



UNIVERSIDADE
ESTADUAL DE LONDRINA

KARINA YANAGUI

**DIVERSIDADE NUCLEOTÍDICA DE GENES ENVOLVIDOS
NA BIOSÍNTESE DE AÇÚCARES E DITERPENOS DE
*Coffea spp.***

Londrina
2012

KARINA YANAGUI

**DIVERSIDADE NUCLEOTÍDICA DE GENES ENVOLVIDOS
NA BIOSÍNTESE DE AÇÚCARES E DITERPENOS DE
*Coffea spp.***

Dissertação 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 Mestre.

Orientador: Dr. Luiz Filipe Protasio Pereira.

Londrina
2012

**Catálogo elaborado pela Divisão de Processos Técnicos da Biblioteca Central da
Universidade Estadual de Londrina**

Dados Internacionais de Catalogação-na-Publicação (CIP)

Y21d Yanagui, Karina.

Diversidade nucleotídica de genes envolvidos na biossíntese de açúcares e diterpenos de *Coffea spp* / Karina Yanagui. – Londrina, 2012.
65 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, 2012.

Inclui bibliografia.

1. Café – Melhoramento genético – Teses. 2. Diterpenos – Teses. 3. Biossíntese – Teses. 4. Marcadores biológicos – Teses. 5. Café – Genética – Teses. 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.

CDU 631.52:633.73

KARINA YANAGUI

**DIVERSIDADE NUCLEOTÍDICA DE GENES ENVOLVIDOS NA
BIOSSÍNTESE DE AÇÚCARES E DITERPENOS DE *Coffea spp.***

Dissertação 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 Mestre.

BANCA EXAMINADORA

Prof. Dr. Luiz Filipe Protasio Pereira
Empresa Brasileira de Pesquisa Agropecuária –
EMBRAPA

Dr. Jorge Mauricio Costa Mondego
Instituto Agrônomo de Campinas – IAC

Prof. Dr. Laurival Antonio Vilas-Boas
Universidade Estadual de Londrina – UEL

Londrina, 24 de fevereiro de 2012.

À minha família e à Priscila Mito por acreditarem incondicionalmente em mim, pois a maior contribuição para realização dos sonhos é acreditar.

Amo vocês.

AGRADECIMENTOS

O Deus, Força Superior ou outras formas que se expresse.

Ao meu orientador, Dr. Luiz Filipe Protasio Pereira, pela longa jornada que trilhamos juntos nos últimos anos, pelos seus conselhos e orientações não apenas acadêmicos e pelo incentivo constante à busca pelo conhecimento.

Ao pesquisador Dr. David Pot pela paciência e atenção em ensinar mesmo a distância e pelo direcionamento e incentivo imprescindíveis na execução deste trabalho.

Ao pesquisador Dr. Luiz Gonzaga Esteves Vieira pela solicitude e sabedoria em discutir e orientar as atividades desenvolvidas.

Ao jovem pesquisador Dr. Douglas Silva Domingues pelas orientações fundamentais e decisivas na finalização do projeto e pelo exemplo de paixão e competência científica.

Ao Prof. Dr. Laurival Antonio Vilas-Boas pela amizade e pelos ensinamentos cultivados ao longo do meu desenvolvimento acadêmico e por mais uma vez contribuir participando da banca examinadora

Ao Dr. Jorge Mauricio Costa Mondego por aceitar prontamente o convite e participar da banca examinadora.

À minha família por fazer de mim o que sou hoje, pelo apoio e amor incondicional que sempre me acompanha e dá forças e sem os quais nada teria sentindo.

À Priscila Mito por incontáveis lutas e superações que passamos juntas, por todas as noites em claro e finais de semana confinados e por estar sempre ao meu lado.

Ao meu melhor amigo, e já irmão, William Xavier de Barros que mesmo a distância continua sendo uma fonte inesgotável de carinho e amizade, e aos outros amigos mais que especiais que nos lembram que o sucesso profissional só é completo quando pode ser dividido com as pessoas que amamos.

À Ruth da Silva e ao Nelson C. de Carvalho por me adotarem em minha segunda família.

Aos amigos do Laboratório de Biotecnologia Vegetal, que participaram de todo meu crescimento científico e do desenvolvimento deste trabalho e dos quais sentirei falta na nova jornada que se iniciará.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pela bolsa concedida, ao Instituto Agronômico do Paraná pela estrutura disponibilizada para realização do trabalho e a Universidade Estadual de Londrina pela formação acadêmica. Ao Consórcio Pesquisa Café e ao FINEP-Genocafé pelo financiamento do projeto.

YANAGUI, Karina, DOMINGUES, Douglas Silva, VIEIRA, Luiz Gonzaga Esteves; POT, David, PEREIRA, Luiz Filipe Protasio. **Diversidade nucleotídica de genes envolvidos na biossíntese de açúcares e diterpenos de *Coffea spp.*** p. 65. Dissertação de Mestrado (Pós Graduação em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Londrina, 2012.

RESUMO

A utilização de ferramentas biotecnológicas, como os marcadores moleculares, pode acelerar os programas de melhoramento através da seleção assistida por marcadores. Para o café, uma planta perene, essa característica é particularmente desejável devido ao tempo e recursos despendidos para o lançamento de uma nova cultivar. Contudo, para *Coffea arabica*, a principal espécie de importância econômica do gênero *Coffea*, o número de marcadores polimórficos disponíveis é muito baixo comparado a outras importantes culturas. Muitos estudos foram realizados para tentar acessar a variabilidade da espécie e a despeito de alguns polimorfismos serem encontrados, estes foram pouco eficientes para a discriminação genotípica e mapeamento genético. Neste contexto, o presente estudo buscou acessar a diversidade nucleotídica pela detecção de SNPs, em uma população do centro de origem de *C. arabica* onde se espera encontrar a maior diversidade da espécie. Foram selecionados oito genes envolvidos na biossíntese de açúcares e diterpenos bem como um gene cloroplastídico e um de álcool desidrogenase (ADH). Fragmentos dos respectivos genes de 12 genótipos de *C. arabica* (oito do centro de origem e quatro comerciais), oito de *Coffea canephora* e um de *Coffea eugenioides* foram amplificados e sequenciados. O alinhamento das sequências e a detecção manual dos polimorfismos foram realizados pelo programa *Codon Code Aligner*. Como resultado, nove kb foram sequenciados e analisados neste trabalho, possibilitando a detecção de 360 polimorfismos entre *C. arabica* e seus ancestrais, *C. canephora* e *C. eugenioides*. Destes, 142 eram polimorfismos intraespecíficos de *C. arabica*: a maioria (82%), como descrito em estudos anteriores, era relacionada a diferenças entre os subgenomas ancestrais. Entretanto, 25 SNPs (18% do total) corresponderam a diferenças intraespecíficas, não observadas em *C. canephora* e *C. eugenioides*: 13 destes eram fixados entre os genótipos e 12 mostraram variabilidade entre os genótipos. Esses 12 SNPs evidenciam a existência de variabilidade intraespecífica, ainda que em frequência reduzida, que pode ser aplicada para genotipagem e estudos de mapeamento. Os resultados demonstram a importância de utilizar genótipos do centro de origem de *C. arabica* para detecção de SNPs e são relevantes para nortear futuros trabalhos de caracterização da variabilidade genética da espécie.

Palavras-chave: Diversidade nucleotídica. *Coffea arabica*. SNPs. *Coffea*.

YANAGUI, Karina, DOMINGUES, Douglas Silva, VIEIRA, Luiz Gonzaga Esteves; POT, David, PEREIRA, Luiz Filipe Protasio. **Nucleotide diversity of genes involved in sugar and diterpens biosynthesis of *Coffea spp.*** p. 65. Dissertação de Mestrado (Pós Graduação em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Londrina, 2012.

ABSTRACT

The utilization of biotechnological tools, such as molecular markers, can speed up crop breeding programs through marker assisted selection. For coffee, a perennial crop, this feature is particularly desirable because of the time and cost to launch a new cultivar. However, for *Coffea arabica*, the main economic importance coffee specie, the number of polymorphic markers currently available is very low compared to other important crops. Many studies were performed in order to access *C. arabica* variability, but the polymorphisms found were barely sufficient for genotypic discrimination and mapping studies. In this context, the present study searched the nucleotide diversity by SNPs detection within a population from *C. arabica* center of origin, where is the highest especie diversity. We selected eight genes involved in sugars and diterpenes biosynthesis, as well as one chloroplastic and one alcohol dehydrogenase (ADH) genes. Fragments of those genes from 12 genotypes of *C. arabica* (eight wild type genotypes and four commercial cultivars), eight of *C. canephora* and one of *C. eugenioides* were amplified and sequenced. The sequences alignment and polimorfisms manual detection were performed by *Codon Code Aligner* program. As a result, nine Kb were sequenced and analyzed in this work, and 360 polymorphisms were found among *C. arabica* and its ancestors, *C. canephora* and *C. eugenioides*. Of these, 142 were intraspecific polymorphisms of *C. arabica*: the majority (82%), as described in previous studies, were related to differences between the ancestral subgenomes. However, 25 SNPs (18% of the total) corresponded to intraspecific differences, not observed in *C. canephora* and *C. eugenioides*: 13 of them were fixed between the genotypes and 12 showed variability between the genotypes. The latest polimorfisms demonstrate the intraespecific variability in *C. arabica* genotypes, and can be applied to genotyping and mapping studies. The data also show the importance of using genotypes from *C. arabica* center of origin for SNPs detection and are important to guide futures studies in characterization genetic variability of *C. arabica*.

Keywords: Nucleotide diversity. *Coffea Arabica*. SNPs. *Coffea*.

LISTA DE ILUSTRAÇÕES

Figura 1 - Continente Africano indicando os centros de origem e diversidade de diversas espécies do gênero <i>Coffea</i>	17
Figura 2 - Tipos de polimorfismos de um único nucleotídeo em um gene	20
Figura 3 - Origem de <i>C. arabica</i>	22
Figura 4 - A, Origem das cultivares modernas de <i>C. arabica</i> . B, Representação da diminuição da diversidade da espécie com a dispersão e desenvolvimento de cultivares	24
Figura 5 - Identificação de SNPs em homeólogos de <i>C. arabica</i> e <i>C. canephora</i>	26
Figura 6 - Metabolismo de sacarose em sementes de café	28
Figura 7 - Fórmula estrutural dos diterpenos Cafestol e Caveol	29
Figura 8 - Início da via de biossíntese de diterpenos	30
Figura 9 - Estrutura das regiões dos 8 genes relacionados as vias biossintéticas de açúcares e diterpenos, da álcool desidrogenase (ADH) e de DNA cloroplastídico de <i>C. arabica</i>	36
Figura 10 - Análise global da diversidade nucleotídica	42
Figura 11 - Polimorfismos detectados em <i>C. arabica</i>	44
Figura 12 - Distribuição dos polimorfismos encontrados em <i>C. arabica</i>	45

LISTA DE TABELAS

Tabela 1 - Lista dos genótipos analisados com descrição do grupo genético que pertecem e da representatividade amostrada	33
Tabela 2 - Sumário dos genes estudados e lista dos oligonucleotídeos iniciadores para cada um deles	37
Tabela 3 - Número de genótipos com sequências amplificadas para cada região	38
Tabela 4 - Análise dos fragmentos contendo: total de pb para cada um deles; total avaliado por região (codante, intron e 3'UTR); número de polimorfismos bem como o tipo (SNP, INDEL e SSR); a região onde foram identificados (codante, intron e 3'UTR) e a classificação dos polimorfismos silenciosos, sinônimos e não sinônimos	40
Tabela 5 - Frequência de polimorfismos observada nas diferentes regiões avaliadas e entre polimorfismos sinônimos e não sinônimos	41
Tabela 6 - Análise de polimorfismos entre os genótipos de <i>C. canephora</i> e <i>C. arabica</i>	43
Tabela 7 - Comparação entre a frequência de SNPs e o número de genótipos de estudos de SNPs em <i>Coffea</i> spp	45
Tabela 8 - Relação dos 25 SNPs encontrados em <i>C. arabica</i>	47
Tabela 9 - Caracterização dos polimorfismos de <i>C. arabica</i>	47
Tabela 10 - Comparação entre os SNPs detectados em fragmentos sequenciados diretamente do produto de PCR e sequências obtidas de produto de clonagem de <i>C. arabica</i>	49

SUMÁRIO

1	INTRODUÇÃO	12
2	OBJETIVO GERAL	14
2.1	OBJETIVOS ESPECÍFICOS.....	14
3	FUNDAMENTAÇÃO TEÓRICA	15
3.1	MARCADORES MOLECULARES E DIVERSIDADE DO CAFEIEIRO	15
3.1.1	Importância Econômica	15
3.1.2	Aspectos Taxonômicos.....	16
3.1.3	Marcadores Moleculares e Estudos Genéticos do Cafeeiro	18
3.1.4	<i>Single Nucleotide Polimorfisms</i> (SNPs).....	20
3.1.5	Diversidade Genética de <i>Coffea arabica</i>	22
3.1.6	Vias de Biossíntese de Açúcares e Diterpenos em Cafeeiros.....	26
4	ARTIGO: ANÁLISE DA DIVERSIDADE NUCLEOTÍDICA DE <i>Coffea arabica</i>	31
4.1	INTRODUÇÃO	31
4.2	MATERIAL E MÉTODO	33
4.2.1	Material Vegetal e Extração de DNA Genômico	33
4.2.2	Desenho de Oligonucleotídeos Iniciadores e Amplificação	34
4.2.3	Sequenciamento.....	34
4.2.4	Análises das Sequências e Detecção dos Polimorfismos	35
4.2.5	Clonagem e Extração Plasmidial	35
4.3	RESULTADOS E DISCUSSÃO	36
4.3.1	Validação de SNPs pela Clonagem e Sequenciamento de Fragmentos	36
4.4	CONCLUSÃO	48
	REFERENCIAS	50
	ANEXOS	57
	ANEXO A - Sequências consenso das regiões dos genes de interesse que foram analisadas.	58

- ANEXO B - Exemplo de sequência interrompida de *C. arabica* devido à presença de INDELS heterozigotos que impossibilitaram a leitura do sequenciamento. As quatro primeiras sequências são de *C.arabica*, seguidas de uma sequência de *C. canephora* e por último *C. eugenioides*. Na posição assinalada (344 do *Contig*) existe um INDEL de um nucleotídeo entre *C. canephora* e *C. eugenioides*, e *C. arabica* possui os dois alelos causando uma sobreposição de fluorescência emitida, que impossibilita a leitura do sequenciamento a partir deste INDEL.62
- ANEXO C - Exemplo de alinhamento no programa *Codon Code Aligner* evidenciando os alelos heterozigotos encontrados em *C. arabica*. Cada linha corresponde a sequência de um genótipo. O SNP assinalado apresenta dois alelos (G e T); observa-se que os genótipos de *C. canephora* e *C. eugenioides* apresentam apenas um dos alelos e os genótipos de *C. arabica* apresentam ambos os alelos.63
- ANEXO D - Tabela de caracterização dos polimorfismos presentes no gene da Sacarose fosfato sintase (SPS). Cada linha corresponde a um SNP. As colunas contém respectivamente: nome do primer, número do SNP no *Contig*, nome do SNP, posição do SNP no *Contig*, sequência anterior e posterior ao polimorfismo, alelos presentes e tipo do polimorfismo (SNP, INDEL ou SSR).64
- ANEXO E - Tabela de caracterização dos polimorfismos presentes em um fragmento do gene da sacarose sintase 2 (P17/Susy2_c4_6_a). Cada linha corresponde a um genótipo e cada coluna numerada corresponde a um polimorfismo. Nas colunas 5 e 12, *C. canephora* apresenta um alelo (amarelo) e *C. eugenioides* outro (laranja), enquanto *C. arabica* possui ambos (vermelho), estes são exemplos de polimorfismos nos subgenomas de *C. arabica* que refletem diferenças entre os ancestrais (Cc e Ce - 40% dos SNPs de *C.arabica*). Na coluna 6 observa-se que os SNPs de *C. arabica* correspondem a diferenças entre os genótipos de *C. canephora* (Cc 22% dos SNPs de *C.arabica*) e nas colunas 3 e 7 os polimorfismos de *C. arabica* podem corresponder as diferenças entre os subgenomas ancestrais e/ou aos polimorfismos dentro de *C. canephora* (Ce e Cc/Cc – 20% dos SNPs de *C.arabica*).....65

1 INTRODUÇÃO

O café é a segunda maior *commodity* mundial tendo seu valor de mercado superado apenas pelo petróleo. O faturamento anual desta cultura no mundo é de aproximadamente US\$ 90 bilhões e sua cadeia produtiva envolve cerca de 500 milhões de pessoas desde o plantio até a comercialização do produto final. (DAMATTA; RAMALHO, 2006). O Brasil responde em média por 30% das exportações no mercado mundial e, em 2010, essa participação chegou a 34,6 %. Na economia nacional esse montante representou 7,5 % das exportações do agronegócio em 2010 e 10,1 % em 2011 (ABIC, 2011a).

Para assegurar a produtividade e potencializar o poder competitivo do país na comercialização do café, os programas de melhoramento precisam constantemente buscar soluções para problemas que podem prejudicar a cafeicultura, como por exemplo, os estresses bióticos e abióticos, além de investir na produção de cultivares com qualidade diferencial da bebida, a fim de conquistar mercados mais exigentes. Dois fatores limitantes ao lançamento de cultivares são o tempo despendido para produzi-las, visto que o melhoramento pode levar dezenas de anos, e a baixa diversidade genética encontrada na principal espécie de interesse econômico, *Coffea arabica*. Diversas ferramentas biotecnológicas têm sido empregadas com o objetivo de diminuir essas limitações, como a cultura de tecidos, a transformação genética e o uso de marcadores moleculares na seleção assistida (CAIXETA et al., 2008).

Os marcadores moleculares são uma opção promissora, pois podem possibilitar o acesso à variabilidade genética dos cafeeiros e a escolha precoce e precisa de genótipos de interesse, permitindo melhor gerenciamento de tempo e recursos para o melhoramento vegetal (FERREIRA; GRATTAPAGLIA, 1998). Dentre os diversos tipos de marcadores moleculares, os SNPs (*Single Nucleotide Polimorfism*) constituem as formas mais abundantes de variações nos genomas dos organismos (RAFALSKI, 2002), apresentando alto potencial informativo por promover uma maior saturação do genoma, característica importante especialmente para *C. arabica*, espécie alotetraplóide com baixa variabilidade.

Estudos de diversidade em *C. arabica* resultaram na detecção de marcas polimórficas que permitem apenas a identificação dos grupos de origem da espécie, não sendo informativos para diferenciação de cultivares ou de genótipos do

centro de origem (LASHERMES et al., 1999, ANTHONY et al., 2001, STEIGER et al., 2002, MONCADA; MCCOUCH, 2004, SILVESTRINI et al., 2005, MALUF et al., 2005). Mesmo a procura por SNPs no transcriptoma da espécie não permitiu a identificação de marcadores moleculares com potencial para discriminação dos genótipos (VIDAL et al., 2010).

Neste contexto, o presente estudo buscou acessar a variabilidade nucleotídica de *C. arabica* em um painel representativo de genótipos comerciais e do centro de origem da espécie, não avaliados em trabalhos anteriores. A utilização dessa população com maior variabilidade genética aumenta a probabilidade de identificação de marcadores polimórficos com potencial para genotipagem, necessários para construção de um mapa físico de *C. arabica* e para nortear os estudos genéticos em cafeeiros.

2 OBJETIVO GERAL

Analisar a variabilidade nucleotídica de genes relacionados às vias de biossíntese de açúcares e diterpenos de *Coffea arabica*.

2.1 OBJETIVOS ESPECÍFICOS

- Avaliar a diversidade nucleotídica de *C. arabica* pela detecção de SNPs em genes de interesse econômico utilizando um grupo representativo dos genótipos comerciais e do centro de origem da espécie.
- Verificar a existência de SNPs intraespecíficos em *C. arabica* não correspondentes a polimorfismos dos genomas ancestrais.
- Identificar os haplótipos pela comparação dos polimorfismos de *C. arabica* e dos ancestrais (*Coffea canephora* e *Coffea eugenioides*).

3 FUNDAMENTAÇÃO TEÓRICA

3.1 MARCADORES MOLECULARES E DIVERSIDADE DO CAFEIRO

3.1.1 Importância Econômica

O café é uma das bebidas mais populares do mundo, entretanto até o século XVII seu consumo era pouco habitual ou desconhecido. A despeito de ser consumido em alguns lugares, como Arábia, desde o século XIV, somente a partir de 1615 iniciou-se a propagação e popularização da bebida na Europa. Os árabes monopolizaram o fornecimento por quase um século, então os europeus realizaram expedições à África para obter sementes ou plantas de *C. arabica* que levou a disseminação mundial das populações de Typica e Bourbon no século XVII (ANTHONY; DUSSERT; DULLOO, 2007).

A produção mundial de café é baseada em duas espécies, *C. arabica* e *C. canephora*, e a primeira é a mais utilizada e importante economicamente (ABIC, 2011b). *C. arabica* é reconhecido por apresentar qualidade e sabor superior da bebida, entretanto é mais suscetível ao ataque de diferentes pragas (fungos, nematódeos e insetos). A dispersão atual da espécie é ampla e seu cultivo é realizado nos continentes americano e asiático, bem como em algumas regiões da África. O desenvolvimento do cafeeiro ocorre em altitudes mais elevadas, entre 1.000 e 2.000 metros, e temperaturas mais amenas, entre 18°C e 21°C (MEDINA FILHO; BORDIGNON; CARVALHO et al., 2008). *C. canephora* por sua vez, proporciona uma bebida de sabor neutro e de qualidade inferior a *C. arabica*, assim é usualmente utilizado em *blends* com cafés de sabor mais ativo ou como café solúvel (MEDINA FILHO; BORDIGNON; CARVALHO, 2008). Todavia a espécie apresenta grande variabilidade, extensa distribuição geográfica, maior resistência a pestes e notável capacidade adaptativa a distintas condições ambientais, tornando-a interessante como fonte de genes para introgressão no melhoramento de *C. arabica* (FERRAO et al., 2007).

O número de apreciadores do café tem aumentado a cada ano, consolidando a popularidade da bebida tanto em âmbito global como nacional. O consumo mundial apresentou um incremento de 2,4 % no ano de 2010, enquanto o

aumento nacional chegou a 3,5 %. O consumo *per capita* de 4,81 quilos por habitante ao ano garantiu ao Brasil o segundo maior mercado consumidor do mundo, atrás apenas dos Estados Unidos (KIST et al., 2011). Além da liderança no mercado consumidor, o país configura-se como maior produtor e exportador mundial de café, seguido dos países Vietnã, Colômbia, Indonésia e Índia na produção e Vietnã, Indonésia, Colômbia e Alemanha na exportação de grãos (FAO, 2009a, FAO, 2009b).

A maior área de produção no Brasil está concentrada na região centro-sul, com Minas Gerais ocupando o posto de maior produtor, seguido pelos estados do Espírito Santo, São Paulo e Paraná. O parque cafeeiro do Paraná possui uma área de 89.480 hectares compostos exclusivamente por *C. arabica*. A produção do Estado, em 2011, foi estimada em 1,705 milhões de sacas, com rendimento aproximado de 22,58 sacas/ha (ABIC, 2011b).

Em termos monetários, as exportações do produto renderam U\$ 2,6 bilhões ao país em 2011; esse montante representa 30,3% do mercado mundial e 10,1% das exportações nacionais no agronegócio (ABIC, 2011a), evidenciando a importância da cultura para a economia brasileira tanto pelo aumento de capital no setor agrícola, quanto pela geração e transferência de renda a outros setores econômicos e expressiva absorção de mão-de-obra (ROMANIELO et al., 2001).

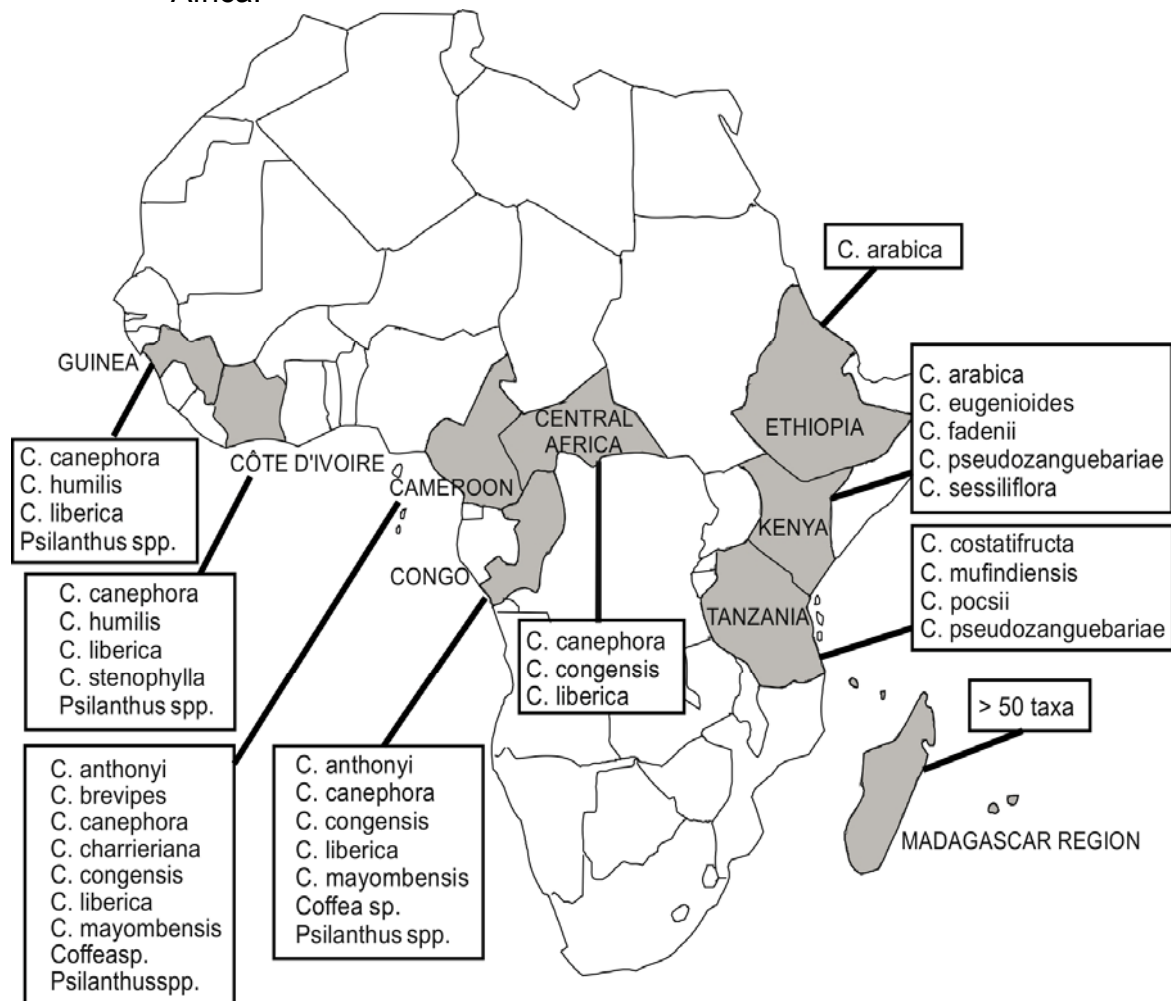
3.1.2 Aspectos Taxonômicos

O cafeeiro (*Coffea* spp) é uma planta dicotiledônea pertencente à família Rubiaceae, tribo Coffeaeae, subtribo Coffeinae. Segundo a revisão taxonômica feita por Davis et al. (2007), essa família é composta por 330 espécies divididas em 11 gêneros, destes, *Psilanthus* e *Coffea* são os mais conhecidos.

O gênero *Coffea* possui 103 espécies divididas em dois subgêneros: *Baracoffea* (oito espécies) e *Coffea* (95 espécies) (MAURIN, et al., 2007, DAVIS et al., 2007), o último compreendendo espécies como *C. arabica*, *C. eugenioides* e *C. canephora* (BERTHOU; MATHIEU; VEDEL, 1983). Todas as espécies de *Coffea* têm como centro de origem as florestas intertropicais da África e das ilhas Madagascar e Mascarenhas e podem ser divididas biogeograficamente em três grupos: região de Madagascar, região leste da África e região central e oeste da África (Figura 1) (LASHERMES et al., 1997). É interessante ressaltar que o centro de origem de uma

espécie é considerado a maior reserva natural de diversidade da mesma, daí a sua importância.

Figura 1 – Continente Africano indicando os centros de origem e diversidade de diversas espécies do gênero *Coffea*. Em destaque as regiões biogeográficas em que podem ser divisados os agrupamentos formados pelos centros de origem: região leste, central e oeste (Madagascar) da África.



Fonte: ANTHONY; DUSSERT; DULLOO (2007, p.15).

As espécies de *Coffea* são diploides ($2n = 2x = 22$) e geralmente autoincompatíveis, à exceção de *C. arabica*, que é tetraploide ($2n = 4x = 44$) e autocompatível (CARVALHO et al., 1991, BERTHOU; MATHIEU; VEDEL, 1983). Estudos indicam que *C. arabica*, é um anfidiplóide originado pela hibridização entre *C. canephora* e *C. eugenioides* (LASHERMES et al. 1999, DAVIS et al., 2007, MAURIN, et al. 2007).

As espécies agronômicas com relações filogenéticas mais próximas a Rubiaceae pertencem à família *Solanaceae*, estando ambas inseridas no clado Asteridae das plantas dicotiledôneas (WIKSTRÖM et al., 2001). A família *Solanaceae* compreende espécies como tomate, batata, pimenta e petúnia, que compartilham grande similaridade em tamanho do genoma e cariótipo cromossômico com o cafeeiro (tomate e maioria das espécies $n=12$, café $n=11$) (LIN et al., 2005). Além das solanáceas, outra espécie utilizada para estudos comparativos com *Coffea spp.* é *Vitis vinifera*, um membro da família *Vitaceae* no clado das Rosídeas. Asterídeas e rosídeas divergiram de um ancestral comum há 114-125 milhões de anos (WIKSTRÖM et al., 2001, YU et al., 2011, MONDEGO et al., 2011).

3.1.3 Marcadores Moleculares e Estudos Genéticos do Cafeeiro

Desde o início das atividades agrícolas a milhares de anos atrás, os primeiros agricultores selecionavam entre os grãos selvagens, os tipos mais desejáveis baseados no tamanho dos grãos, produtividade, sabor, etc. Essa seleção subjetiva levou as primeiras mudanças alélicas dirigidas que influenciaram o processo de domesticação e evolução de diversas culturas modernas (BORÉM; MIRANDA, 2009). Até a descoberta das leis mendelianas o melhoramento era baseado apenas na seleção e recombinação de fenótipos superiores, depois o entendimento das mesmas e a manipulação da hereditariedade passaram a formar as bases dos programas de melhoramento (FERREIRA; GRATTAPAGLIA, 1998, BORÉM; MIRANDA, 2009). Atualmente o melhoramento abrange uma gama de áreas do conhecimento e pode ser sintetizado em três etapas básicas: desenvolvimento de progênies, avaliação de progênies e recombinação dos melhores genótipos, recomeçando o ciclo (BORÉM; MIRANDA, 2009).

Para o cafeeiro, uma planta perene, este processo é particularmente extenso. Um ciclo de melhoramento leva quatro anos da germinação da semente ao florescimento da planta e finalmente a semente de novo. Muitos ciclos e seleções são necessários até a produção de genótipos superiores. As seleções no campo geralmente são realizadas com plantas adultas de seis a oito anos. Além disso, para assegurar a fidelidade da propagação das sementes são requeridos ao menos seis ciclos autofecundação. Considerando um ciclo para hibridização e seleção, e seis ciclos para homozigose da semente, um programa normal de melhoramento do

cafeeiro leva em média 20 a 28 anos, despendendo assim grande soma de tempo e recursos para o desenvolvimento de uma cultivar (SANTOS-BRIONES; HERNÁNDEZ-SOTOMAYOR, 2006).

Para aperfeiçoar e potencializar a eficiência desse processo, diversas ferramentas biotecnológicas vem sendo desenvolvidas e aplicadas no auxílio ao melhoramento tradicional. Dentre ferramentas biotecnológicas utilizadas podemos destacar as técnicas de cultura de tecidos vegetais, os marcadores moleculares e a transformação genética (SANTOS-BRIONES; HERNÁNDEZ-SOTOMAYOR, 2006, MORAIS; MELO, 2011). Os marcadores podem ser aplicados em diversas etapas desde o pré ao pós-melhoramento. No pré-melhoramento eles podem auxiliar na avaliação da diversidade genética da espécie, na caracterização dos bancos de germoplasma, na determinação de genealogia ou paternidade, na construção de mapas genéticos de ligação e na elucidação das relações genéticas com outras espécies do gênero ou próximas filogeneticamente. Na etapa de melhoramento, eles são úteis no mapeamento gênico, na certificação de cruzamentos, na seleção de genitores, na predição de híbridos simples e principalmente na seleção assistida por marcadores. No pós-melhoramento os marcadores podem ser empregados para proteção da propriedade intelectual pela caracterização de cultivares, para certificação da pureza das sementes e para identificação genealogias dos cultivares (BORÉM, 2009, BORÉM; MIRANDA, 2009).

Diferentes tipos de marcadores moleculares estão disponíveis (como *Restriction Fragment Length Polymorphism* - RFLP, *Random Amplification of Polymorphic DNA* - RAPD, *Simple Sequence Repeat* – SSR e *Single Nucleotide Polymorphism* - SNP) e podem variar pela metodologia utilizada para sua identificação, pela frequência com que ocorrem no genoma e pela relação de dominância apresentada. A seleção do marcador mais adequado à espécie deve basear-se no objetivo do estudo a ser realizado, na biologia e na estrutura genética da mesma, visando sempre uma estratégia rápida e precisa de genotipagem (GOMEZ et al., 2009).

Para *C. canephora* foi possível acessar a diversidade da espécie pelo emprego de diversos marcadores como RFLP, SSR e SNPs, e estes se mostraram eficientes em caracterizar não apenas os grupos ou populações (MUSOLI, et al., 2009, GOMEZ et al., 2009) existentes como também discriminar os diferentes genótipos (GOMEZ et al., 2009, VIDAL et al., 2010, MISSIO et al., 2010).

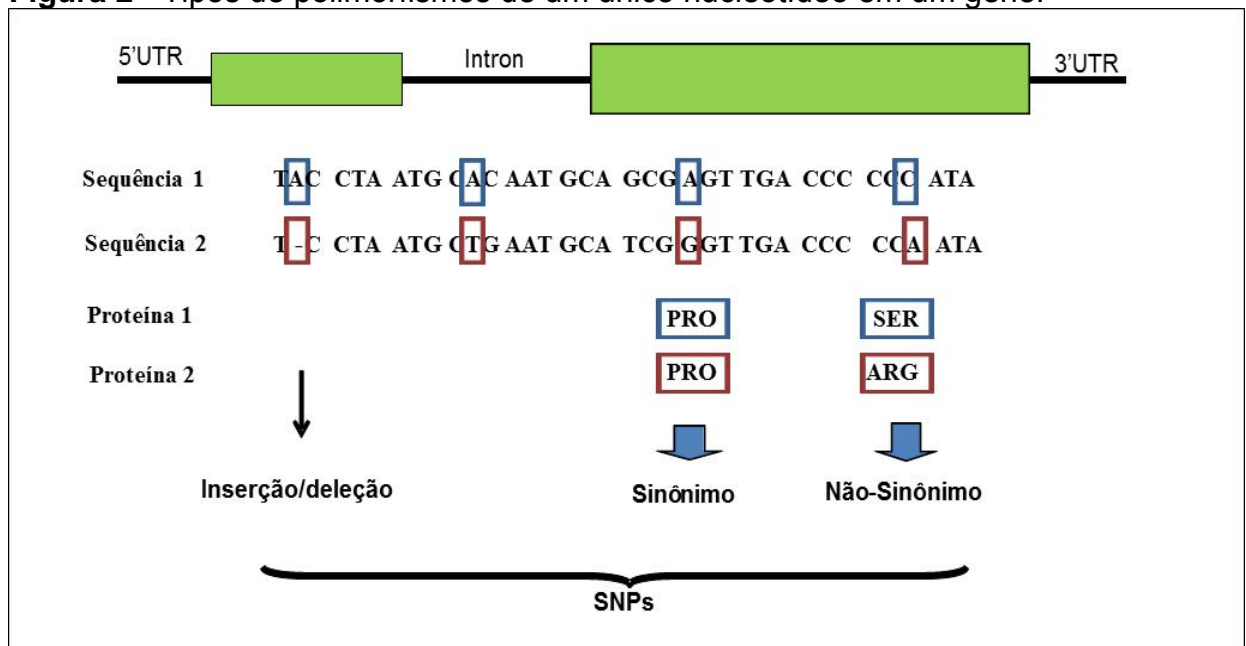
Além disso, existem mapas de ligação (utilizando marcadores RAPD, RFLP e SSR) (PAILLARD; LASHERMES; PÉTIARD, 1996, LASHERMES et al., 2001, COULIBALY et al., 2003), mapas de sintonia (LEFEBVRE-PAUTIGNY et al., 2010) e até mesmo QTLs identificados (*Quantitative trait locus*) (LASHERMES et al., 1996, LEROY et al., 2011) para a espécie.

No caso de *C. arabica* cuja variabilidade é limitada, o emprego de marcadores co-dominantes e multialélicos, com alta incidência em todo o genoma do cafeeiro, como os SNPs, poderá permitir a análise estrutural da população e o desenvolvimento de mapas genéticos, dois pré-requisitos necessários para a identificação de genes responsáveis pela variação natural das características de interesse (LEROY et al. 2006).

3.1.4 Single Nucleotide Polimorfisms (SNPs)

Os SNPs são modificações de um único nucleotídeo e os INDELS são inserções/deleções de poucos pares de base, que caracterizam o conjunto de alelos de um *locus* específico (ao invés deste apresentar um único alelo) (Figura 2) (BROOKES, 1999).

Figura 2 - Tipos de polimorfismos de um único nucleotídeo em um gene.



Os SNPs e INDELS podem ser encontrados em regiões codantes e não codantes dos genes (RAFALSKI, 2002). Quando localizados em regiões não codantes são denominados anônimos, se localizados em regiões codantes podem ser sinônimos (não alteram o aminoácido de tradução) e não-sinônimos (alteram o aminoácido e conseqüentemente a proteína) (Figura 2) (BROOKES, 1999).

As substituições nos polimorfismos não sinônimos podem ser conservativas ou não, dependendo da característica dos aminoácidos trocados, podendo gerar alterações estruturais e funcionais na proteína. Ainda que SNPs sinônimos não modifiquem a sequência proteica, eles podem gerar ou suprimir códons de terminação ou poliadenilação na molécula de mRNA (alterando sua estabilidade), promover o *splicing* alternativo e modificar códons de iniciação de tradução alterando o padrão de expressão gênica (KWOK; GU, 1999).

Os SNPs são causados por modificações de ponto, e podem ser classificados como marcadores de transversão, quando a substituição é de uma purina por uma pirimidina e vice-versa, ou de transição, quando ocorrem trocas entre duas purinas ou duas pirimidinas (BROOKES, 1999). Esses constituem as formas mais abundantes de variações do genoma, fornecendo uma fonte para análise de diversidade e saturação de mapas genéticos, aumentando a precisão dos mesmos e auxiliando em estudos de associação, facilitando a construção de mapas de ligação, seleção de BACs e construção de mapas físicos, bem como na seleção assistida por marcadores (SAM), onde permitem a seleção de populações segregantes (RAFALSKI, 2002, BROOKES, 1999, KWOK; GU, 1999, GUPTA; RUSTGI, 2004). Estes marcadores também são ferramentas eficazes para a determinação de relações filogenéticas, especiação e história demográfica de uma população (BRUMFIELD et al., 2003).

Algumas espécies, como o milho, apresentam alta diversidade nucleotídica; outras como *Arabidopsis thaliana*, sorgo e tomate apresentam diversidade nucleotídicas menores. Dependendo da espécie analisada (autógamas, alógamas) diferentes estratégias para sequenciamento podem ser aplicadas (RAFALSKI, 2002).

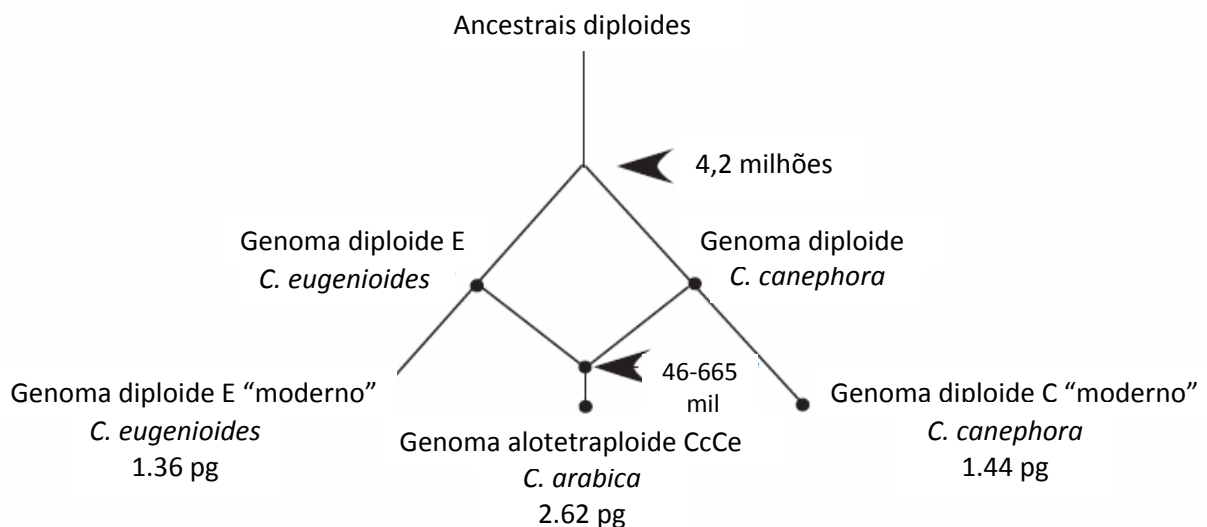
Existem inúmeras estratégias para prospecção e genotipagem de SNPs e a escolha da metodologia mais adequada é baseada no organismo a ser estudado, no número de polimorfismos a genotipar, no tipo de informação a ser obtida, nos materiais e informações disponíveis, bem como no custo e eficácia dos

experimentos (TABASSUM; LAKHANPAUL, 2006). Atualmente, com as novas tecnologias de sequenciamento é possível sequenciar o genoma ou o transcriptoma completo de vários genótipos com baixo custo, potencializando a genotipagem de SNPs nas populações. Nesse estudo foram utilizadas as estratégias de sequenciamento direto do produto de amplificação e sequenciamento de produtos de clonagem, devido a pouca informação existente sobre SNPs em *C. arabica*.

3.1.5 Diversidade Genética de *Coffea arabica*

A espécie *C. arabica*, preponderante economicamente, é um alotetraploide autocompatível que possui dois diferentes subgenomas homeólogos (CaCc e CaCe) derivados de duas espécies diploides de café, *C. canephora* (CC) e *C. eugenioides* (CE) (LASHERMES et al., 1999, DAVIS et al., 2007, MAURIN et al., 2007, YU et al., 2011). A diferenciação dos diploides ancestrais ocorreu há aproximadamente 4.2 milhões de anos, enquanto a formação de *C. arabica* possivelmente se deu entre 46 a 665 mil anos atrás (Figura 3), logo este poliploide é considerado uma espécie recente.

Figura 3 - Origem de *C. arabica*. A diferenciação dos ancestrais (CC e CE) ocorrendo há 4.2 milhões de anos e a formação do alotetraploide (CaCe e CaCc) de 46-665 mil anos. Redução no tamanho do genoma de *C. arabica* em comparação com a soma dos genomas ancestrais.



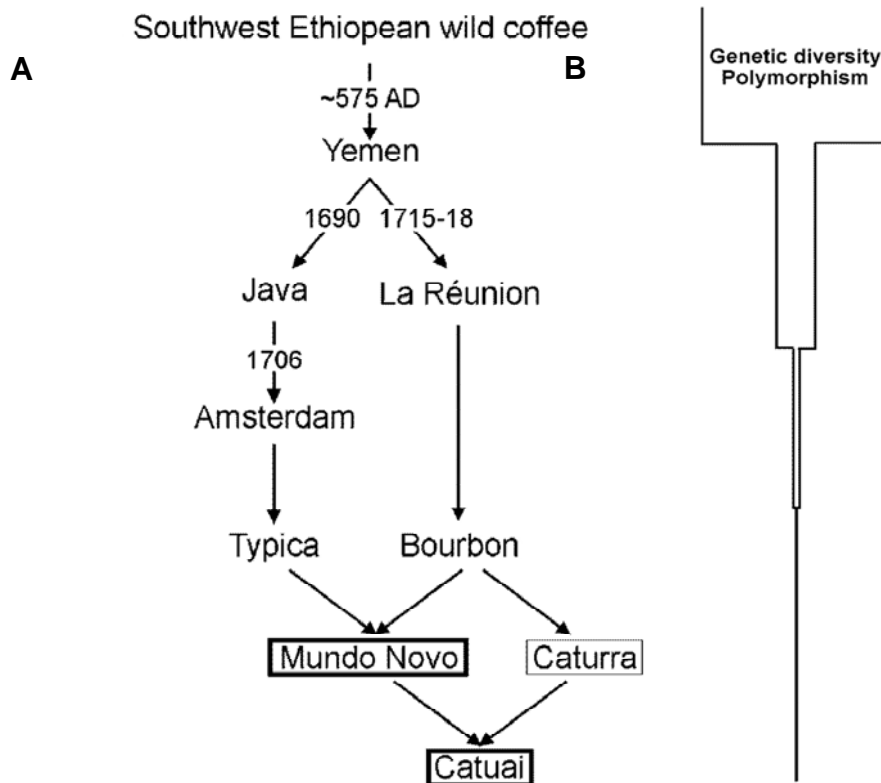
Fonte: traduzido e adaptado de Yu et al. (2011)

Segundo Yu et al. (2011), o evento de poliploidização que originou *C. arabica* não foi seguido de grandes reorganizações nem de rápida divergência nas regiões avaliadas de uma espécie *C. arabica* em comparação a uma espécie de *C. canephora*. Além disso, verificou-se uma diminuição no genoma de *C. arabica*, visto que seu tamanho é menor que a soma do genoma dos diploides (Figura 3), evento verificado também em outros poliploides.

O centro de diversidade genética de *C. arabica*, onde se encontram as plantas selvagens, abrange a Etiópia, o *Boma Plateau* do Sudão e Mount Marsabit do Kênia. Desses locais o café se dispersou para o Yemen no século XIV, onde foi cultivado por um longo tempo; assim o local é considerado como primeiro centro de dispersão da espécie (MONTAGNON et al., 2000, ANTHONY; DUSSERT; DULLOO, 2007).

No final do século XVII os europeus trouxeram sementes do Yemen para ilha de Java, e em 1710 para Amsterdam. No ano de 1717, a cultura foi introduzida na Ilha Bourbon. As populações de Amsterdam e da Ilha Bourbon (atualmente Réunion) originaram duas variedades distintas conhecidas como Typica e Bourbon, respectivamente, que são a base das cultivares modernas (ANTHONY et al., 2001). A cultivar Caturra é um mutante do grupo Bourbon; Mundo Novo foi originada do cruzamento de Bourbon e Typica, enquanto Catuai resulta de hibridização entre Mundo Novo e Caturra (Figura 4) (VIDAL et al., 2010).

Figura 4 – A, Origem das cultivares modernas de *C. arabica*. B, Representação da diminuição da diversidade da espécie com a dispersão e desenvolvimento de cultivares.



Fonte: Traduzido e adaptado de Anthony et al. (2002) e Vidal et al. (2010)

Além das cultivares mencionadas anteriormente, o Híbrido Timor ($2n=4x=44$), originado de um cruzamento natural entre *C. arabica* ($2n=4x=44$) e *C. canephora* ($2n=2x=22$), também é amplamente utilizado no melhoramento como fonte resistência a diversas doenças. Dessa forma, os genótipos de *C. arabica* podem ser agrupados em basicamente três categorias, os indivíduos “selvagens” da região do Sudão e da Etiópia, as cultivares não introgridas como Typica e Bourbon e as variedades introgridas (em sua maior parte proveniente de genótipos do híbrido Timor) (LEROY et al. 2006).

O estudo de Lashermes et al. (1999) sobre a diversidade de *C. arabica*, utilizando 23 marcadores RFLP confirmou a presença do genoma, ou dos cromossomos, de ambos os possíveis progenitores em *C. arabica*. Entretanto observou-se baixa variabilidade em *C. arabica*, diferentemente de *C. canephora* que apresentou a maior diversidade entre as espécies analisadas.

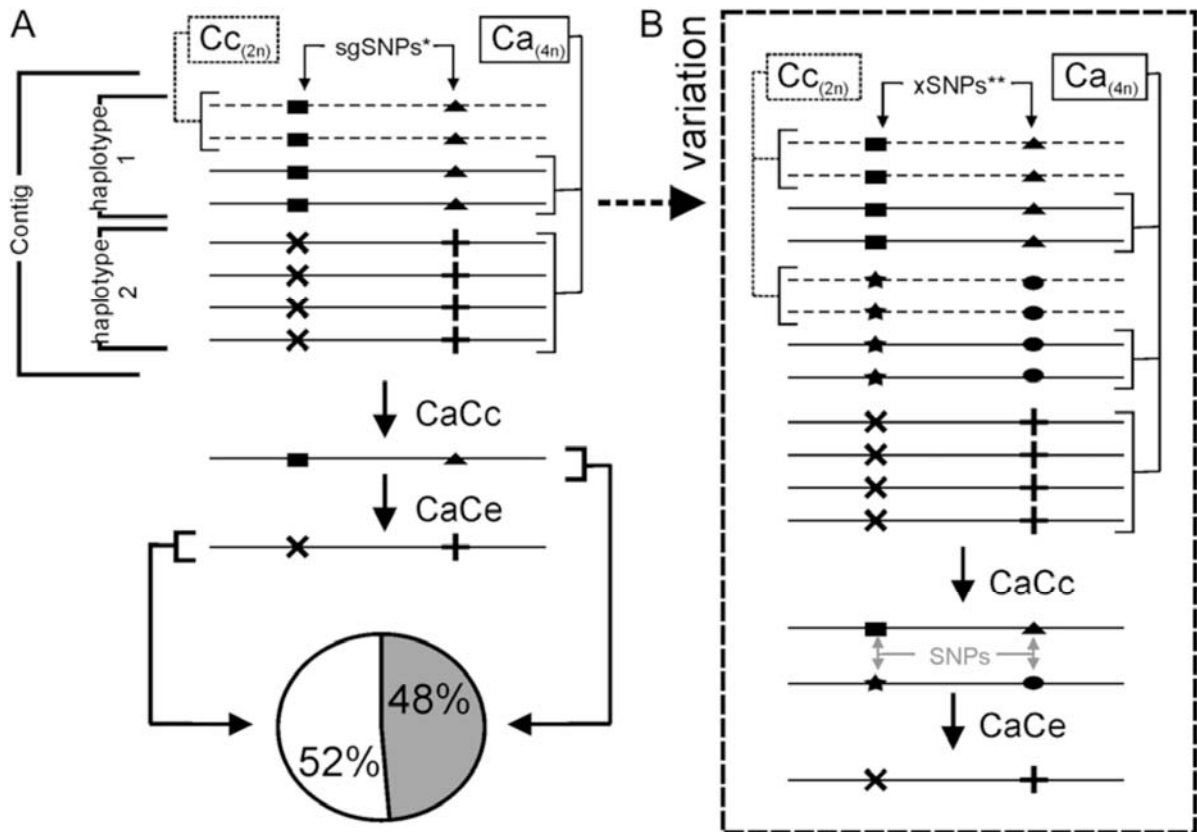
Anthony et al. (2001) analisou a relação filogenética entre 119 indivíduos de *C. arabica* utilizando 29 marcadores polimórficos (RAPD). Os

polimorfismos detectados dividiram os genótipos em três grupos (Typica, Bourbon e Etiópia) e o último grupo foi subdividido em quatro classes (Etiópia 1, 2, 3 e 4). Apesar da separação entre os grupos, nenhuma diferenciação em nível individual foi possível mesmo entre os genótipos do centro de origem.

Outros trabalhos desenvolvidos com marcadores AFLP, RAPD e SSR mostraram que esses têm potencial para identificação dos grupos de origem, mas não para discriminação de indivíduos, tanto para cultivares (STEIGER et al., 2002, MONCADA; MCCOUCH, 2004, SILVESTRINI et al., 2005, MALUF et al., 2005, VIDAL et al., 2010) quanto para genótipos do centro de origem (STEIGER et al., 2002, SILVESTRINI et al., 2005, MONCADA; MCCOUCH, 2004), e que a maior diversidade da espécie é observada entre os genótipos do centro de origem que podem atuar como fonte de genes de interesse para introgressão em programas de melhoramento. Missio et al. (2010) utilizando SSRs em *C. arabica* e *C. canephora* conseguiram identificar marcadores para realizar o *fingerprint* dos genótipos de *C. canephora* analisados e de cultivares de *C. arabica* resistentes a pragas, entretanto esses cultivares sofreram introgressão de genes de *C. canephora* do híbrido Timor e apresentavam maior variabilidade como reflexo desta introgressão.

Recentes estudos aplicaram os SNPs para acessar a variabilidade de *C. arabica* em 267,533 ESTs (*Expressed Sequence Tags*) disponibilizadas pelo projeto Genoma Café (VIEIRA et al., 2006) e por Lin et al. (2005). O banco de ESTs analisado era composto por mais de 21,5 mil *contigs* e 39 mil *singletons*. Como resultado, 23.062 SNPs e 2.165 INDELS foram encontrados e atribuídos aos subgenomas de *C. eugenioides* e *C. canephora* presentes em *C. arabica* (Figura 5).

Figura 5 – Identificação de SNPs em homeólogos de *C. arabica* e *C. canephora*.



A, *Contig* contendo dois padrões de polimorfismos (haplótipo 1 e 2) para *C. arabica* (Ca): o primeiro referente ao ancestral *C. canephora* (Cc) e o segundo atribuído ao outro ancestral, *C. eugenioides* (Ce). Do total de SNPs identificados no trabalho, 52 % correspondem ao subgenoma CaCe e 48% ao subgenoma CaCc. B, Alguns genes podem apresentar mais de um haplótipo para cada subgenoma devido à variação existente entre os genótipos de um ancestral, neste caso *C. canephora*.

Fonte: Vidal et al. (2010)

A despeito da grande quantidade de SNPs detectados nenhum deles apresentou potencial para genotipagem. Para base desse estudo foram utilizados apenas dois cultivares muito próximos de *C. arabica*, o que explica em parte os resultados observados (VIDAL et al., 2010).

3.1.6 Vias de Biossíntese de Açúcares e Diterpenos em Cafeeiros

Fatores como o aumento do mercado consumidor do café e o excedente de produção mundial da cultura (LEROY et al., 2006), têm impulsionado novas demandas para o melhoramento do cafeeiro: cada vez mais a qualidade da bebida tem se destacado pela possibilidade de agregar mais valor ao produto e

atender a mercados mais exigentes. A qualidade da bebida está diretamente relacionada aos seus atributos sensoriais (sabor, aroma, corpo, fragrância, entre outros).

As características sensoriais da bebida do café são determinadas por inúmeras substâncias. Por exemplo, o sabor é composto por uma combinação de mais de 800 compostos aromáticos, já a acidez, o amargor e a adstringência são determinadas por vários compostos não voláteis (BUFFO; CARDELLI-FREIRE, 2004; LEROY et al., 2006). Apesar dessa complexidade, alguns componentes, como os açúcares e diterpenos, exercem grande influência nos atributos finais da bebida e podem ser aplicados como indicadores de qualidade (SPILLER, 1998; KY et al., 2001; GEROMEL et al., 2006; LEROY et al., 2006).

C. arabica e *C. canephora* apresentam características distintas na composição dos grãos e na qualidade da bebida: o primeiro apresenta maior quantidade de sacarose nos grãos e é conhecido pela qualidade superior (menor amargor e melhor sabor), enquanto o segundo contém maior concentração de cafeína e ácidos clorogênicos e é amplamente utilizado na indústria para aromatização do café solúvel (KY et al., 2001; LEROY et al., 2006). Devido à diferença na composição de açúcares e diterpenos nos grãos e a sua influência na bebida das espécies de *Coffea* comercialmente utilizadas, estas se destacam como importantes alvos para o melhoramento.

Os açúcares, majoritariamente representados pela sacarose, podem originar diversas substâncias que influenciam as características sensoriais da bebida do café (KY, 2001). A sacarose está presente nos grãos verdes do café e é degradada rapidamente durante a torragem, resultando na produção de açúcares e outros compostos que interagem com aminoácidos para formar ácidos, alifáticos, furfurais, etc., que influenciam no sabor e no aroma da bebida (GEROMEL et al., 2006).

O metabolismo da sacarose desempenha papel central para a planta e está relacionado a resposta a estímulos ambientais, à partição de carbono, à síntese de parede celular, à tolerância à seca, entre outros (GONZÁLEZ et al., 1995; GEROMEL et al., 2006). Diversas enzimas participam do metabolismo da sacarose (Figura 6) e são alvos potenciais para o melhoramento, visto que podem influenciar a produção, acúmulo e transporte dessa substância no grão e conseqüentemente os atributos finais da bebida.

4 ARTIGO: ANÁLISE DA DIVERSIDADE NUCLEOTÍDICA DE *Coffea arabica*

4.1 INTRODUÇÃO

O café é a segunda maior *commodity* mundial tendo seu valor superado apenas pelo petróleo. Mundialmente, o faturamento anual desta cultura é de aproximadamente US\$ 90 bilhões e sua cadeia produtiva envolve cerca de 500 milhões de pessoas, desde o plantio até a comercialização do produto final. (DAMATTA; RAMALHO, 2006). A produção da bebida é baseada em duas espécies, *C. arabica* e *C. canephora*, e a primeira é preponderante economicamente (ABIC, 2011b). O café arábica é reconhecido por apresentar qualidade e sabor superior da bebida, entretanto é mais suscetível ao ataque de diferentes pragas (fungos, nematóides e insetos). Por outro lado, *C. canephora* apresenta grande variabilidade, extensa distribuição geográfica, maior resistência a pestes e notável capacidade adaptativa a distintas condições ambientais. O café canéfora, conhecido, como robusta, é comumente utilizado em *blends* com cafés de sabor mais ativo ou como café solúvel (CARVALHO et al., 2008).

O cafeeiro (*Coffea* spp) é uma planta dicotiledônea pertencente à família Rubiaceae. As espécies de *Coffea* são diplóides ($2n = 2x = 22$) e autoincompatíveis, à exceção de *C. arabica*, que é tetraplóide ($2n = 4x = 44$) e autocompatível (CARVALHO et al., 1991, BERTHOU; MATHIEU; VEDEL, 1983). Estudos indicam que esta espécie, é um anfidiplóide que possui dois diferentes subgenomas homeólogos (CaCc e CaCe) derivados de duas espécies diplóides de café, *C. canephora* (CC) e *C. eugenioides* (CE) (LASHERMES et al., 1999, DAVIS et al., 2007, MAURIN et al., 2007, YU et al., 2011).

A introgressão de novos caracteres em cafeeiros leva em média 28 anos em um programa de melhoramento tradicional, despendendo assim grande soma de tempo e recursos para o desenvolvimento de uma cultivar (SANTOS-BRIONES; HERNÁNDEZ-SOTOMAYOR, 2006). Esse processo é ainda mais demorado e custoso para *C. arabica* devido a sua limitada diversidade. Ferramentas biotecnológicas como os marcadores moleculares podem diminuir essas limitações, possibilitando o acesso à variabilidade genética dos cafeeiros e podendo permitir a escolha precoce e precisa de genótipos de interesse, otimizando o gerenciamento de tempo e recursos para o melhoramento vegetal.

Diversos estudos foram desenvolvidos para *Coffea canephora* empregando RFLP, SSR e SNPs. Esses marcadores se mostraram eficazes em caracterizar não apenas os grupos ou populações existentes (MUSOLI, et al., 2009, GOMEZ et al., 2009), como também discriminar os diferentes genótipos (GOMEZ et al., 2009, VIDAL et al., 2010, MISSIO et al., 2010). Além disso, existem mapas de ligação (utilizando marcadores RAPD, RFLP e SSR) (PAILLARD; LASHERMES; PÉTIARD, 1996, LASHERMES et al., 2001, COULIBALY et al., 2003), mapas de sintenia (LEFEBVRE-PAUTIGNY et al., 2010) e até mesmo QTLs identificados (*Quantitative trait locus*) (LASHERMES et al., 1996, LEROY et al., 2011) para a espécie. O emprego de diferentes marcadores (AFLP, RAPD e SSR) para acessar a diversidade de *C. arabica* mostrou que esses têm potencial para identificação dos grupos de origem, mas não para discriminação de indivíduos, tanto para cultivares (LASHERMES et al., 1999, ANTHONY et al., 2001, STEIGER et al., 2002, MONCADA; MCCOUCH, 2004, SILVESTRINI et al., 2005, MALUF et al., 2005, VIDAL et al., 2010) quanto para genótipos do centro de origem (LASHERMES et al., 1999, ANTHONY et al., 2001, STEIGER et al., 2002, MONCADA; MCCOUCH, 2004, SILVESTRINI et al., 2005), e que a maior diversidade da espécie é observada entre os genótipos do centro de origem, que podem atuar como fonte de genes para introgressão no melhoramento.

Um estudo com marcadores SNPs realizado por VIDAL et al., (2010) identificou 23.062 destes e 2.165 INDELS em ESTs de *C. arabica* e *C. canephora*. Apesar da grande quantidade de SNPs identificados não foi possível detectar variabilidade dentro dos subgenomas de *C. arabica* evidenciando a dificuldade de encontrar marcadores com potencial para genotipagem e mapeamento da espécie. Para base desse estudo foram utilizados apenas dois cultivares muito próximos de *C. arabica*, o que explica em parte esses resultados (VIDAL et al., 2010).

Neste contexto, tendo em vista a necessidade de identificação de marcadores intraespecíficos com potencial para o melhoramento *C. arabica* e a baixa variabilidade observada em estudos recentes, esse trabalho buscou avaliar a diversidade nucleotídica pela busca de SNPs presente em um painel de genótipos representativos de *Coffea spp*, pela detecção manual dos polimorfismos em fragmentos amplificados de DNA cloroplastídico, da álcool desidrogenase (ADH) e de oito genes envolvidos na biossíntese de açúcares e diterpenos, importantes compostos relacionados à qualidade da bebida.

4.2 MATERIAL E MÉTODO

4.2.1 Material Vegetal e Extração de DNA Genômico

Para acessar a diversidade nucleotídica de *Coffea* spp. foram utilizados 12 genótipos de *C. arabica*, oito de *C. canephora* e um de *C. eugenioides* (Tabela 1). A escolha dos genótipos foi baseada em Anthony et al. (2001) para *C. arabica* e Montagnon (2000) para *C. canephora*. DNA genômico foi extraído da folha do cafeeiro utilizando um protocolo MATAB (RISTERUCCI et al., 2000).

Tabela 1 – Lista dos genótipos analisados com descrição do grupo genético que pertencem e da representatividade amostrada.

Espécie	Genótipo	Grupo genético ^a	Representatividade/Importância ^b
<i>C. arabica</i>	E-007	Ethiopian2	Centro de origem da espécie
	E-017	Ethiopian3	Centro de origem da espécie
	E-123A	Ethiopian4	Centro de origem da espécie
	E-123B	Ethiopian4	Centro de origem da espécie
	E-237	Ethiopian2	Centro de origem da espécie
	E-238	Ethiopian4	Centro de origem da espécie
	E-464	Ethiopian1	Centro de origem da espécie
	E-516	Ethiopian1	Centro de origem da espécie
	Bourbon	Bourbon	Programas de Melhoramento
	Typica	Typica	Programas de Melhoramento
	Catuai	Typica/Bourbon	Utilizados para construção de Bibliotecas de ESTs e BACs
	Mundo Novo	Typica/Bourbon	Utilizados para construção de Bibliotecas de ESTs e BACs
<i>C. canephora</i>	C1007	Congolese_B	Centro de origem da espécie
	C2011	-	Centro de origem da espécie
	C3001	Congolese_SG1	Centro de origem da espécie
	C4001	Congolese_C	Centro de origem da espécie
	G2020	Guinea	Centro de origem da espécie
	G21	Congolese_SG1	Utilizados para construção de Bibliotecas de ESTs e BACs
	UW002	Uganda_Selvagem	Centro de origem da espécie
	UW099	Uganda_Selvagem	Centro de origem da espécie
<i>C. eugenioides</i>	<i>C. eugenioides</i>	-	Espécie ancestral de <i>C. arabica</i>

^abaseado em Anthony et al. (2001) para *C. arabica* e Montagnon (2000) para *C. canephora*.

^bMateriais selecionados de acordo com a representatividade genética e/ou importância em estudos genéticos do cafeeiro

4.2.2 Desenho de Oligonucleotídeos Iniciadores e Amplificação

Foram avaliadas regiões de 8 genes candidatos relacionados às vias de biossíntese de açúcares e diterpenos. Como controle foram analisadas duas regiões: uma do gene da álcool desidrogenase (ADH) e uma de DNA cloroplastídico (Figura 9). As sequências dos dez genes foram recuperadas dos bancos de ESTs de cafeeiro (LIN et al., 2005; VIDAL et al., 2010; MONDEGO et al. 2011) pela identificação dos ortólogos e parálogos do cafeeiro em comparação com sequências de *A. thaliana* (disponíveis em <http://www.arabidopsis.org>) através da ferramenta BLAST (ALTSCHUL et al., 1997).

As sequências recuperadas foram alinhadas no programa *Codon Code Aligner* 3.7.1, gerando uma sequência consenso para cada fragmento de gene (Anexo 1). Foram desenhados pares de oligonucleotídeos iniciadores (*primers*) para cada sequência consenso utilizando o software Primer3, levando em consideração as posições dos introns (ROZEN; SKALETSKY, 2000).

Todas as regiões foram amplificadas pela técnica da Reação da Polimerase em Cadeira (PCR) em reações contendo 25 ng de DNA, 1U de *Taq* DNA polimerase, 2 mM de $MgCl_2$, 2,5 mM de cada dNTP e 10 mM de cada primer, com volume final de 20 μ L. A amplificação foi realizada por PCR *touchdown* consistindo em um ciclo inicial de desnaturação a 94 °C/3', seguido de 10 ciclos de desnaturação a 94 °C/45'', pareamento de 60 a 50 °C /1' e extensão a 72 °C/3'30''. Os últimos 25 ciclos foram de 94 °C/45'', 50 °C /1', 72 °C/3'30'', seguido pela extensão final de 5'. A eficiência da reação foi verificada em gel de agarose 1%. A purificação dos produtos de PCR foi realizada adicionando 10% do volume do produto de acetato de sódio 3M com pH 5,2 e três vezes o volume de etanol 100%. Após centrifugação a 4000 RPM, por 30' a 4 °C, a fase líquida foi eliminada e o DNA foi eluído em 20 μ L de água ultrapura autoclavada.

4.2.3 Sequenciamento

A reação de sequenciamento foi realizada com o *kit BigDye® terminator v3.1 – Cycle Sequencing Kit* (Applied Biosystems), segundo instruções do fabricante, utilizando 20 ng do produto de PCR. Os produtos da amplificação foram precipitados com a adição de 80 μ L de isopropanol 80% incubado por 15 a

temperatura ambiente. Posteriormente a uma centrifugação de 45' a 4000 RPM e descarte da fase líquida, adicionou-se 100 µL de etanol 70%, seguindo nova centrifugação de 20' a 4000 RPM. Após a eliminação de todo etanol adicionou-se 10 µL de formamida e a amostra foi desnaturada a 95 °C por 2'. O sequenciamento dos produtos de PCR purificados foi realizado no ABI3130 XL *Genetic analyzer*.

4.2.4 Análises das Sequências e Detecção dos Polimorfismos

O alinhamento das sequências e a detecção manual dos polimorfismos foram realizados no programa *Codon Code Aligner* 3.7.1. Observou-se a presença ou ausência e o tipo de polimorfismos nos diferentes genótipos, bem como a relação deste entre as espécies e a frequência em que foi encontrado. As informações obtidas foram anotadas em tabelas do programa Excel, contendo: o número de polimorfismos encontrados; a sua classificação quanto ao tipo (SNP, INDELS ou SSR – alguns faziam parte de regiões de repetição); em que região estes se encontravam (Codante, Intrônica ou 3'UTR); quais eram silenciosos e quais estavam localizados em regiões codantes sinônimas ou não. Também foram calculadas a frequência total de polimorfismos em cada fragmento e as frequências de polimorfismos em cada região; além do cálculo da frequência entre as regiões sinônimas e não sinônimas, e a relação entre polimorfismos sinônimos e não sinônimos (NS/S).

4.2.5 Clonagem e Extração Plasmidial

Muitos genótipos de *C. arabica*, possuíam dois alelos (bases) diferentes para um SNP, logo, as sequências produzidas pelo sequenciamento do produto direto de PCR apresentavam picos duplos nessas posições. Visando isolar os alelos para validar os resultados encontrados, foram clonados três fragmentos de Sacarose Sintase 1 para dois genótipos de *C. arabica* (E007 e E238). Os referidos fragmentos foram amplificados e purificados pelo *kit Wizard® SV Gel and PCR Clean-Up System* (Promega, EUA) e clonados no vetor pGEM (Promega, EUA) conforme as instruções do fabricante. Bactérias *E. coli*, linhagem DH10b, foram preparadas pelo método de Cloreto de Cálcio e transformadas por choque térmico (HANAHAN, 1983). A confirmação da inserção dos fragmentos nas colônias foi

realizada por reações de PCR de colônia (segundo o protocolo descrito anteriormente) e a extração dos plasmídeos selecionados foi realizada segundo protocolo de Sambrook, Fritsch e Maniatis (1989) com modificações. Para cada um dos fragmentos, foram sequenciados dez clones de cada genótipo, a fim de prospectar todos os alelos existentes em *C. arabica* (espécie tetraplóide).

4.3 RESULTADOS E DISCUSSÃO

Para a análise da diversidade nucleotídica foram amplificadas e sequenciadas 17 regiões de 8 genes relacionados às vias de biossíntese de açúcares e de diterpenos. Como controle utilizou-se uma região do gene ADH e uma do gene trnL-trnF (cloroplastídico) (Tabela 2). Na figura 9 estão representadas as regiões dos genes analisados neste estudo.

Figura 9 - Estrutura das regiões dos 8 genes relacionados as vias biossintéticas de açúcares e diterpenos, da álcool desidrogenase (ADH) e de DNA cloroplastídico de *C. arabica*. As regiões representadas (com exceção dos GAPs) foram utilizadas na análise de diversidade nucleotídica.

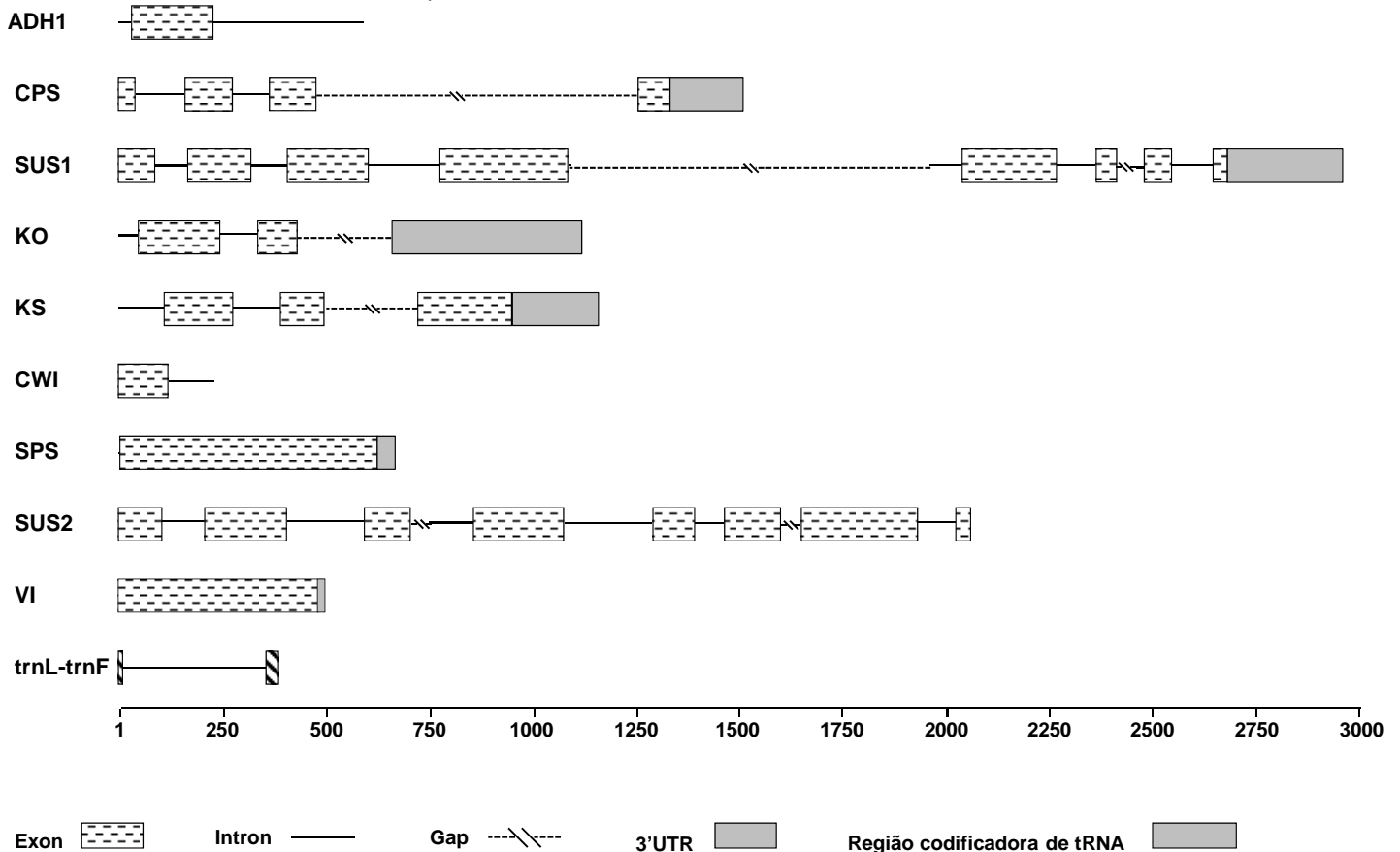


Tabela 2 - Sumário dos genes estudados e lista dos oligonucleotídeos iniciadores para cada um deles.

Gene	Número do GenBank ^a	Associação funcional via BLAST ^b	Sequência genômica completa do gene (bp) ^b	Fragmentos sequenciados (bp)	Nome Oligonucleotídeo	Pares de oligonucleotídeos iniciadores
ADH	HQ834787.1	Álcool desidrogenase 1	2024 (AT1G77120)	594	ADH1.1	F GGAAGCAGGGAAGCCACTA R AGAGGAGCAAGGGGATTGAT
CPS	FJ409842	Copalil difosfato sintase	7047 (AT4G02780.1)	475 + 254	CPS_C3_1 CPS_C3_3	F TCAAGCCAGAAATCGAGGAG R GCAAGAATTCCGACGAGAAC F CACAAGCGTCAGCTCCATAG R TGCCAATATTTTCTAAGGCACA
SUS1	AJ880768	Sacarose sintase 1	3895 (AT5G20830.1)	543+561+448+475	CSP_SUS1_4 CPS_SUS1_5 CSP_SUS1_8 P23/Susy1_c435_e	F TTTTGGCTGAGTTTGAAGAAA R AGAGAGGTGCCTGTTGAGGA F ACATAGGCGACGGAGTTGAG R ACGTTTTCTGGGCAAAGTA F GAGCAAATTTGATGAGAAGTACCA R CCAGTCTTGCCATGGTGAAT F CGCCGTTATCTGGAAATGTT R TCGAAATGGAAGGTAAATTGAT
KO	FJ409844	Caureno oxidase	2389 (AT5G25900.1)	440+457	Ko_C1_1 Ko_C1_2	F GAATGGGCCATGTACGAACT R ATTGTAATGCACGCGATCAA F GCCTCGACCACATCTTTGTT R GGCAGGAGAACAATCAAGC
KS	FJ409845	Caureno sintase	4196 (AT1G79460.1)	498+483	Ks_C1_1 Ks_C1_2	F GGGAGGTTCTGAGGAGGAAC R AGCCGACCACAATTGCTAAC F TTGCAKAGGTGAAATTGCTG R TCCAACGTTTAAACCCRAGC
CWI	DQ834314.1	Invertase de parede celular	2957 (AT3G13790.1)	220	P5/CWI_c5a_PP3	F CATCCCAGTTGGAGCAAGTT R TCTGGCAAATACGTACCTTTTCTAT
SPS	DQ842233.1	Sacarose fosfato sintase	5541 (AT5G20280.1)	668	P7/SPS_c1_PP2	F TCGATGGGGTATGGATTTGT R CAAGCGTAATCCCACACTCA
SUS2	AM087676.1	Sacarose sintase 2	4441 (AT5G49190.1)	700+423+836	P12/Susy2_c4_b P14/Susy2_c6_b P17/Susy2_c4_6_a	F GCCTGGTGTGGGAGTATG R GTGGGAGCTTGGCAAGATAA F GCTGGATCGGGTAAAAACA R TAGGAAGGCCACAGGTCATC F TCCAGGCTTGAATCTGCTTT R TCCAGGTAAGGCCAAACATC
VI	DQ834315.1	Invertase vacuolar	2469 (AT3G52600.1)	500	P25/Inv_c1_c	F TCCTCTCCTCTTGACTGGA R TACCCACGTCAGCCTTTAC
TRNL -TRNF	EF044213.1	precursor de tRNA (cloroplastídico)	597 (ATCG00400.1)	420	trnL_trnF	F GGTTCAAGTCCCTCTATCCC R ATTTGAACTGGTGACACGAG

^aNúmero de acesso as sequências depositadas no banco de dados do NCBI (*National Center for Biotechnology Information*)^bBaseado em sequências de *A. thaliana* (<http://www.arabidopsis.org>).

Alguns genes tiveram mais de uma região amplificada. Os *gaps* representam a distância entre as regiões estudadas de um mesmo gene, visto que na maioria dos casos estas não compunham fragmentos/regiões contíguas. A escala da figura está em pb e a soma do tamanho dos gaps e fragmentos analisados é organizada de forma a representar o tamanho total do gene (Figura 9 e Tabela 2). A sequência dos *primers* utilizados para as amplificações, encontram-se na tabela 2.

A Tabela 3 mostra a relação dos genótipos que foram sequenciados para cada fragmento dos genes.

Tabela 3 - Número de genótipos com sequências amplificadas para cada região.

Fragmentos (regiões)*	Número de genótipos com sequências amplificadas			
	Total	<i>C. arabica</i>	<i>C. canephora</i>	<i>C. eugenioides</i>
ADH1 (alcool_desidrogenase)	21	12	8	1
CPS_C3_1 (Copalil_difosfato_sintase)	16	8	7	1
CPS_C3_3 (Copalil_difosfato_sintase)	21	12	8	1
CSP_SUS1_4 (Sacarose_sintase_1)	21	12	8	1
CSP_SUS1_5 (Sacarose_sintase_1)	21	12	8	1
CSP_SUS1_8 (Sacarose_sintase_1)	21	12	8	1
Ko_C1_1 (Caureno_oxidase)	21	12	8	1
Ko_C1_2 (Caureno_oxidase)	21	12	8	1
Ks_C1_1 (Caureno_sintase)	21	12	8	1
Ks_C1_2 (Caureno_sintase)	19	12	6	1
P5/CWI_c5a_PP3 (Invertase_de_parede_celular)	17	10	6	1
P7/SPS_c1_PP2 (Fosfato_sacarose_sintase)	21	12	8	1
P12/Susy2_c4_b (Sacarose_sintase_2)	21	12	8	1
P14/Susy2_c6_b (Sacarose_sintase_2)	18	10	7	1
P17/Susy2_c4_6_a (Sacarose_sintase_2)	20	12	7	1
P23/Susy1_c435_e (Sacarose_sintase_2)	15	7	7	1
P25/Inv_c1_c (Invertase_vacuolar)	10	5	4	1
trnL_trnF (cloroplasto)	21	12	8	1

*Nomes correspondem aos oligonucleotídeos iniciadores utilizados para amplificação dos fragmentos.

A estratégia de sequenciamento direto dos produtos de PCR das regiões selecionadas foi eficiente para grande parte dos genes. Entretanto não foi possível amplificar completamente algumas regiões nos genótipos de *C. arabica* (e estes apresentaram sequências mais curtas), devido à presença de INDELS heterozigotos que impossibilitaram a leitura do sequenciamento a partir destes

polimorfismos (Anexo B). Por isso uma etapa posterior de clonagem dos produtos de amplificação antes do sequenciamento foi realizada para permitir um acesso mais preciso aos haplótipos.

As sequências obtidas foram analisadas e as que apresentavam boa qualidade foram alinhadas no programa *Codon Code Aligner*, com a ferramenta de alinhamento local, utilizando o mínimo percentual de identidade de 80%. A partir dos alinhamentos, os SNPs eram identificados e verificados manualmente a fim de evitar falsos polimorfismos. Os SNPs heterozigotos de *C. arabica* apresentaram picos duplos nos cromatogramas, correspondentes a cada um dos alelos, nas sequências obtidas diretamente do produto de PCR (Anexo C) demandando uma etapa posterior de clonagem para sua validação.

Para facilitar a visualização dos resultados dos alinhamentos, esses foram anotados em tabelas do *Excel*. As tabelas contém a caracterização do SNP, ou seja, nome do polimorfismo, posição em que se encontra no alinhamento, as sequências anteriores e posteriores ao mesmo, os alelos encontrados e o tipo de polimorfismo (Anexo D), bem como a descrição dos alelos presentes em cada genótipo (Anexo E). Tabelas como essas foram elaboradas para cada um dos fragmentos estudados e os resultados gerais foram sumarizados na Tabela 4 e 5.

Tabela 4 – Análise dos fragmentos contendo: total de pb para cada um deles; total avaliado por região (codante, intron e 3'UTR); número de polimorfismos bem como o tipo (SNP, INDEL e SSR); a região onde foram identificados (codante, intron e 3'UTR) e a classificação dos polimorfismos silenciosos, sinônimos e não sinônimos.

Fragmentos*	Total Analisado (pb)	Região			Total de Polimorfismos	Tipo de polimorfismo			Número polimorfismo/região			Polimorfismos		
		Codante	Intron	3'UTR		SNP	INDEL	SSR	Codante	Intron	3'UTR	Silenciosos**	Sinônimos	Não Sinônimos
ADH1 (alcool_desidrogenase)	594	198	396	NA	23	21	1	1	5	18	NA	20	2	3
CPS_C3_1 (Copalil_difosfato_sintase)	475	355	120	NA	16	13	2	1	11	5	NA	10	5	6
CPS_C3_3 (Copalil_difosfato_sintase)	254	77	NA	177	14	12	NA	2	2	NA	12	12	NA	2
CSP_SUS1_4 (Sacarose_sintase_1)	543	359	180	NA	15	14	NA	1	5	10	NA	14	4	1
CSP_SUS1_5 (Sacarose_sintase_1)	561	394	167	NA	15	13	2	NA	3	12	NA	14	2	1
CSP_SUS1_8 (Sacarose_sintase_1)	448	300	148	NA	8	7	1	NA	3	5	NA	7	2	1
Ko_C1_1 (Caureno_oxidase)	440	340	100	NA	24	23	NA	1	9	15	NA	21	6	3
Ko_C1_2 (Caureno_oxidase)	457	NA	NA	457	28	21	3	4	NA	NA	28	28	NA	NA
Ks_C1_1 (Caureno_sintase)	498	228	270	NA	15	13	2	NA	4	11	NA	12	1	3
Ks_C1_2 (Caureno_sintase)	483	227	NA	207	23	22	NA	1	4	19	NA	22	3	1
P5/CWI_c5a_PP3 (Invertase_de_parede_celular)	220	118	102	NA	11	11	NA	NA	7	4	NA	8	4	3
P7/SPS_c1_PP2 (Fosfato_sacarose_sintase)	668	620	5	43	19	18	NA	1	19	NA	NA	6	6	13
P12/Susy2_c4_b (Sacarose_sintase_2)	700	411	298	NA	28	26	NA	2	6	22	NA	27	5	1
P14/Susy2_c6_b (Sacarose_sintase_2)	423	332	91	NA	22	19	1	2	11	11	NA	17	6	5
P17/Susy2_c4_6_a (Sacarose_sintase_2)	838	453	382	NA	37	33	3	1	8	29	NA	35	6	2
P23/Susy1_c435_e (Sacarose_sintase_2)	475	80	97	318	31	23	5	3	2	8	21	29	NA	2
P25/Inv_c1_c (Invertase_vacuolar)	500	482	18	NA	18	18	NA	NA	17	1	NA	10	9	8
trnL_trnF (cloroplasto)	420	82	338	NA	13	13	NA	NA	NA	13	NA	13	NA	NA
Total	8997	5056	2712	1202	360	320	20	20	116	183	61	305	61	55

* Nomes correspondem aos oligonucleotídeos iniciadores utilizados para amplificação dos fragmentos.

**Soma dos polimorfismos das regiões 3'UTR e intrônicas, bem como dos polimorfismos sinônimos das regiões codantes. NA = Não analisado

Tabela 5 - Frequência de polimorfismos observada nas diferentes regiões avaliadas e entre polimorfismos sinônimos e não sinônimos.

Fragmentos*	Frequência**						
	Total	Codante	Intron	3'UTR	Sinônimos	Não sinônimos	NS/S***
ADH1 (alcoool_desidrogenase)	3,87	2,53	4,55	NA	1,01	1,52	1,5
CPS_C3_1 (Copalil_difosfato_sintase)	3,37	3,1	4,17	NA	1,41	1,69	1,2
CPS_C3_3 (Copalil_difosfato_sintase)	5,51	2,6	NA	6,78	NA	2,6	NA
CSP_SUS1_4 (Sacarose_sintase_1)	2,76	1,39	5,56	NA	1,11	0,28	0,25
CSP_SUS1_5 (Sacarose_sintase_1)	2,67	0,76	7,19	NA	0,51	0,25	0,5
CSP_SUS1_8 (Sacarose_sintase_1)	1,79	1	3,38	NA	0,67	0,33	0,5
Ko_C1_1 (Caureno_oxidase)	5,45	2,65	15	NA	1,76	0,88	0,5
Ko_C1_2 (Caureno_oxidase)	6,13	NA	NA	6,13	NA	NA	NA
Ks_C1_1 (Caureno_sintase)	3,01	1,75	4,07	NA	0,44	1,32	3,00
Ks_C1_2 (Caureno_sintase)	4,76	1,76	NA	NA	1,32	0,44	0,33
P5/CWI_c5a_PP3 (Invertase_de_parede_celular)	5	5,93	3,92	NA	3,39	2,54	0,75
P7/SPS_c1_PP2 (Fosfato_sacarose_sintase)	2,84	3,06	NA	NA	0,97	2,1	2,17
P12/Susy2_c4_b (Sacarose_sintase_2)	4	1,46	7,38	NA	1,22	0,24	0,2
P14/Susy2_c6_b (Sacarose_sintase_2)	5,2	3,31	12,09	NA	1,81	1,51	0,83
P17/Susy2_c4_6_a (Sacarose_sintase_2)	4,42	1,77	7,59	NA	1,32	0,44	0,33
P23/Susy1_c435_e (Sacarose_sintase_2)	6,53	2,5	8,25	6,6	NA	2,5	NA
P25/Inv_c1_c (Invertase_vacuolar)	3,6	3,53	5,56	NA	1,87	1,66	0,89
trnL_trnF (cloroplasto)	3,1	NA	3,85	NA	NA	NA	NA
Total	4	2,29	6,75	5,07	1,21	1,09	0,9

*Nomes correspondem aos oligonucleotídeos iniciadores utilizados para amplificação dos fragmentos.

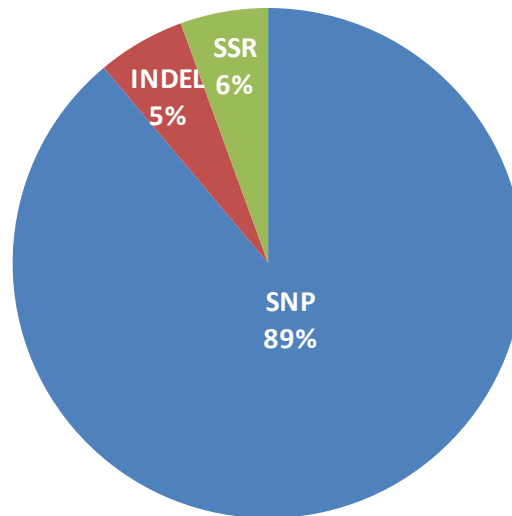
** A frequência foi estimada em 100 pb.

***Razão entre polimorfismos sinônimos (S) e não sinônimos (NS).

NA = Não analisado

No total foram analisados aproximadamente 9 kb resultando na identificação de 360 SNPs intra ou inter específico para *C. canephora*, *C. eugenoides* e *C. arabica* (Tabela 4). Destes 320 são SNPs, 20 são INDELS e 20 são microsatélites (SSRs) (os polimorfismos classificados como SSR fazem parte de regiões repetivas curtas) (Figura 10).

Figura 10 - Análise global da diversidade nucleotídica.



A maior parte dos SNPs localiza-se em regiões não codantes assim como observado em estudos anteriores em cafeeiros (MONDEGO et al., 2011; IVAMOTO et al., 2012) e em outras espécies como *Pinus pinaster*, *Pinus radiata* (POT et al., 2005) e *Lolium perenne* (BRAZAUSKAS et al., 2010; FILL et al., 2011). Os polimorfismos observados nas regiões codantes são majoritariamente sinônimos, resultado descrito em estudos de diversidade nucleotídica de cafeeiro (MONDEGO et al., 2011; IVAMOTO et al., 2012), de girassol (LIU; BURKE, 2005, FUSARI et al., 2008) e de *Lolium perenne* (BRAZAUSKAS et al., 2010).

A razão entre os polimorfismos não sinônimos e sinônimos (NS/S) é um indicativo de pressão seletiva em nível de sequência: quando esta razão é >1 , a taxa de evolução é maior que a taxa de seleção neutra e quando é <1 , a taxa de evolução é menor que a neutra. Dessa forma, a relação NS/S entre o painel de genes estudados está de acordo com estudos anteriores em cafeeiro e outras espécies (FUSARI et al., 2008; BRAZAUSKAS et al., 2010; MONDEGO et al., 2011; IVAMOTO et al., 2012; LIU;), onde a maior parte das sequências de regiões codantes são conservadas, diminuindo a possibilidade de mutações para variantes deletérias (MONDEGO et al., 2011). Os genes com a relação NS/S >1 teoricamente sofrem uma pressão seletiva maior e que podem ser utilizados para mapeamento associativo. Nesse estudo quatro fragmentos apresentaram esse perfil: um fragmento de ADH1, um de CPS, um de KS e um de SPS. Como os fragmentos analisados correspondem a regiões desses genes, seria necessário ampliar o tamanho da região avaliada (se possível, o gene completo), para confirmar essa

tendência e verificar o potencial desses genes para estudos de mapeamento associativo. No caso de KS por exemplo, o outro fragmento analisado para este gene apresenta a relação NS/S <1.

A frequência geral observada entre os genótipos analisados foi de 1 SNP a cada 25 pb. Xing et al. (2007) avaliaram a diversidade nucleotídica de 11 genes em 20 genótipos de *Lolium perenne* e observaram uma frequência similar de 1 SNP/ 33 bp. Em outro trabalho com *L. perenne* utilizando cinco genes e 20 genótipos, a frequência foi de 1 SNP/ 24 bp (BRAZAUSKAS et al. 2010). De acordo com esse último trabalho, a frequência só pode ser comparada com estudos que utilizem número similar de genótipos. Por exemplo no trabalho de Liu & Burke (2005) com 16 genótipos selvagens e 16 genótipos cultivados, foi observada uma frequência geral de: 1 SNP/ 16pb.

Dentre os polimorfismos detectados, 188 estão presentes entre os genótipos de *C. canephora* e 142 entre os genótipos de *C. arabica* (Tabela 6).

Tabela 6 - Análise de polimorfismos entre os genótipos de *C. canephora* e *C. arabica*.

Espécie	Total	Exons	Introns	3'UTR	Silenciosos*	Frequência total	NS/S**
<i>C. canephora</i>	188	60	93	35	158	2,09	0,77
<i>C. arabica</i>	142	56	70	16	123	2,13	0,57

*Soma dos polimorfismos das regiões 3'UTR e intrônicas, bem como dos polimorfismos sinônimos das regiões codantes.

**NS = não sinônimos, S= Sinônimos.

Os SNPs encontrados para *C. canephora*, são importantes, pois além de serem úteis para o melhoramento desta espécie, podem auxiliar no monitoramento da introgressão de genes na espécie arábica. Por exemplo, *C. canephora* possui maior resistência a estresses bióticos e abióticos, entretanto a qualidade da bebida dessa espécie é considerada inferior. Dessa forma, o monitoramento seria importante para garantir a introgressão apenas dos genes de resistência, assegurando que a qualidade da bebida não seja prejudicada pela introgressão de genes da espécie canéfora para esta característica.

A frequência de polimorfismos é próxima para ambas as espécies, 2 SNPs a cada 100 pb, entretanto *C. arabica* continua apresentando menor

variabilidade com potencial para genotipagem, pois a maioria de seus polimorfismos (82%) correspondem a diferenças entre *C. canephora* e *C. eugenioides* refletidas nos subgenomas de *C. arabica* (CaCc e CaCe) (Figura 11, a), ou a polimorfismos intraespecíficos de *C. canephora* (CaCc) (Figura 11, b) também refletidas nos subgenomas de *C. arabica*. Os 18% restantes são polimorfismos encontrados apenas nos genótipos de *C. arabica* (Ca), e correspondem a diferenças fixadas na população (todos os genótipos apresentam os mesmos alelos) (Figura 11, c), ou a polimorfismos variáveis entre os genótipos de *C. arabica* (com potencial aplicação para genotipagem) (Figura 11, d).

Dos 82% de polimorfismos que refletem diferenças entre *C. canephora* e *C. eugenioides*, 40% são diferenças fixadas entre a população dos ancestrais (CaCe e CaCe), 22% refletem diferenças intraespecíficas de *C. canephora* (CaCc) e 20% podem ser diferenças entre os genomas de *C. canephora* e *C. eugenioides* ou a variabilidade de *C. canephora* (CaCc e CaCe/CaCc) (Figura 12).

Figura 11 - Polimorfismos detectados em *C. arabica*. Polimorfismo entre os subgenomas que: a. reflete diferenças entre os alelos dos ancestrais, *C. canephora* e *C. eugenioides*, (CaCc e CaCe) ou b. reflete polimorfismos intraespecíficos de *C. canephora* (CaCc). SNPs detectados apenas na espécie arábica que: c. são fixados entre a população, ou d. variáveis entre os genótipos da espécie (com potencial para genotipagem).

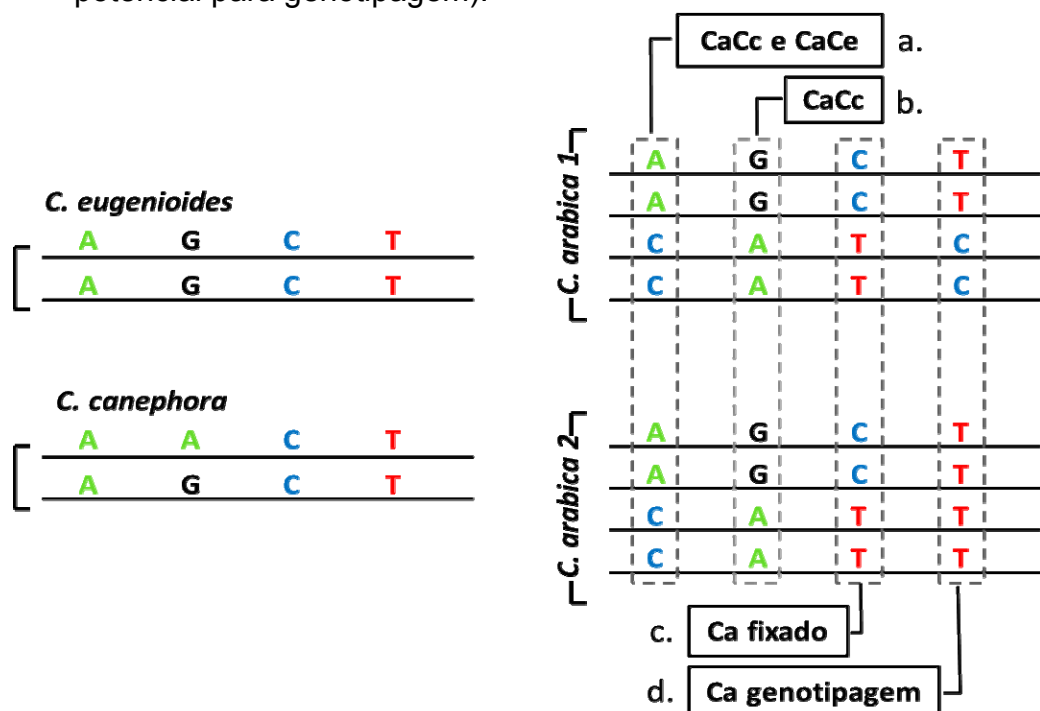
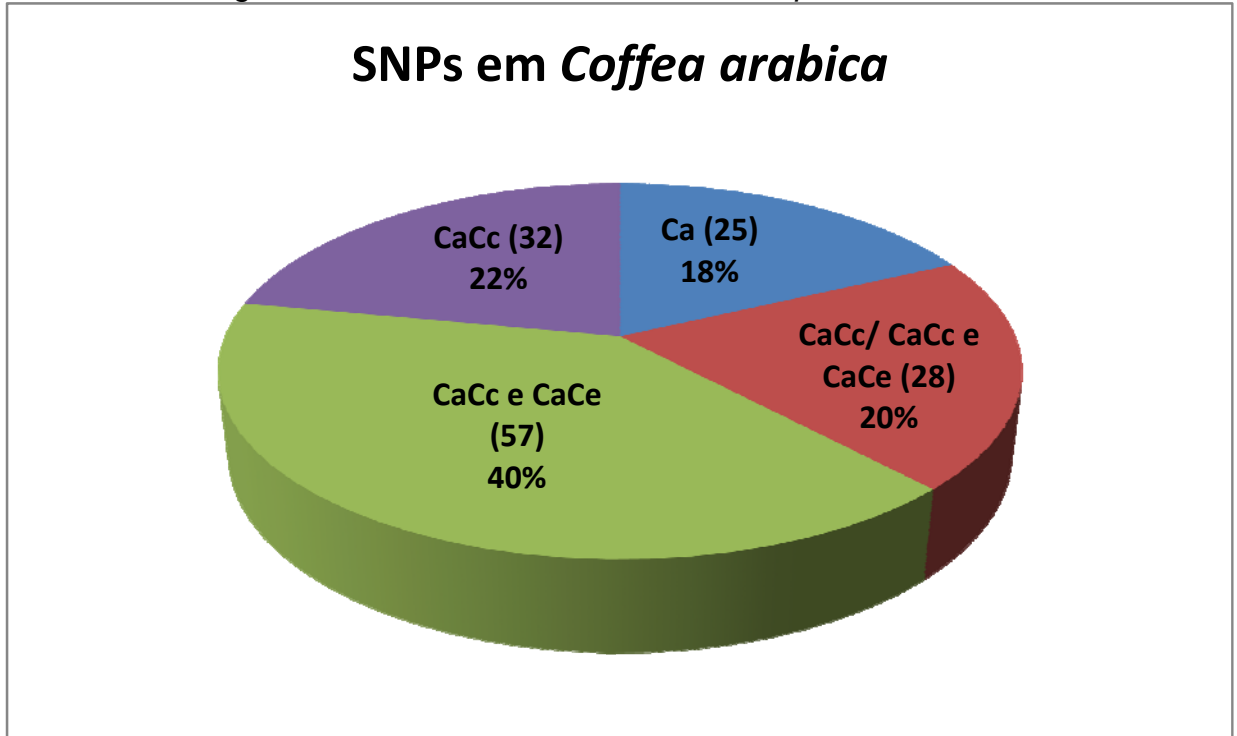


Figura 12 - Distribuição dos polimorfismos encontrados em *C. arabica*. Ca= diferenças entre os subgenomas de *C. arabica*. CaCc e CaCe= polimorfismos entre Cc e Ce para os quais *C. arabica* possui ambos os alelos. Cc= diferenças observadas em *C. arabica*, reflexo da variabilidade de *C. canephora*. Cc/ Cc e Ce = polimorfismos de *C. arabica* podem refletir diferenças entre os genomas de *C. canephora* e *C. eugenioides* ou a variabilidade de *C. canephora*.



Em estudos anteriores, Vidal et al. (2010) e Mondego et al. (2011) observaram uma menor frequência total de SNPs em *C. arabica*, 0,39 e 0,35 SNP/100pb, e em *C. canephora*, 0,17 e 0,19 SNP/100pb respectivamente. Ivamoto et al. (2012) ao analisarem SNPs em genes de ácidos clorogênicos, observaram uma frequência de 1,4 SNP / 100pb (Tabela 7).

Tabela 7 - Comparação entre a frequência de SNPs e o número de genótipos de estudos de SNPs em *Coffea* spp.

	Frequência (/100pb)		Número de Genótipos		
	<i>C. arabica</i>	<i>C. canephora</i>	<i>C. arabica</i>	<i>C. canephora</i>	<i>C. eugenioides</i>
Presente Estudo	2,09	2,13	12	8	1
Vidal et al. (2010)	0,39	0,17	2*	6	0
Mondego et al. (2010)	0,35	0,19	2*	6	0
Ivamoto et al. (2012)	1,4		2*	1	0

Nos referidos trabalhos observou-se que o número de SNPs de *C. arabica* é aproximadamente o dobro de polimorfismos de *C. canephora*. No presente estudo o número total de polimorfismos entre as duas espécies e a frequência estão próximos e o número de SNPs/100 pb é significativamente maior que os observados anteriormente. Essa diferença pode ser explicada pela amostra populacional analisada que continha 8 genótipos de *C.canephora* e 12 de *C. arabica* (incluindo os genótipos do centro de origem da espécie) em comparação aos 2 genótipos comerciais de *C. arabica* e de *C.canephora* utilizados por Vidal et al. (2010) e Mondego et al. (2011) e os 2 genótipos comerciais de *C.arabica* e 1 genótipo de *C. canephora* utilizados por Ivamoto et al. (2012). A separação dos subgenomas contendo os SNPs correspondentes aos observados nos ancestrais, *C. eugenioides* e *C.canephora* foi observada assim como descrito por Vidal et al. (2010), Mondego et al. (2011) e Ivamoto et al. (2012).

Diversos estudos com a aplicação de diferentes marcadores moleculares em *C. arabica* não detectaram marcas polimórficas que permitissem a discriminação entre os genótipos da espécie (LASHERMES et al., 1999, ANTHONY et al., 2001, STEIGER et al., 2002, MONCADA; MCCOUCH, 2004, SILVESTRINI et al., 2005, MALUF et al., 2005), incluindo os SNPs observados por Vidal et al. (2010). No presente trabalho foi possível verificar a presença de 25 polimorfismos (Tabela 8 e Tabela 9), aproximadamente 18% do total para a espécie, que permitem a diferenciação entre *C. arabica* e as outras espécies avaliadas. Entre esses SNPs, 13 são diferenças fixadas na população (ex: colunas 2, 8, 9, 12 e 13 da tabela 8), enquanto 12 demonstram variabilidade (ex; coluna 1, 3, 4 e 5 da tabela 8). Estes últimos polimorfismos, encontrados a uma frequência de 0,17 a cada 100 pb, são interessantes para identificação dos genótipos e apresentam potencial mapeamento e, por consequência, aplicação nos trabalhos de melhoramento.

Tabela 8 – Relação dos 25 SNPs encontrados em *C. arabica*. As colunas 1 a 25 representam os SNPs. Cada linha corresponde a um genótipo analisado. Os pontos representam polimorfismos que não foram acessados.

Genótipo	Especie	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
E-007	<i>C. arabica</i>	.	G/A	.	A/T	T	T	.	C/T	.	.	T	C/A	G/T	T/5	C	A
E-017	<i>C. arabica</i>	C	G/A	C/T	A/T	T/A	T/G	.	C/T	.	.	T	.	G/T	.	.	.	C	C	A	A/C
E-123A	<i>C. arabica</i>	C	G/A	C/T	.	T	T/G	.	C/T	G/C	.	T	C/A	G/T	C	A	A/C
E-123B	<i>C. arabica</i>	C	G/A	C	A/T	T/A	T	.	C/T	G/C	.	T	C/A	G/T	.	C/G	A	T	C/G	A/G	G/C	G/A	G/A	A/G	A/T	A/C
E-237	<i>C. arabica</i>	C	.	.	A/T	T	T	.	C/T	.	.	T	C/A	G/T	C	A
E-238	<i>C. arabica</i>	C	.	.	T	T/A	T/G	.	C/T	.	.	C	G/T	T/C	C	A	.	G/A	G	A/G	A/T	A/C
E-464	<i>C. arabica</i>	C	.	.	T	T	T/G	.	C/T	.	.	T	G/T	T/5	C	.	.	T/C	C	A	A/C
E-516	<i>C. arabica</i>	A/C	.	.	T	T/A	T/G	.	C/T	G/C	.	NA	C/A	G/T	G/A	G/A	A/G	A/T	A/C
Bourbon	<i>C. arabica</i>	C	.	C/T	A/T	T	T	.	C/T	G/C	.	T	G/T	C	A	A/C
Typica	<i>C. arabica</i>	C	G/A	C/T	T	T	T	.	C/T	G/C	T/C	T	G/T	T/C	C	A
Catuai	<i>C. arabica</i>	C	G/A	C/T	A/T	T	T	.	C/T	.	.	T	G/T	A/C
Mundo Novo	<i>C. arabica</i>	C	G/A	.	.	T	T/G	T/A	C/T	G/C	.	T	G/T	C	C	A	.	G/A	G/A	A/G	A/T	.
C1007	<i>C. canephora</i>	.	G	T	T	T	T	A	T	.	T	T	A	T	8 (T)	C	T	C	C	A	G	G	A	G	A	A
C2011	<i>C. canephora</i>	C	G	T	T	T	T	A	T	.	T	T	T	T	6 (T)	C	T	C	C	A	G	.	.	G	A	.
C3001	<i>C. canephora</i>	C	G	T	T	T	T	A	T	.	T	T	T	T	.	C	T	C	C	A	.	.	.	G	A	A
C4001	<i>C. canephora</i>	.	G	.	T	.	T	A	T	G	.	T	T	T	6 (T)	C	T	C	C	A	A
G2020	<i>C. canephora</i>	C	G	T	T	T	T	A	T	.	T	T	A	T	.	C	T	C	C	A	G	G	A	G	A	.
G21	<i>C. canephora</i>	C	G	T	T	T	T	.	T	.	T	T	T	T	A
UW002	<i>C. canephora</i>	.	.	.	T	T	T	A	T	G	.	T	T	T	A
UW099	<i>C. canephora</i>	C	G	T	T	T	T	A	T	.	T	T	A	T	8 (T)	C	T	C	C	A	G	G	A	G	A	A
<i>C. eugenioides</i>	<i>C. eugenioides</i>	C	G	.	T	T	T	A	T	.	T	T	T	T	.	C	T	C	C	A	G	G	.	.	.	A
<i>C. racemosa</i>	<i>C. racemosa</i>	C	G	T	T	T	T	.	T	.	T	T	T	T	6 (T)	C	T	C	C	A	G	G/A	G	A	A/T	A
<i>P. bengalensis</i>	<i>P. bengalensis</i>	C	G	.	T	T/A	T	A	T	.	T	T	T	T	.	C	T	C	C	A	G	A

Tabela 9 - Caracterização dos polimorfismos de *C. arabica*. Cada linha corresponde a descrição de um dos 25 SNPs de *C. arabica*.

Gene	Primer	SNP <i>C. arabica</i>	SNP	Sequência Anterior ao Polimorfismo	Sequência Posterior ao Polimorfismo	Alelo 1	Alelo2	Alelo3	Tipo	Região
Álcool desidrogenase	ADH1.3	1	3	TTTGAATGACGACATTCTTACACT	CCCTGTGAAAAACAGGGAGC	C	A		SNP	ExonNS
Copaliil Difosfato Sintase	CPS_C3_1	2	17	GTAGTATGCGACTCTCTGATAA	GGATAAGGCGACGTTGTAAT	G	A		SNP	ExonS
Copaliil Difosfato Sintase	CPS_C3_1	3	19	GACTCTCTGATAAGGGAT	AGGCGACGTTGTAATTGACA	A	G		SNP	ExonS
Sacarose Sintase 1	CSP_SUS1_4	4	35	TTAACAGGTTAGTCATAGAC	GTAAAAACAATTACCTGT	T	A		SNP	Intron
Sacarose Sintase 1	CSP_SUS1_5	5	15	TCTTGACTTCTCAGAACATTTT	ACAATAAGAAAAACAATTATGAG	T	A		SNP	Intron
Sacarose Sintase 1	CSP_SUS1_8	6	5	ATAATGATTGCCTTTTTTGT	TCTTTTTTTTTTT	T	C		SNP	Intron
Caureno sintase	Ko_c1_1	7	11	GACAGTTATGCAAGTAT	ATATAAAGAAGAAATGTC	A	T		SNP	Intron
Caureno sintase	Ko_c1_1	8	24	GTAGGAATATGATATCTCTCT	ATCTGTATCTTCATGCAC	T	C		SNP	ExonS
Caureno sintase	Ko_c1_1	9	32	CAACTAAAATCaTACAGGA	ACTATGCATTATGGAATCA	G	C		SNP	ExonNS
Caureno oxidase	Ko_C1_2	10	41	TAAATTACAGGACCAGGT	CACCGTAAAAATATTGGCG	T	C		SNP	3'UTR
Caureno synthase	Ks_c1_1	11	4	TATTCTGAAATCTGAACAT	CTTAGCCCTATAC	T	C		SNP	Intron
Invertase de Parece Celular	P5 (CWI_c5a_PP3)	12	1		TTTGGACTAAA	C	A		SNP	ExonS
Sacarose Sintase 2	P12 (Susy2_c4_b)	13	32	TCTTTATTGCGGAACATGAT	GAAGAGAGGTGACGATTAAGA	T	G		SNP	Exon S
Sacarose Sintase 2	P14 (Susy_c6_b)	14	6	AAAGAAGTCAAGAA	CAATTTAATCCGTTAAG	8 (T)	6 (T)	5 (T)	SSR	Intron
Sacarose Sintase 2	P14 (Susy_c6_b)	15	11	AGAAAAACATCCTG	ACGCAAATTGCTGATCC	C	G		SNP	Intron
Sacarose Sintase 2	P14 (Susy_c6_b)	16	13	TCCTGCACGCAAAATTGc	GATCCACGGGAAA	T	A		SNP	Intron
Sacarose Sintase 2	P14 (Susy_c6_b)	17	17	ACTGTACCTGAACAAAAAT	CCCCTTTTGTACGCAAT	C	T		SNP	ExonS
Sacarose Sintase 2	P14 (Susy_c6_b)	18	19	GCTCTATTGTTTGGG	AGCTATCCAACGGAATTGG	C	G		SNP	ExonNS
Sacarose Sintase 2	P14 (Susy_c6_b)	19	20	GCTCTATTGTTTGGGc	GCTATCCAACGGAATTGG	A	G		SNP	ExonS
Sacarose Sintase 2	P14 (Susy_c6_b)	20	31	CCTGTGATGTTTTAAACCC	CTCCACCACTGGCCGTCGT	G	C		SNP	ExonNS
Invertase vacuolar	P25 (Inv3_c1_c)	21	8	ggGGTTTTACACGCGGTTCCCC	CACGGGCATGTGGGAGTGC	G	A		SNP	ExonNS
Invertase vacuolar	P25 (Inv3_c1_c)	22	9	TTTACACGCGGTTCCCGCAC	GGCATGTGGGAGTGCCTGG	G	A		SNP	ExonS
Invertase vacuolar	P25 (Inv3_c1_c)	23	12	AGCACGTCTGAAAGGCAAGTTT	gACGACAATAAGCACG	G	A		SNP	Intron
Invertase vacuolar	P25 (Inv3_c1_c)	24	15	CACCTGGGCACCTATGATCCGAA	AACAATAAGTGGACCCCTGACGA	A	T		SNP	ExonNS
trnL trnF -The <i>trnL-trnF</i>	trnL trnF	25	9	TTCAATTTATATGATATATC	TAAATAACACAAGATTTG	A	C		SNP	Intron

4.3.1 Validação de SNPs pela Clonagem e Sequenciamento de Fragmentos

A busca baseada no sequenciamento de produtos de PCR para uma população representativa de genótipos de *C. arabica* (dezesesseis) permitiu a detecção de polimorfismos interespecíficos importantes, entretanto a presença de INDELS heterozigotos dificultou a identificação de alguns homeólogos (Anexo B). Portanto, para validação dos polimorfismos e determinação dos haplótipos de *C. arabica*, foram clonados três fragmentos de Sacarose Sintase para dois genótipos (E238 e E007) (Tabela 10).

A despeito de apresentar algumas sequências incompletas das regiões de interesse, foram confirmados 15 polimorfismos com alelos heterozigotos para *C. arabica*: três para Sus1_4 (polimorfismos 1, 5 e 17) , nove para Sus1_5 (todos os polimorfismos apenas para E-007) e três Sus1_8 (polimorfismos 12, 14 e 15). Estes SNPs correspondem às diferenças entre os subgenomas dos ancestrais e são importantes para diferenciar os alelos destes presentes em *C. arabica*, podendo, por exemplo, ser utilizados para estudos de expressão de homeólogos (MARRACCINI et al. 2011). Vale destacar que para Sus1_8 foi possível verificar um polimorfismo intraespecífico segregante para os genótipos da espécie que não havia sido detectado anteriormente que apresenta potencial para estudos de genotipagem.

Tabela 10 – Comparação entre os SNPs detectados em fragmentos sequenciados diretamente do produto de PCR e sequências obtidas de produto de clonagem de *C. arabica*.

Genótipo	Espécie	Origem do fragmento Sequenciado	Sus1_4						Sus1_5								Sus1_8							
			1	5	17	28	32	35	4	5	6	11	13	14	15	16	18	3	5	9	12	14	15	35
E-007	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	A/T	T/C	T/C	G/C	T/C	G/A	T/C	T	A/G	T/G	C/G	T	NA	NA	NA	NA	NA
E-017	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	A/T	T/C	T	G	T/C	G/A	T/C	T/A	A	T/G	C/G	T/G	T	A/G	NA	NA	NA
E-123A	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	NA	T/C	T/C	G/C	T/C	G/A	T/C	T	A/G	T/G	C/G	T/G	T	NA	NA	NA	NA
E-123B	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	A/T	T/C	T	G	T/C	G/A	T/C	T/A	A	T/G	C/G	T	T/C	NA	NA	NA	NA
E-237	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	A/T	T/C	T/C	G/C	T/C	G/A	T/C	T	A/G	T/G	C/G	T	NA	NA	NA	NA	NA
E-238	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	T	T/C	T	G	T/C	G/A	T/C	T/A	A	T/G	C/G	T/G	T	A/G	A/T	C/T	NA
E-464	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	T	T/C	T/C	G/C	T/C	G/A	T/C	T	A/G	T/G	C/G	T/G	NA	NA	NA	NA	NA
E-516	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	T	T/C	T	G	T/C	G/A	T/C	T/A	A	T/G	C/G	T/G	T/C	NA	NA	NA	NA
Bourbon	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	A/T	T/C	T/C	G/C	T/C	G/A	T/C	T	A/G	T/G	C/G	T	NA	NA	NA	NA	NA
Typica	<i>C. arabica</i>	Direto do Produto de PCR	A/G	T/G	C/A	A/G	T/G	T	T/C	T/C	G/C	T/C	G/A	T/C	T	A/G	T/G	C/G	T	NA	NA	NA	NA	NA
Catuai	<i>C. arabica</i>	Direto do Produto de PCR	NA	T/G	C/A	A/G	T/G	A/T	T/C	T/C	G/C	T/C	G/A	T/C	T	A/G	NA	C/G	T	NA	NA	NA	NA	NA
Mundo Novo	<i>C. arabica</i>	Direto do Produto de PCR	NA	NA	C/A	A/G	T/G	NA	T/C	T/C	G/C	T/C	G/A	T/C	T	A/G	T/G	C/G	T/G	T	NA	NA	NA	NA
E-007	<i>C. arabica</i>	Produto de Clonagem	NA	NA	NA	NA	NA	NA	T	C	C	T	G	C	T	G	T	NA	NA	NA	G	A	C	(TCCTG)
E-007	<i>C. arabica</i>	Produto de Clonagem	NA	NA	NA	NA	NA	NA	T	C	NA	T	NA	C	T	G	T	C	T	T	A	T	T	INDEL
E-007	<i>C. arabica</i>	Produto de Clonagem	NA	NA	NA	NA	NA	NA	NA	C	NA	T	G	C	T	G	T	NA	NA	NA	A	A	T	INDEL
E-007	<i>C. arabica</i>	Produto de Clonagem	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	T	A	G	C	T	T	A	T	T	INDEL
E-007	<i>C. arabica</i>	Produto de Clonagem	A	T	C	NA	NA	NA	NA	NA	NA	T	G	C	T	G	T	C	T	T	A	A	T	(TCCTG)
E-007	<i>C. arabica</i>	Produto de Clonagem	A	G	C	NA	NA	NA	NA	NA	NA	G	T	T	A	G	C	T	T	A	T	T	INDEL	
E-007	<i>C. arabica</i>	Produto de Clonagem	G	G	A	A	NA	NA	C	T	G	C	A	T	T	A	G	NA	NA	NA	A	T	T	INDEL
E-007	<i>C. arabica</i>	Produto de Clonagem	G	G	A	A	T	NA	C	T	G	C	A	T	T	A	G	NA	NA	NA	A	T	T	INDEL
E-007	<i>C. arabica</i>	Produto de Clonagem	A	NA	NA	NA	NA	NA	C	T	G	C	A	T	T	A	G	NA	NA	NA	G	A	C	(TCCTG)
E-007	<i>C. arabica</i>	Produto de Clonagem	A	T	C	NA	NA	NA	C	T	G	C	A	T	T	A	G	C	T	T	A	T	T	INDEL
E-238	<i>C. arabica</i>	Produto de Clonagem	G	G	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	A	A	C	(TCCTG)
E-238	<i>C. arabica</i>	Produto de Clonagem	G	G	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	T	T	G	A	C	(TCCTG)
E-238	<i>C. arabica</i>	Produto de Clonagem	A	T	C	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	A	T	T	(TCCTG)
E-238	<i>C. arabica</i>	Produto de Clonagem	G	G	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	A	C	(TCCTG)
E-238	<i>C. arabica</i>	Produto de Clonagem	G	G	A	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	T	T	(TCCTG)
E-238	<i>C. arabica</i>	Produto de Clonagem	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	A	T	T	(TCCTG)
E-238	<i>C. arabica</i>	Produto de Clonagem	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	A	T	T	(TCCTG)
E-238	<i>C. arabica</i>	Produto de Clonagem	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	A	C	(TCCTG)
E-238	<i>C. arabica</i>	Produto de Clonagem	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	A	C	(TCCTG)
E-238	<i>C. arabica</i>	Produto de Clonagem	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	T	T	(TCCTG)

4.4 CONCLUSÃO

A análise da diversidade nucleotídica de fragmentos de genes em um painel representativo de genótipos de *C. arabica* permitiu a identificação de vários polimorfismos a nível inter e intraespecífico. A maioria dos polimorfismos de *C. arabica* correspondeu a diferenças observadas entre as espécies ancestrais, *C. canephora* e *C. eugenioides*. Esses polimorfismos podem ser utilizados para estudos de diferenciação de expressão de homeólogos em *C. arabica*, para monitoramento da introgressão de genes nos programas de melhoramento, dentre outras aplicações.

Foram identificados 12 SNPs que corresponderam variações apenas entre os subgenomas de *C. arabica*, evidenciando a existência de variabilidade intraespecífica, ainda que em frequência reduzida, com potencial para genotipagem e estudos de mapeamento.

A clonagem de fragmentos amplificados permitiu a confirmação de alguns polimorfismos heterozigotos de *C. arabica*, e a identificação de um novo polimorfismo segregante apenas para os subgenomas da espécie.

Os dados obtidos também mostram a importância de utilizar genótipos representativos da diversidade da espécie para detecção de SNPs. Os resultados deste estudo são relevantes para nortear futuros trabalhos na caracterização da variabilidade genética de *C. arabica*.

REFERENCIAS

- ABIC - ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ. Estatísticas - **Indicadores de Desempenho da Cafeicultura Brasileira**, 2011a. Disponível em: <<http://www.abic.com.br/publique/cgi/cgilua.exe/sys/start.htm?sid=52>> Acesso em: 7 out. 2011.
- ABIC - ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DO CAFÉ. Estatísticas – **Parque Cafeeiro, Produção e Produtividade**, 2011b. Disponível em: <<http://www.abic.com.br/publique/cgi/cgilua.exe/sys/start.htm?sid=48>> Acesso em: 7 out. 2011.
- ANTHONY, F. et al. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. **Euphytica**, v.118, p.53-65, mar. 2001.
- ANTHONY, F.; DUSSERT, S.; DULLOO, E. Coffee genetic resources. In: Engelmann, M. E. et al. (Ed.). **Conserving coffee genetic resources. Maccarese**. Biodiversity International, p.12-22, 2007 (Topical Reviews in Agricultural Biodiversity)
- ALTSCHUL, S.F., et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research** v.25, p.3389-3402, 1997.
- BERTHOU, F.; MATHIEU, C.; VEDEL, F. Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L. **Theoretical and Applied Genetics**, v.65, n.1, p.77-84, Apr. 1983.
- BORÉM, A. Aplicação dos Marcadores Moleculares no Melhoramento. In **Marcadores Moleculares**. Ed. Borém e Caixeta, 2 ed. p. 95-102, Viçosa: Editora UFV, 2009.
- BORÉM, A.; MIRANDA, G. V. **Melhoramento de Plantas**. 5 ed. Viçosa: Editora UFV, 2009.
- BRAZAUSKAS, G. et al. Nucleotide diversity and linkage disequilibrium in five *Lolium perenne* genes with putative role in shoot morphology. **Plant Science**, v. 179, p. 194-201, abr. 2010.
- BROOKES, A. J. The essence of SNPs. **Gene**, v.234, p.177-186, Jul. 1999.
- BRUMFIELD R.T et al. The utility of single nucleotide polymorphisms in inferences of population history. **Trends in Ecology and Evolution**, v.18, n.5, p.249-256, maio 2003.
- BUFFO, R. A.; CARDELLI-FREIRE, C. Coffee flavour: an overview. **Flavour and Fragrance Journal** v.19, p.99–104, 2004.

CAIXETA, E.V. et al. Biotecnologia Aplicada ao Melhoramento Genético do Cafeeiro. In: CARVALHO, A.H.S. (Org.). **Cultivares de Café: origem, características e recomendações**. Brasília: Embrapa Café, 2008. p. 103-128

CARVALHO, a. et al. Aspectos genéticos do cafeeiro. **Revista Brasileira de Genética**, v.14, p.135-138, 1991.

CARVALHO, C.H.S. (Org.). **Cultivares de café: origem, características e recomendações**. Brasília: Embrapa Café, 2008.

CHENG, A. X. et al. Plant Terpenoids: Biosynthesis and Ecological Functions. **Journal of Integrative Plant Biology**, v. 49 (2), p.179–186, 2007.

COULIBALY, I. et al. AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. heterocalyx* x *C. canephora*) x *C. canephora*]. **Theoretical and Applied Genetics**, v.107, p.1148–1155, jul. 2003.

DAMATTA, F.M., RAMALHO, J.D.C. Impacts of drought and temperature stress on coffee physiology and production: a review. **Brazilian Journal of Plant Physiology**, v.18, n.1, p. 55-81, 2006.

DAVIS, A. P. et al. Searching for the relatives of *Coffea* (Rubiaceae, Ixoroideae): the circumscription and phylogeny of *Coffeae* based on plasmid sequence data and morphology. **American Journal of Botany**, v.94, p.313-329, 2007.

FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Top production**. 2009a. Disponível em: <<http://faostat.fao.org/site/339/default.aspx>>. Acesso em: 08 out. 2011

FAO – FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **Top exports**. 2009b. Disponível em:< <http://faostat.fao.org/site/342/default.aspx>>. Acesso em: 08 out. 2011

FERRÃO, M.A.G. et al. Técnicas Moleculares e Biotecnológicas Aplicadas ao Café. In: FERRÃO, M.A.G. et al. (Ed.). **Café Conilon**. Vitória, ES: Incaper, 2007. p. 177-197

FERREIRA, M.E.; GRATTAPAGLIA, D. Aplicações de Marcadores Moleculares na Genética e Melhoramento de Plantas. In: _____ **Introdução ao uso de marcadores moleculares em análise genética**. 3ªed. Brasília: EMBRAPA-CERNAGEN, 1998. p. 69-116 (EMBRAPA-CERNAGEN Documento 20)

FILL, A. et al. Nucleotide diversity and linkage disequilibrium of nine genes with putative effects on flowering time in perennial ryegrass (*Lolium perenne* L.). **Plant Science**, v. 180, p. 228-237, 2011.

FUSARI, C.M. et al. Identification of Single Nucleotide Polymorphisms and analysis of Linkage Disequilibrium in sunflower elite inbred lines using the candidate gene approach. **BMC Plant Biology**, v.8, 7, jan. 2008.

- GEROMEL, C. et al. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. **Journal of Experimental Botany**, v. 57, n.12, p.3243–3258, abr. 2006.
- GEROMEL, C. et al. Sucrose metabolism during fruit development in *Coffea racemosa*. **Annals of Applied Biology**, v.152, n.2, p.179-187, nov. 2007.
- GOMEZ, C. et al. Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. **BMC Evolutionary Biology**, v.9, 167, jul. 2009.
- GONZÁLEZ, E.M et al. The role of sucrose synthase in the response of soybean nodules to drought. *Journal of Experimental Botany*, v.46, n.10, p. 1515-1523, maio 1995.
- GUPTA, P.K.; RUSTGI, S. Molecular markers from the transcribed/expressed region of the genome in higher plants. **Functional e integrative genomics**. v.4, p.139-162, abr. 2004.
- HANAHAN D. Studies on transformation of *Escherichia coli* with plasmids. **Journal of Molecular Biology** 166: 557-580, 1983.
- KIST, B.B. et al., **Anuário brasileiro do café 2011**, Santa Cruz do Sul: Editora Gazeta Santa Cruz do Sul, 2011.
- IVAMOTO, S.T. et al. (in Press) Diversidade nucleotídica dos genes envolvidos na biossíntese de ácidos clorogênicos de cafeeiros. *Coffee Science*, aceito em fev. 2012.
- JOËT, T. et. al. Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. **New Phytologist** v.182, p. 146–162, 2009.
- KWOK P.Y.; Gu Z. Single nucleotide polymorphism libraries: why and how are we building them? **Molecular Medicine Today**, v.5, n12, p.538-543, dez. 1999.
- KY, C.L. et al. Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. **Food Chemistry**, v. 75 p. 223-230, abr. 2001.
- LASHERMES, P. et al. Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. **Theoretical and Applied Genetics**, v. 93, p. 458-462, 1996.
- LASHERMES, P. et al. Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. **Theoretical and Applied Genetics**. V.94, p.947-955, out. 1997.

- LASHERMES, P. et al. Molecular characterisation and origin of the *Coffea arabica* L. genome. **Molecular and General Genetics**, v.261, p.259-266, 1999.
- LASHERMES, P. et al. Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. **Genome**, v.44, p. 589–596, jul. 2001.
- LEFEBVRE-PAUTIGNY, F. et al. High resolution synteny maps allowing direct comparisons between the coffee and tomato genomes. **Tree Genetics & Genomes**, v.6, p. 565–577, fev. 2010.
- LEROY, T. et al. Genetics of coffee quality. **Brazilian Journal of Plant Physiology**, v.18, p.229-242, 2006.
- LEROY, T. et al. Improving the quality of African robustas: QTLs for yield- and quality-related traits in *Coffea canephora*. **Tree Genetics & Genomes**, v.7, n. 4, p. 781-798, 2011.
- LIN, C. et al., Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. **Theoretical and Applied Genetics**, v. 112, p.114–130, nov. 2005.
- LIU, A.; BURKE, J.M. Patterns of Nucleotide Diversity in Wild and Cultivated Sunflower. **Genetics**, v. 173, p. 321-330, maio 2006.
- MALUF, M.P., Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. **Scientia Agricola**, v.62, n.4, p.366-373, Jul/Aug. 2005.
- MARRACCINI, P. et al. RBCS1 expression in coffee: *Coffea* orthologs, *Coffea arabica* homeologs, and expression variability between genotypes and under drought stress. **BMC Plant Biology**, v.11, art.85, maio 2011.
- MAURIN, O. et al. Towards a Phylogeny for *Coffea* (Rubiaceae): Identifying Well-supported Lineages Based on Nuclear and Plastid DNA Sequences. **Annals of Botany**, v.100, p.1565–1583, 2007.
- MEDINA FILHO, H.P.; BORDIGNON, R.; CARVALHO, A.H.S. Desenvolvimento de Novas Cultivares de Café Arábica. In: CARVALHO, A.H.S. (Org.). **Cultivares de Café: origem, características e recomendações**. Brasília: Embrapa Café, 2008. p. 79-102
- MISSIO, R.F. et al. Polymorphic information content of SSR markers for *Coffea* spp. **Crop Breeding and Applied Biotechnology**. V.10, p.89-94, 2010.
- MOCADA, P.; MCCOUCH, S. Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome** v.47, p. 501–509, maio 2004.

- MONDEGO, J.M.C. et al. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. **BMC Plant Biology**, v.11, art.30, fev. 2011.
- MONTAGNON, C. **Optimisation des gains génétiques dans le schéma de sélection récurrente réciproque de *Coffea canephora* Pierre**. 2000. Tese (Doutorado em Melhoramento de Plantas) - CIRAD, França.
- MORAIS, T.P., MELO, B. Biotecnologia aplicada ao melhoramento genético do cafeeiro. **Ciência Rural**, Santa Maria, v.41, n.5, p.753-760, mai. 2011.
- MUSOLI, P. et al. Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda. **Genome**, v. 52, p.634-646, 2009.
- PAILLARD, M.; LASHERMES, P.; PÉTIARD, V. Construction of a molecular linkage map in coffee. **Theoretical and Applied Genetics**,v.93, p.41-47, fev. 1996.
- POT, D. et al. Nucleotide variation in genes involved in wood formation in two pine species. **New Phytologist**, v. 167, p. 101-112, fev. 2005.
- RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Plant Biology**, v.5, p.94-100, abr. 2002.
- RISTERUCCI, A.M. et al. A high-density linkage map of *Theobroma cacao* L. **Theoretical and Applied Genetics**, v.101, p. 948-955, out. 2000.
- ROITSCH T.; GONZÁLEZ M.C.. 2004. Function and regulation of plant invertases: sweet sensations. *Trends in Plant Science* v.9, n.12, p. 606–613, dez. 2004.
- ROMANIELO, M.M. et al. Transferência de tecnologia em cafeicultura para diferentes regiões produtoras no estado de Minas Gerais. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 2, 2001, Vitória, ES. **Anais do Congresso**. Brasília. DF : EMBRAPA/CBP&D-Café, 2001. p. 1-4.
- ROZEN, S.; SKALETSKY, H.J. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S.; Misener S. **Bioinformatics Methods and Protocols: Methods in Molecular Biology**. Totowa: Humana Press, p. 365-386, 2000.
- SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T.; **Molecular cloning: a laboratory manual**. 2ed. New York: Cold Spring Harbor Laboratory Press; 1989.
- SANTOS-BRIONES, C.; HERNÁNDEZ-SOTOMAYOR, S.M.T. Coffee biotechnology. **Brazilian Journal of Plant Physiology**, 18(1):217-227, 2006.
- SILVESTRINI, M.; et al. Genetic diversity and structure of ethiopian, Yemen and brazilian *Coffea arabica* accessions using microsatellites markers. **Genetic Resources and Crop Evolution**, v. 54, p. 1367-1379, 2007.

SPEER, K.; KÖLLING-SPEER, I. The lipid fraction of the coffee bean. **Brazilian Journal of Plant Physiology**, v18, p. 201–216, 2006.

SPILLER, M. A. The Chemical components of coffee. In: SPILLER, G. A. **Caffeine**. New York: CRC Press, p.98-161, 1998.

STEIGER, D.L. et al. AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. **Theoretical and Applied Genetics**, v.105, p. 209–215, 2002.

TABASSUM, J.; LAKHANPAUL, S. Single polymorfism (SNP) – Methods and applications in plant genetics: A review. **Indian Journal of Biotechnology**, v.5, p.453-459, 2006.

VIDAL, R. et al. 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**, p. 1053-1066, 2010.

VIEIRA, L.G.E. et. al. Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, v.18, n.1, p.95-108, jan./mar. 2006.

WANG, Q. et al. CYP701A8: A Riceent-Kaurene Oxidase Paralog Diverted to More Specialized Diterpenoid Metabolism. **Plant Physiology**, v. 158, p. 1418–1425, 2012.

Wikström, N.; Savolainen, V.; Chase, M.W. Evolution of the angiosperms: calibrating the family tree. *Proceedings of Royal Society*, v. 268, n. 1482, p. 2211-2220, 2001.

XING, Y., et al. Nucleotide diversity and linkage disequilibrium in 11 expressed resistance candidate genes in *Lolium perenne*. **BMC Plant Biology**, v.7, 43, ago. 2007.

YU, Q. et al., Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (*Coffea*). **The Plant Journal**, v. 67, p.305–317, maio 2011.

ANEXOS

ANEXO A

Sequências consenso das regiões dos genes de interesse que foram analisadas.

ADH1_1

TGTGGAAGTCCCAACAAAATGATAAATTGGTTTTCCCATTAATTGAGAACCTTGATTTCCCATCATTAAAGCATCA
CACCCCTGTCTGTGTTGATCCTGAGGAGATCACACATGTTACTTTCTTCTGATTTGCAATGACGGCATTCCCT
ACACTCCCCTGTGAAAACAGGGAGCACATGGTCTCCGGGCTGAAGTTCTGTCACGCCTTCCCCTACACTCTC
CACAATCCTAATCGCCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
TATGAGTAACAGTATAACGAGAATCAGTGAAATCATGGCTCAAACATTAACATACCCTGATGCCTCGTGCCCG
AGAATTCTAGGAAACACGGGGTGTGGCCCTGAAAATATTGAAAAGTACGCAACCTCAGATCAGCAAAAATTTG
CTAAAGGACACTATGATCCATC---
AGGTTGATTTGACATAAGACACATTGCATTGATTATATACTATGCAAAAAGTGTGTGATTGTTGTTGTTACCTTC
GCTTCCCAGAAATAAACATCAGTATGGCAGAGGGAAGTGAAGAGAATCTTGAGCCGGA

CPS_C3_1

TCAAGCCAGAAATCGAGGAGTGTATCGATTATGTTACCAGGTAATCTCTGATCAAACCTCGACTACCTACCC
GAAACTCAAACACATTAATTCTTCCGTAACACAGACTAACATCAACCGTCAAAATTAATACTGATCATCGATCC
TACATAATTGTCAGATATTGGACGGAGAAAGGTATCTGTTGGGCTAGAAAATACCCACGTTTATGACATTGATG
ACACAGCAATGGGTTTCAGGCTCCTAAGGCAGCACGGCTACACGGTTTCTGCAGGTTAATTACTTGCCGTTT
ATAATTTGTTGCAGTACATGGTAGTATGCGACTCTCTGATAAGGGATAAGGCGACGTTGTAATTGACAGATG
TCTTCCGTAACCTTTGAGAAGGATGGTGGATTCTTTGCCCTTTGCCGACAGTCGA-
CCAGGCTGTGACCGGGATGTATAACCTGTACAGAGCTTGTGAGGTGATGTTCT

CPS_C3_3

TCACAAGCGTCAGCTCCATAGAACGAGACATGCAAGAACTAGCGAAGCTTGTCTCACC-AGTAG-
CCCGAGCGATCT-G-
GATGCTGATGTCAAGCAAACATTTCTCACAGTGGCAAGGAGCTTCTACTACACCGCCATTGCAATCCAGGA
ACCGTCAATTTCCATATGGCCAAAGTTCTCTTTGAGAGAGTACTTTGATTTAA-----
CATGTCCATTTCTTTTGTCTCGTCTGCTAAATGGCACCCAGTGAAAAGTTTATTCTCATACTGTTGTAGTAGTA
ATTAAGTGCTCACAAATGTATAGAAACGAATAATTATATA-
GTAATTTTTCTTTAATCTTCAATGTGCCTTAAAAAATATTGGCAA

CSP_SUS1_4

GGAAACAGCTAATACCATGAGAGAGGTGCCTGTTGAGGAACTCAACTCCGTCGCCTATGTAAGTTAGTTAAAG
TTGGCTTGGGAAAAGATGCTGTAAATGGTTCGAAGTCCAGTTCAAAACAAAATTTCCATTTTTGCTGCATTA
ACACATCGCATTAAATTAAGCAGAAAAA-
CAAAGGGATTGATGTAGTCATTAACCGAGTAAAAAGTACCATATAAAGTCCAAGACCCCTTATTTA--
CCTTCCATCAACGAGTTCTTCTTCAAATGCAGGTAAGTCTGGCACGGTTAACTCCTCAACAACGAGTGCATG
GACATTGACTCGAACATACTCCCAGACACCAGGTCTGAGACGAATAGCAAGTGAACCCAGGGGGGCAACA
CAATTGCTTCTGGAATAGTAAAGACCGTTTAAACAGAGTTGCTTTCTAGTAATTAACATGTTAGTCATAGACTG
TAAAAACAATTCTACCTGTGTGGACTTCAGGACTTCTTCAAAGGCATGATCATGAATTTTTGTTTACCATCCT
TGTTAATTTCTTCAAACCTCAGCCAAAAACTGGCCGGTCTTTTTACA

CSP_SUS1_5

GACATGACGTTTTCTGGGCAAAGTATCCATGGGGGAAAGAATAACAACATTGAATACCATAGGGATTCTC
CCTAGGAATTTCTTAGTGTGACGAGTCCAGGAGCCTCCAGAAGATCCAGAAGCATGCAGATCATTCCAAAG
ACACG-
CTCAGCAGTATCACCCCAACCTCTCTCCAGTCCAATTTCTTGGAAATTTGTGCTCGAATTCAGAGTATGGTGT
TCTGCAGAGAGTGTGTTAGGTAAGTCTCTGCCTTCCCTCAGAAGTCTTGGAGAGTGTAAAGGTCCTTGATCC
TGTCGTTAAGCATCATCGTCTGCAGAAGGATCCAATCTATCTATTAATAATATGCTGAAAACATGTCAAAAAT
ATATAATTCTAAACAAGCAAATATAAATTAAGTAGTGCACATCTTGACTTTCTTGAACATTTTACAATAAA
GAAAACAATTATGAGACAGTGAAGTTATGAGAAAGGCTACCTTGCCCTTGATTGGTGAACACGGAGAAAATCA
AGGAGAGGGGCCATGCTCTCCTGTGATGGAACATTTTGGCAGAGAGGTGCCTGTTGAGGAACTCAACTCC
GTCGCCTATGTAAGTGGCCGTCGTTTTAC

CSP_SUSY_8

TCACTGCGGATCTTATCGCAATGAACCATACAGATTTTATTATCACTAGCACTTTCCAAGAAATAGCTGGAAG
GTACATTGAAGTTCCTTGTGTTGTAACATAATTAATCTTTGCATCTTAATCCCTCTGATAATCATTGCCCTTTTGT
TTCTTTTTTTTTTTTTTTCAGCAAGGACACTGTTGGCAATATGAAAGCCATATGGCCTTCACAATGCCAGGAT
TATACAGAGTTGTACATGGCATTGATGTTTTGATCCAAAATTAACATTGTCTCACCTGGAGCTGATACAAAC

CTCTACTACCCACACACAGAGAAGGAAAAGAGATTGACATCCTTCCATCCTGAAATTGAGGAGTTGCTTTTCA
 GCGATGTGGAGAATGAGGAACACCTGTAATATCTATCTCTTCTGTACTAATGGATTAGTTACTACTAAGTT
 CTCTGAAGTACTCATCAACATGTTTTAT

Ko_C1_1

CCTGAGAGCACCTGCACATACCCTCTTCCACCTCCATCAGGCAATCGTCTTGTGCAATTCCATGTAATCATG
 CAAGTCATCAACAAATCTCTCAGGAATCCACTGTTCCAGGGTTTTCCCAAGTTTTTTGTCCATGTTACACCCGT
 AAAGGTTGATAGCAATCTACACATTGCAAGAGACAGTTCAGTAAGTATATAATATAACAAGGTAGAAATGTCT-
 AGCAAGGATAGCAAACATGCAACAGCTCATGGTTAGGCATTAACCTCAGTTCAGCAGGAATATGATATCCTC
 CTAACTCTGTATCTTCATGCACATATCTCAGAGGAACTACGGGCACAGGACTGTGTCTCCTCAACGTTTCATG
 AAATATGGCAGACAATAAGGAAGTTGACCCAGGTTCTCCTCGGTAACCTTGTTCGGTCCACAAACATTTTTG
 ATATCCCGAAACAGACGGTCTGTGGATTATGGTGAATCAGAGCAACTAAAATCATAACAGGAGACTATGCAT
 TATGGAATCACATCGGATCAAGTGTGACAACCT

Ko_C1_2

TGCCTCGACCACATCTTTGTTAATGTTCCAGCTTATCCTTATCTTCAGACACATTAT---
 GTAATCCTTATCTTCTCTTT-ATCTCCG-----
 GAGTCATTGGAAGTAGGGAATTTTATGTTCCAGCTTCTTGAACCGAGCCCCATAT--CGTCTGTATGGTACTT-
 GAAACAAGGGCTGAAAGATGTTGCCGTTGTTAATGTCTGTATTAGG-AAGTCTT-GAGTAATCA----
 CTAGAGAGTAGCAGCAGCTTGTATGAATGAAACAAATAAATCATGCTGTGGCGCTATTTGGTTTTCAATAAAC
 GGTTTTTCTAAATTACAGGACCAGGGTGCACCGTAAAAATTGGCGTAGTCAAAGAATTTACAGGGTTGCTA
 TAGAGAGTGGGACGTTTAGCCCTCTTCTTGTTTTTATAGATAGATGGCCGCTGCCTTTTGCTTGAATTGTTCT
 CC-TGCC

Ks_C1_1

GTGAGTCTCCCTGATCCTCTTCTCACTATTCTGAAATCTGAACATTCTTAGCCCCATACTATGTA-
 TATAATTTCTAAGTAATAAGATAATGTCTGCAAACACTACAGGTGGGATGTGAATATCAGCAGTGATTATTGCTC
 AGATATCAGCAGTGATTGTTGCTCAGAGAATGTCCAAATCGTATTTTCAGCACTGCGTAGTACCATCTCTGAG
 ACTGGAGAGAAAGCAAGCATATTGCAGGGACGCAATGTAACAAGCCACACGAAAGAGATTGTGAGATGTCTA
 CTTCTTTCATCTTCTTGCT-----
 ATTGTACTCTGATGTATGCGTATGGGTTATCCAGTTCGCTCTCTATAATATATGATGAC----
 GATTCTTTAAATGGTTTATTACAGTGGCTAGAATTGCTCTACGGATTTTTTAGGGAGTTTGAATGGACTAGAGA
 CAAGACGGTGCCAACATTGGATGAATATATGGCAAATGGTTATGTATCATTCCGC

Ks_C1_2

ATTTTTTTTCCAACGTTTAAACCCGAGCTGTGTGCGAGCCCACTATAGTCCAGGGATACACGCAGCTTGGCCCT
 AATTATTTGGAGAAGCTACATTGAAAACCTGTACACGAGACGGCAAATTTCAAATCCTAGTTATTTGGATCTC
 CATTGCGGGTGTCTAACAATCATCACCTCCTCCCATAACCGGACT-
 CGTTTGATCCTTAGAGAAGCTATTCAAATCCTCCAAAGGTATTGGCTTATGAATAATTGCTTTTACGAGATCA
 ACCATCTCCACTTCTGATGTAATCCATCATTGTTGAAGTAAATTTGATGTAGTATTCTGCTTAGATTCCAAAA
 CATGTCCTTGAAGCTCTTGAAGTACACTACCATCCTTCTGCACAACCAATCTTAGCAGTCTCCTCCTCTCA
 TTGTCTATCAGCAATTTACCTCTGCAAAA

P5 / CWI_c5a_PP3_

ACCTTTGGACTAAAAGTTTTGGCTTCAAAGGAGCTGCAGGAATACACGGCCGTCTTCTTTAGAATTTTCAA
 GCCCGAACAATATGTGGTGCTGATGTGCAGTGATCAGAGCAGGTAACATTAATCACCACACTTCGCATGT
 GATCCTAATCTAAAAGTATTTTTGCTTCTTCTTGTAAAGAAATTTCCAGAATTTCTCATCATCCACGTGACG
 TC

P7 / SPS_c1_PP2

GAAACAGCTATGACCATGCAAGCGTAATCCCACACTCAACCTGGAAAGGAACAACCTCATTGATCCTATAGCTT
 TCAGCATTACAACGATGTCTCCCTCAACTACGATACAAAAAGGAATTTACAACACCTATACGGCATAAAGTTAGT
 TGGCCACTTTAGTTAAACCCATTTATCAACAGGTACGAAAGCACAACCGTGCCAGTTTCAAAGCAAGAACGAA
 ATTAACGAATAGAATCATACACTGATGGTTTTTATCAGGGTATCATAACAAGGTAGCAAAAACTTTTGCTTCGA
 GGATCTTGGCAACTCTACGAGCAACGGAGAAACAGGCCAATTATTAACACGATATATAGTTTCATGTCAAAC
 TTTTCTGTGATGTGCGGAAGTGCAAATTTCTAGCTTTTGTGAAACCCCTAGCTTTTCCAACGAGTCGCGCAGCT
 CAGCACTGCTGCAGTCTTCACTTGTCTGAATAAGGTTCCGATTGTCAAATGCCACCACATCGGTAAGTGGGT
 AGCTTCTGTTGGCATGAAGTTGGCTGCTTCTCCACTGCAAACCTCTTTCAGTATCACAGACTTATGCACACC
 ACCAAGTAATCCTTCATAATCAGTGTCCCCACTTTCTCCAACAAAAACCACCCTTTTGACAAATCCATACCCC

ATCGAACTGGCCGTCGTTTTA

P12 / Susy2_c4_b

ACAGCTATGACCATGGTGGGAGCTTGGCAAGATAATCCTCTGCCTTGGACAAAGCAGATTCAAGCCTGGATA
 TCGTGTATCCTATCATTCAACATTAATACCTGCATGGGAACAATGGAATCAAATCAAATTCAGAGGAAAA
 GGTTCAGAAATTCGATAGCTGTGCCACCTAAATGGAGTTGGTATATACAATGCAGTATAAACAGTCAGGTT
 AACCACAAAGCCAGGGAGCAATAAAATCTGTGCAACAGTAGACATAATCAAATAAGAACGTGAAGTTCAGAA
 GAACTTACATGCCCTTTGTGTTTATGTGCACGAAGGAAATCAAGCAATGGCTCCAAAGAATCTTTATTGCGGA
 ACATGATTGAAGAGAGGTGACGATTAAGAAATGAAGCTCCATTACCAATATAAGAAGATCTGGTTGGGCGAG
 GAAATGTTGCATTGAATGGTTCAAATCCAGCTCAAGAACTAAATGGTCTTCAGAGCTAGAACACAGTGAAAA
 GCAATCAAATATTCATAATATCCAGAATTGGTGACCCTGAATCAGAAAACCTATTCAACTCAATTAAGACCAC
 ACTAATTACCGGCCATCCACAAGTTCTTCTTAAGATGGAGATATTCTGAAATTGACAACCTGATCAACACTAA
 GTTCGTATACATTAACGCGTACATACTCCAAACACCAGGCACTGGCCG

P14 / Susy2_c6_b

TGGAACAGCTATGACCATGTAGGAAGGCCACAGGTCATCGCCTCCACAACAGTAAGCCCCAAAGCTTCATA
 AAATGCAGGCTGAAAAGAAGTTCAACAACTTTTTCAATTTAATCCGTTAAGAAAACATCCTGCACGCAAATTG
 CTGATCCACGGGAAACATGATTACTGTACCTGAACAAAAATCCCCCTTTTGTGAGCAATATAACGATATAGCT
 CACCATTCGCGCTCTATTTGTTTGGGAGCTATCCAACGGAATTGGCCATCCAAATTATTTCTTTCATAAG
 CATGTGCATCTTCTATTTCTGAAATTTCTTCCCTGTCACTTGATTTTTTACATCGTTGTAACCTGCCACCAC
 AACAAGTTTGCTAATCCCGTAATTCAGCATTTTTAGCATAGCACTCTACCAGGCCTGTGATGTTTTTAACCC
 GATCCAGCACTGGCCGTCGTTTTACAAAAACAA

P17 / Susy2_c4_6_a

TCAACCTCGAGATCCATTTGCGGAGAATTCCTTCTCAGTTCTAAAAGGGACACGTAAAAATGCTTGTATACTC
 AGTTCCACTAACTCTTCCAACCTCTGGTTGCAAGTAGTGCCTTTTGCATCAGGAATTAACCGAGTCACCTAG
 TAAGGAAGACAAGTCATATCAACAGTAATATTAGATGGTTTCAGATTTCAAAAAGGATATCACTTACAATAAG
 AATCTGGGAGTGACATTCAGCCCTTGCTGCTTTATTTCTTAAAAGCATTTCATTCTCCAAGGCTCGAAGTTGAT
 CCAGTATATAGACAATCTGTATGTCATCCAAGAGGAACAATGATTTCTAAGCACCTTTAGCAGCATAAACTA
 CCATTACATAATAAGGAGAACTTTGGTAAAAAACGATGCCACAGTTAAAGACAAAAAAGAAGCTGAAATCTT
 TAGAAAAAGTACAACCTCTAGTTACAAGTATAAAATTTCTGCCCTCAATTTAAAATGATTTGGACCAAGACAACA
 AACATACCTGGCCACCAGTATCAGGCAAGCCCAAGACATTTGCTTGACCAAAATATCCATGAACAGACAAAAT
 GACAACGTTAAACACCATAGGTATTCTACCAAGAAAAGTTCCAGAGTAGAAGGATCAGGTGCTTGAAGGAT
 GTCAGAAAAGAAGATGCATCATATTCAGAACCCTTGCTGCGGTGTCACCCCAACCCCTCTCAAACCCAGTTC
 TTGCAATCTGCATAGATAACAACCTGTAGGTAACAGCCCTAAAAAAAATTAATCTAGCTAAATGTATATATTCTA
 TTTCCATAACTACAAAAATCATCTTAGGGTTGGTACTTGTGAAGAAACATACGCATATTCAAAGTCCGAATAAA
 G

P23 / Susy1_c435_e

TGAAACAGCTATGACCATGTTGAAATGGAAGGTAATTTGATCTTATTACAAAATGACATTTGAAATGCCCAAT
 TCCTCGCACTACCAAAAAGCAGGAAAAGAAAAGGAAATGGCAAACAACCCAACGAAAATGACCTACAAGAAG
 GACCTAATTGACACTTGAACAGGGACTAAAAAGACACAACCAACCAACAATAAAGAAAAGGCAAGACTGGCAG
 AAGTAACAAAAACACAATCCTCTTTTACACGATCTCAAGCTCTTTTATTTTATTTGCTCCCTCCCGCTGC
 TTCCAACCTTTTCTGCCGCAAACCTTACTGATCAACAGCCAATGGAACAGCTTCAGCCTGGAACCCAAAACA
 GAGAACACGATTGTTAGAAGAAATGCTTTTGAAGAGGAGAGCTGAACAATGCAGATTCATCAACTCAAGAC
 CAGATAACTTACCAACTTGCGATACTTGAGAGCATAAAACATTTCCAGATAACGGCGACTGGCCGTCGTTTTA
 CAAA

P25 / Inv_c1_c

GGGGGGGGGGCAGGCCCCCGATGGGACCAAGTGGTTGGAAACCCTCGGGTCCAAGGTCAATAAAACGGGC
 ATTGCCCTGGTTTATGAAACCAGCGACTTCAAGGGCTACCGGCTCTTGGATGGGGTTTTACACGCGGTTCCC
 CGCACGGGCATGTGGGAGTGCCTGGATTTCTACCCGTTTCCACCACGGGGGATAATGGGCTGGATACATC
 GGCTAATGGGCCTGGCACTAAGCACGTCCTGAAGGCAAGCTTGGACGACAACAAGCACGATTACTACGCAC
 TGGGCACCTATGATCAAAAAACAATAAGTGGACCCTGACGATCCGGAACCTGGATGTGGGTATCGGGTTC
 GTCTGGACTATGGCAAGTATTACGCGTCAAAGACATTTTACAACCAAAAACAAGAAGGAGGATTTTGTGGG
 GTTGGATTGAAAAACTGATACGAAGCTGCTGACCTCATGAAAGGATGGGCATCGGTTCCAGGTATCATTATTA
 TTATTAGTATTATTGTTCTCCTCATTACAAGTGCATAATGCATTGCCTTGACTTGAATTC

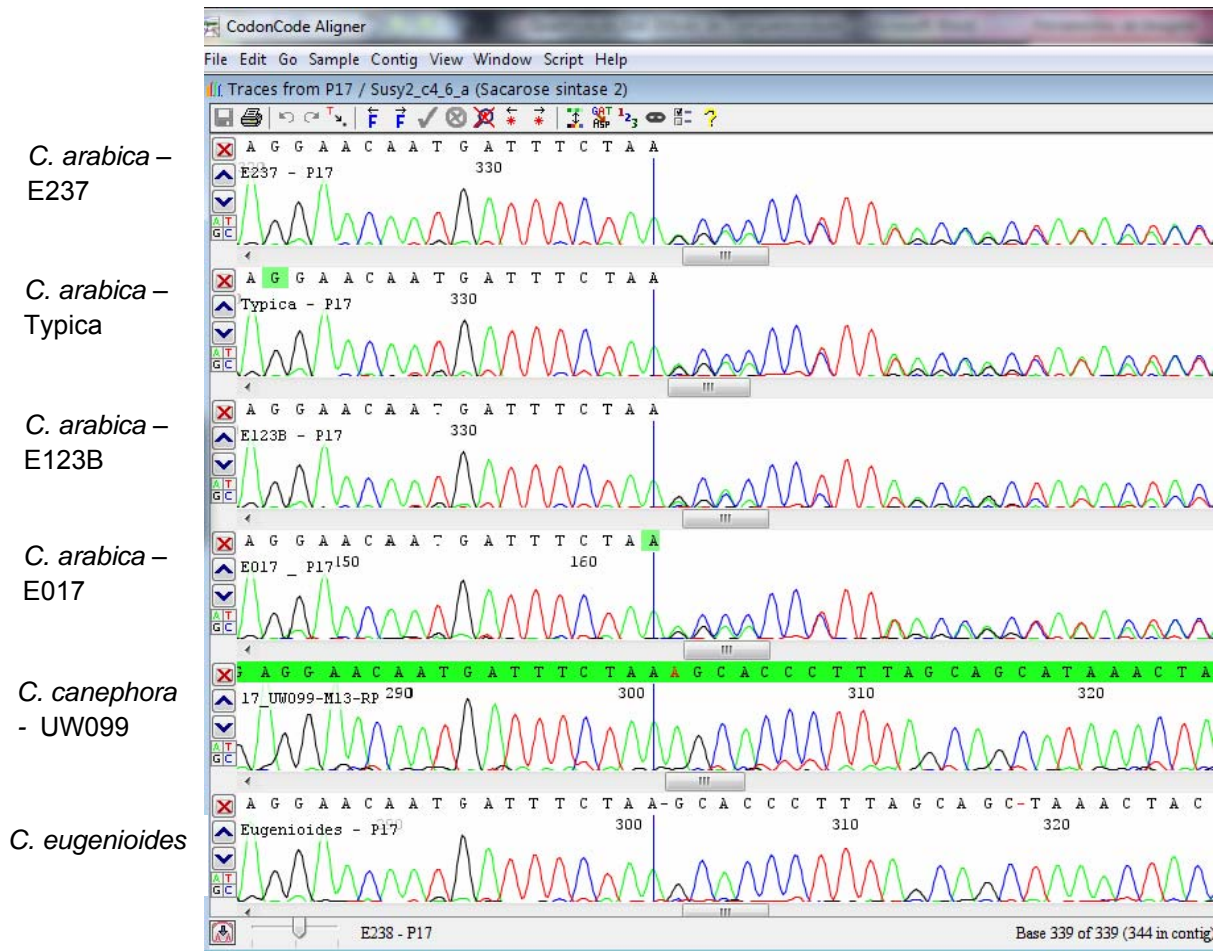
trnL_trnF

TTATTTGAACTGGTGACACGAGGATTTTTCAGTCTCTGCTCTACCATCTGAGCTATCCCGACCAGTCCCAATG

TAGCATCCTCATTTCATGAGAAAAGGGGGGCTATGTCAATTAAGAAGGAAAGGGGATTACAAAGTTTATC
TAGGTACTAGAATTTCTGGATCTTAAAAGAGAGTACTTTGTAAGTTTCAGTATGAATAGTTATATCGATTGTAA
ACCATTCAAATGGGTATTTCTTGCTCAAAGATGTTCAATTTATATGTATATCATAAATAACACAAGATTTGTGATA
AGAAAAGAGCTTTCCGTCCAGATCGATTTCTAAGAGAATAGAGTAAATGAATAAGGTTTTTTGAACCGCTAA
CGAAAGGGGGATAGGATAAATAG—GGGGGATCAAATATGTTTTTTGGGGATAGAGGGACTTGAACCAA

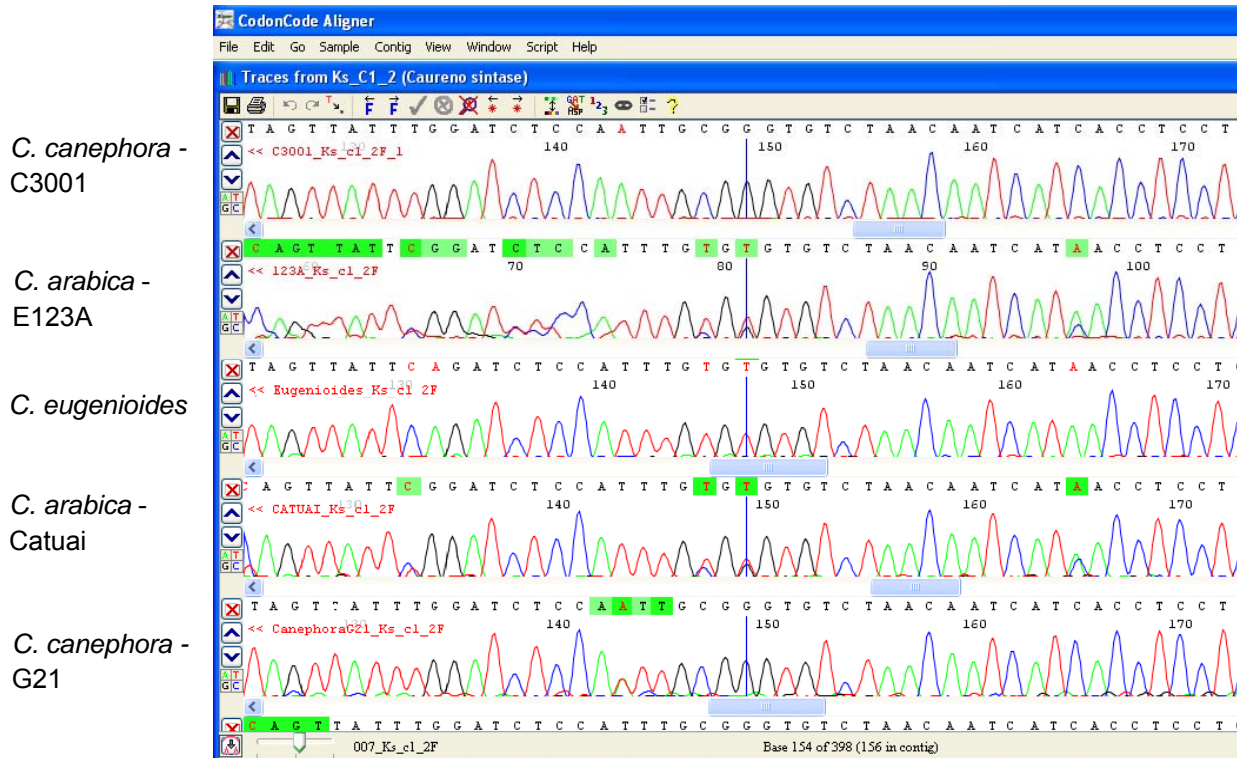
ANEXO B

Exemplo de sequência interrompida de *C. arabica* devido à presença de INDELS heterozigotos que impossibilitaram a leitura do sequenciamento. As quatro primeiras sequências são de *C. arabica*, seguidas de uma sequência de *C. canephora* e por último *C. eugenioides*. Na posição assinalada (344 do *Contig*) existe um INDEL de um nucleotídeo entre *C. canephora* e *C. eugenioides*, e *C. arabica* possui os dois alelos causando uma sobreposição de fluorescência emitida, que impossibilita a leitura do sequenciamento a partir deste INDEL.



ANEXO C

Exemplo de alinhamento no programa *Codon Code Aligner* evidenciando os alelos heterozigotos encontrados em *C. arabica*. Cada linha corresponde a sequência de um genótipo. O SNP assinalado apresenta dois alelos (G e T); observa-se que os genótipos de *C. canephora* e *C. eugenioides* apresentam apenas um dos alelos e os genótipos de *C. arabica* apresentam ambos os alelos.



ANEXO D

Tabela de caracterização dos polimorfismos presentes no gene da Sacarose fosfato sintase (SPS). Cada linha corresponde a um SNP. As colunas contêm respectivamente: nome do primer, número do SNP no Contig, nome do SNP, posição do SNP no Contig, sequência anterior e posterior ao polimorfismo, alelos presentes e tipo do polimorfismo (SNP, INDEL ou SSR).

Primer	SNP	SNP_name	contig posicion	sequence before	sequence after	allele1	allele2	allele3	Tipo
P7 (SPS_c1_PP2)	1	SPS_P7_1	135	AAAAGGAATTTACAACACCTATAC	GGCATAACTTAGTTGGCCACTTTA	A	G		SNP
P7 (SPS_c1_PP2)	2	SPS_P7_2	145	AATTTACAACACCTATACGGCATAACT	GTTCGCCACTTTAGTTTAA	A	C		SNP
P7 (SPS_c1_PP2)	3	SPS_P7_3	164	CTTAGTTGGCCACTTTAGTTTA	CCCATTATCAACAGGTCGAAAGCAC	A	G		SNP
P7 (SPS_c1_PP2)	4	SPS_P7_4	212	CAACC GTGCAGTTTCAAAGC	AGAACGAAATTAACGAATAGAATCATA	A	G		SNP
P7 (SPS_c1_PP2)	5	SPS_P7_5	218	CCGTGCCAGTTTCAAAGCAAAGAAC	AAATTAACGAATAGAATCATA	C	G		SNP
P7 (SPS_c1_PP2)	6	SPS_P7_6	249	ATAGAATCATAACACGTATGG	TTTTATCAGGGTAT	C	T		SNP
P7 (SPS_c1_PP2)	7	SPS_P7_7	261	TACACTGATGGTTTTATCAGGG	ATCATAACAAGGTAGC	C	T		SNP
P7 (SPS_c1_PP2)	8	SPS_P7_8	264	GTTTTATCAGGGTAT	ATAACAAGGTAGCAAACCTTTT	C	T		SNP
P7 (SPS_c1_PP2)	9	SPS_P7_9	347-354	CAGGCCAATTATTAACAAGA	GTTCATGTCAAACCTTTCTGT	(AT)3	(AT)4		SSR
P7 (SPS_c1_PP2)	10	SPS_P7_10	377	TCATGTCAAACCTTTCTGTG	TGTGCGGAAGTGCAAATTTCTAGCTTTT	A	C		SNP
P7 (SPS_c1_PP2)	11	SPS_P7_11	388	AAACTTTTCTGTGATGTCCGGAAG	GCAAATTTCTAGCTTTTGAG	T	G		SNP
P7 (SPS_c1_PP2)	12	SPS_P7_12	424	TAGCTTTTGAGAACCCCTAGCTTTT	AACGAGTCGCGCAGCTCAGCAC	C	T		SNP
P7 (SPS_c1_PP2)	13	SPS_P7_13	427	AGAACCCCTAGCTTTTCCAA	GAGTCGCGCAGCTCAGCAC	C	T		SNP
P7 (SPS_c1_PP2)	14	SPS_P7_14	430	TGAGAACCCCTAGCTTTTCCAA	TGCGCAGCTCAGCACTGCTG	C	G		SNP
P7 (SPS_c1_PP2)	15	SPS_P7_15	439	AGCTTTTCCAAAGGTCGCGCAG	TCAGCACTGCTGCACTTTCACT	C	T		SNP
P7 (SPS_c1_PP2)	16	SPS_P7_16	442	TTTCTAATGAGTCGCGCAGCTC	GCACGCTGCTGCACTTTCa	A	G		SNP
P7 (SPS_c1_PP2)	17	SPS_P7_17	445	TTTCAACGAGTCGCGCAGCTCAGC	CTGCTGCACTTTCACTTTGTC	A	T		SNP
P7 (SPS_c1_PP2)	18	SPS_P7_18	447	CAACGAGTCGCGCAGCTCAGCAC	GCTGCACTTTCACTTTGTCTGAA	A	T		SNP
P7 (SPS_c1_PP2)	19	SPS_P7_19	448	TCCAACGAGTCGCGCAGCTCAGCACT	CTGCAGCTTTCACT	G	T		SNP
P7 (SPS_c1_PP2)	20	SPS_P7_20	463	AGCTCAGCACGCTGCACTTTCACT	GTCTGAATAAGGTTCCGATTGTCA	C	T		SNP
P7 (SPS_c1_PP2)	21	SPS_P7_21	490	ATAAGGTTCCGATTGTCAA	GCCACCACATCGGTAA	T	C		SNP
P7 (SPS_c1_PP2)	22	SPS_P7_22	508	TGCCACCACATCGGTAAAG	GGTAGCTTTCTGTTGCATGAAGTTGG	C	T		SNP

ANEXO E

Tabela de caracterização dos polimorfismos presentes em um fragmento do gene da sacarose sintase 2 (P17/Susy2_c4_6_a). Cada linha corresponde a um genótipo e cada coluna numerada corresponde a um polimorfismo. Nas colunas 5 e 12, *C. canephora* apresenta um alelo (amarelo) e *C. eugenioides* outro (laranja), enquanto *C. arabica* possui ambos (vermelho), estes são exemplos de polimorfismos nos subgenomas de *C. arabica* que refletem diferenças entre os ancestrais (Cc e Ce - 40% dos SNPs de *C. arabica*). Na coluna 6 observa-se que os SNPs de *C. arabica* correspondem a diferenças entre os genótipos de *C. canephora* (Cc 22% dos SNPs de *C. arabica*) e nas colunas 3 e 7 os polimorfismos de *C. arabica* podem corresponder as diferenças entre os subgenomas ancestrais e/ou aos polimorfismos dentro de *C. canephora* (Ce e Cc/Cc – 20% dos SNPs de *C. arabica*).

Genótipo	Especie	Grupo genético	1	2	3	4	5	6	7	8	9	10	11	12	13
E-007	<i>C. arabica</i>		G	T	G/T	A	C/A	C/T	G/A	G/A	G	A	G	C/T	2A/3A
E-017	<i>C. arabica</i>	Ethiopian3	G	NA	NA	NA	NA	NA	G/A	G/A	G	A	G	C/T	2A/3A
E-123A	<i>C. arabica</i>	Ethiopian4	NA	T	G/T	A	C/A	C/T	G/A	G/A	G	A	G	C/T	2A/3A
E-123B	<i>C. arabica</i>	Ethiopian4	G	T	G/T	A	C/A	C/T	G/A	G/A	G	A	G	C/T	2A/3A
E-237	<i>C. arabica</i>	Ethiopian2	G	T	G/T	A	C/A	C/T	G/A	G/A	G	A	G	C/T	2A/3A
E-238	<i>C. arabica</i>	Ethiopian4	G	T	G/T	A	C/A	C/T	G/A	G/A	G	A	G	C/T	2A/3A
E-464	<i>C. arabica</i>	Ethiopian1	G	T	G/T	A	C/A	C/T	G/A	G/A	G	A	G	C/T	2A/3A
E-516	<i>C. arabica</i>	Ethiopian1	NA	NA	T	NA	NA	NA	G/A	G/A	G	A	G	C/T	2A/3A
Bourbon	<i>C. arabica</i>	Bourbon	NA	NA	NA	NA	NA	NA	G/A	G/A	G	A	G	C/T	2A/3A
Typica	<i>C. arabica</i>	Typica	G	T	G/T	A	C/A	C/T	G/A	G/A	G	A	G	C/T	2A/3A
Catuai	<i>C. arabica</i>	Typica/Bourbon	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mundo Novo	<i>C. arabica</i>	Typica/Bourbon	G	T	G/T	A	C/A	C/T	G/A	G/A	G	A	G	C/T	2A/3A
C1007	<i>C. canephora</i>	Congolese_B	G	G/T	G	A/C	C	C/T	G	G	G	A	G	C	3A
C2011	<i>C. canephora</i>		G	T	G	A	C	T	G	G	G	A	G	C	3A
C3001	<i>C. canephora</i>	Congolese_SG1	G	G/T	G	A/C	C	T	G	G/T	C/G	A	G	C	3A
G2020	<i>C. canephora</i>	Guinea	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G21	<i>C. canephora</i>	Congolese_SG1	NA	T	G/T	A	C	NA	G/A	G	G	A	G/A	C	3A
UW002	<i>C. canephora</i>	Uganda_Selvagem	G	G	G	C	C	T	G	T	G	A	G	C	3A
UW099	<i>C. canephora</i>	Uganda_Selvagem	G	T	G	A	C	C	G	G	G	A	G	C	3A
<i>C. eugenioides</i>	<i>C. eugenioides</i>		T	T	G/T	A	A	T	G/A	A	G	A	G	T	2A