259-266, 1999.
- Lee, K.J.; Choi, J.H.; Jeong, H.G. **Hepatoprotective and antioxidant effects of the coffee diterpenes kahweol and cafestol on carbon tetrachloride-induced liver damage in mice**. Food and Chemical Toxicology, Oxford, v. 45, n. 11, p. 2118-2125, 2007.
- Li, F.; Fan, G.; Lu, C.; Xiao, G.; Zou, C.; Kohel, R. J.; Liang, X. **Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution**. Nature biotechnology, 33(5), 524-530, 2015.
- López-Gartner, G.; Cortina, H.; McCouch, S.R.; Moncada, M.D.P. **Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers**. Tree Genetics & Genomes, 5:435-446, 2009.
- Lu, F.; Lipka, A.E.; Glaubitz, 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 protocol**. PLoS Genet, 9(1), e1003215, 2013.
- Mamanova L; Coffey A.J; Scott C.E; Kozarewa I; Turner E.H; Kumar, A.; Howard, E.; Shendure, J.; Turner, D.J. **Target enrichment strategies for next-generation sequencing**. Nature Methods 7: 111–118, 2010.

Marquetti, I.; Link, J.V.; Lemes, A.L.G.; dos Santos Scholz, M.B.; Valderrama, P.; Bona, E. **Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee.** Computers and Electronics in Agriculture, 121, 313-319, 2016.

Martins, A.L. **História do café.** São Paulo: Contexto. 2008. 316p.

Maughan, P.J.; Yourstone, S.M.; Jellen, E.N.; Udall, J.A. **SNP discovery via genomic reduction, barcoding, and 454-pyrosequencing in amaranth.** The Plant Genome 2: 260-270, 2009.

Mazzafera, P.; Eskes, A.B.; Parvais, J.; Carvalho A. **Stérilité mâle détectée chez *Coffea arabica* et *C. canephora* au Brésil.** In: XIII Colloque Scientifique International du Café, vol. 1. Paipa, Colombia, pp466-473, 1990.

Metzker, M. L. **Applications of next-generation sequencing technologies e the next generation.** [Review]. Nature Reviews Genetics, 11(1), 31-46, 2010.

Miranda, E.M.; Pereira, R.C.A.; Bergo, C.L. **Comportamento de seis linhagens de café (*Coffea arabica* L.) em condições de sombreamento e a pleno sol no estado do Acre.** Ciência e Agrotecnologia, v. 23, p. 62-69, 1999.

Montagnon, C.; Bouharmont, P. **Multivariate analysis of phenotypic diversity of *Coffea arabica*.** Genet Res Crop Evol 43:221-227, 1996.

Moncada, M.D.P.; Tovar, E.; Montoya, J.C.; González, A.; Spindel, J.; McCouch, S. **A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size.** Tree Genetics & Genomes, 12(1), 5, 2016.

Moreira, R. F. A.; Trugo, L. C.; De Maria, C. A. B. **Componentes voláteis do café torrado. Parte II. Compostos alifáticos, alicíclicos e aromáticos.** Química Nova, São Paulo, v. 23, n. 2, p. 195-203, 2000.

Morgano, M.A.; Faria, C.G.; Ferrão, M.F.; Ferreira, M.M.C. **Determinação de açúcar total em café cru por espectroscopia no infravermelho próximo e regressão por mínimos quadrados parciais.** Química Nova, São Paulo, v. 30, n. 2, p. 346-350, 2007.

Nathanson, J. A. **Caffeine and related methylxanthines: possible naturally occurring pesticides.** Science 226, 184-187, 1984.

Okuda, T.; Yoshida, T.; Hatano, T. **Ellagitannins as Active Constituents of Medicinal Plants.** Planta Medica, n. 55, p. 117-122, 1989.

Okuda, T.; Yoshida, T.; Hatano, T. **Classification of oligomeric hydrolysable tannins and specificity of their occurrence in plants.** Phytochemistry, n. 32, p. 507-521, 1993.

Pagiatto, N. F. **Análise de diterpenos e cafeína em uma coleção da Etiópia de *Coffea arabica*.** [Dissertação de Mestrado] Londrina: Pós-Graduação em Biotecnologia – Universidade Estadual de Londrina, 2013.

Pasquini, C. **Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications.** Journal of the Brazilian Chemical Society, São Paulo, v. 14, n. 2, p. 198-219, 2003.

Pereira, G.S.; Padilha, L.; Von Pinho, E.V.R.; Teixeira, R.K.S.; Carvalho, C.H.S.; Maluf, M.P.; Carvalho, B.L. **Microsatellite markers in analysis of resistance to coffee leaf miner in Arabica coffee**. *Pesq. agropec. Bras* 46,12: 1650-1656, 2011.

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

Pierce, B.A. **Genética: um enfoque conceitual**. Rio de Janeiro: Guanabara Koogan, 2011.

Poncet, V.; Hamon, P.; Minier, J.; Carasco, C.; Hamon, S.; Noirot, M. **SSR cross-amplification and variation within coffee trees (*Coffea* spp.)**. *Genome*, 47(6), 1071-1081, 2004.

Prakash, N. S.; Combes, M. C.; Somanna, N.; Lashermes, P. **AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid**. *Euphytica*, v. 124, p. 265-271, 2002.

Rafalski, A. **Applications of single nucleotide polymorphisms in crop genetics**. *Currents Opinion in Plant Biology* 5: 94-100, 2002.

Ricketts, M.L.; Boekschoten, M.V.; Kreeft, A.J.; Hooiveld, G.J.E.J.; Moen, C.J.A.; Müller, M.; Frants, R.R.; Hasanmoentalib, S.; Post, S.M.; Princen, H.M.G.; Porter, J.G.; Katan, M.B.; Hofker, M.H.; Moore, D.D. **The cholesterol-raising factor from coffee beans, cafestol, as an agonist ligand for the farnesoid and pregnane x receptors**. *Molecular Endocrinology*, Bethesda, v. 21, n. 7, p. 1603-1616, 2007.

Scalbert, A. **Antimicrobial properties of tannins**. *Phytochemistry*, n. 30, p. 3875-3883, 1991.

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 (2):429-438, 2016.

Shimizu, M. M.; Mazzafera, P. **Compositional changes of proteins and amino acids in germinating coffee seeds**. *Brazilian Archives of Biology and technology*, 43(3), 259-265, 2000.

Silva, A. C. P. **Informações Estatísticas do Café**. Cafés do Brasil. Ano 17, 2016.

Silvarolla, M.B.; Mazzafera, P.; Fazuoli, L.C. **Plant biochemistry: a naturally decaffeinated Arabica coffee**. *Nature* 429, 826-429, 2004.

Silvestrini S.; Junqueira M.G.; Favarin A.C.; Guerreiro-Filho O.; Maluf M.P.; Silvarolla M.B.; Colombo C.A. **Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers**. *Genetic Resources Crop Evolution* 54, 6: 1367-1379, 2007.

Speer, K.; Kölling-Speer, I. **The lipid fraction of the coffee bean**. *Brazilian Journal of Plant Physiology*, 18, 201-216, 2006.

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. Theoretical and Applied Genetics, 126(11), 2699-2716, 2013.

Su, J.; Pang, C.; Wei, H.; Li, L.; Liang, B.; Wang, C.; Mao, G. **Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton.** BMC genomics, 17(1), 687, 2016.

Teixeira, A. L. **Quantificação da cafeína e seleção precoce para produção em *Coffea arabica* L.** 2011. 87 f. [Tese de Doutorado] Lavras: Pós-Graduação em Genética e Melhoramento de Plantas – Universidade Federal de Lavras, 2011.

Tessema, A.; Alamerew, S.; Kufa, T.; Garedew, W. **Genetic Diversity Analysis for Quality Attributes of Some Promising *Coffea arabica* Germplasm Collections in Southwestern Ethiopia.** J Biol Sci, 11, 236-244, 2011.

Tfouni S.A.V.; Carreiro L.B.; Teles C.R.A.; Furlani R.P.Z.; Cipolli K.M.; Camargo M.C.R. **Caffeine and chlorogenic acids intake from coffee brew: influence of roasting degree and brewing procedure.** International Journal of Food Science and Technology 49:747-752, 2014.

The Coffee and Caffeine Genetics Consortium (Cornelis, M. C.) et al. **Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption.** Molecular psychiatry, v. 20, n. 5, p. 647-656, 2015.

Urgert, R.; Weg, G.; Kosmeijer-Schuil, T.G.; Bovenkamp, P.; Hovenier, B.; Katan, M.B. **Levels of the cholesterol-elevating diterpenes cafestol and kahweol in various coffee brews.** Journal of Agricultural and Food Chemistry, Easton, v. 43, n. 8, p. 2157-2172, 1995.

Urgert, R.; Katan, M.B. **The cholesterol-raising factor from coffee beans.** Annual Review of Nutrition, Palo Alto, v. 17, p. 305-323, 1997.

Van der Graaff, N. **Selection of Arabica coffee types resistant to coffee berry disease in Ethiopia.** PhD Thesis Wageningen University, 1981.

Van der Vossen, H.A.M.; Walyaro, D.J. **Breeding for resistance to coffee berry disease in *Coffea arabica* L.** Euphytica 29(3): 777-791, 1980.

Velmurugan, J.; Mollison, E.; Barth, S.; Marshall, D.; Milne, L.; Creevey, C.J.; Lynch, B.; Meally, H.; McCabe, M.; Milbourne, D. **An ultra-high density genetic linkage map of perennial ryegrass (*Lolium perenne*) using genotyping by sequencing (GBS) based on a reference shotgun genome assembly.** Annals of botany, 118(1), 71-87, 2016.

Vidal, R. O.; Mondego, J. M. C.; Pot, D.; Ambrósio, A. B.; Andrade, A. C.; Pereira, L. F. P.; Colombo, C.A.; Vieira, L.G.E.; Carazzolle, M.F.; Pereira, G.A.G. **A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*.** Plant Physiology, 154(3), 1053-1066, 2010.

Villegas H., A.M.; Pérez H., C.; Arana, V.A.; Sandoval, T.A.; Posada S., H.E.; Garrido V., A.; Guerrero G., J.E.; Pérez M., D.; García O., J. **Identificación de origen y calibración para tres compuestos químicos en café, por espectroscopia de infrarrojo cercano.** Revista Cenicafé 65 (1):7-16, 2014.

Wagemaker, T.A.L.; Carvalho, C.R.L.; Maia, N.B.; Baggio, S.R.; Guerreiro Filho, O. **Sun protection factor, content and composition of lipid fraction of green coffee beans.** Industrial Crops and Products, 33, 469-473, 2011.

Wang, L.; Ware, D.; Lushbough, C.; Merchant, N.; Stein, L. **A genome-wide association study platform built on iPlant cyber-infrastructure.** *Concurrency and Computation: Practice and Experience*, 27(2), 420-432, 2015.

Wiedmann, R.T.; Smith T.P.L.; Nonneman, D.J. **SNP discovery in swine by reduced representation and high throughput pyrosequencing.** *BMC Genetics* 9: 81, 2008.

Yu, Q.; Guyot, R.; Kochko, A.; De Byers, A.; Navajas-Pérez, R.; Langston B, J.; Dubreuil-Tranchant, C. D.; Paterson, A. H.; Poncet, V.; Nagai, C.; Ming, R. **Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (*Coffea*).** *Plant Journal* 67: 305-317, 2011.

Zhang, T.; Hu, Y.; Jiang, W.; Fang, L.; Guan, X.; Chen, J.; Hulse-Kemp, A. M. **Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement.** *Nature biotechnology*, 33(5), 531-537, 2015.

3 OBJETIVOS

3.1 Geral

Contribuir para identificação de novos marcadores moleculares e associar estes marcadores com características agronômicas de interesse.

3.2 Específicos

- Fenotipar genótipos da coleção da Etiópia do IAPAR pela tecnologia de NIR;
- Caracterizar a composição química dos grãos de café verde dessa coleção;
- Genotipar a coleção da Etiópia do IAPAR pela tecnologia GBS;
- Identificar e caracterizar os polimorfismos de nucleotídeos únicos (SNPs);
- Avaliar as associações entre os dados de genotipagem e fenotipagem, aplicando a estratégia de GWAS.

4 CAPÍTULO I: Análise geral da composição química de genótipos de café arábica da Etiópia

4.1 RESUMO

Genótipos de *Coffea arabica* originados da Etiópia concentram boa parte da variabilidade encontrada para esta espécie e podem contribuir para estudos de melhoramento do café. Técnicas modernas de fenotipagem e genotipagem permitem novas abordagens de estudos genéticos, onde a associação de fenótipos com marcadores moleculares podem ser realizadas com precisão cada vez maior. Neste estudo, utilizando a análise NIR nós avaliamos a composição de sete compostos bioquímicos do café verde de 66 genótipos de *C. arabica* originados da Etiópia. Correlação positiva e significativa entre proteínas e cafeína se manteve entre os dois anos avaliados, assim como correlação negativa significativa entre lipídios e sacarose. A análise do agrupamento hierárquico identificou três grupos entre os genótipos, dentre eles um que apresenta baixos níveis de ácidos clorogênicos e cafeína, além de um teor maior de açúcar, tanto em 2012 como em 2015. Diferenças encontradas entre os teores de ácidos clorogênicos nos anos avaliados e as correlações desse componente com os demais demonstraram a importância da coleta de frutos com maturação homogênea, a fim de se obter resultados semelhantes em avaliações por mais de um período com o mesmo material. Os resultados obtidos também podem ser utilizados em estudos de associação genética com marcadores moleculares visando facilitar a seleção de cafés com maior qualidade de bebida.

Palavras-Chave: *Coffea arabica*; Genótipos da Etiópia; Fenotipagem; Conteúdos químicos; NIR.

General chemical composition analysis of Arabica coffee genotypes from Ethiopia

4.2 ABSTRACT

Coffea arabica genotypes originating from Ethiopia concentrate part of the variability observed for this species and can contribute for coffee breeding studies. Modern techniques for phenotyping and genotyping allow new approaches to genetic studies, where their association can be performed with increasing accuracy. In this study, using NIRs analysis we evaluate composition of seven biochemical compounds in green coffee beans of 66 *C. arabica* genotypes from Ethiopia. Significant positive correlation between protein and caffeine remained between the two years evaluated, as well as significant negative correlation between lipids and sucrose. The analysis of hierarchical clustering identified 3 groups among genotypes one of them who with low levels of both chlorogenic acids and caffeine and higher sugar content, both in 2012 and 2015. Differences between chlorogenic acid contents in the evaluated years and correlations of this component with the others demonstrated the importance of fruits collecting with homogeneous maturation, in order to obtain similar results in evaluations for more than one period with the same material. The results presented can also be used in genome wide association studies to identify markers that can be used in coffee breeding towards cup quality.

Keywords: *Coffea arabica*; Ethiopian genotypes; Phenotyping; Chemical contents; NIR.

4.3 INTRODUÇÃO

O café é uma das bebidas mais populares mundialmente e representa uma das mais importantes fontes de renda agrícola em países tropicais. O gênero *Coffea* compreende 124 espécies, no entanto somente 10 são cultivadas em todo mundo (Bekele 2005; Davis et al. 2011). *Coffea arabica* e *Coffea canephora* são as mais importantes economicamente, responsáveis por 80 e 20% da produção brasileira (Conab 2016). *C. arabica* tem origem na Etiópia (Anthony et al. 2002), enquanto *C. canephora* origina-se na África Ocidental e Central (Davis et al., 2011; Jeszka-Skowron et al. 2016). Dentro dos cultivares de *C. arabica* há uma base genética estreita, devido à sua origem. A espécie é autógama e a dispersão de poucos genótipos originou a maioria dos cultivares, utilizados em diferentes sistemas de produção, com um gargalo da diversidade genética de genótipos cultivados (Steiger et al. 2002; Anthony et al. 2002; Vidal et al. 2010).

Compostos como cafeína e ácidos clorogênicos (ACG) estão entre uma grande variedade de substâncias biologicamente ativas nos grãos de café, que lhe confere um sabor único e aromático (Johnston et al. 2003). Estes compostos secundários são responsáveis pela acidez e adstringência e em atribuir amargor, o que influencia o sabor da bebida e consequentemente sua qualidade (Tfouni et al. 2014). O acúmulo de ambos nas sementes em maturação (Clifford & Kazi 1987), a contribuição à resistência do fruto contra estresses biológico e mecânico (Macheix et al. 1990) e também à qualidade da bebida (Clifford 1985; Farah et al. 2006) são características conhecidas desde o início do século XIX (Anthony et al. 1993).

A cafeína é o principal alcaloide presente nos grãos de café verde e torrado e um potente estimulante do sistema nervoso central – SNC (Esquivel & Jimenez 2012). Nas plantas atua como pesticida natural (Nathanson, 1984; Hollingsworth, Armstrong & Campbell, 2003; Guerreiro Filho & Mazzafera, 2003) e seu conteúdo na bebida final contribui para o amargor (Farah et al., 2006). Os ácidos clorogênicos são ésteres de ácido quínico e ácido caféico, classificados como polifenóis encontrados em frutas e no café (Garambone &

Rosa 2007). Durante o processo de torra o ACG é degradado em compostos fenólicos, o que também proporciona amargor à bebida (Campa et al. 2005).

Além da cafeína e ACG, outros componentes importantes determinam a qualidade do café, como os açúcares, lipídios, proteínas e taninos. Os açúcares contribuem para a formação do aroma, reagindo com aminoácidos e determinando o perfil dos compostos voláteis (Ky et al. 2000). O conteúdo lipídico é determinado de acordo com o tipo de preparo da bebida, (Urgert et al. 1995; Urgert & Katan 1997; Ricketts et al. 2007), influenciando nos níveis de diterpenos, uma das principais frações lipídicas do café. As proteínas também estão envolvidas com o aroma, pois durante a torra aquelas que ainda estão estocadas nas sementes sofrem desnaturação, e com a hidrólise das ligações peptídicas são liberadas aminas e carbonilas (Fernandes et al., 2001). Já os taninos apresentam a propriedade de adstringência, interferindo no sabor do café (Ferrão et al. 2003).

O Instituto Agrônomo do Paraná (IAPAR) mantém uma coleção de 132 genótipos de *C. arabica* provenientes da Etiópia, resultado de uma pesquisa organizada pela FAO na década de 60 (FAO 1968). Modelos de previsão para aquisição de espectros no infravermelho próximo (NIR) de cada componente foram desenvolvidos anteriormente, criando um banco de dados com base na análise de amostras de café, dentre cultivares tradicionais e genótipos selvagens, coletadas em várias estações na região cafeeira do Estado do Paraná. (Scholz et al 2014a, b). Variações morfológicas e agronômicas, além de diferenças bioquímicas e moleculares desta coleção têm sido avaliadas em menor escala (Scholz et al. 2016). Não há descrição abrangente da maioria dos genótipos desta coleção por múltiplos anos. Neste trabalho, analisamos a composição química dos grãos de café verde dessa coleção, obtida através da Espectroscopia de Infravermelho Próximo (NIRS), considerando a diversidade do material estudado.

4.4 MATERIAIS E MÉTODOS

4.4.1 *Material Vegetal*

Nos anos de 2012 e 2015 foram coletados frutos de 66 genótipos (Suplementar 1) dentre os 132 genótipos de *C. arabica* da coleção do IAPAR. Os genótipos utilizados foram aqueles que mantiveram produtivos naqueles anos. Os frutos foram coletados e secados naturalmente em caixas com fundo de malha. Após secos, os frutos foram descascados mecanicamente e então, armazenados em sacos de papel em local seco e escuro. Para análise, os grãos de café verde foram congelados com nitrogênio líquido, moídos em moinho de disco PERTEN 3600 e passados em peneiras com malha 0,5 mm. Após a moagem, as amostras foram colocadas em frascos plásticos identificados e foram mantidas congeladas a -18 C até posterior análise.

4.4.2 *Análise no NIRS*

A composição química geral das amostras coletadas dos 2 anos (2012 e 2015) foi analisada utilizando a Espectroscopia no Infravermelho Próximo (NIRS), no espectrofotômetro NIRSystem 6500, com reflectância difusa e região espectral de 400 a 2500 nm. Modelos de previsão desenvolvidos anteriormente (Scholz et al 2014a, b) foram empregados para cada componente. Foram realizadas varreduras entre 1100 e 2300 nm em intervalos de 2 nm, utilizando-se uma cubeta retangular contendo 6 g de café moído e em temperatura ambiente. Todos os cálculos foram feitos usando os *softwares* ISIScan e WinISI 4.5 (Infrasoft Internacional, Porto Matilda, Pennsylvania - EUA).

4.4.3 *Análise Estatística*

Toda a análise estatística foi realizada pelo software XLSTAT (Addinsoft 2010). Análise de Componentes Principais (ACP) e Análise de Agrupamento Hierárquico (AAH) foram aplicadas à matriz formada por genótipos e componentes químicos. Os dados utilizados para HCA foram primeiramente normalizados, e em seguida, aplicado o algoritmo de Ward com distância Euclidiana em conjunto para definir os grupos. A correlação de

Pearson ($\alpha = 0,05$) mostra uma estimativa da relação bilateral entre os conteúdos químicos.

4.5 RESULTADOS E DISCUSSÃO

Foram analisados sete componentes bioquímicos de grãos de café verde de 66 genótipos da coleção do IAPAR coletados nos anos de 2012 e de 2015. Além dos ácidos clorogênicos e da cafeína, também foram avaliados os teores de açúcares totais, sacarose, lipídeos, proteínas e taninos. As tabelas 1 e 2 mostram os valores médios, máximo e mínimo para essas variáveis. Devido a diferenças significativas de um ano para outro encontradas para ACG e proteínas (Suplementar 2), não utilizou-se a matriz contendo a média dos dois anos.

Tabela 1 Valores mínimos, máximos e médios e coeficiente de variação para os compostos químicos dos genótipos de café da Etiópia coletados em 2012.

Compostos ^a	Mínimo	Máximo	Média	Desvio padrão	CV %
ACG	6,44	10,92	8,06	0,87	10,79
AT	3,26	8,47	5,46	1,05	19,32
Sacarose	3,1	8,26	5,34	1,05	19,63
Cafeína	1,03	1,61	1,26	0,13	10,01
Lipídios	11,32	15,34	13,79	0,87	6,31
Proteínas	12,59	17,62	14,89	1,08	7,28
Taninos	4,17	6,19	5,34	0,38	7,15

ACG ácidos clorogênicos CV coeficiente de variação

AT açúcares totais

^a Expresso g100⁻¹

Tabela 2 Valores mínimos, máximos e médios e coeficiente de variação para os compostos químicos dos genótipos de café da Etiópia coletados em 2015.

Compostos ^a	Mínimo	Máximo	Média	Desvio padrão	CV %
ACG	1,66	6,92	4,15	1,00	24,13
AT	3,41	7,58	5,44	0,88	16,16
Sacarose	2,65	7,1	4,99	0,91	18,27
Cafeína	0,88	1,6	1,28	0,15	12,15
Lipídios	12,63	17,05	14,73	0,91	6,17
Proteínas	11,68	16,4	14,09	1,11	7,89
Taninos	4,03	6,6	5,45	0,59	10,78

ACG ácidos clorogênicos CV coeficiente de variação

AT açúcares totais

^a Expresso g100⁻¹

Os ácidos clorogênicos são encontrados em maior quantidade em grãos de café verde, estando assim relacionados também com a maturação do grão (Jeszka-Skowron et al. 2016; Farah et al. 2006). Estudos anteriores com *C. arabica* demonstraram que o teor em cafés cultivados pode variar entre 6 a 12 g100⁻¹ (Farah et al. 2005). Nesse estudo o conteúdo de ACG apresentou valores variando de 6,44 a 10,92 g100⁻¹ em 2012 e de 1,66 a 6,92 g100⁻¹ em 2015. Na coleta de 2012 os valores foram próximos aos anteriormente encontrados. No entanto na coleta de 2015 foram encontrados valores mais baixos e com alto CV (24,13 %). Essa discrepância pode ser explicada devido à presença de frutos verdes na coleta de 2012 e frutos com maturação mais uniforme na coleta de 2015.

Bebidas com teor de sacarose mais elevado são preferidas a aquelas que não o apresentam. O teor de sacarose em outros genótipos da Etiópia pode variar de 6,40 a 9,06 g100⁻¹ (Tessema et al. 2011). Nos genótipos avaliados aqui este teor teve valores variando de 3,10 a 8,26 g100⁻¹ em 2012 e de 2,65 a 7,10 g100⁻¹ em 2015, como observado nas tabelas 1 e 2.

Em relação aos teores de cafeína, para acessos da Etiópia cultivados no IAC foram encontrados valores de 0,00076 g100⁻¹, considerado um café descafeinado, a 1,65 g100⁻¹ (Silvarolla et al. 2004). Em um estudo prévio da coleção do IAPAR foi encontrado teores de cafeína variando de 0,72 a 1,23 g100⁻¹ (Scholz et al 2016). Os teores de cafeína encontrados neste trabalho tiveram valores variando de 1,03 a 1,61 g100⁻¹ em 2012 e de 0,88 a 1,6 g100⁻¹. Esse aumento pode ser explicado devido a processos de adubação que foram aplicados a partir de 2011.

C. arabica apresenta maior teor lipídico que *C. canephora* (Mazzafera et al. 1998). Durante a torra, compostos lipossolúveis ficam retidos na fase lipídica do café, o que confere aroma à bebida (Scholz et al. 2016). Para os genótipos deste estudo, os valores do teor lipídico variaram de 11,32 a 15,34 g100⁻¹ em 2012 e de 12,63 a 17,05 g100⁻¹ em 2015, valores semelhantes aos encontrados anteriormente (Scholz et al. 2016).

A proteína encontrada nos grãos de café é ainda pouco estudada, sendo relacionada com o acúmulo de compostos nitrogenados (Acuña et al.

1999). Para cultivares de café já foram encontradas variações de 14,5 a 17,0 g100⁻¹ (Scholz et al. 2011) mas também variando entre 16,10 a 18,0 g100⁻¹ (Kitzberger et al. 2013). Os teores de proteínas totais encontrados aqui variaram de 12,59 a 17,62 g100⁻¹ em 2012 e de 11,68 a 16,4 g100⁻¹ em 2015 (Tabelas 1 e 2).

Taninos são substâncias complexas, classificadas como polifenóis. Quando hidrossolúveis apresentam sabor adstringente, o que interfere na qualidade da bebida (Ferrão et al. 2003). Os teores de taninos do presente estudo variaram de 4,17 a 6,19 g100⁻¹ em 2012 e de 4,03 a 6,6 g100⁻¹ em 2015, sofrendo pouca variação entre os anos estudados (Tabelas 1 e 2).

4.5.1 Correlação entre as variáveis

Tabela 3 Correlação de Pearson entre os componentes nos genótipos de café da Etiópia coletados em 2012.

Variáveis	ACG	AT	Sacarose	Cafeína	Lipídios	Proteínas	Taninos
ACG	1						
AT	-0,090	1					
Sacarose	-0,058	0,992	1				
Cafeína	0,396	-0,061	-0,064	1			
Lipídios	-0,144	-0,339	-0,388	-0,128	1		
Proteínas	0,431	-0,287	-0,277	0,631	-0,030	1	
Taninos	0,424	-0,233	-0,236	0,481	0,036	0,110	1

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

Tabela 4 Correlação de Pearson entre os componentes nos genótipos de café da Etiópia coletados em 2015.

Variáveis	ACG	AT	Sacarose	Cafeína	Lipídios	Proteínas	Taninos
ACG	1						
AT	0,238	1					
Sacarose	0,197	0,992	1				
Cafeína	0,171	-0,237	-0,216	1			
Lipídios	-0,157	-0,381	-0,369	0,208	1		
Proteínas	0,286	-0,257	-0,250	0,769	0,146	1	
Taninos	-0,284	-0,409	-0,373	0,444	0,629	0,159	1

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

A análise de correlação entre os compostos foi realizada, mostrando padrão similar nas correlações entre os compostos nos dois anos avaliados (Tabelas 3 e 4).

Em 2012 seis correlações positivas e quatro correlações negativas foram observadas entre os compostos analisados (Tabela 3). A maior correlação positiva foi entre cafeína e proteínas, resultado já esperado e observado, visto que ambos componentes dependem dos compostos nitrogenados que se acumulam durante a formação do fruto (Scholz et al. 2016). A aplicação de fertilizantes durante a época anterior a colheita também pode explicar essa forte correlação. A maior correlação negativa significativa foi entre sacarose e lipídios ($r = -0,388$ – Tabela 2). Isto é bastante interessante uma vez que confirma um padrão observado no desenvolvimento do grão onde o acúmulo de lipídeos cessa enquanto o acúmulo de açúcares é contínuo até o final da sua maturação (Joët et al., 2010).

A análise de 2012 também demonstrou uma correlação significativa positiva entre ACG e cafeína ($r = 0,396$). Anthony et al. (1993) mostraram esse tipo de correlação em um estudo interespecífico dentro do gênero *Coffea*. Outros estudos já haviam relatado essa correlação devido à existência de complexo cafeína-ACG nos grãos (Campa et al. 2005, D'Amelio et al. 2009; Scholz et al. 2016). Além disso, em testes com cafés cafeinados e naturalmente descafeinados, foi observado que: o conteúdo de ACG em cafés cafeinados é cinco vezes maior que em cafés descafeinados (Ky et al. 1999). No entanto essa correlação é mais bem compreendida quando se estuda os isômeros de ACG separadamente (Ky et al. 1999, Campa et al. 2005, Ky et al. 2013).

Em 2015 foram encontradas cinco correlações positivas e sete correlações negativas entre os compostos analisados (Tabela 4). A correlação entre cafeína e proteínas se manteve e foi a maior correlação positiva encontrada ($r = 0,769$). Também foi mantida a correlação negativa entre sacarose e lipídios ($r = -0,369$). No entanto, a maior correlação negativa apresentada foi entre taninos e açúcares totais ($r = -0,409$). Também em relação aos taninos, uma correlação negativa significativa com ACG ($r = -$

0,284) foi encontrada em 2015 e que diferiu de 2012, onde essa correlação foi positiva ($r = 0,424$). A coleta de frutos de maturação incompleta influencia as concentrações desses compostos (Vaast et al, 2006; Scholz et al, 2011) e a correlação positiva de 2012 entre taninos e ACG pode ser explicada devido aos altos valores de ACG encontrados nesse ano.

A correlação positiva significativa entre ACG e cafeína encontrada anteriormente em 2012 não se manteve em 2015. Essas diferenças podem estar relacionadas à discordância encontrada nos conteúdos de ACG entre os anos de 2012 e 2015, mostrando a importância da coleta de frutos com maturação homogênea. Além disso, essas variações podem ser atribuídas às diferentes respostas ao ambiente, considerando os dois anos estudados.

Outros estudos que demonstram correlações entre os componentes químicos do café trabalharam com um número reduzido de amostras. Este é o primeiro estudo que utiliza uma amostragem superior, com 66 genótipos, para a melhor compreensão dessas correlações.

4.5.2 Análise de Componentes Principais (ACP) e Análise de Agrupamento Hierárquico (AAH)

A ACP foi utilizada para analisar a estrutura dos dados nos dois anos, a fim de dar uma noção dos grupos de genótipos em relação à sua composição química. A projeção de 2012 dos genótipos e as variáveis podem ser observadas na biplot da Figura 1, onde os dois primeiros eixos explicam 65,48 % da variação total (37,6 % para PC1 e 27,88 % para PC2). Os genótipos localizados do lado esquerdo da Figura 1 são caracterizados por apresentar altos níveis de cafeína, ACG, proteínas, taninos e lipídeos; apenas os níveis de açúcares totais e sacarose foram baixos para estes genótipos. Por outro lado podemos observar que houve a separação dos genótipos de acordo com o teor de açúcar.

Já a projeção de 2015 dos genótipos e as variáveis podem ser observadas na biplot da Figura 2, onde os dois primeiros eixos explicam 67,55 % da variação total (42,96 % para PC1 e 24,60 % para PC2). Os genótipos localizados do lado esquerdo da Figura 2 são caracterizados por apresentar

altos níveis de cafeína, proteínas, taninos e lipídeos. Níveis de açúcares totais, sacarose e ácidos clorogênicos foram baixos para estes genótipos.

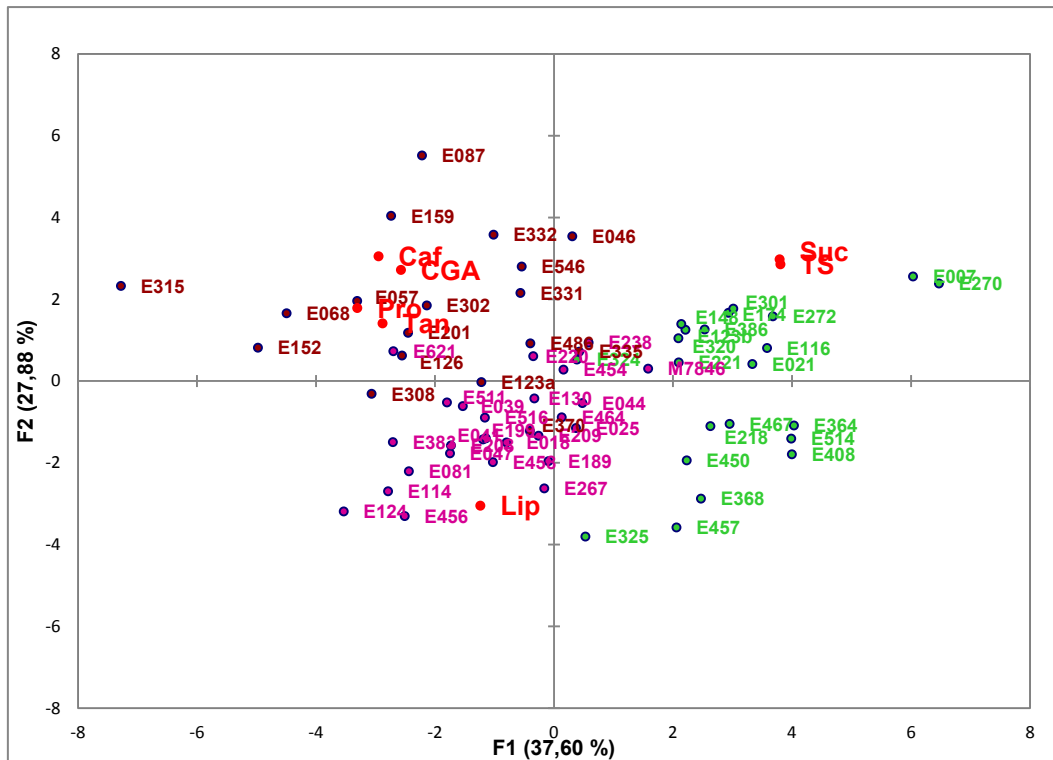


Figura 1 ACP dos compostos bioquímicos de 66 genótipos de café da Etiópia de 2012. Os grupos destacados em diferentes cores correspondem aos formados no dendograma a seguir.

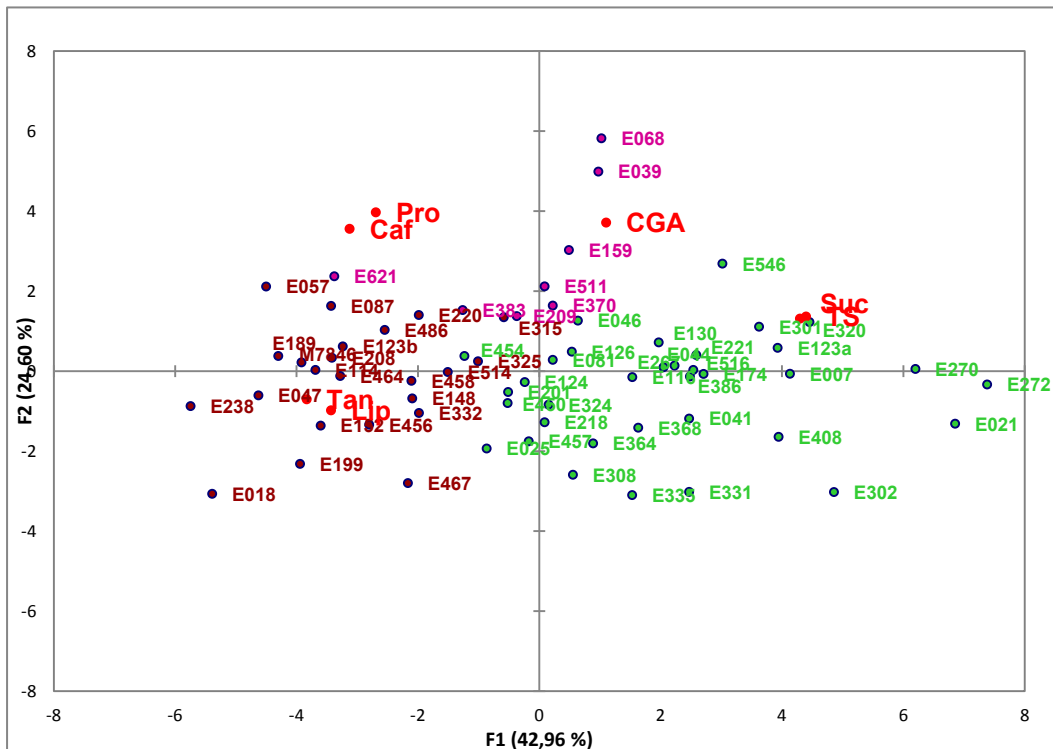


Figura 2 ACP dos compostos bioquímicos de 66 genótipos de café da Etiópia de 2015. Os grupos destacados em diferentes cores correspondem aos formados no dendograma a seguir.

Após a estrutura sugerida pelo ACP, observando a existência de grupos entre as amostras, foi aplicada também a análise de agrupamento hierárquico (AAH), classificando os grupos com base em uma matriz de dissimilaridade (Figuras 3 e 4).

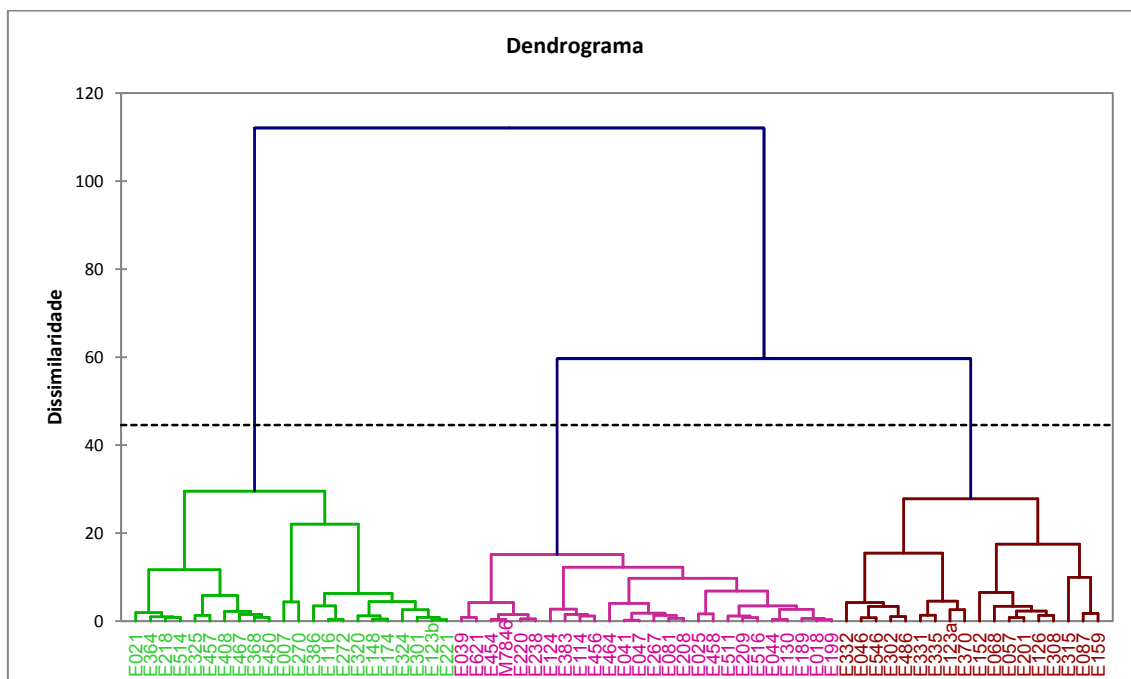


Figura 3 Dendrograma com base na análise de agrupamento hierárquico para os componentes químicos de 66 genótipos de café da Etiópia de 2012.

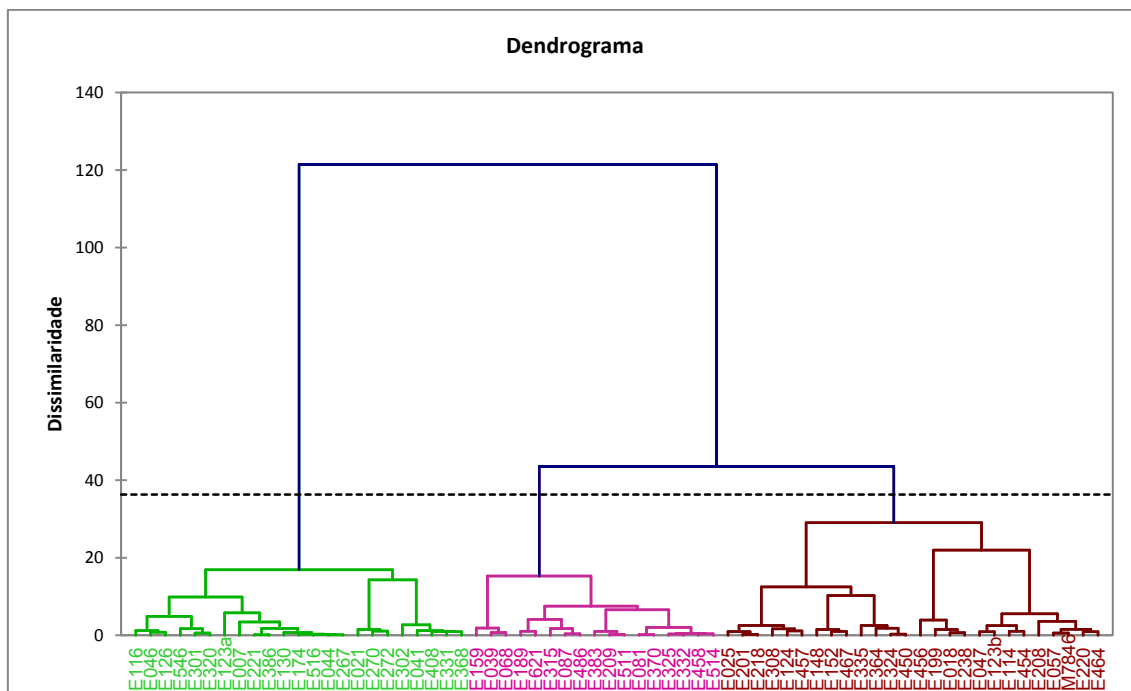


Figura 4 Dendrograma com base na análise de agrupamento hierárquico para os componentes químicos de 66 genótipos de café da Etiópia de 2015.

Nós calculamos os valores médios de cada composto para cada grupo separado no AAH (Tabelas 5 e 6). Para os dados de 2012, no grupo 1 estão os genótipos com baixo conteúdo de ACG, baixa cafeína e maior teor de açúcares, principalmente a sacarose, dentre os grupos formados. Este grupo de genótipos pode ser particularmente importante para a produção de café relacionado à qualidade. Baixos teores de ACG e cafeína estão relacionados a uma melhor qualidade do café, bem como ao maior teor de açúcar.

Tabela 5 Valores médios para sete compostos do café verde de três grupos de genótipos definidos pelo AAH entre 66 genótipos de café da Etiópia em 2012.

Grupo	ACG ^a	AT ^a	Sacarose ^a	Cafeína ^a	Lipídios ^a	Proteínas ^a	Taninos ^a
G1	7,67	6,29	6,14	1,19	13,73	14,10	5,20
G2	8,00	4,50	4,40	1,21	14,20	15,10	5,19
G3	8,62	5,22	5,10	1,40	13,51	15,74	5,64

ACG ácidos clorogênicos CV coeficiente de variação

AT açúcares totais

^aExpresso g 100⁻¹

No grupo 2 estão os genótipos que apresentaram conteúdos intermediários de ACG e cafeína e baixo conteúdo de açúcares. No grupo 3 ficaram os genótipos que apresentaram altos conteúdos de ACG, conteúdo intermediário de açúcares e alto conteúdo de cafeína.

Para os dados de 2015 no grupo 1 também estão os genótipos com conteúdo intermediário de ACG, baixa cafeína e maior teor de açúcares, dentre os grupos formados.

Tabela 6 Valores médios para sete compostos do café verde de três grupos de genótipos definidos pelo AAH entre 66 genótipos de café da Etiópia em 2015.

Grupo	ACG ^a	AT ^a	Sacarose ^a	Cafeína ^a	Lipídios ^a	Proteínas ^a	Taninos ^a
G1	4,21	6,24	5,81	1,17	14,13	13,44	4,99
G2	3,89	4,86	4,38	1,27	15,47	13,93	5,83
G3	4,48	5,23	4,81	1,41	14,40	15,21	5,47

ACG ácidos clorogênicos CV coeficiente de variação

AT açúcares totais

^aExpresso g 100⁻¹

No grupo 2 estão os genótipos que apresentaram baixo conteúdo de ACG, conteúdo intermediário de cafeína e baixo conteúdo de açúcares. No

grupo 3 ficaram os genótipos que apresentaram altos conteúdos de ACG e cafeína, conteúdo intermediário de açúcares. Lembrando que nessa coleta de 2015, os conteúdos de ACG sofreram uma grande mudança em relação ao ano de 2012, por apresentarem ausência de frutos verdes, que compromete no teor final de ACG.

Apesar de haver diferenças entre os genótipos dos grupos entre os dois anos estudados, uma parte deles se mantém nas duas avaliações (Tabela 7).

Tabela 7 Classificação de agrupamento hierárquico a partir de 66 genótipos de café da Etiópia em 2012 e em 2015.

2012			2015				
Classe	1	2	3	Classe	1	2	3
Objetos	27	18	21	Objetos	23	26	17
	E007	E018	E039		E007	E018	E039
	E021	E025	E046		E021	E025	E068
	E116	E041	E057		E041	E047	E081
	E123b	E044	E068		E044	E057	E087
	E148	E047	E087		E046	E114	E159
	E174	E081	E126		E116	E123b	E189
	E218	E114	E152		E123a	E124	E209
	E220	E123a	E159		E126	E148	E315
	E221	E124	E201		E130	E152	E325
	E238	E130	E302		E174	E199	E332
	E270	E189	E308		E221	E201	E370
	E272	E199	E315		E267	E208	E383
	E301	E208	E331		E270	E218	E458
	E320	E209	E332		E272	E220	E486
	E324	E267	E335		E301	E238	E511
	E325	E370	E383		E302	E308	E514
	E364	E456	E486		E320	E324	E621
	E368	E458	E511		E331	E335	
	E386		E516		E368	E364	
	E408		E546		E386	E450	
	E450		E621		E408	E454	
	E454				E516	E456	
	E457				E546	E457	
	E464					E464	
	E467					E467	
	E514					M7846	
	M7846						

Mesmo com a necessidade de realizar análises sensoriais desses grupos, essa caracterização bioquímica pode auxiliar a escolha de genótipos a serem utilizados em programas de melhoramento genético.

4.6 CONCLUSÃO

A análise geral da composição química de genótipos da coleção da Etiópia localizada no IAPAR revela e ainda corrobora com dados anteriores a variabilidade genética existente neste material. Nos dois anos avaliados, um grupo de genótipos com baixos níveis de ácidos clorogênicos e cafeína, além de um teor maior de açúcar foi apresentado podendo ser de utilidade aos programas de melhoramento.

Destaca-se a importância de uma coleta mais homogênea de frutos maduros, a fim de não haver diferenças entre os teores químicos avaliados no mesmo material e em diferentes anos. Apesar disso, algumas variações podem ser atribuídas também às diferentes respostas ambientais, nos períodos avaliados.

Os dados de fenotipagem também serão utilizados em estudos de associação com marcadores moleculares, visando à produção de ferramentas biotecnológicas que facilitem o melhoramento de cafeeiros.

4.7 REFERÊNCIAS

Acuña R, Bassüner R, Beilinson V, Cortina H, Cadena-Gómez G, Montes V, Nielsen NC (1999) Coffee seeds contain 11S storage proteins. *Physiologia Plantarum* 105(1):122-131.

Addinsoft (2010) XLStat: software for statistical analysis. Versão 2008.4.02 (2008). Paris. 1 CD-ROM

Anthony F, Clifford MN, Noirot M (1993) Biochemical diversity in the genus *Coffea* L.: chlorogenic acids, caffeine and mozambioside contents. *Genetic Resources and Crop Evolution* 40(2):61-70.

Anthony F, Combes MC, Astorga C, Bertrand B, Grazosi G, Leshermes P (2002) The origin of cultivated *Coffea Arabica* L. varieties revealed by AFLP and SSR markers. *Theor Appl Genet* 104:894–900.

Bekele YD (2005) Assessment of cup quality, morphological, biochemical and molecular diversity of *Coffea arabica* L. genotypes of Ethiopia (Doctoral dissertation, University of the Free State).

Campa C, Doulebeau S, Dussert S, Hamon S, Noirot M (2005) Qualitative relationship between caffeine and chlorogenic acid contents among wild *Coffea* species. *Food chemistry* 93(1):135-139.

Clifford MN (1985) Chlorogenic acids. In *Coffee* (pp. 153-202). Springer Netherlands.

Clifford MN & Kazi T (1987) The influence of coffee bean maturity on the content of chlorogenic acids, caffeine and trigonelline. *Food Chemistry* 26:59-69.

COMPANHIA NACIONAL DE ABASTECIMENTO (Brasil). Acompanhamento da safra brasileira de café: Segundo levantamento – Safra 2016. Brasília: CONAB, 2016.

D'Amelio N, Fontanive L, Uggeri F, Suggi-Liverani F, Navarini L (2009) NMR reinvestigation of the caffeine–chlorogenate complex in aqueous solution and in coffee brews. *Food Biophys* 4:321–330.

Davis AP, Tosh J, Ruch N, Fay MF (2011) 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.

Esquivel P & Jiménez VM (2012) Functional properties of coffee and coffee by-products. *Food Research International* 46(2) 488-495.

FAO Food and Agriculture Organization (1968) Coffee mission to Ethiopia 1964–65. FAO, Rome.

Farah A, Depaulis T, Trugo LC, Martin PR (2005) Effect of roasting on the formation of chlorogenic acid lactones in coffee. *J Agric Food Chem* 53:1505–1513.

Farah A, Monteiro MC, Calado V, Franca AS, Trugo LC (2006) Correlation between cup quality and chemical attributes of Brazilian coffee. *Food Chemistry* 98(2):373-380.

Fernandes SM & Pinto N (2001) Teores de polifenóis, ácido clorogênico, cafeína e proteína em café torrado. *Current Agricultural Science and Technology* 7(3):197-199.

Ferrão MF, Furtado JC, Neumann LG, Konzen PHA, Morgano MA, Bragagnolo N, Ferreira MMC (2003) Técnica não destrutiva de análise de tanino em café empregando espectroscopia no infravermelho e algoritmo genético. *Tecno-lóg* 7:9-26.

Garambone E & Rosa G (2008) Possíveis benefícios do ácido clorogênico à saúde. *Alimentos e Nutrição Araraquara* 18(2):229-235.

Jeszka-Skowron M, Sentkowska A, Pyrzyn´ska K, De Peña MP (2016) Chlorogenic acids, caffeine content and antioxidant properties of green coffee extracts: influence of green coffee bean preparation. *Eur Food Res Technol* 242:1403–1409.

Joët T, Laffargue A, Descroix F, Doubeau S, Bertrand B, Kochko A, Dussert S (2010) Influence of environmental factors, wet processing and their interactions on the biochemical composition of green coffee beans. *Food Chem* 118:693–701.

Johnston KL, Clifford MN & Morgan LM (2003) Coffee acutely modifies gastrointestinal secretion and glucose tolerance in humans: glycemic effects on chlorogenic acid and caffeine. *The American Journal of Clinical Nutrition* 78:728–733.

Kitzberger CSG, Scholz MBS, Pereira LFP, Benassi MT (2013) Composição química de cafés arábica de cultivares tradicionais e modernas. *Pesquisa Agropecuária Brasileira* 48:1498–1506.

Ky CL, Louarn J, Guyot B, Charrier A, Hamon S, Noirot M (1999) Relations between and inheritance of chlorogenic acid contents in an interspecific cross between *Coffea pseudozanguebariae* and *Coffea liberica* var 'dewevrei'. *Theoretical and applied genetics* 98(3-4):628-637.

Ky CL, Louarn J, Dussert S, Guyot B, Hamon S, Noirot M (2001) Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. *Food chemistry* 75(2):223-230.

Ky CL, Barre P, Noirot M (2013) Genetic investigations on the caffeine and chlorogenic acid relationship in an interspecific cross between *Coffea liberica dewevrei* and *C. pseudozanguebariae*. *Tree genetics & genomes* 9(4):1043-1049.

Macheix JJ, Fleuriet A & Billot J (1990) *Fruit phenolics*. CRC Press.

Mazzafera P, Soave D, Zulo MAT, Guerreiro Filho O (1998) Oil content of green beans from some species. *Bragantia* 57(1): 45-48.

Scholz MBS, Figueiredo VRG, Silva JVN, Kitzberger CSG (2011) Características físico-químicas de grãos verdes e torrados de cultivares de café do Iapar. *Coffee Science* 6:245–255.

Scholz MBS, Pagiatto NF, Kitzberger CSG, Pereira LFP, Davrieux F, Charmetant P, Leroy T (2014a) Validation of nearinfrared spectroscopy for the quantification of cafestol and kahweol in green coffee. *Food Res Int* 61:176–182.

Scholz MBS, Kitzberger CSG, Pereira LFP, Davrieux F, Pot D, Charmetant P, Leroy T (2014b) Application of near infrared spectroscopy for green coffee biochemical phenotyping. *JNear Infrared Spectrosc* 22:411–421.

Scholz MBS, Kitzberger CSG, Pagiatto NF, Pereira LFP, Davrieux F, Pot D, Charmetant P, Leroy T (2016) Chemical composition in wild ethiopian Arabica coffee accessions. *Euphytica* 209(2):429-438.

Silvarolla MB, Mazzafera P, Fazuoli LC (2004) Plant biochemistry: a naturally decaffeinated Arabica coffee. *Nature* 429:826.

Steiger DL, Nagai C, Moore PH, Morden CW, Osgood RV, Ming R (2002) AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theor Appl Genet* 105:209–215.

Tessema A, Alamerew S, Kufa T, Garedew W (2011) 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.

Tfouni SAV, Carreiro LB, Teles CRA, Furlani RPZ, Cipolli KM, Camargo MCR (2014) Caffeine and chlorogenic acids intake from coffee brew: influence of roasting degree and brewing procedure. *International Journal of Food Science and Technology* 49:747–752.

Vaast P, Bertrand B, Perriot JJ, Guyot B, Genard M (2006) 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* 86(2), 197-204.

4.8 MATERIAL SUPLEMENTAR

Suplementar 1a – Conteúdo de ácidos clorogênicos (ACG), açúcares totais (AT), sacarose, cafeína, lipídios, proteínas e taninos encontrados nos

Genótipos	ACG	AT	Sacarose	Cafeína	Lipídios	Proteínas	Taninos	Genótipos	ACG	AT	Sacarose	Cafeína	Lipídios	Proteínas	Taninos
E007	6,87	7,99	7,88	1,26	12,37	15,88	4,17	E267	8,05	4,89	4,63	1,11	15,10	14,42	5,25
E018	8,09	4,65	4,56	1,20	14,14	15,32	5,04	E270	8,19	8,47	8,26	1,15	13,38	14,15	4,63
E021	7,85	6,46	6,44	1,14	13,22	15,08	4,66	E272	7,66	7,35	7,18	1,30	13,79	14,11	5,08
E025	8,66	4,96	5,08	1,13	13,97	15,32	4,70	E301	7,84	7,11	6,93	1,25	13,33	13,84	5,56
E039	7,83	5,10	4,72	1,37	14,58	15,16	5,37	E302	8,64	5,21	5,21	1,35	12,97	15,56	5,94
E041	7,88	4,81	4,60	1,27	14,59	14,49	5,51	E308	9,17	4,73	4,41	1,38	14,92	14,75	5,59
E044	7,97	5,17	5,24	1,24	13,62	14,44	5,16	E315	10,92	3,62	3,72	1,42	13,23	17,62	5,83
E046	8,59	6,93	6,87	1,34	13,20	15,98	5,80	E320	8,49	6,30	6,20	1,18	13,07	13,28	5,66
E047	7,77	4,59	4,37	1,24	14,70	14,89	5,61	E324	8,58	6,01	5,94	1,30	14,48	13,60	5,71
E057	9,02	5,35	5,32	1,48	14,36	16,38	5,68	E325	7,16	4,15	4,24	1,06	14,33	13,42	5,31
E068	10,03	5,24	5,15	1,39	15,21	16,14	6,14	E331	8,40	5,09	5,03	1,37	11,32	14,94	5,36
E081	8,87	4,10	4,13	1,17	15,06	15,06	5,36	E332	9,39	5,73	5,83	1,44	12,05	14,16	5,92
E087	9,72	6,36	6,42	1,61	12,55	17,31	5,27	E335	7,55	5,30	5,18	1,30	12,28	15,03	5,22
E114	7,77	3,98	3,88	1,28	15,34	15,43	5,30	E364	7,78	6,24	6,16	1,03	13,64	14,02	4,68
E116	7,13	6,91	6,87	1,32	13,78	13,77	5,07	E368	7,27	5,28	5,09	1,03	14,19	12,84	5,30
E123a	9,32	4,43	4,54	1,13	12,68	15,57	5,11	E370	7,82	4,15	4,04	1,18	12,27	14,29	5,23
E123b	8,20	6,90	6,69	1,20	13,92	14,34	5,57	E383	6,93	4,53	4,29	1,40	14,97	15,94	5,53
E124	7,77	3,26	3,10	1,25	14,66	16,12	5,01	E386	6,74	6,35	6,25	1,32	12,43	14,08	5,47
E126	9,47	4,89	4,69	1,28	13,84	15,18	5,69	E408	6,82	5,89	5,85	1,14	13,68	12,77	4,92
E130	8,61	4,92	4,94	1,21	13,52	14,57	5,23	E450	7,10	5,54	5,46	1,17	14,26	12,59	5,48
E148	8,66	6,41	6,28	1,19	12,97	14,15	5,29	E454	7,49	5,92	5,73	1,29	14,19	15,82	5,33
E152	8,09	4,21	4,00	1,56	13,83	15,43	6,19	E456	6,70	3,60	3,57	1,31	14,80	15,07	5,36
E159	9,55	5,71	5,58	1,44	12,38	17,23	5,55	E457	6,44	4,71	4,52	1,06	13,82	13,84	4,92
E174	8,32	6,49	6,38	1,19	12,24	13,85	5,25	E458	7,97	4,40	4,30	1,20	14,28	16,45	4,51
E189	7,28	5,04	4,73	1,23	14,63	15,05	5,09	E464	8,41	5,34	5,04	1,07	13,83	13,98	5,87
E199	7,71	4,85	4,52	1,25	14,40	15,59	5,18	E467	8,06	6,30	6,02	1,09	14,37	12,67	5,46
E201	9,31	5,26	5,12	1,42	14,36	15,35	5,47	E486	7,68	5,47	5,45	1,26	12,93	15,25	5,83
E208	8,39	4,39	4,36	1,18	14,31	14,95	5,49	E511	7,20	4,71	4,55	1,28	13,54	16,24	5,56
E209	7,32	4,72	4,62	1,22	13,51	14,80	5,32	E514	7,31	6,49	6,16	1,11	14,34	13,41	4,88
E218	7,47	5,96	6,05	1,11	14,19	14,64	4,93	E516	7,13	4,73	4,56	1,36	13,82	15,24	5,29
E220	7,94	5,94	5,66	1,37	14,45	15,51	5,29	E546	8,34	6,20	6,21	1,41	13,15	15,46	5,79
E221	7,68	6,53	6,42	1,19	14,00	14,48	5,50	E621	8,07	5,17	4,80	1,47	14,27	16,15	5,46
E238	8,44	6,36	5,95	1,32	14,32	14,98	5,22	M7846	7,32	6,35	6,15	1,24	14,03	15,52	5,18

cafés

da

Etiópia

coletados

em

2012

Suplementar 1b – Conteúdo de ácidos clorogênicos (ACG), açúcares totais (AT), sacarose, cafeína, lipídios, proteínas e taninos encontrados nos

Genótipos	ACG	AT	Sacarose	Cafeína	Lipídios	Proteínas	Taninos	Genótipos	ACG	AT	Sacarose	Cafeína	Lipídios	Proteínas	Taninos
E007	3,24	5,91	5,64	1,16	13,16	14,27	4,03	E267	4,07	5,93	5,32	1,26	14,03	13,71	5,01
E018	1,66	3,73	3,36	1,29	15,86	14,09	6,13	E270	4,25	6,73	6,32	1,12	12,63	13,04	4,56
E021	2,95	7,08	6,79	0,99	13,72	13,09	4,18	E272	3,43	7,58	7,10	1,15	13,09	12,57	4,51
E025	4,03	4,78	4,36	1,06	15,66	13,61	5,46	E301	5,31	6,46	6,02	1,22	13,53	13,60	5,33
E039	6,40	6,21	5,71	1,60	13,64	15,87	5,13	E302	3,76	6,12	5,84	0,88	13,92	11,68	5,23
E041	3,99	5,46	5,07	1,16	13,47	12,74	5,11	E308	3,92	4,96	4,35	1,08	15,37	12,33	5,02
E044	4,29	5,78	5,24	1,24	14,09	13,70	4,76	E315	4,12	5,93	5,40	1,36	14,32	15,26	6,12
E046	5,35	5,91	5,48	1,29	14,89	14,30	5,64	E320	4,96	6,98	6,48	1,24	13,98	13,62	5,01
E047	4,12	4,35	3,83	1,28	16,52	14,91	6,10	E324	4,40	5,91	5,67	1,25	15,58	12,83	6,27
E057	5,20	4,82	4,28	1,55	16,10	15,49	6,06	E325	3,79	5,37	4,87	1,32	14,66	14,69	5,73
E068	6,92	6,65	6,23	1,60	14,32	16,40	5,30	E331	3,03	5,63	5,22	1,06	13,97	11,87	5,53
E081	4,43	5,32	4,88	1,29	14,58	14,08	5,19	E332	2,26	5,12	4,90	1,39	14,95	14,18	5,88
E087	3,83	5,00	4,68	1,46	14,87	16,20	6,13	E335	3,57	5,71	5,45	1,05	15,47	11,68	5,83
E114	3,38	4,55	4,05	1,37	16,02	15,45	5,35	E364	2,88	6,03	5,54	1,08	15,08	13,81	5,88
E116	3,29	6,65	6,34	1,30	15,49	13,81	5,57	E368	3,78	5,73	5,28	1,20	14,63	12,60	5,36
E123a	6,17	6,38	5,94	0,95	15,52	14,10	4,61	E370	6,33	4,97	4,69	1,29	14,25	14,23	5,24
E123b	4,95	5,11	4,67	1,35	16,44	14,73	6,22	E383	4,39	4,78	4,37	1,43	14,28	15,28	4,88
E124	5,30	4,55	3,89	1,14	15,09	14,02	4,59	E386	4,02	6,20	5,68	1,26	14,60	13,31	4,94
E126	3,24	6,33	5,87	1,27	15,39	15,22	5,40	E408	4,06	6,27	5,65	1,08	14,33	12,14	4,96
E130	5,07	5,42	5,06	1,19	14,00	14,24	4,70	E450	3,91	5,67	5,33	1,31	15,40	13,14	6,10
E148	3,74	5,11	4,59	1,27	14,85	14,22	6,40	E454	4,98	5,13	4,52	1,26	15,80	14,47	5,28
E152	3,11	4,34	3,96	1,39	14,52	13,63	6,49	E456	4,51	3,41	2,65	1,15	14,16	14,14	5,19
E159	5,07	5,69	5,19	1,48	13,97	15,59	4,80	E457	3,51	4,81	4,11	1,17	14,37	13,31	5,18
E174	4,20	5,74	5,39	1,23	13,56	13,40	4,89	E458	2,84	5,01	4,68	1,40	15,08	14,74	5,52
E189	3,23	4,18	3,75	1,44	14,62	15,68	5,82	E464	3,76	4,84	4,53	1,45	15,42	14,12	6,25
E199	3,47	3,61	3,00	1,21	15,60	13,71	5,47	E467	2,91	4,59	4,10	1,24	14,44	12,53	6,52
E201	4,72	5,13	4,61	1,22	15,26	13,56	5,50	E486	3,77	5,05	4,68	1,44	14,80	15,38	5,86
E208	4,43	5,67	5,16	1,50	17,05	13,76	6,60	E511	5,85	5,05	4,39	1,26	13,70	15,48	5,06
E209	4,78	4,73	4,31	1,32	13,44	15,20	5,01	E514	3,72	5,32	4,98	1,33	15,53	14,53	5,68
E218	3,92	5,33	4,85	1,15	14,98	13,51	5,61	E516	4,02	6,03	5,57	1,20	13,92	13,89	5,05
E220	4,66	5,61	5,15	1,46	15,69	14,75	5,99	E546	5,25	7,00	6,66	1,37	14,23	14,73	5,39
E221	5,00	6,19	5,70	1,20	14,77	13,54	5,03	E621	4,48	4,47	4,07	1,60	13,86	15,72	5,68
E238	3,36	3,55	3,07	1,40	15,48	14,83	5,96	M7846	2,61	5,16	4,92	1,45	15,94	15,66	6,22

cafés da Etiópia coletados em 2015.

Suplementar 2 – ANOVA para componentes do café de 66 genótipos apresentou diferenças significativas para ACG e proteínas entre os dois anos analisados (2012 e 2015).

		ACG	AT	Sacarose	Cafeína	Lipídios	Proteína	Taninos
Ano	F	1611,130	0,240	62,349	4,273	21,628	261,794	2,317
	Pr > F	0,016	0,710	0,080	0,287	0,135	0,039	0,370
Genótipo	F	3,325	23,017	23,639	17,279	0,771	23,201	1,979
	Pr > F	0,415	0,164	0,162	0,189	0,741	0,164	0,520
Ano*Genótipo	F	2,286	8,365	7,214	7,682	0,406	6,505	0,937
	Pr > F	0,489	0,269	0,289	0,281	0,879	0,304	0,695

ACG ácidos clorogênicos

AT açúcares totais

5 CAPÍTULO II: Estudo de associação genômica ampla (GWAS) em uma coleção de *Coffea arabica*

5.1 RESUMO

Responsável por 70% do café consumido mundialmente, o café arábica apresenta melhor qualidade de bebida, sendo preferido dentre as demais espécies. Entre os fatores que determinam sua qualidade está a composição química dos grãos do café cru, que indica as características desejáveis da bebida. Para avaliar a diversidade genética dessa espécie de grande importância comercial e descobrir marcas associadas às características de qualidade, uma coleção constituída por 157 genótipos de *Coffea arabica* foi genotipada utilizando a abordagem de genotipagem por sequenciamento (GBS). Um subconjunto de 66 genótipos que representaram a diversidade dentro dessa coleção foi caracterizado fenotipicamente para teores bioquímicos relacionados à qualidade da bebida. Utilizando o modelo linear generalizado (GLM) foi conduzido o estudo de associação genômica ampla (GWAS). Um total de 6210920 tags produzidas pelos *reads single-end*, a partir das bibliotecas formadas pelo GBS, foi mapeado no genoma diploide de *C. canephora*, e foram identificados 6.696 SNPs, com profundidade média de 39X. Os SNPs foram filtrados para MAF >0.05, Call Rate >0.8 e heterozigosidade < 0.9, fornecendo 2460 marcas que foram usados nas análises de associação da coleção da Etiópia do IAPAR. O GWAS foi realizado utilizando o software TASSEL 5.2.20, em um conjunto de 66 genótipos de *C. arabica* analisados quanto à composição química do grão de café verde, incluindo cafeína, ácidos clorogênicos, açúcares totais, sacarose, lipídios, proteínas e taninos. Foram encontrados 44 SNPs significativos associados a esses componentes, considerando aqueles que foram comuns em três diferentes perfis de análises. Nesse trabalho também foram identificadas marcas que estariam associadas às vias de ácidos clorogênicos, cafeína, lipídios, proteínas e sacarose, responsáveis pela qualidade da bebida, dentro de uma coleção de genótipos com maior variabilidade genética.

Palavras-chave: Café Arabica; genotipagem por sequenciamento; fenotipagem; qualidade de bebida.

Genome-Wide Association Study (GWAS) in a *Coffea arabica* collection

5.2 ABSTRACT

Responsible for 70% of the coffee consumed worldwide, the Arabica coffee presents better cup quality, being preferred among the other species. Among the factors that determine its quality is the chemical composition of the green coffee beans, which indicates the desirable characteristics of the beverage. To evaluate the genetic diversity of this species of great commercial importance and discover marks associated with quality characteristics, a collection consisting of 157 genotypes of *Coffea arabica* was genotyped using the *genotyping-by-sequencing* approach (GBS). A subset of 66 genotypes that represented diversity within this collection was phenotypically characterized for biochemical levels related to the cup quality. Using the generalized linear model (GLM) the *genome-wide association study* (GWAS) was conducted. A total of 6210920 tags produced by single-end read, from libraries formed by GBS, was mapped in the diploid genome of *C. canephora*, and 6.696 SNPs were identified, with an average depth of 39X. SNPs were filtered for MAF > 0.05, Call Rate > 0.8 and heterozygosity < 0.9, providing 2460 marks that were used in the association analysis of the IAPAR Ethiopian collection. The GWAS was performed using the software TASSEL 5.2.20, in a set of 66 *C. arabica* genotypes analyzed for the chemical composition of the green coffee bean, including caffeine, chlorogenic acids, total sugars, sucrose, lipids, proteins and tannins. We found 44 significant SNPs associated with these components, considering those that were common in three different analysis profiles. In this work, we also identified marks that would be associated with chlorogenic acids, caffeine, lipids, proteins and sucrose, biochemical compounds responsible for the cup quality, within a collection of genotypes with greater genetic variability.

Keywords: Arabica coffee; Genotyping-By-Sequencing; Phenotyping; Cup Quality.

5.3 INTRODUÇÃO

O café é um dos principais produtos agrícolas mais comercializados e consumidos no mundo, e uma das *commodities* mais valorizada nos mercados internacionais. Atualmente o Brasil é responsável por pouco mais de 30% da produção mundial de café e lidera o mercado produtor, seguido de Vietnã, Colômbia e Indonésia. Além disso, é também o maior exportador e recentemente foi considerado o maior mercado consumidor dentro desse cenário internacional (Silva, 2016).

Embora o gênero *Coffea* compreenda 124 espécies, somente 10 são cultivadas em todo mundo (Bekele, 2005; Davis et al. 2011). Cerca de 70% da produção brasileira é de *Coffea arabica*, enquanto *Coffea canephora* representa aproximadamente 30%, sendo as espécies mais importantes economicamente (Conab 2016).

Por apresentar melhor qualidade na bebida, o café arábica é preferido entre as demais espécies (De Castro & Marraccini, 2006) e por isso é responsável por 70% do café consumido no mundo, liderando o mercado consumidor (Conab, 2017). Desde o grão até a bebida final, muitos componentes importantes estão presentes no café. A composição química dos grãos do café cru faz parte de muitos fatores que determinam a qualidade e a identificação desses componentes faz-se importante para indicar as características desejáveis da bebida. (Ferrão et al., 2003).

Compostos como a cafeína, os ácidos clorogênicos, a sacarose, os lipídios, as proteínas e os taninos são responsáveis por contribuir com o sabor e o aroma do café e assim estariam relacionados à sua qualidade (Ky et al., 2000; Fernandes et al., 2001; Ferrão et al., 2003; Campa et al. 2005; Farah et al., 2006; Esquivel & Jiménez, 2012). Cafeína e ácidos clorogênicos atribuem ao café acidez e adstringência, conferindo uma bebida mais amarga em altas concentrações (Tfouni et al. 2014). Os taninos, juntamente com esses compostos, também apresentam a propriedade de adstringência (Ferrão et al. 2003). Os açúcares, no processo de torra, reagem com aminoácidos e determinam o perfil dos compostos voláteis, assim como as proteínas, que na desnaturação liberam aminas e carbonilas, contribuindo para o aroma da

bebida (Ky et al. 2000; Fernandes et al., 2001). Já os lipídios influenciam tanto no sabor, quanto no aroma, e de acordo com o preparo do café, influenciam os níveis de diterpenos, uma das principais frações lipídicas do grão (Urgert & Katan 1997; Ricketts et al. 2007).

A hibridização natural entre os progenitores diploides *Coffea eugenioides* e *Coffea canephora* originou *C. arabica* (Lashermes et al., 1999), espécie alotetraploide ($2n = 4x = 44$ cromossomos) e que apresenta sistema reprodutivo predominantemente autógamo (Prakash et al., 2002). Além disso, *C. arabica* apresenta uma origem recente (Lashermes et al., 1999; Vidal et al., 2010), o que confere a essa espécie uma baixa diversidade genética (Lashermes et al., 1996). No entanto, uma maior variabilidade é encontrada no centro de origem de *C. arabica*, no sudoeste da Etiópia (Anthony et al., 2001; Silvestrini et al., 2007).

Em busca dessa variabilidade, prospecções foram organizadas para coletar genótipos originários dessa região (FAO, 1968) e as amostras coletadas foram encaminhadas para seis institutos: na Índia, na Tanzânia, na Etiópia, na Costa Rica, no Peru e em Portugal (Lashermes et al., 1996; Silvestrini et al., 2007). O Instituto Agrônomo de Campinas (IAC) obteve parte destes genótipos do *Centro Agronómico Tropical de Investigación y Enseñanza* (CATIE – Costa Rica), e desses, 132 genótipos foram fornecidos ao Instituto Agrônomo do Paraná (IAPAR). As plantas vêm sendo cultivadas em campo e, nos últimos anos, caracterizadas fenotípica e genotipicamente (Ferreira, 2013; Pagiatto, 2013; Scholz et al. 2016).

Os recursos genéticos de *C. arabica* são restritos devido a sua natureza aloploide e a sua baixa diversidade compromete a busca por novos marcadores moleculares para estudos da espécie. Atualmente há disponível apenas o genoma de um de seus progenitores diploides, *C. canephora* (Denoëud et al., 2014; Dereeper et al., 2015), servindo como base aos estudos genéticos. Apesar das pesquisas, não há o genoma de referência para *C. arabica* relatado na literatura. Recentemente, um mapa genético para a espécie foi descrito, utilizando diferentes tipos de marcadores relacionados à

QTLs, a partir de uma população F_2 de progenitores comerciais (Moncada et al., 2016).

Esses recursos são importantes para os estudos de associação genômica ampla (GWAS), que verificam se um conjunto de variantes genéticas, como os SNPs, em diferentes indivíduos estaria associado a alguma característica de interesse (Wang et al., 2015). Alguns avanços com espécies poliploides e que não apresentam genoma de referência também tem sido mostrados (Jia et al., 2014; Li et al., 2015; Zhang et al., 2015; Su et al., 2016). Em café estudos preliminares tem revelado resultados satisfatórios, tanto para *C. canephora* (Andrade et al., 2017) como também para *C. arabica* (Pereira et al., 2016).

Novas tecnologias de sequenciamento têm permitido um avanço no melhoramento vegetal em geral (Metzker, 2010), o que tem contribuído para os estudos de associação. A genotipagem por sequenciamento (GBS) caracteriza-se por ser uma metodologia de genotipagem em larga escala, em um sistema multiplexado, direto e que assegura o enriquecimento ou a redução da complexidade de genomas grandes por enzimas de restrição, garantindo sobreposição suficiente na cobertura das sequências, visando a identificação de polimorfismos (Elshire et al., 2011). O baixo custo dessa metodologia tem atraído a busca por marcas e mapas mais saturados, tanto em espécies diploides (Spindel et al., 2013; Celik et al., 2016; Henning et al., 2016; Velmurugan, et al., 2016) como também em poliploides (Lu et al., 2013; Li et al., 2015; Zhang et al., 2015; Moncada et al., 2016). O grande número de dados gerados é em parte resolvido, por programas que auxiliam nas análises (Bradbury et al, 2007; Glaubitz et al., 2014).

Neste trabalho, para uma melhor compreensão da diversidade genética de uma coleção de genótipos de *C. arabica* da Etiópia obteve-se SNPs por GBS, utilizando como genoma de referência um de seus ancestrais diploides, *C. canephora*. Além disso, a estratégia de GWAS foi utilizada para identificar marcas associadas aos conteúdos bioquímicos dos genótipos dessa coleção.

5.4 MATERIAIS E MÉTODOS

5.4.1 Material Vegetal

No presente estudo foram analisados 66 genótipos de *C. arabica* coletados pela missão da FAO (FAO, 1968) no centro primário de origem da espécie, a Etiópia, e mantidos a campo no IAPAR (23°23'00"S e 51°11'30"W).

5.4.2 Genotipagem por Sequenciamento

Os DNAs de alto peso molecular foram extraídos de folhas de plantas individuais usando o protocolo CTAB com modificações (Doyle & Doyle, 1990; Brasileiro & Carneiro, 1998). A genotipagem por sequenciamento foi realizada pela *Genomic Diversity Facility LIMS*, em Cornell University (Ithaca, NY – EUA). Para a preparação das bibliotecas a enzima de restrição *Pst*I foi utilizada, seguindo os protocolos descritos em Elshire et al. (2011). Um total de 157 genótipos de *C. arabica*, incluindo os genótipos do centro de origem, cultivares e outros genótipos introduzidos na coleção, provenientes da República de Camarões, (Suplementar 1), foram distribuídos em duas placas de 96 poços e enviados para o sequenciamento *single-end* usando o Illumina HiSeq2000. O TASSEL-GBS *pipeline* descrito em Glaubitz et al. (2014) foi utilizado para a identificação dos SNPs (TASSEL versão 3.0.166).

5.4.3 Fenotipagem para Compostos Bioquímicos

Frutos maduros foram coletados em 2012 e 2015, entre os meses de abril e junho, período de maior maturação dos frutos. Um painel de 66 genótipos da coleção de *C. arabica* do IAPAR foi utilizado (Suplementar 2), representando a sua diversidade. Os frutos maduros foram secos naturalmente em caixas com fundo de malha. Após secos, os frutos foram descascados mecanicamente e então, armazenados em sacos de papel em local seco e escuro por um período de três meses. Os grãos de café verde foram congelados com nitrogênio líquido, moídos em moinho de disco PERTEN 3600 e passados em peneiras com malha 0,5 mm. Após a moagem, as amostras foram colocadas em frascos plásticos identificados e foram mantidas congeladas a -18 C até posterior análise.

A composição química geral das amostras coletadas foi analisada utilizando a Espectroscopia no Infravermelho Próximo (NIRS), no espectrofotômetro NIRSystem 6500, com reflectância difusa e região espectral de 400 a 2500 nm. Modelos de previsão desenvolvidos anteriormente (Scholz et al 2014a, b) foram empregados para cada componente. Foram realizadas varreduras entre 1100 e 2300 nm em intervalos de 2 nm, utilizando-se uma cubeta retangular contendo 6 g de café moídos e em temperatura ambiente. Os *softwares* ISIScan e WinISI 4.5 foram utilizados para os cálculos. A análise estatística foi realizada pelo *software* XLSTAT (Addinsoft, 2010).

5.4.4 Estrutura Genética da Coleção e Desequilíbrio de Ligação

Para analisar a estrutura genética da população foi utilizado o método proposto por Alexander et al. (2009). O número mais provável de grupos em que a população pode ser estruturada foi estimado com base nos valores de entropia cruzada (Zhou et al. 2009). Essa análise foi realizada com auxílio do *software* Sniplay (Dereeper et al., 2011).

A estrutura populacional também foi analisada por meio da análise de componentes principais com base nos 2460 SNPs utilizando o programa TASSEL 5.2.20 (Bradbury et al., 2007).

O desequilíbrio de ligação (DL) foi estimado entre cada par de SNPs de um mesmo cromossomo com menos de 20% de dados perdidos e frequência do menor alelo maior que 0.05 (MAF > 0.05). O DL foi estimado de acordo com Weir (Weir & Cockerham, 1984) pelo cálculo de r^2 para cada pare de SNPs usando o *software* TASSEL 5.2.20 (Bradbury et al., 2007). De acordo com os padrões de DL obtidos em todo o painel, estimou-se o número de marcadores necessários para o mapeamento associativo considerando o limiar $r^2 = 0,2$.

As distâncias físicas correspondentes a esses limiares foram obtidas a partir da regressão não-linear da decomposição de DL com a distância. O número de SNPs necessários para os estudos de associação considerando estes limiares foi calculado dividindo o tamanho do genoma de *C. canephora*

(710 Mb, Deneud et al., 2014) pela distância de decaimento DL para todo o painel de associação.

5.4.5 Associação Genômica Ampla

A análise de associação entre os 2460 marcadores SNPs e os teores de ácidos clorogênicos, cafeína, açúcares totais, sacarose, lipídios, proteínas e taninos foi realizada utilizando modelos lineares generalizados (GLM). Para correção do efeito da estrutura populacional, uma matriz de estrutura populacional considerando os três primeiros componentes da PCA foi incluída no modelo. Todas essas análises foram realizadas utilizando o *software* TASSEL 5.2.20 (Bradbury et al, 2007). Para verificar o ajuste do modelo para cada característica, foram gerados os gráficos *Quantile-Quantile* (Q-Q) *plots*, mostrando a distribuição dos *p-values* observados em relação aos *p-values* esperados. Foram considerados significativamente associados as características fenotípicas analisadas os SNPs que apresentaram *p-value* < 0,005 e com MAF (*minor allele frequency*) maior que 0,005. O banco de dados *Coffee Genome Hub* (Dereeper et al., 2015) foi usado para mapear os SNPs no genoma de *C. canephora*. Genes localizados no intervalo de 100 Kb foram considerados como candidatos a associação com as características fenotípicas. Para comparar a distribuição dos valores fenotípicos dos indivíduos com diferentes genótipos para os SNPs significativos (homozigotos para os diferentes alelos e os heterozigotos) foram construídos gráficos box plots com auxílio do programa R.

5.5 RESULTADOS E DISCUSSÃO

5.5.1 Análise Fenotípica

Foram analisados os teores de sete componentes bioquímicos de grãos de café verde de 66 genótipos da coleção do IAPAR (Suplementar 2), coletados nos anos de 2012 e de 2015: ácidos clorogênicos, cafeína, açúcares totais, sacarose, lipídeos, proteínas e taninos.

Tabela 1 Valores mínimos, máximos e médios e coeficiente de variação para os compostos químicos dos genótipos de café da Etiópia coletados em 2012 e 2015

Compostos ^a	2012				2015				
	Mínimo	Máximo	Média	CV %	Mínimo	Máximo	Média	CV %	
ACG	6,44	10,92	8,06	10,79	ACG	1,66	6,92	4,15	24,13
AT	3,26	8,47	5,46	19,32	AT	3,41	7,58	5,44	16,16
Sacarose	3,1	8,26	5,34	19,63	Sacarose	2,65	7,1	4,99	18,27
Cafeína	1,03	1,61	1,26	10,01	Cafeína	0,88	1,6	1,28	12,15
Lipídios	11,32	15,34	13,79	6,31	Lipídios	12,63	17,05	14,73	6,17
Proteínas	12,59	17,62	14,89	7,28	Proteínas	11,68	16,4	14,09	7,89
Taninos	4,17	6,19	5,34	7,15	Taninos	4,03	6,6	5,45	10,78

ACG ácidos clorogênicos CV coeficiente de variação

AT açúcares totais

^a Expresso $g100^{-1}$

Na análise dos dois anos observou-se uma alta variabilidade dos genótipos para as características estudadas. Foi observada uma diferença expressiva dos valores de ACG entre 2012 e 2015. Os ácidos clorogênicos são encontrados em maiores concentrações em grãos de café verde, e esta concentração é sabidamente relacionada também com o grau de maturação do grão (Farah et al. 2006; Jeszka - Skowron et al. 2016). Em 2012 concentração de ACG variou de 6.44 a 10.92 $g100^{-1}$, valores próximos aos encontrados por outros autores que analisaram os teores de ACG em grãos verdes de cafés cultivados (Farah et al. 2005). A redução da concentração desse composto no ano de 2015 deve-se a coleta de frutos com maturação mais uniforme, com a separação de frutos verdes dos frutos cereja.

Correlações positivas significativas entre cafeína e proteínas se mantiveram nos dois anos estudados ($r_{2012} = 0,631$ e $r_{2015} = 0,769$) (Tabela 2). Isso pode ser explicado pelo fato de que no metabolismo de ambos componentes, participam compostos nitrogenados que se acumulam durante a formação do fruto (Scholz et al. 2016). Nos dois anos foram observadas correlações negativas entre sacarose e lipídios ($r_{2012} = -0,388$ e $r_{2015} = -0,369$). No desenvolvimento do grão, o acúmulo de lipídeos cessa enquanto o acúmulo de açúcares é contínuo até o final da maturação (Joët et al., 2010).

Tabela 2 Correlação de Pearson entre os componentes nos genótipos de café da Etiópia coletados em 2012 e 2015.

2012							
Variáveis	ACG	AT	Sacarose	Cafeína	Lipídios	Proteínas	Taninos
ACG	1						
AT	-0,090	1					
Sacarose	-0,058	0,992	1				
Cafeína	0,396	-0,061	-0,064	1			
Lipídios	-0,144	-0,339	-0,388	-0,128	1		
Proteínas	0,431	-0,287	-0,277	0,631	-0,030	1	
Taninos	0,424	-0,233	-0,236	0,481	0,036	0,110	1

2015							
Variáveis	CGA	AT	Sacarose	Cafeína	Lipídios	Proteínas	Taninos
ACG	1						
AT	0,238	1					
Sacarose	0,197	0,992	1				
Cafeína	0,171	-0,237	-0,216	1			
Lipídios	-0,157	-0,381	-0,369	0,208	1		
Proteínas	0,286	-0,257	-0,250	0,769	0,146	1	
Taninos	-0,284	-0,409	-0,373	0,444	0,629	0,159	1

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

As correlações significativas entre ACG e cafeína e entre ACG e taninos foram diferentes entre os anos estudados. Enquanto em 2012 houve uma correlação significativa positiva entre ACG e cafeína ($r = 0,396$), em 2015 ela não foi significativa. Já a correlação entre ACG e taninos apresentou resultados opostos: em 2012 ela foi positiva ($r = 0,424$) e em 2015 ela foi negativa ($r = - 0,284$). A coleta de frutos de maturação incompleta influencia as concentrações desses compostos (Vaast et al, 2006; Scholz et al, 2011). A presença de frutos verdes, conseqüentemente, de um maior teor de ACG (Jeszka- Skowron et al. 2016) pode explicar em como esse teor aumentado influenciou nas correlações entre os compostos, o que não foi observado em 2015, onde houve a coleta de frutos com maturação homogênea.

5.5.2 GBS e Detecção dos SNPs

Os *reads single-end* produzidos pelas bibliotecas GBS de 157 genótipos de *C. arabica* renderam 6.210.920 tags. Dessas, apenas 20% alinharam-se no genoma de referência, o ancestral diploide *C. canephora*, em posições únicas. Contudo, foram identificados 6.696 SNPs, com profundidade média de 39X.

Após a filtragem para $MAF > 0.05$, $Call\ Rate > 0.8$ e heterozigosidade < 0.9 , foram obtidos 2.460 marcadores SNPs de alta qualidade que foram usadas nas análises de estrutura populacional e de associação. O trabalho de identificação de marcas em espécies alotetraplóides como *C. arabica* é mais complexo do que em espécies diploides, já que existe uma dificuldade em identificar os SNPs próprios da espécie daqueles que mostram as diferenças entre os dois subgenomas diploides (Vidal et al. 2010, Logan-Young et al. 2015). Para isso a filtragem pela heterozigosidade < 0.9 permitiu que esses apenas SNPs segregantes entre os homozigotos fossem identificados e utilizados nas análises de estrutura populacional e de associação, pois aqueles que demonstravam as diferenças entre os subgenomas (heterozigosidade = 1) foram descartados.

5.5.3 Análise da Estrutura Genética da Coleção e Desequilíbrio de Ligação

Os valores da entropia cruzada da análise da estrutura genética populacional indicaram que o número mais provável de grupos em que a população pode ser dividida (K) é 2, embora a estruturação em 3 grupos ($K = 3$) também apresentou um baixo valor de entropia cruzada (Figura 1A).



Figura 1 - Valor de entropia cruzada em função do número de grupos (A) e estrutura populacional dos genótipos de *C. arabica* com $K = 3$ (B).

Considerando a estrutura populacional em três grupos (Figura 1B), 13 genótipos do Oeste do Vale do Rift e um do Leste (E238) foram alocados no grupo Q1. O grupo Q2 foi formado por sete genótipos, todos do Oeste. Já no grupo Q3, formado por 45 genótipos, ficaram alocados o maior número de genótipos do Leste analisados (E007, E018, E021), dentre os demais genótipos do Oeste do Vale do Rift.

Na análise do decaimento do DL foi observado um valor de $r^2 = 0,2$ entre SNPs posicionadas a 4.3 Mb de distância entre si (Figura 2). Com base nesse valor de decaimento, estimamos que o número de marcadores necessários para realizar GWAS dividindo o tamanho do genoma (710Mb – genoma de *C. canephora*) pela distância em que $r^2 = 0,2$ (4.3 Mb). Assim, verificou-se que 1.651 SNPs são suficientes para realizar GWAS com cobertura razoável do genoma. No presente trabalho foram utilizados 2.460 SNPs distribuídos por 11 cromossomos do genoma de referência, o que indica que uma boa cobertura do genoma.

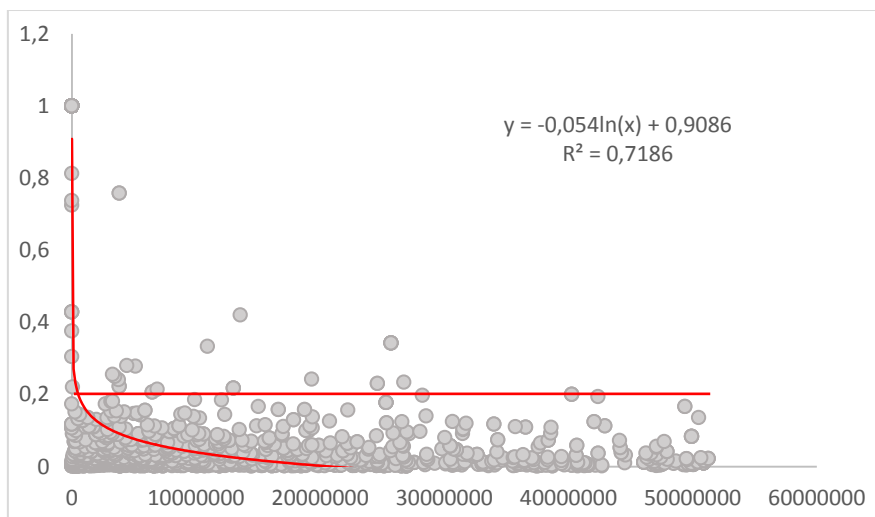


Figura 2 Decaimento do DL (medida por r^2) em função da distância genética em pb para o painel de *C. arabica*.

5.5.4 Estudos de associação para os componentes bioquímicos do café verde.

As análises de associação genômica foram realizadas considerando os resultados das fenotipagens realizadas em dois anos (2012 e 2015) e também considerando a média dos dois anos. Após detecção dos SNPs significativos nas diferentes análises (2012, 2015 e média) foi possível identificar aqueles comuns entre elas.

Ácidos Clorogênicos (ACG)

Para ácidos clorogênicos, foram encontrados 27, 12 e 20 SNPs significativamente associados em 2012, 2015 e média, respectivamente (Figura 3), com p -values entre 9.79^{-7} a 0.005. Foram detectados sete SNPs comuns entre 2012 e a média (S2_4604235, S2_27727131, S2_43079004, S5_11612082, S7_18668686, S8_29163163 e S9_2563555) (Tabela 3). A Figura 4 mostra a distribuição dos valores fenotípicos entre diferentes genótipos para o SNP S8_29163163, significativamente associado a ACG.

Figura 3 Manhatan e Q-Q plots para os teores de ácidos clorogênicos em 2012, 2015 e média.

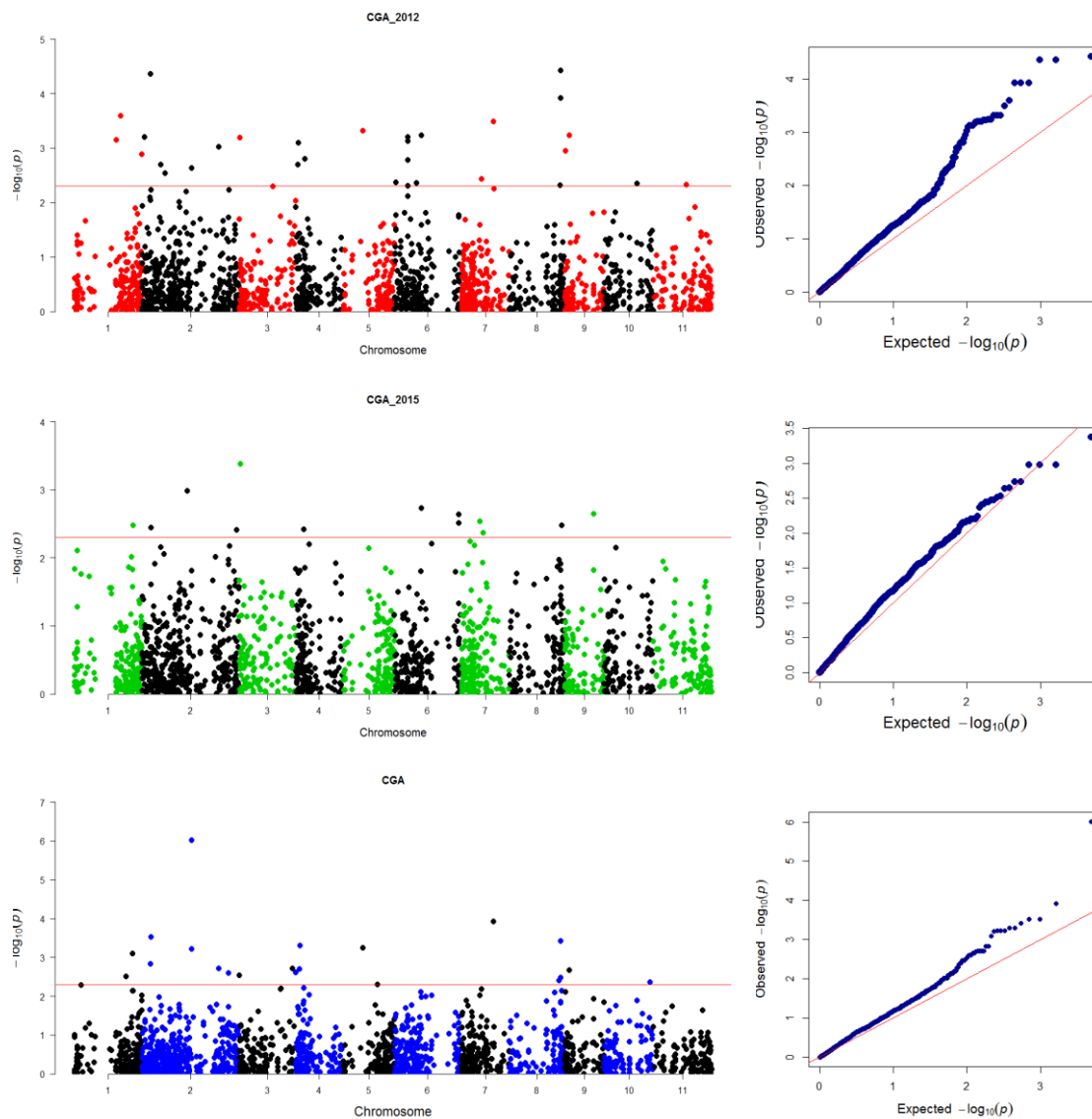
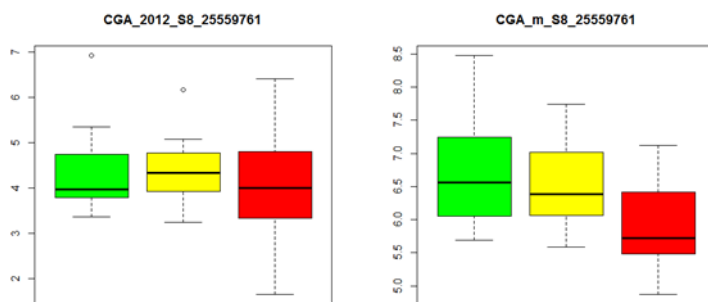


Figura 4 Valores fenotípicos entre diferentes genótipos para o SNP S8_29163163, significativamente associado ao teor de ACG nos grãos. Os homozigotos para o alelo referência e para o alelo alternativo estão representados em verde e vermelho, respectivamente, e os heterozigotos em amarelo.



Para o SNP S8_29163163, indivíduos com genótipo com alelo referência (verde) mostraram uma concentração de ACG superior à dos indivíduos heterozigotos (amarelo) e à dos indivíduos com alelo alternativo (vermelho). O SNP S8_29163163 localiza-se dentro do gene Cc08_g14110 (Tabela 3), que codifica a proteína contendo o domínio PAS da família tirosina quinase. Essa proteína atua como regulador negativo na via de resposta ao etileno (Kieber et al, 1993), envolvido na maturação do fruto do cafeeiro, e que promove um aumento da atividade respiratória no início da maturação. Nesse período, além de outros eventos, ocorre a síntese de etileno endógeno, afetando o metabolismo de açúcares e de ácidos clorogênicos, além de promover a degradação da clorofila e aumento da síntese de carotenoides, antocianinas e antocianidinas, flavonoides que darão coloração aos frutos maduros (Martinez et al, 2014).

Cafeína

Foram detectados 5, 15 e 14 SNPs significativamente associados ao teor de cafeína em 2012, 2015 e média, respectivamente (Figura 5), com *p-values* entre 1.24^{-5} a 0.005. Foram observados um SNP comum entre 2012 e a média (S4_1614149), três SNPs comuns entre 2015 e a média (S2_25412500, S2_5043332 e S8_25559580) e um SNP comum entre todas as análises (S9_2563555) (Tabela 3). A Figura 6 mostra a distribuição dos valores fenotípicos entre diferentes genótipos para os SNPs significativamente associados a cafeína.

Figura 5 Manhattan e Q-Q plots para os teores de cafeína em 2012, 2015 e média.

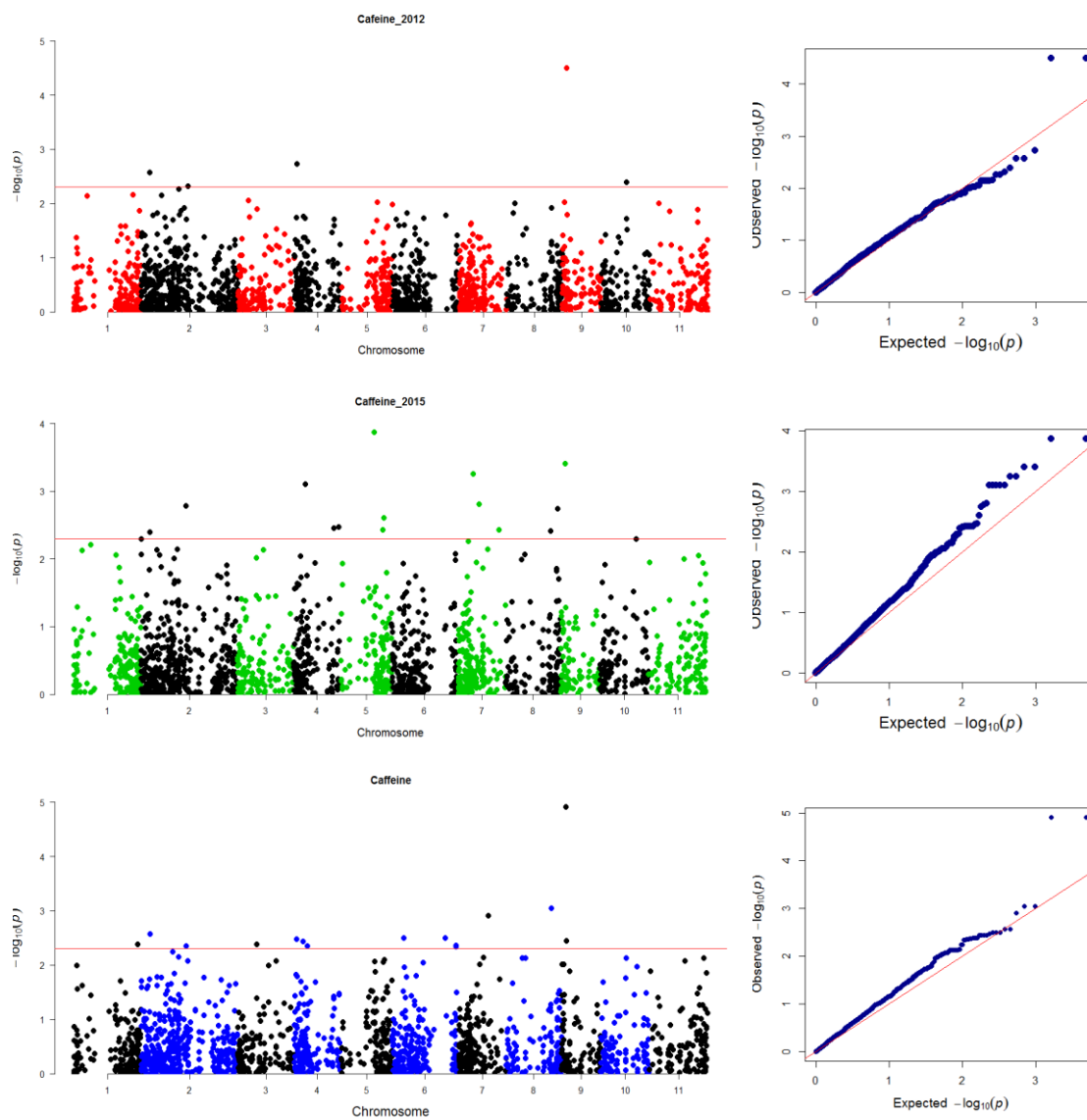
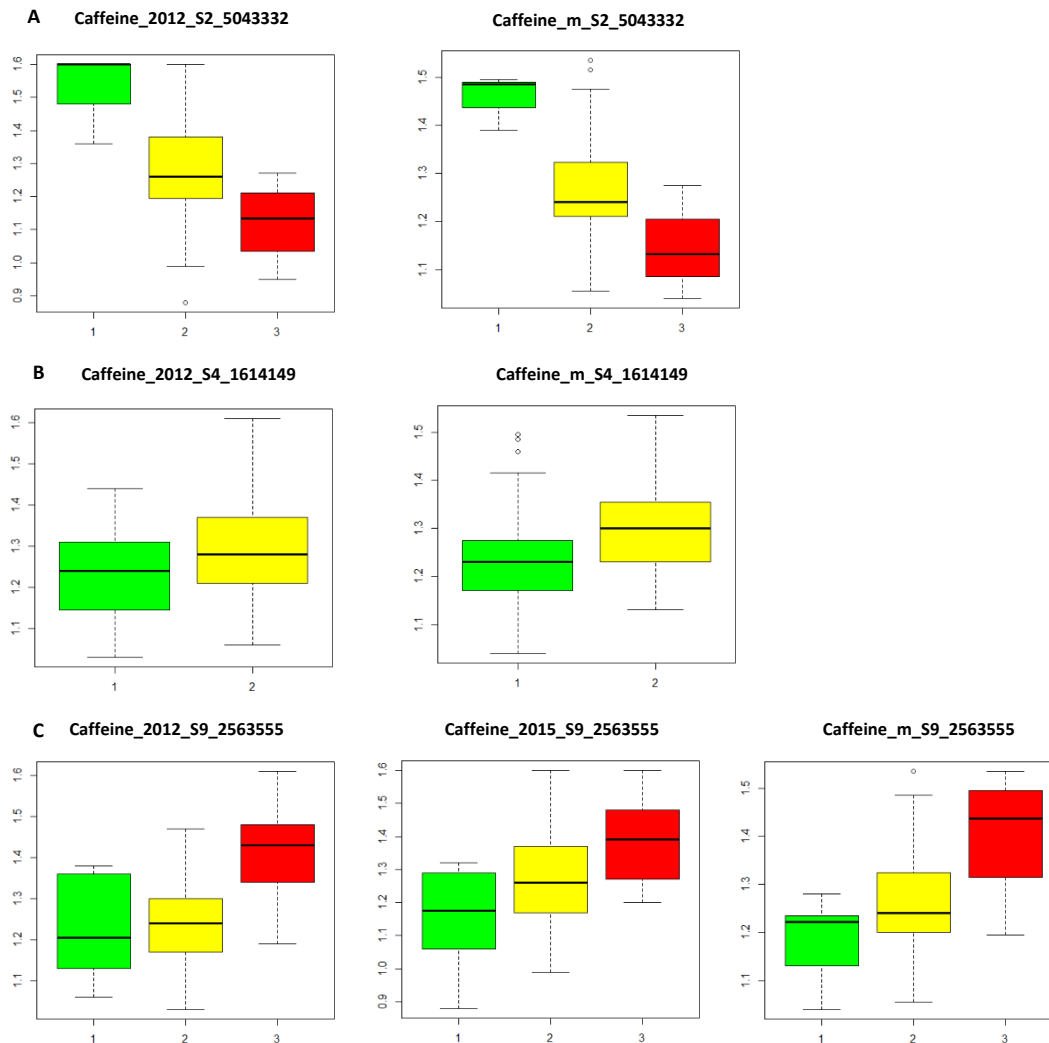


Figura 6 Valores fenotípicos dos diferentes genótipos para os SNPs S2_5043332, S4_1614149 e S9_2563555, significativamente associados ao teor de cafeína nos grãos.



Para o SNP S2_5043332 (Figura 6A), os indivíduos com genótipo com alelo referência (verde) mostraram uma concentração de cafeína significativamente maior do que indivíduos heterozigotos (amarelo), seguido de indivíduos com alelo alternativo (vermelho). Para o SNP S4_1614149 (Figura 6B) os indivíduos homozigotos para o alelo referência apresentaram teor de cafeína superior à dos indivíduos heterozigotos. Já para o SNP S9_2563555 (Figura 6C), indivíduos com genótipo com alelo referência (vermelho) mostraram uma concentração de cafeína significativamente maior do que indivíduos heterozigotos (amarelo), seguido de indivíduos com alelo alternativo (verde).

Apesar de não se terem sido identificados SNPs ligados a genes especificamente relacionados as etapas finais da via de biossíntese da cafeína, foram detectados SNPs posicionados próximos a genes relacionados ao metabolismo de proteínas. Como foi observado na análise fenotípica, o teor de cafeína nos grãos é fortemente correlacionado ao teor de proteínas ($r_{2012} = 0,631$ e $r_{2015} = 0,769$; Tabela 2). Assim, encontramos os SNPs S2_5043332, S4_1614149 e S9_2563555, relacionados a genes envolvidos com a via de ubiquitinação das proteínas.

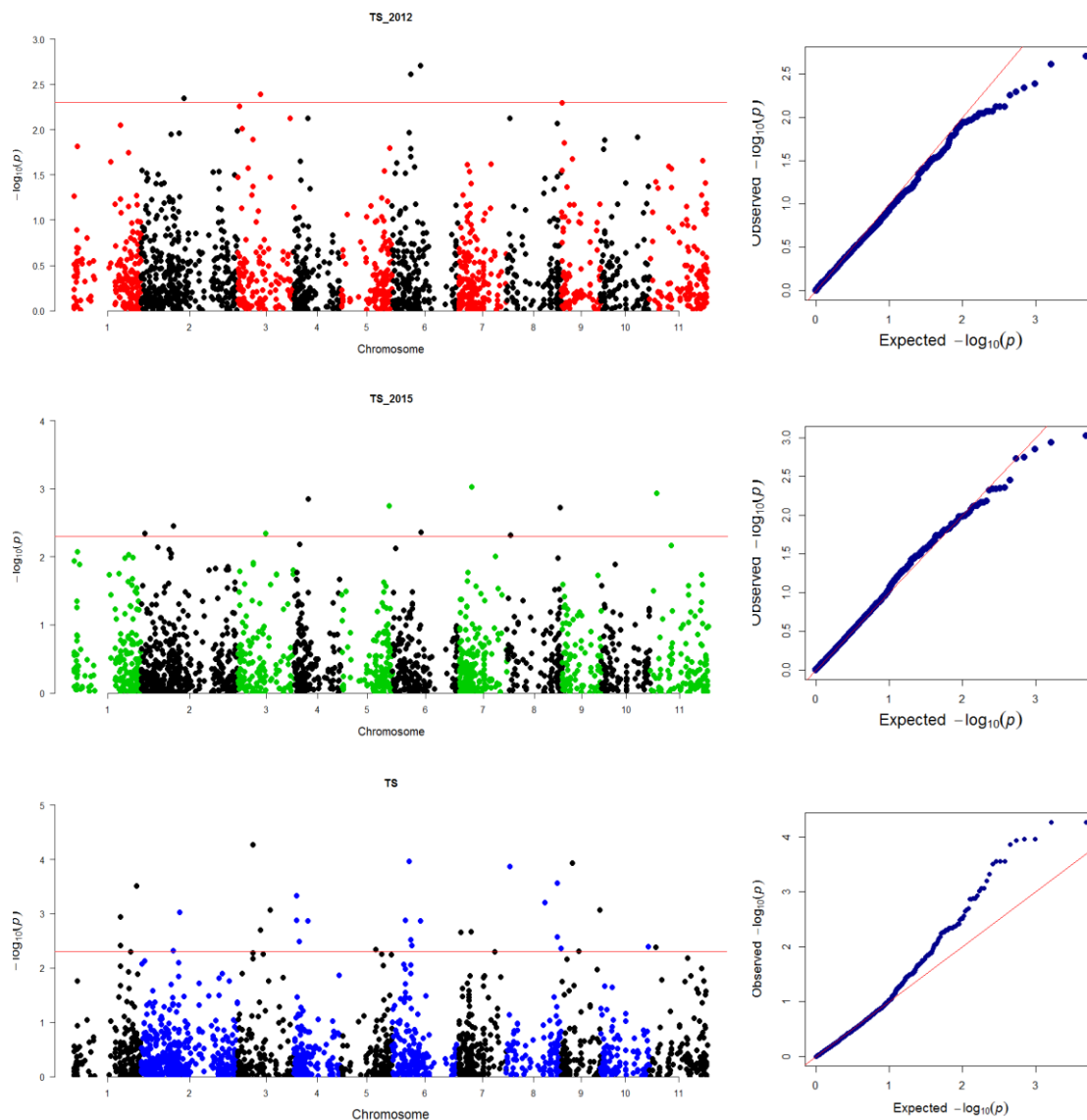
O SNP S2_5043332 está posicionado próximo (78536 pb *downstream*) ao gene Cc02_g06280, envolvido com a proteína componente do complexo de ubiquitina-proteína ligase APC / C, que controla vários passos chave no ciclo celular mediando a ubiquitinação e subsequente degradação de proteínas alvo tais como ciclinas (Heyman et al, 2011). O SNP S4_1614149 está posicionado próximo (71740 pb *upstream*) ao gene Cc04_g02170, que codifica a proteína ubiquitina carboxil-terminal hidrolase 6, envolvida no reconhecimento e hidrólise da ligação peptídica no C-terminal Gly da ubiquitina e no processamento de precursores de poli-ubiquitina, bem como o de proteínas ubiquitinadas (Moon et al, 2005). Já o SNP S9_2563555 está posicionado próximo (89587 pb *upstream*) ao gene Cc09_g03100, envolvida na estabilidade da proteína pós-traducional (Bostick al., 2004), além do gene Cc09_g03020 (8804 pb *downstream*), um provável transportador de amônio, envolvido na absorção de amônio do solo (Gazzarrini et al., 1999). Esse SNP apresentou o menor *p-value* e explica 30% da variação fenotípica ($R_{sq} = 0,30$) do teor de cafeína nos grãos, o que corrobora com a correlação positiva observada entre os teores de cafeína e proteínas.

Açúcares Totais (AT)

Para açúcares totais foram encontrados 5, 10 e 31 SNPs significativamente associados em 2012, 2015 e média, respectivamente (Figura 7), com *p-values* entre 5.46^{-5} a 0.005. Foram detectados dois SNPs comuns entre 2012 e a média (S3_12834205 e S6_10652258), quatro SNPs comuns entre 2015 e a média (S2_18205445, S7_7540264, S8_2154851 e

S11_3798869) e 2 SNPs comuns entre todos os perfis de análises (S4_7902771 e S6_16281119) (Tabela 3).

Figura 7 Manhattan e Q-Q plots para os teores de AT em 2012, 2015 e média.



Sacarose

Para sacarose foram encontrados 6, 11 e 35 SNPs significativamente associados em 2012, 2015 e média, respectivamente (Figura 8), com *p-values* entre 3.15^{-5} a 0.005. Foram encontrados dois SNPs comuns entre 2012 e a média (S3_12834205 e S6_10652258), quatro SNPs comuns entre 2015 e a média (S2_18205445, S7_7540264, S8_2154851 e S11_3798869) e dois SNPs comuns entre todos os perfis de análises

(S4_7902771 e S6_16281119) (Tabela 3), as mesmas marcas encontradas para açúcares totais. Essa coincidência corrobora com a correlação positiva encontrada para o teor desses dois compostos, nas análises fenotípicas, tanto em 2012 como em 2015 ($r = 0,992$). A Figura 9 mostra a distribuição dos valores fenotípicos entre diferentes genótipos para SNPs significativamente associados a sacarose.

Figura 8 Manhattan e Q-Q plots para os teores de sacarose em 2012, 2015 e média.

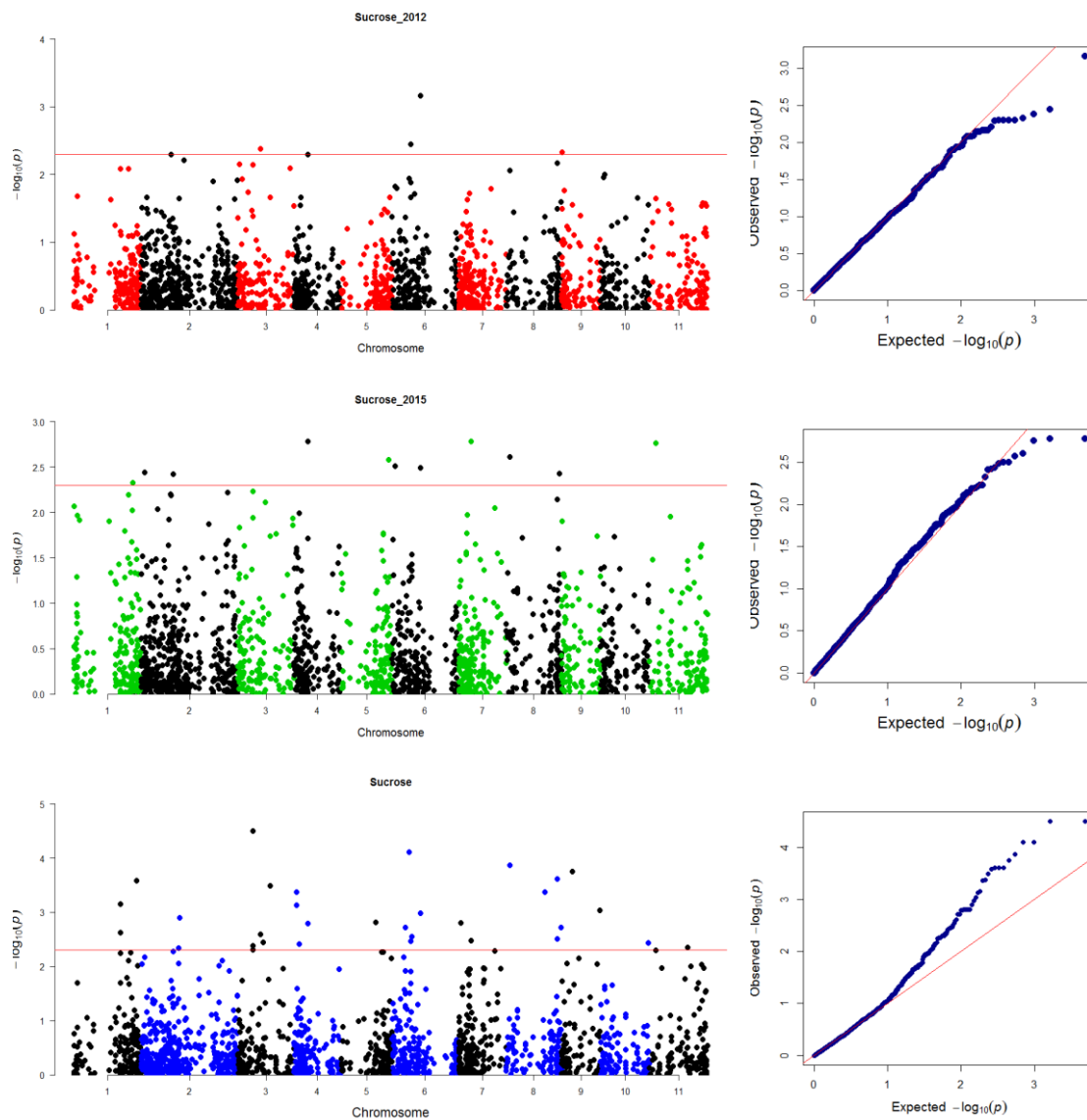
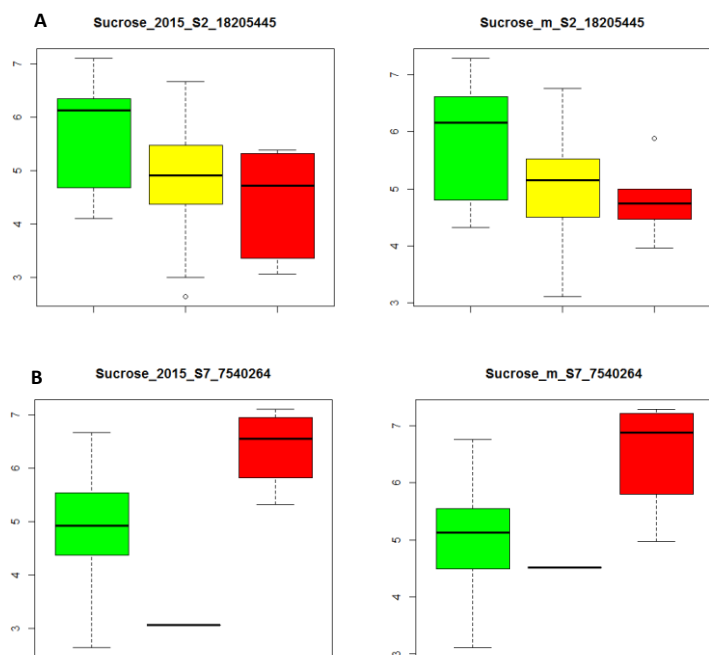


Figura 9 Valores fenotípicos entre diferentes genótipos para os SNPs S2_18205445 e S7_7540264, significativamente associados ao teor de sacarose nos grãos.



Para o SNP S2_18205445 (Figura 9A), os indivíduos com genótipo com alelo referência (verde) mostraram uma concentração de sacarose significativamente maior do que indivíduos heterozigotos (amarelo), seguido de indivíduos com alelo alternativo (vermelho). Para o SNP S7_7540264 (Figura 9B) os indivíduos homozigotos para o alelo referência apresentaram teor de cafeína superior à dos indivíduos heterozigotos.

O SNP S2_18205445 está posicionado próximo (82862 pb *upstream*) ao gene Cc02_g20460, que codifica a proteína Pirofosfatase inorgânica solúvel 1, que controla o equilíbrio das reações gliconeogênicas na fase de estabelecimento de plântulas. Além disso, determina a taxa de glicólise citosólica, disponibilizando carbono para o acúmulo de lipídios armazenados nas sementes (Meyer e al., 2012), lembrando que há uma correlação significativa negativa entre esses dois compostos durante o desenvolvimento do fruto. Já o SNP S7_7540264 está posicionado próximo (82862 pb *upstream*) ao gene Cc07_g10450, que codifica a proteína fructose-bisfosfatase aldolase 2, que desempenha um papel fundamental na glicólise e na gliconeogênese. Esta proteína está relacionada com a síntese do D-gliceraldeído 3-fosfato e fosfato de glicerona a partir de D-glucose (Lu et al., 2012).

Lipídios

Para lipídios foram encontrados 28, 2 e 16 SNPs significativamente associados em 2012, 2015 e média, respectivamente (Figura 10), com p -values entre 3.12^{-5} a 0.005. Foram encontrados cinco SNPs comuns entre 2012 e a média (S6_7634389, S6_14809031, S8_27153678, S10_3632868 e S11_6285049) e um SNP comum entre 2015 e a média (S2_251546) (Tabela 3). A Figura 11 mostra a distribuição dos valores fenotípicos entre diferentes genótipos para os SNPs significativamente associados a lipídios.

Figura 10 Manhattan e Q-Q plots para os teores de lipídios em 2012, 2015 e média.

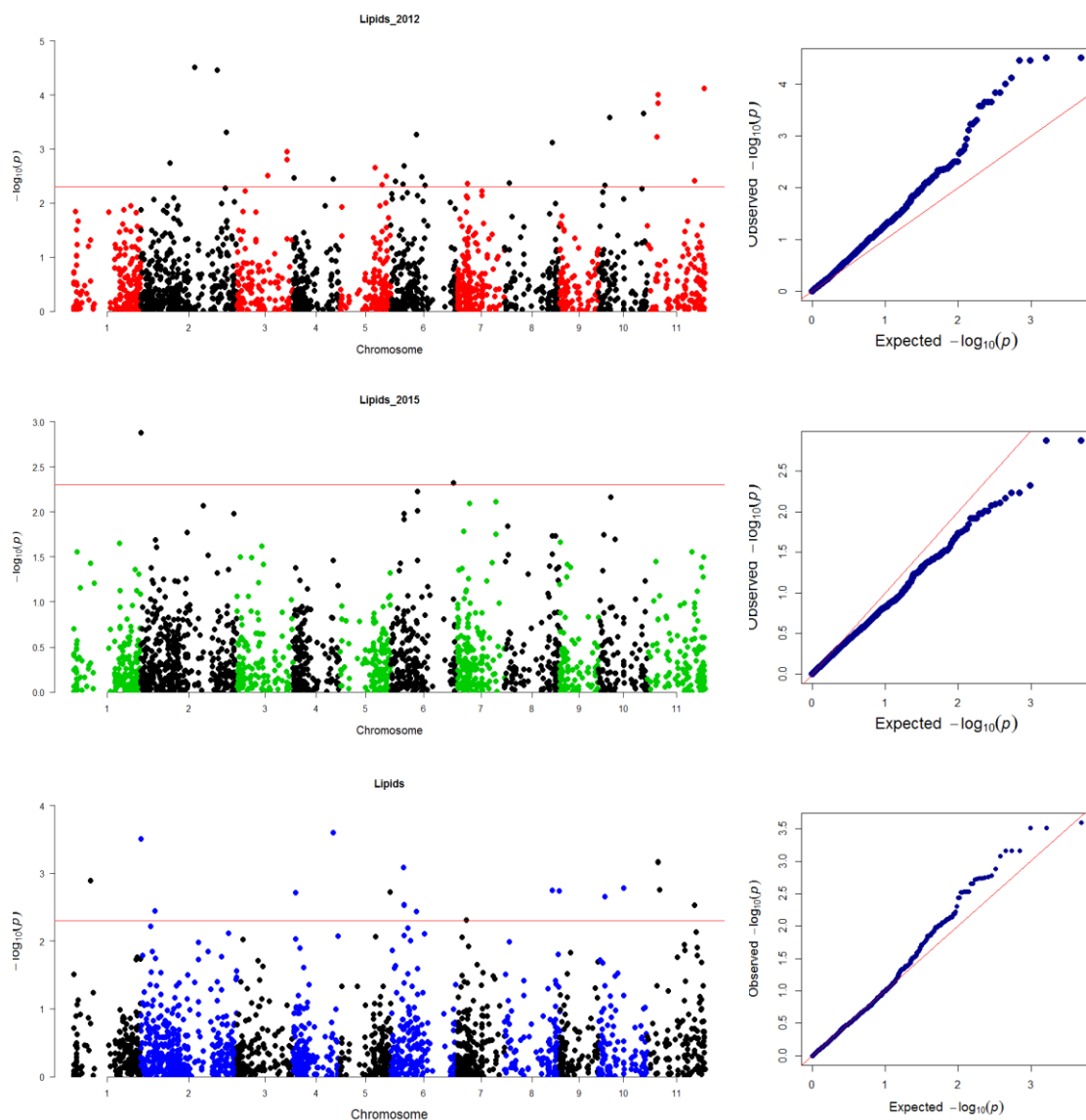
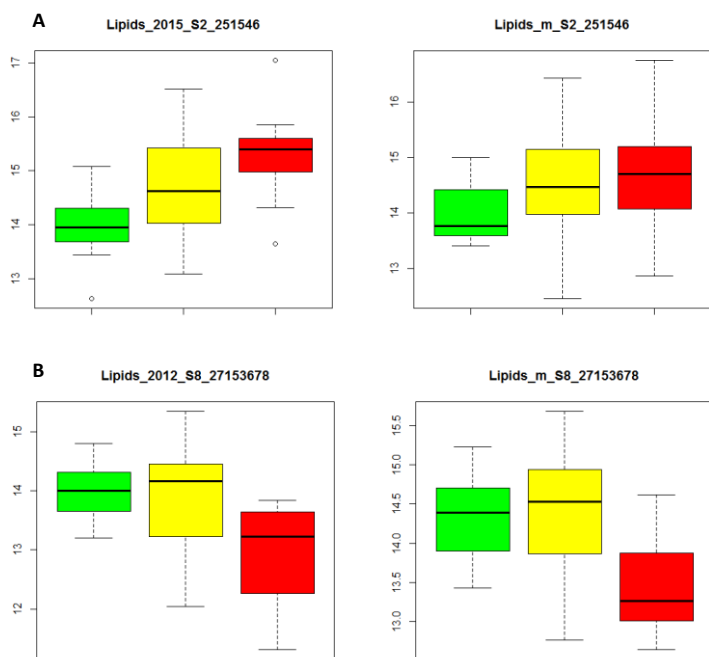


Figura 11 Valores fenotípicos entre diferentes genótipos para os SNPs S2_251546 e S8_27153678, significativamente associados ao teor de lipídios nos grãos.



Para o SNP S2_251546 (Figura 11A), indivíduos com genótipo com alelo referência (vermelho) mostraram uma concentração de lipídios superior à dos indivíduos heterozigotos (amarelo) e à dos indivíduos com alelo alternativo (verde). Para o SNP S8_27153678 (Figura 11B) indivíduos com genótipo heterozigoto (amarelo) mostraram uma concentração de lipídios superior aos indivíduos homozigotos (verde e vermelho).

O SNP S2_251546 está posicionado próximo (28872 pb *downstream*) ao gene Cc02_g00250, relacionado ao fator de transcrição MYB30, alvo da proteína Myb 1. Esse fator regula a expressão dos genes que codificam as quatro enzimas que formam o complexo de acil-coA elongase, que sintetiza os ácidos graxos de cadeia muito longa. (Raffaele et al., 2008), relacionando-se assim a biossíntese de lipídios. Já o SNP S8_27153678 está posicionado próximo (814 pb *upstream*) ao gene Cc08_g11900, relacionado a oleosina, podendo ter um papel estrutural para estabilizar o corpo lipídico durante a dessecação da semente, evitando a coalescência do óleo. Pode também proporcionar sinais de reconhecimento para ancoragem de lipase específica em lipólise durante o crescimento de plântulas. (Lee, Ratnayake &

Huang, 1995). Genes relacionados apresentam expressão diferencial no desenvolvimento e germinação das sementes de café (Simkin et al., 2006).

Proteínas

Para proteínas foram encontrados 5, 12 e 5 SNPs significativamente associados em 2012, 2015 e média, respectivamente (Figura 12), com *p-values* entre 2.20^{-4} a 0.005. Foram encontrados dois SNPs comuns entre 2015 e a média (S6_772355 e S6_36332515) e um SNP comum entre todos os perfis de análises (S2_4604235) (Tabela 3). A Figura 13 mostra a distribuição dos valores fenotípicos entre diferentes genótipos para os SNPs significativamente associados as proteínas.

Figura 12 Manhattan e Q-Q plots para os teores de proteínas em 2012, 2015 e média.

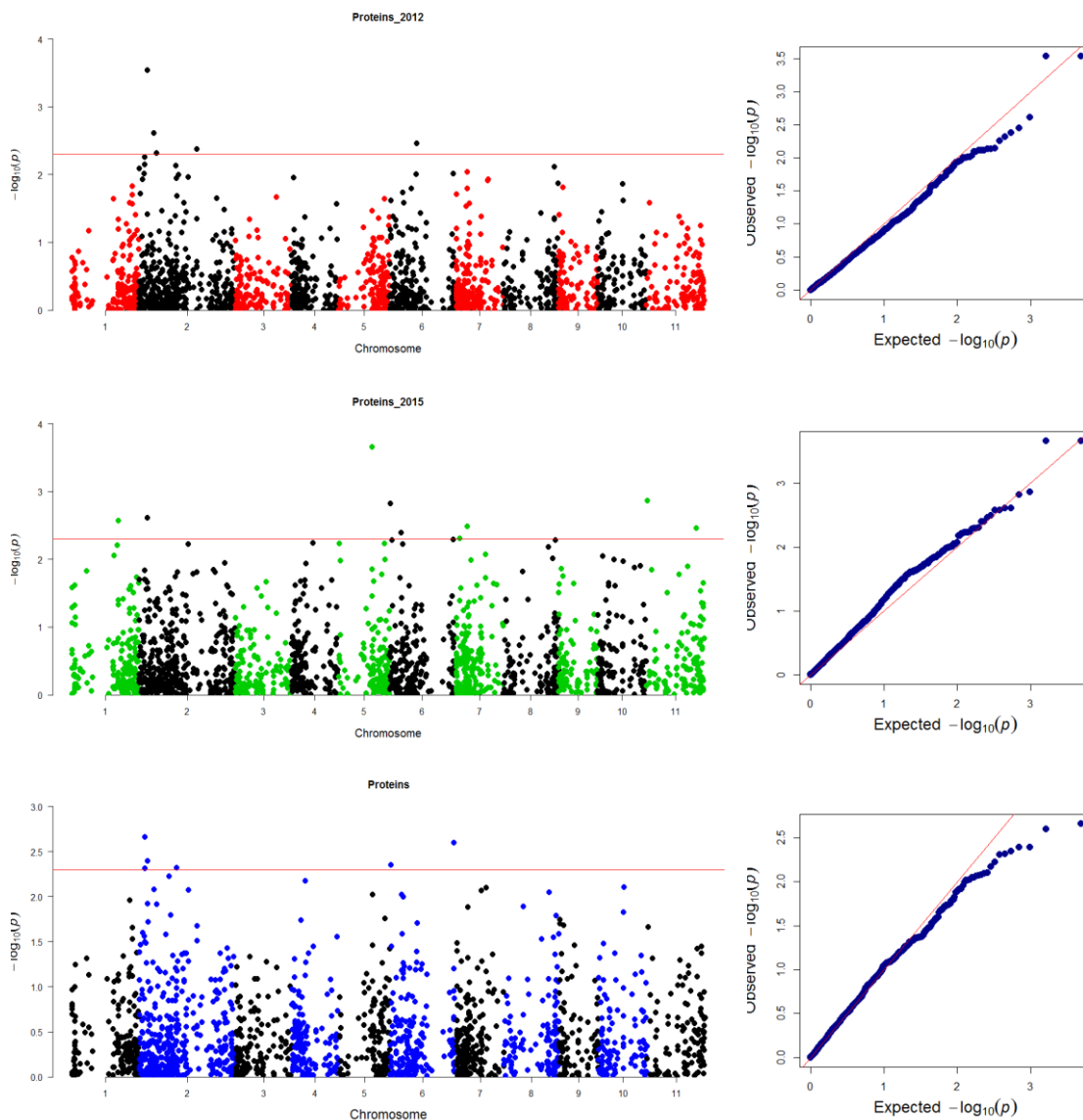
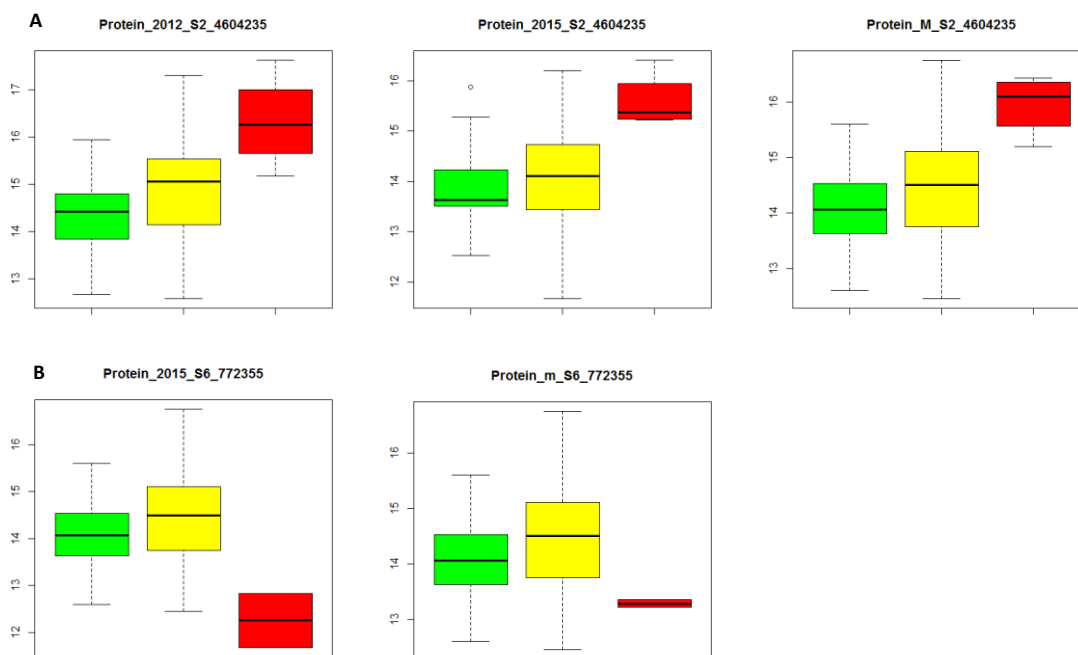


Figura 13 Valores fenotípicos entre diferentes genótipos para os SNPs S2_4604235 e S6_772355, significativamente associados ao teor de proteínas nos grãos.



Para o SNP S2_4604235 (Figura 13A), indivíduos com genótipo com alelo referência (vermelho) mostraram uma concentração de proteínas superior à dos indivíduos heterozigotos (amarelo) e à dos indivíduos com alelo alternativo (verde). Para o SNP S6_772355 (Figura 13B) indivíduos com genótipo heterozigoto (amarelo) mostraram uma concentração de lipídios superior aos indivíduos homozigotos (verde e vermelho).

O SNP S2_4604235 está posicionado próximo (21210 pb *downstream*) ao gene Cc02_g05840, que codifica a serina protease, essencial para a liberação de mucilagem de cascas de sementes. Para isso desencadeia o acúmulo e/ou ativação de enzimas modificadoras da parede celular, necessárias para o afrouxamento da parede celular externa primária e para auxílio no inchaço da mucilagem (Rautengarten et al., 2008). Já o SNP S6_772355 está posicionado próximo (55654 pb *upstream*) ao gene Cc06_g00980, que codifica a proteína de sobreprodução de etileno 1, regulador essencial da via do etileno, além de estar envolvida na via da ubiquitinação, relacionada à modificação das proteínas (Christians et al., 2009).

Taninos

E para taninos foram encontrados 11, 8 e 15 SNPs significativamente associados em 2012, 2015 e média, respectivamente (Figura 14), com p -values entre $7.44 \cdot 10^{-5}$ a 0.005. Foram encontrados três SNPs comuns entre 2012 e a média (S2_54222076, S9_1240734 e S10_3633116) e quatro SNPs comuns entre 2015 e a média (S1_9508368, S4_25726475, S6_5679004 e S7_4132022) (Tabela 3). Para taninos não foi encontrada alguma causalidade entre os SNPs significativos, nos diferentes perfis de análises, com genes referentes ao metabolismo desse componente.

Figura 15 Manhattan e Q-Q plots para os teores de taninos em 2012, 2015 e média.

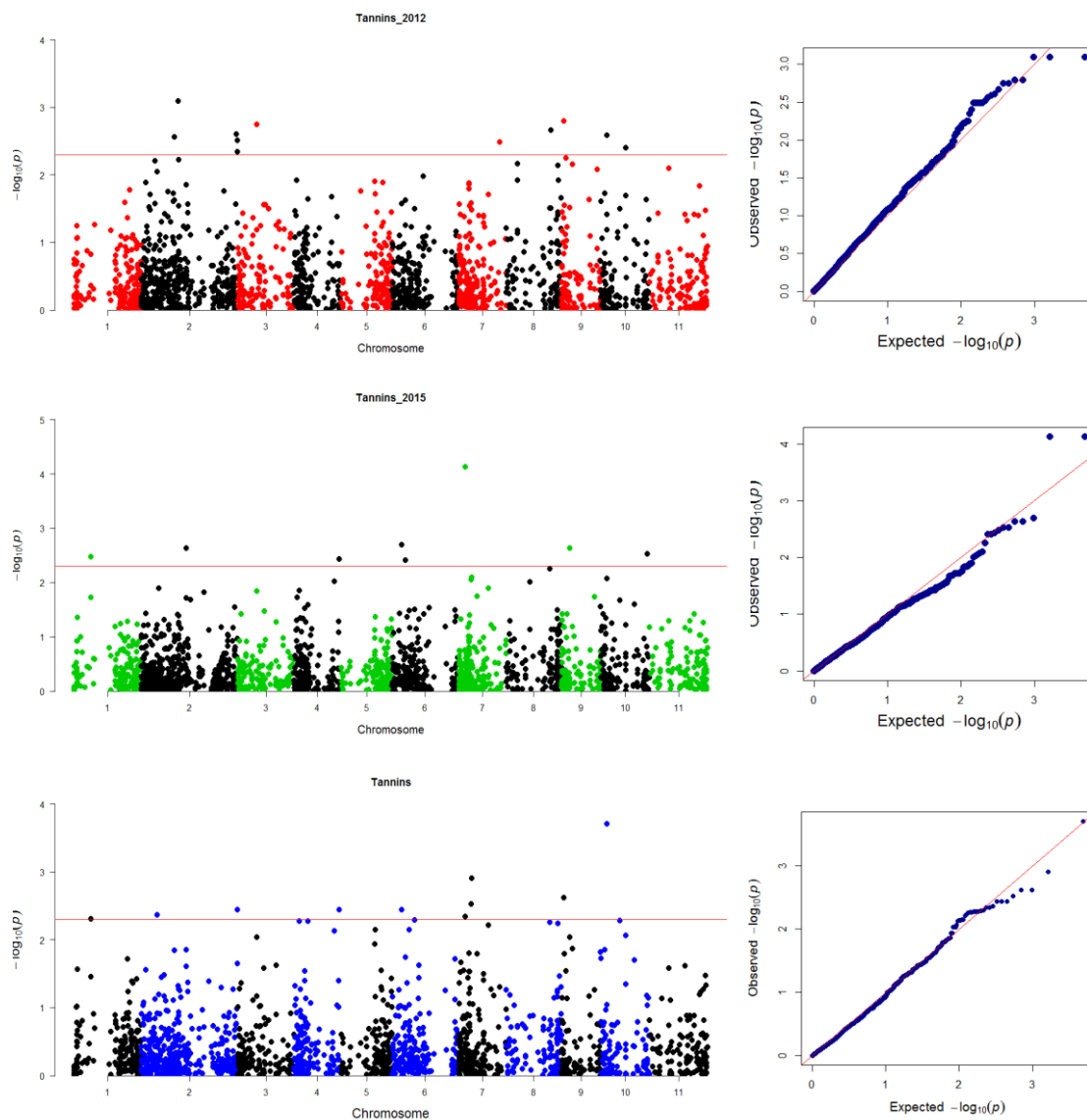


Tabela 3 SNPs associados com as concentrações de ACG, cafeína, AT/sacarose, lipídios, proteínas e taninos posicionados dentro (HIT) ou perto de genes candidatos relacionados com o metabolismo desses componentes.

Componente	Análise	SNP significativo	<i>p</i> -value	Genes associados	Distância
ACG	12/m	S2_4604235	0,001		
	12/m	S2_27727131	0,001		
	12/m	S2_43079004	0,001		
	12/m	S5_11612082	0,001		
	12/m	S7_18668686	1,20E-04		
	12/m	S8_29163163	3,80E-04	Cc08_g14110	HIT
	12/m	S9_2563555	0,002		
Cafeína	15/m	S2_25412500	0,005		
	15/m	S2_5043332	0,003	Cc02_g06280	4964796
	12/m	S4_1614149	0,003	Cc04_g02170	1693439
	15/m	S8_25559580	9,10E-04		
	12/15/m	S9_2563555	1,24E-05	Cc09_g03100 Cc09_g03020	89587 8804
	15/m	S2_18205445	0,005	Cc02_g20460	82862
	12/m	S3_12834205	0,003		
AT/Sacarose	12/15/m	S4_7902771	0,002		
	12/m	S6_10652258	0,003		
	12/15/m	S6_16281119	0,001		
	15/m	S7_7540264	0,003	Cc07_g10450	85270
	15/m	S8_2154851	1,36E-04		
	15/m	S11_3798869	0,005		
	15/m	S2_251546	3,11E-04	Cc02_g00250	28872
Lipídios	12/m	S6_7634389	0,003		
	12/m	S6_14809031	0,004		
	12/m	S8_27153678	0,002	Cc08_g11900	814
	12/m	S10_3632868	0,002		
	12/m	S11_6285049	0,001		
	12/15/m	S2_4604235	0,004	Cc02_g05840	21210
Proteínas	15/m	S6_772355	0,004	Cc06_g00980	55654
	15/m	S6_36332515	0,003		
	15/m	S1_9508368	0,005		
Taninos	12/m	S2_54222076	0,004		
	15/m	S4_25726475	0,004		
	15/m	S6_5679004	0,004		
	15/m	S7_4132022	0,005		
	15/m	S1_9508368	0,005		

12/m	S9_1240734	0,002
12/m	S10_3633116	1,96E-04

5.6 CONCLUSÃO

Nesse trabalho foram identificados 44 marcadores SNPs associados aos teores de ácidos clorogênicos, cafeína, açúcares totais, sacarose, lipídios, proteínas e taninos, considerando aqueles que foram comuns nas análises fenotípicas de 2012 e de 2015.

Além disso, o mapeamento desses SNPs do genoma de referência de *C. canephora* permitiu encontrar marcas dentro e próximas a genes que estariam associados às vias metabólicas de ácidos clorogênicos, cafeína, sacarose, lipídios e proteínas, responsáveis pela qualidade da bebida.

Esses resultados também demonstraram a grande diversidade entre os genótipos da coleção de *C. arabica* da Etiópia, podendo ser amplamente explorados para o melhoramento genético do cafeeiro.

5.7 REFERÊNCIAS

- Addinsoft. **XLStat: software for statistical analysis**. Versão 2008.4.02 (2008). Paris. 1 CD-ROM, 2010.
- Alexander, D.H.; Novembre, J. & Lange, K. **Fast model-based estimation of ancestry in unrelated individuals**. *Genome Research*, 19:1655-1664, 2009.
- Andrade, A.C., De A. Carneiro, F.; Da Silva Jr, O.B.; Marraccini, P.; Grattapaglia, D. **Towards GWAS and genomic prediction in coffee: development and validation of a 26K SNP chip for *Coffea canephora***. In: Proceedings Plant and Animal Genome XXV Conference. San Diego, 2017.
- 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.
- Bekele, Y.D. **Assessment of cup quality, morphological, biochemical and molecular diversity of *Coffea arabica* L. genotypes of Ethiopia** (Doctoral dissertation, University of the Free State), 2005.
- Bostick, M.; Lochhead, S.R.; Honda, A.; Palmer, S.; Callis, J. **Related to ubiquitin 1 and 2 are redundant and essential and regulate vegetative growth, auxin signaling, and ethylene production in *Arabidopsis***. *The Plant Cell*, 16(9), 2418-2432, 2004.
- Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. **TASSEL: Software for association mapping of complex traits in diverse samples**. *Bioinformatics* 23:2633-2635, 2007.
- Brasileiro, A.C.D.M. & Carneiro, V.D.C. **Manual de transformação genética de plantas**. Brasília, Embrapa-SPI/Embrapa-Cenargen, 1998.
- Campa C.; Doulebeau S.; Dussert S.; Hamon S.; Noirot M. **Qualitative relationship between caffeine and chlorogenic acid contents among wild *Coffea* species**. *Food chemistry* 93(1):135-139, 2005.
- Celik, I., Bodur, S., Frary, A.; Doganlar, S. **Genome-wide SNP discovery and genetic linkage map construction in sunflower (*Helianthus annuus* L.) using a genotyping by sequencing (GBS) approach**. *Molecular Breeding*, 36(9), 133, 2016.
- Christians, M.J.; Gingerich, D.J.; Hansen, M.; Binder, B.M.; Kieber, J.J.; Vierstra, R.D. **The BTB ubiquitin ligases ETO1, EOL1 and EOL2 act collectively to regulate ethylene biosynthesis in *Arabidopsis* by controlling type-2 ACC synthase levels**. *The Plant Journal*, 57(2), 332-345, 2009.
- CONAB – Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de café: Segundo levantamento – Safra 2016**. Brasília: CONAB, 2016.
- CONAB – Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira – Café**. V. 4 - safra 2017 - N. 1 - Primeiro levantamento, 2017.
- 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.

De Castro, R. D.; Marraccini, P. **Cytology, biochemistry and molecular changes during coffee fruit development.** Brazilian Journal of Plant Physiology, v.18, n.1, p.175-199, 2006.

Denoeud, F.; Carretero-Paulet, L.; Dereeper, A. et al. **The coffee genome provides insight into the convergent evolution of caffeine biosynthesis.** Science 345:1181-1184, 2014.

Dereeper, A.; Nicolas, S.; Lecunff, L.; Bacilieri, R.; Doligez, A.; Peros, J.P.; Ruiz, M.; This, P. **SNiPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects.** BMC Bioinformatics 5;12(1):134, 2011.

Dereeper, A.; Bocs, S.; Rouard, M. et al. **The coffee genome hub: a resource for coffee genomes.** Nucleic Acids Res 43: 1028-1035, 2015.

Doyle, J.J. & Doyle, J.L. **Isolation of plant DNA from fresh tissue.** Focus 12: 13-15, 1990.

Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S. Mitchell, S.E. **A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species.** PloS one, 6(5), e19379, 2011.

Esquivel, P.; Jiménez, V.M. **Functional properties of coffee and coffee by-products.** Food Research International 46(2) 488-495, 2012.

FAO – **Food and Agriculture Organization. Coffee mission to Ethiopia 1964–65.** FAO, Rome, 1968.

Farah, A.; Depaulis, T.; Trugo, L.C.; Martin, P.R. **Effect of roasting on the formation of chlorogenic acid lactones in coffee.** J Agric Food Chem 53:1505–1513, 2005.

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.

Fernandes, S.M.; Pinto, N.A.V.D.; Thé, P.M.P.; Pereira, R.G.F.A.; Carvalho, V.D. **Teores de polifenóis, ácido clorogênico, cafeína e proteína em café torrado.** Rev. Bras. de Agrociência, v.7 n 3, p.197-199, 2001.

Ferrão, M.F.; Furtado, J.C.; Neumann, L.G.; Konzen, P.H.A.; Morgano, M.A.; Bragagnolo, N.; Ferreira, M.M.C. **Técnica não destrutiva de análise de tanino em café empregando espectroscopia no infravermelho e algoritmo genético.** Tecnol. Santa Cruz do Sul, v.7, n.1, p. 9-26, 2003.

Ferreira, R. V. **Genotipagem de uma coleção de *Coffea arabica* utilizando marcadores microssatélites.** [Dissertação de Mestrado] Londrina: Pós-Graduação em Genética e Biologia Molecular – Universidade Estadual de Londrina, 2013.

Gazzarrini, S.; Lejay, L.; Gojon, A.; Ninnemann, O.; Frommer, W.B.; Von Wirén, N. **Three functional transporters for constitutive, diurnally regulated, and starvation-induced uptake of ammonium into *Arabidopsis* roots.** The Plant Cell, 11(5), 937-947, 1999.

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, 9(2), e90346, 2014.

- Henning, J.A.; Gent, D.H., Twomey, M.C., Townsend, M.S., Pitra, N.J.; Matthews, P.D. **Genotyping-by-sequencing of a bi-parental mapping population segregating for downy mildew resistance in hop (*Humulus lupulus* L.)**. *Euphytica*, 208(3), 545-559, 2016.
- Heyman, J.; Van den Daele, H.; De Wit, K.; Boudolf, V.; Berckmans, B.; Verkest, A.; De Veylder, L. ***Arabidopsis* ULTRAVIOLET-B-INSENSITIVE4 maintains cell division activity by temporal inhibition of the anaphase-promoting complex/cyclosome**. *The Plant Cell*, 23(12), 4394-4410, 2011.
- Jeszka-Skowron, M.; Sentkowska, A.; Pyrzyńska, K.; De Peña, M.P. **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.
- Jia Y, Sun X, Sun J, Pan Z, Wang X, et al. **Association Mapping for Epistasis and Environmental Interaction of Yield Traits in 323 Cotton Cultivars under 9 Different Environments**. *PLoS ONE* 9(5): e95882, 2014.
- Joët, T.; Laffargue A.; Descroix, F.; Doubeau, S.; Bertrand, B.; Kochko, A.; Dussert, S. **Influence of environmental factors, wet processing and their interactions on the biochemical composition of green coffee beans**. *Food Chem* 118:693–701, 2010.
- Kieber, J.J.; Rothenberg, M.; Roman, G.; Feldmann, K.A.; Ecker, J.R. **CTR1, a negative regulator of the ethylene response pathway in *Arabidopsis*, encodes a member of the raf family of protein kinases**. *Cell*, 72(3), 427-441, 1993.
- Ky, C.L.; Doubeau, B.; Guyot, B.; Akaffou, S.; Charrier, A.; Hamon, S.; Louarn, J.; Noiro, M. **Inheritance of coffee bean sucrose content in the interspecific cross *Coffea pseudozanguebariae* × *Coffea liberica* 'dewevrei'**. *Plant Breeding*, 119: 165-168, 2000.
- Lashermes, P.; Trouslot, P.; Anthony, F.; Combes, M.C.; Charrier, A. **Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica***. *Euphytica* 87:59-64, 1996.
- 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 Gen Genet* 261:259-266, 1999.
- Lee, K.; Ratnayake, C.; Huang, A.H. **Genetic dissection of the co-expression of genes encoding the two isoforms of oleosins in the oil bodies of maize kernel**. *The Plant Journal*, 7(4), 603-611, 1995.
- Li, F.; Fan, G.; Lu, C.; Xiao, G.; Zou, C.; Kohel, R. J.; Liang, X. **Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution**. *Nature biotechnology*, 33(5), 524-530, 2015.
- Logan-Young, C.J.; Yu, J.Z.; Verma, S.K.; Percy, R.G.; Pepper, A.E. **SNP discovery in complex allotetraploid genomes (*Gossypium* spp., Malvaceae) using genotyping by sequencing**. *Appl. Plant Sci* 3:1400077. doi:10.3732/apps.1400077, 2015.
- Lu, F.; Lipka, A.E.; Glaubitz, 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 protocol**. *PLoS Genet*, 9(1), e1003215, 2013.

Lu, W.; Tang, X.; Huo, Y.; Xu, R.; Qi, S.; Huang, J.; Wu, C.A. **Identification and characterization of fructose 1, 6-bisphosphate aldolase genes in *Arabidopsis* reveal a gene family with diverse responses to abiotic stresses.** *Gene*, 503(1), 65-74, 2012.

Martinez, H.E.P.; Clemente, J.M.; Lacerda, J.S.; Neves, Y.P.; Pedrosa, A.W. **Nutrição mineral do cafeeiro e qualidade da bebida.** *Rev. Ceres*, 61, 838-848, 2014.

Metzker, M. L. **Applications of next-generation sequencing technologies e the next generation.** [Review]. *Nature Reviews Genetics*, 11(1), 31-46, 2010.

Meyer, K.; Stecca, K.L.; Ewell-Hicks, K.; Allen, S.M.; Everard, J.D. **Oil and protein accumulation in developing seeds is influenced by the expression of a cytosolic pyrophosphatase in *Arabidopsis*.** *Plant physiology*, 159(3), 1221-1234, 2012.

Moncada, M.D.P.; Tovar, E.; Montoya, J.C.; González, A.; Spindel, J.; McCouch, S. **A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size.** *Tree Genetics & Genomes*, 12(1), 5, 2016.

Moon, B.C.; Choi, M.S.; Kang, Y.H.; Kim, M.C.; Cheong, M.S.; Park, C.Y.; Cho, M.J. ***Arabidopsis* ubiquitin-specific protease 6 (AtUBP6) interacts with calmodulin.** *FEBS letters*, 579(18), 3885-3890, 2005.

Pagiatto, N.F. **Análise de diterpenos e cafeína em uma coleção da Etiópia de *Coffea arabica*.** [Dissertação de Mestrado] Londrina: Pós-Graduação em Biotecnologia – Universidade Estadual de Londrina, 2013.

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

Pot, D.; Scholz, M.B.D.S.; Lannes, S.D.; Del Grossi, L.; Perreira, L.F.P.; Vieira, L.G.E.; Tumoru, S. **Phenotypic analysis of *Coffea arabica* accessions from Ethiopia: Contribution to the understanding of *Coffea arabica* diversity.** In: 22nd International Conference on Coffee Science – ASIC. Campinas, Brasil, 2008.

Prakash, N. S.; Combes, M. C.; Somanna, N.; Lashermes, P. **AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid.** *Euphytica*, v. 124, p. 265-271, 2002.

Raffaele, S.; Vaillieu, F.; Léger, A.; Joubès, J.; Miersch, O.; Huard, C.; Blee E.; Mongrand S.; Domergue F.; Roby, D. **A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in *Arabidopsis*.** *The Plant Cell*, 20(3), 752-767, 2008.

Rautengarten, C.; Usadel, B.; Neumetzler, L.; Hartmann, J.; Büssis, D.; Altmann, T. **A subtilisin-like serine protease essential for mucilage release from *Arabidopsis* seed coats.** *The Plant Journal*, 54(3), 466-480, 2008.

Ricketts, M.L.; Boekschoten, M.V.; Kreeft, A.J.; Hooiveld, G.J.E.J.; Moen, C.J.A.; Müller, M.; Frants, R.R.; Hasanmoentalib, S.; Post, S.M.; Princen, H.M.G.; Porter, J.G.; Katan, M.B.; Hofker, M.H.; Moore, D.D. **The cholesterol-raising factor from coffee beans, cafestol, as an agonist ligand for the farnesoid and pregnane x receptors.** *Molecular Endocrinology*, Bethesda, v. 21, n. 7, p. 1603-1616, 2007.

- 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 do lapar.** *Coffee Science* 6:245–255, 2011.
- Scholz, M.B.S.; Pagiatto, N.F.; Kitzberger, C.S.G.; Pereira, L.F.P.; Davrieux, F.; Charmetant, P.; Leroy, T. **Validation of nearinfrared spectroscopy for the quantification of cafestol and kahweol in green coffee.** *Food Res Int* 61:176–182, 2014a.
- 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, 2014b.
- 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 (2):429-438, 2016.
- Silva, A.C.P. **Informações Estatísticas do Café.** Cafés do Brasil. Ano 17, 2016.
- Silvestrini S.; Junqueira M.G.; Favarin A.C.; Guerreiro-Filho O.; Maluf M.P.; Silvarolla M.B.; Colombo C.A. **Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers.** *Genetic Resources Crop Evolution* 54, 6: 1367-1379, 2007.
- Simkin, A.J.; Qian, T.; Caillet, V.; Michoux, F.; Amor, M.B.; Lin, C.; Tanksley, S.; McCarthy, J. **Oleosin gene family of *Coffea canephora*: quantitative expression analysis of five oleosin genes in developing and germinating coffee grain.** *Journal of plant physiology*, 163(7), 691-708, 2006.
- 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.** *Theoretical and Applied Genetics*, 126(11), 2699-2716, 2013.
- Su, J.; Pang, C.; Wei, H.; Li, L.; Liang, B.; Wang, C.; Mao, G. **Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton.** *BMC genomics*, 17(1), 687, 2016.
- Tfouni S.A.V.; Carreiro L.B.; Teles C.R.A.; Furlani R.P.Z.; Cipolli K.M.; Camargo M.C.R. **Caffeine and chlorogenic acids intake from coffee brew: influence of roasting degree and brewing procedure.** *International Journal of Food Science and Technology* 49:747-752, 2014.
- Urgert, R.; Katan, M.B. **The cholesterol-raising factor from coffee beans.** *Annual Review of Nutrition*, Palo Alto, v. 17, p. 305-323, 1997.
- Velmurugan, J.; Mollison, E.; Barth, S.; Marshall, D.; Milne, L.; Creevey, C.J.; Lynch, B.; Meally, H.; McCabe, M.; Milbourne, D. **An ultra-high density genetic linkage map of perennial ryegrass (*Lolium perenne*) using genotyping by sequencing (GBS) based on a reference shotgun genome assembly.** *Annals of botany*, 118(1), 71-87, 2016.
- Vaast, P.; Bertrand, B.; Perriot, J.J.; Guyot, B.; Genard, 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* 86(2), 197-204, 2006.

Vidal, R.O.; Mondego, J.M.C.; Pot, D.; Ambrósio, A.B.; Andrade, A.C.; Pereira, L.F.P.; Colombo, C.A.; Vieira, L.G.E.; Carazzolle, M.F.; Pereira, G.A.G. **A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*.** *Plant Physiology*, 154(3), 1053-1066, 2010.

Wang, L.; Ware, D.; Lushbough, C.; Merchant, N.; Stein, L. **A genome-wide association study platform built on iPlant cyber-infrastructure.** *Concurrency and Computation: Practice and Experience*, 27(2), 420-432, 2015.

Weir, B.S. & Cockerham, C.C. **Estimating F-statistics for the analysis of population structure.** *Evolution*, 1358-1370, 1984.

Zhang, T.; Hu, Y.; Jiang, W.; Fang, L.; Guan, X.; Chen, J.; Hulse-Kemp, A. M. **Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement.** *Nature biotechnology*, 33(5), 531-537, 2015.

Zhou, H.; Alexander, D.H.; Lange, K. **A quasi-Newton method for accelerating the convergence of iterative optimization algorithms.** *Statistics and Computing*, 2009.

5.8 MATERIAL SUPLEMENTAR

Suplementar 1 – Lista dos genótipos de *C. arabica* enviados para a genotipagem por sequenciamento (GBS).

Genótipos	Nº na Coleção	Genótipos	Nº na Coleção	Genótipos	Nº na Coleção
1	BA10_057	54	E209_031	107	E456_062
2	E007_087	55	E213_211	108	E457_477
3	E012_136	56	E218_581	109	E458_097
4	E016_298	57	E220_127	110	E464_417
5	E017_419	58	E221_214	111	E466_125
6	E018_494	59	E233_015	112	E467_045
7	E021_011	60	E237_071	113	E478_408
8	E022_163	61	E238_022	114	E481_238
9	E025_308	62	E254_284	115	E486_189
10	E030_075	63	E261_052	116	E490_516
11	E037_676	64	E265_101	117	E494_173
12	E038_043	65	E267_090	118	E505_140
13	E039_434	66	E268_067	119	E511_157
14	E041_079	67	E270_044	120	E514_129
15	E044_122	68	E272_143	121	E516_069
16	E046_021	69	E279_618	122	E534_036
17	E047_267	70	E283_096	123	E546_118
18	E055_005	71	E287_029	124	E552_323
19	E057_496	72	E298_382	125	E565_010
20	E061_126	73	E301_111	126	E571_072
21	E068_014	74	E302_083	127	E621_139
22	E071_258	75	E308_049	128	M7846_67
23	E080_584	76	E315_081	129	SEL_106
24	E081_041	77	E320_145	130	Catuai_V26
25	E085_085	78	E324_093	131	M_Novo38
26	E087_194	79	E325_522	132	Typica
27	E089_391	80	E326_124	133	Bourbon
28	E114_447	81	E327_032	134	IAPAR59
29	E116_061	82	E331_280	135	L1C1
30	E118_213	83	E332_023	136	L3C3
31	E123a_231	84	E333_104	137	<i>C. canephora</i>
32	E123b_121	85	E335_219	138	<i>C. eugenioides</i>
33	E124_245	86	E338_218	139	Java1
34	E126_359	87	E340_179	140	Java6
35	E130_169	88	E344_008	141	Java15
36	E131_018	89	E351_248	142	Java26
37	E146_012	90	E363_735	143	Et34_1
38	E148_254	91	E364_059	144	Et34_2
39	E152_017	92	E368_600	145	Et34_5

40	E159_180	93	E370_196	146	Et34_6
41	E164_417	94	E383_142	147	Et34_8
42	E169_180	95	E386_131	148	Et34_12
43	E174_164	96	E389_133	149	Et34_14
44	E179_650	97	E401_643	150	Et34_17
45	E180_070	98	E404_135	151	Et19_14
46	E181_358	99	E408_001	152	IPR99
47	E183_138	100	E409_114	153	IPR100
48	E189_119	101	E419_098	154	IPR101
49	E190_013	102	E428_109	155	IPR102
50	E196_117	103	E439_094	156	IPR103
51	E199_011	104	E442_279	157	IPR104
52	E201_134	105	E450_235	158	IPR105
53	E208_193	106	E454_107	159	IPR107

Suplementar 2 – Amostragem de 66 genótipos da coleção de *C. arabica* do IAPAR utilizada nas análises de fenotipagem e nos estudos de associação (GWAS).

Genótipos	Nº na coleção	Origem	Genótipos	Nº na coleção	Origem	Genótipos	Nº na coleção	Origem
1	E007	L	23	E159	O	45	E332	O
2	E018	L	24	E174	O	46	E335	O
3	E021	L	25	E189	O	47	E364	O
4	E025	O	26	E199	O	48	E368	O
5	E039	O	27	E201	O	49	E370	O
6	E041	O	28	E208	O	50	E383	O
7	E044	O	29	E209	O	51	E386	O
8	E046	O	30	E218	O	52	E408	O
9	E047	O	31	E220	O	53	E450	O
10	E057	O	32	E221	O	54	E454	O
11	E068	O	33	E238	L	55	E456	O
12	E081	O	34	E267	O	56	E457	O
13	E087	O	35	E270	O	57	E458	O
14	E114	O	36	E272	O	58	E464	O
15	E116	O	37	E301	O	59	E467	O
16	E123a	O	38	E302	O	60	E486	O
17	E123b	O	39	E308	O	61	E511	O
18	E124	O	40	E315	O	62	E514	O
19	E126	O	41	E320	O	63	E516	O
20	E130	O	42	E324	O	64	E546	O
21	E148	O	43	E325	O	65	E621	O
22	E152	O	44	E331	O	66	M7846	O

6 CONCLUSÃO GERAL

Nesse trabalho foram realizadas a fenotipagem e a genotipagem de uma coleção de genótipos de *Coffea arabica* provenientes da Etiópia. A caracterização fenotípica de dois anos, para diferentes componentes bioquímicos indicou, nas duas análises, um grupo de genótipos que pode ser útil aos programas de melhoramento, com baixos níveis de ácidos clorogênicos e cafeína, além de um teor maior de açúcar. A importância de uma coleta mais homogênea de frutos maduros foi destacada, a fim de não haver diferenças entre os teores químicos avaliados no mesmo material e em diferentes anos.

A realização da genotipagem dessa coleção pela técnica de GBS permitiu a descoberta de novos SNPs para *C. arabica*, mesmo a partir do genoma de referência diploide de *Coffea canephora*. Com os resultados de fenotipagem associados aos resultados de genotipagem realizamos o GWAS, identificando SNPs ligados aos teores de ácidos clorogênicos, cafeína, açúcares totais, sacarose, lipídios, proteínas e taninos. O mapeamento desses SNPs no genoma de *C. canephora* permitiu identificar vários genes envolvidos nas vias metabólicas de ácidos clorogênicos, cafeína, sacarose, lipídios e proteínas, responsáveis pela qualidade da bebida.

A utilização dessa estratégia também desenvolverá resultados expressivos para avaliação da diversidade em genótipos nessa coleção. Além da caracterização de grupos de genótipos com características fenotípicas similares, as informações geradas aqui podem auxiliar na identificação precoce de materiais genéticos de interesse. A partir dessas informações, os SNPs encontrados juntamente com os dados de fenotipagem serão utilizados futuramente na construção de mapas genéticos de associação de *C. arabica*.

Deste modo, abre-se uma imensa oportunidade de se utilizar tais ferramentas para auxiliar os programas de melhoramento do cafeeiro, com redução de custos e tempo.