



UNIVERSIDADE
ESTADUAL DE LONDRINA

HENRIQUE MARQUES DE OLIVEIRA

**VALIDAÇÃO DE UMA COLEÇÃO DE *COFFEA ARABICA*
COM MARCADORES MOLECULARES SNPs**

Londrina
2014

HENRIQUE MARQUES DE OLIVEIRA

**VALIDAÇÃO DE UMA COLEÇÃO DE *COFFEA ARABICA*
COM MARCADORES MOLECULARES SNPs**

Dissertação apresentada ao Programa de Pós-Graduação, em Biotecnologia da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

Orientador: Dr. Luiz Filipe Protasio Pereira



Londrina
2014

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

Dados Internacionais de Catalogação-na-Publicação (CIP)

O48v Oliveira, Henrique Marques de.

Validação de uma coleção de *Coffea arabica* com marcadores moleculares SNPs / Henrique Marques de Oliveira. – Londrina, 2014.
51 f. : il.

Orientador: Luiz Filipe Protasio Pereira.

Dissertação (Mestrado em Biotecnologia) – Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Biotecnologia, 2014.

Inclui bibliografia.

1. Biotecnologia agrícola – Teses. 2. Polimorfismo (Genética) – Teses. 3. Café – Biotecnologia – Teses. 4. Genética vegetal – Teses. 5. Marcadores biológicos – Teses. 6. *Coffea arabica* – Teses. I. Pereira, Luiz Filipe Protasio. II. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Biotecnologia. III. Título.

CDU 663.8/9

HENRIQUE MARQUES DE OLIVEIRA

**VALIDAÇÃO DE UMA COLEÇÃO DE *COFFEA ARABICA*
COM MARCADORES MOLECULARES SNPs**

Dissertação apresentada ao Programa de Pós-Graduação, em Biotecnologia da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

BANCA EXAMINADORA

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

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

Prof. Dr. Douglas Silva Domingues
Instituto Agrônômico do Paraná – IAPAR

Londrina 28 de maio de 2014.

OLIVEIRA, Henrique Marques. **Validação de uma coleção de *Coffea arabica* com marcadores moleculares SNPs**. 2014. 51 f. Dissertação (Mestrado em biotecnologia – Universidade Estadual de Londrina, Londrina. 2014.

RESUMO

Como constituem as formas mais abundantes de variações do genoma, os polimorfismos de nucleotídeo único (SNPs) podem auxiliar na caracterização de recursos genéticos, inclusive de espécies que possuem variabilidade genética reduzida como *Coffea arabica*. Com aumento da eficiência e acessibilidade das tecnologias de sequenciamento, uma grande quantidade de dados está sendo produzida. Partindo deste princípio o presente trabalho visa otimizar a técnica de validação de SNPs através do método Snapshot. Foram selecionados os SNPs, 4, 5, 6, 15 e 17, observados por YANAGUI, 2010 que foram testados em 11 genótipos. Dos 5 (cinco) SNPs testados apenas o SNP17 não apresentou diferenças alélicas intraespecíficas, enquanto o SNP6 apresentou 3 (três) perfis polimórficos e o restante apresentou 2(dois) perfis polimórficos. O SNP5 foi selecionado para ser testado em 143 genótipos de uma coleção da Etiópia, apresentando 3 (três) perfis polimórficos diferentes, evidenciando a existência de variabilidade intraespecífica.

Palavras-chave: SNP; *Coffea arabica*; Snapshot; Polimorfismo.

OLIVEIRA, Henrique Marques. **Validation of a *Coffea arabica* Collection with SNPs as molecular markers**. 2014. 51 p. Dissertation (Master's Degree in Biotechnology) – Universidade Estadual de Londrina, Londrina. 2014.

ABSTRACT

Single Nucleotide Polymorphisms (SNPs) are the most abundant form of genomic variations and can be used in the characterization of genetic resources, even in species with reduced genetic variability as *Coffea arabica*. With the increasing efficiency and accessibility of sequencing technologies, a large amount of data to identify SNPs is being produced. The aim of this research was to validate previous observed SNPs in *Coffea arabica* and to optimize the Snapshot method for SNP genotyping. Were selected the SNPs 4, 5, 6, 15 and 17, observed by YANAGUI, 2010. They were tested in 11 *C. arabica* genotypes. Considering the five (5) SNPs tested, only SNP17 allele showed no intraspecific differences, while SNP6 presented three (3) polymorphic and the remaining SNPs showed two (2) polymorphic profiles. The SNP5 was selected to be tested on a collection of 143 *C. arabica* Ethiopian genotypes, and presented three (3) different polymorphic profiles, evidencing the existence of intraspecific variability.

Key Words: SNPs; *Coffea arabica*; Snapshot; Polymorphism.

LISTA DE FIGURAS

Figura 1.	Produção de café por estado (CONAB, 2013).	4
Figura 2.	Ilustração científica de <i>Coffea arabica</i> (KOHLE, 1887).....	6
Figura 3.	Representação esquemática da origem das cultivares de cafeeiro e redução da variabilidade genética (ANTHONY et al., 2002)	9
Figura 4.	Distribuição geográfica das espécies de café coletadas (ANTHONY; DUSSERT; DULLOO, 2007).....	10
Figura 5.	Alinhamento de sequências correspondentes ao mesmo segmento genômico de três indivíduos, mostrando um SNP não sinônimo A/G, que causa a substituição de glutamina para arginina na sequência proteica (GUIMARÃES; COSTA, 2002).....	13
Figura 6.	Representação das categorias de variação de sequência dentro e entre subgenomas de uma planta aloploide e genomas diploides correspondentes (KAUR; FRANCKI; FORSTER, 2011).....	15
Figura 7.	Fluxograma de trabalho para análise de SNPs usando kit SNaPshot™ Multiplex.....	22
Figura 8.	Método SNaPshot Multiplex. Introdução de dinoculeotídeos fluorescentes baseados nos SNPs dos fragmentos de DNA amplificados.	25
Figura 9.	Determinação de padrão através do software GeneScan-120, mostrando tamanho do oligo seguido nucleotídeo complementar e abaixo tamanho relativo após.....	28
Figura 10.	Exemplo de um eletroferograma gerado pelo GeneMapper	30
Figura 11.	Eletroferograma do genótipo 365\652 analisado no software Peak Scanner Software v1.0 para 5 SNPs em multiplex.....	31
Figura 12.	Eletroferograma do genótipo 365\652 analisado no software Peak Scanner Software v1. As reações foram feitas individualmente e a corrida em multiplex.	32
Figura 13.	Resultado do teste de otimização do Volume e concentração dos Oligos SNaPshot e parâmetros modificados.....	33
Figura 14.	Eletroferograma da purificação após reação SNaPshot, em conjunto e individualmente.....	34

Figura 15. Eletroferograma exibido no software Peak scanner v0.1. dos oligos do grupo A.	37
Figura 16. Perfil alélico da população de 143 genótipos da população da Etiópia para o SNP5. Não foi possível analisar os SNPs em 14% das amostras (X).....	39

LISTA DE TABELAS

Tabela 1. Relação dos 25 SNPs encontrados em <i>C. arabica</i> Colunas 1 a 25 representam os SNPs. Linha representa o genótipo analisado. NA = informação não disponível (YANAGUI, 2010).	20
Tabela 2. Oligos controle ABI PRISM® SNaPshot™ Multiplex.....	28
Tabela 3. Preparação das reações do controle positivo e negativo	29
Tabela 4. Desenho dos Oligos SNaPShot.....	36
Tabela 5. Polimorfismos encontrados no grupo de genótipos teste.	38

SUMÁRIO

1	INTRODUÇÃO	1
2	REVISÃO DA LITERATURA	2
2.1	CAFÉ NO BRASIL.....	2
2.2	TAXONOMIA E GENÉTICA DE <i>COFFEA ARABICA</i>	5
2.3	ORIGEM E VARIABILIDADE GENÉTICA	7
2.4	MARCADORES MOLECULARES E SNPs EM CAFÉ	11
3	OBJETIVOS	18
4	MATERIAS E MÉTODOS	18
4.1	MATERIAL VEGETAL.....	18
4.2	ISOLAMENTO DO DNA GENÔMICO	18
4.3	IDENTIFICAÇÃO DE SNPs E DESENHO DO OLIGOS SNAPSHOT	19
4.4	VALIDAÇÃO DOS SNPs	22
4.5	LEITURA E ANÁLISE DA AMOSTRA	27
4.6	REAÇÕES CONTROLE	27
4.7	ANÁLISE DOS DADOS COM <i>GeneMapper</i>	29
5	RESULTADOS	30
5.1	OTIMIZAÇÃO DO PROTOCOLO.....	30
5.2	OLIGO SNAPSHOT	35
5.3	VALIDAÇÃO NO GRUPO TESTE	37
6.4	VALIDAÇÃO DO SNP5 NA POPULAÇÃO DA ETIÓPIA.....	38
7	CONCLUSÃO	39
8	REFERÊNCIAS	40

1-INTRODUÇÃO

O Brasil é o maior produtor mundial de café e o segundo maior consumidor da bebida. Os principais grãos comercialmente produzidos são das espécies, *Coffea arabica* e *Coffea canephora* (MAPA, 2010). A cafeicultura se fixou, inicialmente, no Sudeste e depois se expandiu para o Paraná e Bahia. Atualmente, é produzido em 14 estados, com área plantada de 2,3 milhões de hectares, o equivalente a cerca de seis bilhões de pés (MAPA, 2010). Em 2012, as exportações de café renderam ao Brasil US\$ 7,841 bilhões, resultado recorde. O número representa crescimento de 5% em relação no ano anterior (EMBRAPA, 2012). Para a colheita de 2014 é esperada a produção entre 46,53 e 50,15 milhões de sacas. Se confirmada será a maior safra de ciclo de baixa bionalidade já produzida no país (CONAB, 2014). Apesar da produção de *Coffea canephora* ter aumentado nos últimos anos, *Coffea arabica* ainda representa 74,9% da produção do País (ABIC, 2013).

Originário da Etiópia, o gênero *Coffea* pertence à família Rubiaceae e é formado por cerca de 100 espécies (DAVIS et al., 2006). A maioria das espécies conhecidas do gênero são diploides ($2n = 2x = 22$ cromossomos) e predominantemente alógamas. *C. arabica*, no entanto, é classificada como alotetraploide ($2n = 4x = 44$ cromossomos), oriunda dos progenitores diploides *Coffea eugenioides* e *Coffea canephora* (LASHERMES et al., 1999) com sistema reprodutivo predominantemente autógamo.

Cultivares de *C.arabica* derivam, na grande maioria, das variedades Típica e Bourbon, originalmente introduzidas no Brasil em 1727 e

1859, respectivamente (ANTHONY et al., 2002). Com isso, a base genética dos cultivares comerciais é bastante limitada e a diversidade molecular observada entre eles é baixa (ANTHONY et al., 2002). Como constituem as formas mais abundantes de variações do genoma, os polimorfismos de nucleotídeo único (SNPs) podem auxiliar na caracterização de recursos genéticos, inclusive de espécies que possuem variabilidade genética reduzida e surgem a partir de erros de incorporação de bases durante a replicação do DNA; em outros casos são causadas por lesões no DNA por agentes ambientais (GUIMARÃES; COSTA, 2002).

Os SNPs estão, cada vez mais, se tornando o marcador de escolha para uma ampla gama de aplicações incluindo mapeamento genético, seleção assistida de marcadores e análises de diversidade (JEHAN; LAKHANPAUL, 2006). A partir da análise de diversidade genética, é possível inferir relações filogenéticas entre linhagens e variedades ou unidades taxonômicas mais abrangentes (JEHAN; LAKHANPAUL, 2006).

2- REVISÃO DA LITERATURA

2.1- CAFÉ NO BRASIL

No Brasil, as primeiras sementes e mudas de café chegaram em Belém do Pará na década de 1720, trazidas das Guianas pelo Sargento-Mor Francisco de Melo Palheta, após missão oficial. Elas deram início ao cultivo do café no Pará, e posteriormente foi difundido para os estados vizinhos do Amazonas e Maranhão (SIQUEIRA, 2005).

Na década de 1770, o café chegou ao Rio de Janeiro, onde foi cultivado nos arredores da cidade, dando início ao ciclo agrícola de maior impacto econômico no País desde o descobrimento (SIQUEIRA, 2005).

Em meados da década de 1880, São Paulo e Minas Gerais ultrapassaram o Rio de Janeiro, tornando-se os maiores produtores de café do País. Em 1928, o Espírito Santo assumiu a terceira posição, deslocando o Rio para a quarta colocação (SIQUEIRA, 2005). No Paraná a cultura do café chegou ao início do século XX. Seu crescimento foi lento até 1945, em consequência das duas guerras mundiais e também da quebra da bolsa de Nova York, em 1929, quando, em função dos grandes estoques, a forte crise econômica mundial e os baixos preços do produto, milhões de sacas tiveram que ser queimadas.

Na década de 1950, o Paraná assumiu a liderança por vários anos até meados da década de 1980. A partir daí, Minas Gerais passou a liderar, tornando-se o maior produtor nacional de café (SIQUEIRA, 2005).

Segundo a CONAB (2012), Minas Gerais continua como maior produtor e com 55 % do café produzido, seguido do Espírito Santo (25%), São Paulo (8 %). O Paraná ocupa a quinta posição com produção de 3% do café produzido no país como apresentado na figura 1.

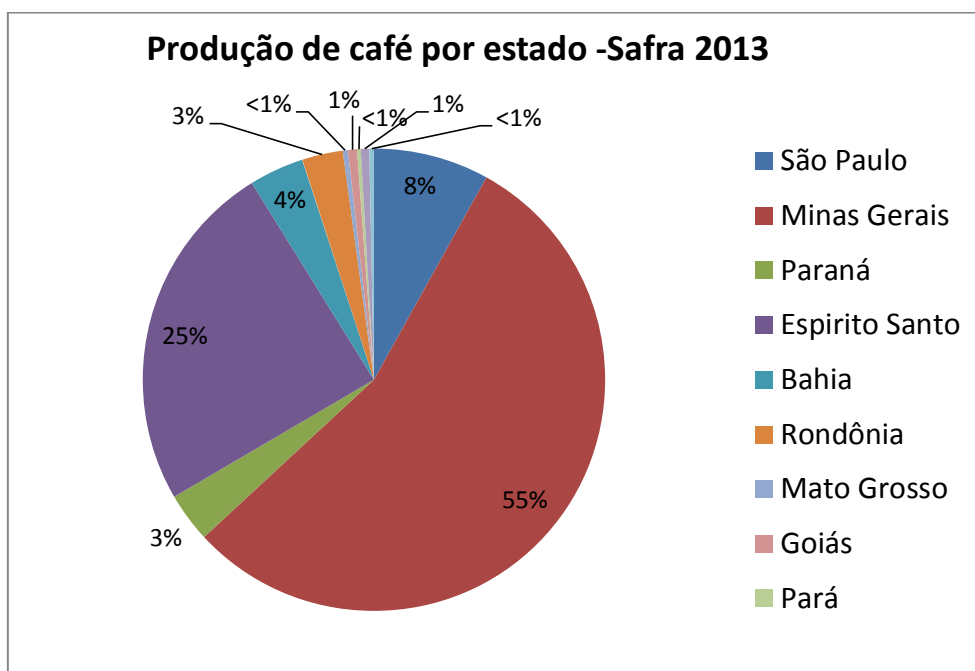


Figura 1. Produção de café por estado (CONAB, 2013).

Atualmente, o Brasil é o maior produtor mundial de café e o segundo maior consumidor da bebida. Os principais grãos são das espécies, *Coffea arabica* e *Coffea canephora* (MAPA, 2010). O café é produzido em 14 estados, com área plantada de 2,3 milhões de hectares, o equivalente a cerca de seis bilhões de pés (MAPA, 2010). (MAPA, 2010).

Em 2012, as exportações de café renderam ao Brasil US\$ 7,841 bilhões, resultado recorde. O número representa crescimento de 5% em relação no ano anterior (CONAB, 2012). Para a colheita de 2013/2014 é esperada a produção de 48,59 milhões de sacas. Se confirmada será a maior safra de ciclo de baixa bienalidade já produzida no país (CONAB, 2013). Apesar da produção de *Coffea canephora* ter aumentado nos últimos anos, *Coffea arabica* ainda representa 74,9% da produção do País (ABIC, 2013).

2.2- TAXONOMIA E GENÉTICA DE *Coffea arabica*

A família Rubiaceae é a maior família de plantas lenhosas nos trópicos úmidos, e é a quarta maior família de plantas com flores. Contém aproximadamente 13.000 spp. em 600 gêneros. Dentre estes, o gênero *Coffea* é o mais economicamente valioso (DAVIS et al., 2009). A primeira descrição botânica de uma árvore de café foi realizada sob o nome *Jasminum arabicanum*, feita em 1713 Antoine de Jussieu, que estudou uma única planta proveniente do jardim botânico de Amsterdã. No entanto, Linnaeus (1737) classificou-a como um gênero separado *Coffea* com somente uma espécie conhecida na época, *Coffea arabica* (CHARRIER; BERTHAUD, 1985). Muitas espécies novas de *Coffea* foram descobertas durante a exploração das florestas tropicais da África desde a segunda metade do século XIX (CHARRIER; BERTHAUD, 1985).

O gênero *Coffea* possui cerca de 100 espécies e é dividido em dois subgeneros, *Coffea* subgen. *Coffea* e *Coffea* subgen. *Baracoffea* (DAVIS et al., 2009). A maioria das espécies de *Coffea* pertence ao subgen. *Coffea*, incluindo aquelas mais utilizadas para a produção da bebida, *C. arabica* e *C. canephora*.



Figura 2. Ilustração científica de Coffea arabica (KOHLE, 1887).

Coffea arabica é a única espécie de seu gênero, que é auto-compatível. É uma espécie poliploide natural com $2n = 4x = 44$ cromossomos e considerada um alopoliploide segmentar porque apresenta uma herança dissômica e comportamento meiótico regular. Todas as outras espécies do gênero são diplóides e auto-incompatíveis com $2n = 2x = 22$ (PINTO-MAGLIO, 2006).

Alopoliploides são originados pela duplicação de genomas diferentes, normalmente após um evento de hibridação. São chamados segmentares aqueles originados pela duplicação dos genomas de espécies

próximas o suficiente para apresentarem certa homeologia cromossômica (SCHIFINO-WITTMANN, 2004).

O estudo de Lashermes et al. (1999) utilizando GISH e marcadores RFLP, determinou a origem anfidiplóide de *C. arábica*. Os resultados mostraram que *C. arabica* provavelmente deriva de hibridação natural entre os diploides *C. eugenioides* e *C. canephora*.

2.3- ORIGEM E VARIABILIDADE GENÉTICA

O gênero *Coffea* é caracterizado por um grande número de espécies cuja diferenciação ocorreu há relativamente pouco tempo, cerca de 5 a 25 milhões de anos (ANTHONY et al., 2002). A diversidade de espécies de café e a diversidade genética dentro delas foi o resultado de um processo de especiação rápida e radiação adaptativa (LASHERMES et al., 1997).

Coffea é um grupo endêmico dos Trópicos Africanos; mais de 100 espécies selvagens são encontrados na região Afrotropical, incluindo as ilhas Comores e Mascarenhas (ANTHONY; DUSSERT; DULLOO, 2007). Essa região, juntamente com os campos que produzem café antigo e tradicional da Etiópia, representa a última fonte de diversidade genética do café, em que o futuro do melhoramento do café depende (ANTHONY; DUSSERT; DULLOO, 2007).

Algumas autoridades acreditam que o primeiro cultivo de café foi feito no Iêmen e remonta a 575 AD, quando a invasão persa pôs fim ao reinado de Caleb Negus na Etiópia. Certamente a descoberta da bebida resultou no cultivo da planta na Etiópia e na Arábia, mas seu progresso foi lento

até os séculos XV e XVI, quando seu cultivo passou a ser exercido intensamente no Iêmen, península arábica (UKERS, 2009).

A história do cultivo do café é caracterizada pelas reduções sucessivas da diversidade dentro das duas sub-populações de *Coffea arabica* selvagem originárias na Etiópia e introduzidas no Iêmen. A redução ocorreu pela primeira vez quando o café foi introduzido no Iêmen, entre 1.500 e 300 anos atrás. Esta foi, sem dúvida, a mais severa em termos de perda de diversidade mesmo com as prováveis introduções sucessivas da Etiópia para o Iêmen (ANTHONY et al. 2002).

Novas reduções na diversidade foram produzidas durante a introdução de café em Java, em seguida, Amsterdã e Ilha de Reunião, no início do século 18. Estas introduções foram realizadas com um número limitado de sementes e plantas, uma única planta para Amsterdã e vários descendentes para Ilha de Reunião. Para a origem do café cultivado na Ilha de Reunião, Haarer (1956) apud ANTHONY et al. 2002) afirmou que a primeira introdução em 1708 foi composta de sementes e 60 pequenas plantas, as quais nenhuma sobreviveu. Apenas duas plantas da segunda introdução em 1715 sobreviveram. Estas informações demonstram as difíceis condições de cultivo que foram enfrentadas pelas primeiras plantas de café introduzidas.

Muitas das falhas iniciais para introduzir a cultura do cafeeiro em outras terras também podem ser atribuídas ao fato, descoberto mais tarde, que as sementes logo perdiam o seu poder germinativo (UKERS,1999). Os árabes com intuito de proteger o seu produto lucrativo, e por um tempo com sucesso impediram a sua propagação a outros países, não permitindo qualquer um dos frutos de café deixassem o país, ao menos que, tivessem sido

previamente mergulhados em água fervente ou secos, de modo a destruir seu poder de germinação (UKERS,1999).

Esta seleção intensa, sem dúvida, contribuiu para a redução da diversidade genética das plantas conforme o introduzido na figura 3.

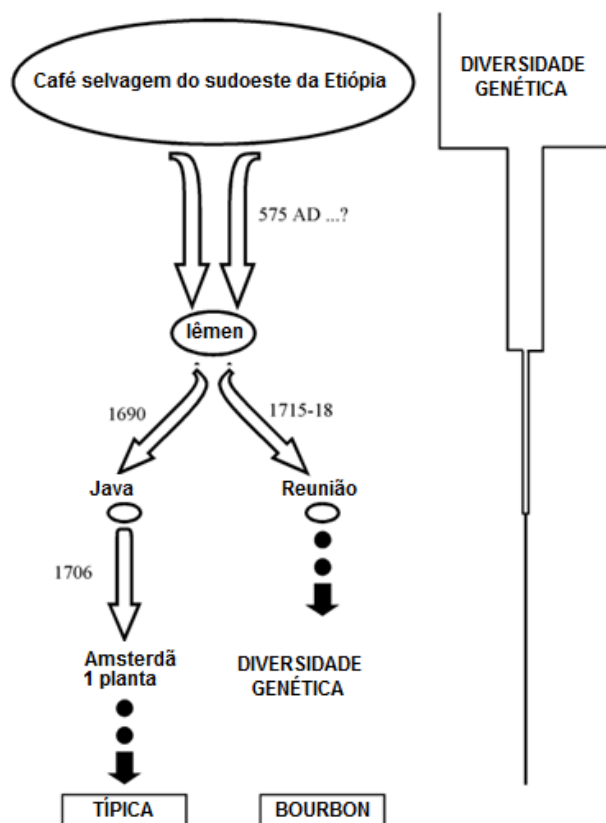


Figura 3. Representação esquemática da origem das cultivares de café e redução da variabilidade genética (ANTHONY et al., 2002)

A cultura de café, em seguida, espalhou-se rapidamente para o continente americano e Indonésia a partir de sementes produzidas por autofecundação de algumas árvores de café sobreviventes (ANTHONY et al., 2002). Considerando a importância socioeconômica do café e o aumento do desmatamento nas áreas nativas da planta, grandes expedições foram organizadas por diferentes grupos de pesquisa para coleta de material e pesquisa de espécies de *Coffea* nas prováveis regiões de origem (Figura 4) (ANTHONY; DUSSERT; DULLOO, 2007). A FAO (*Food and Agriculture Organization of United Nations*) foi a pioneira, durante o período de 1964 a

1965, amostras de sementes de diferentes locais da Etiópia, que foram distribuídas a centros de pesquisa de diversos países (ANTHONY; DUSSERT; DULLOO, 2007). O Instituto Agrônomo de Campinas (IAC) teve acesso a esta coleção e uma parte foi repassada para o Instituto Agrônomo do Paraná (IAPAR).

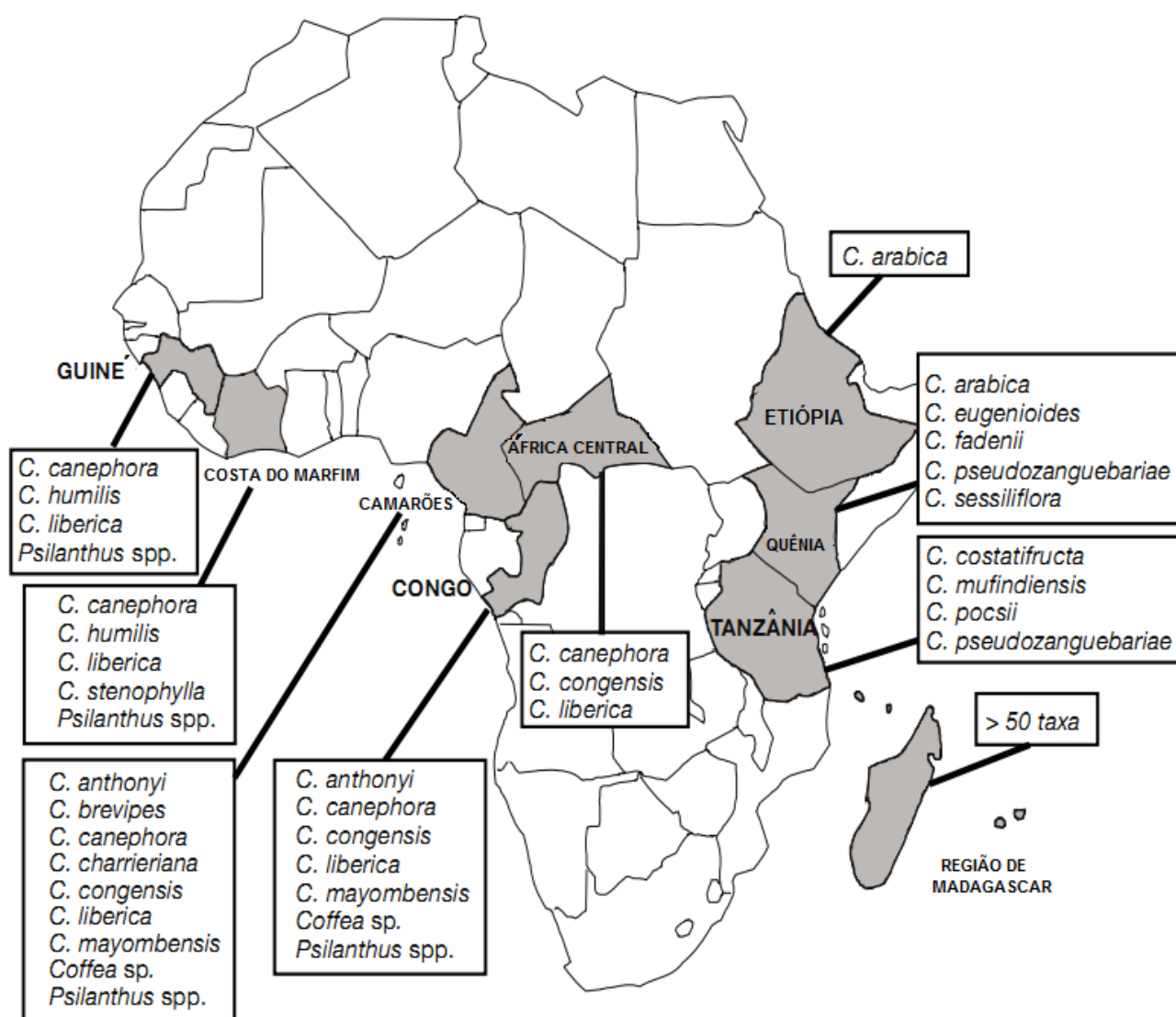


Figura 4. . Distribuição geográfica das espécies de café coletadas (ANTHONY; DUSSERT; DULLOO, 2007).

2.4- MARCADORES MOLECULARES E SNPs em CAFÉ

Marcadores moleculares além de aumentarem a eficiência no melhoramento genético podem ser utilizadas, na certificação de origem e qualidade de produtos agrícolas (GUIMARÃES; COSTA, 2002). Com relação à identificação e caracterização de variedades, dispõe-se atualmente de tecnologias capazes de revelar diferenças mínimas entre genótipos. Essa tecnologia é denominada de marcadores moleculares (MORAIS; MELO, 2011). A partir dessas técnicas, é possível, por exemplo, detectar ambiguidades em coleções de cafeeiros, estabelecer relações de parentesco entre linhagens, cultivares ou espécies diferentes e, sobretudo, identificar com precisão a identidade de uma determinada cultivar em casos de dúvidas (LASHERMES; CROS; COMBES, 1996; OROZCO-CASTLLO et al., 1994).

Como constituem as formas mais abundantes de variações do genoma, os polimorfismos de nucleotídeo único (SNPs) podem auxiliar na caracterização de recursos genéticos, inclusive de espécies que possuem variabilidade genética reduzida (RAFALSKI, 2002). Os marcadores SNP tem como base as alterações mais elementares da molécula de DNA, ou seja, mutações em bases únicas da cadeia de bases nitrogenadas (Adenina, Citosina, Timina e Guanina), surgem a partir de erros de incorporação de bases durante a replicação do DNA, em outros casos são causadas por lesões no DNA por agentes ambientais (GUIMARÃES; COSTA, 2002). Em uma abordagem mais restrita, somente são considerados SNPs os polimorfismos

que possuem uma frequência alélica mínima de 1% em uma dada população (BROOKES, 1999).

Os SNPs ocorrem tanto em regiões codificadoras como em não codificadoras dos genomas. Em regiões codificadoras, quando resultam em uma substituição de aminoácido na sequência proteica, são denominados não sinônimos, podendo a substituição ser conservativa ou não conservativa em função das características dos aminoácidos envolvidos na troca (Figura 5). Nesses, casos, pode haver modificações estruturais e funcionais na proteína. Embora SNPs sinônimos não alterem a sequência proteica, eles podem modificar a estrutura e a estabilidade do RNA mensageiro e conseqüentemente afetar a quantidade de proteína produzida. Além disso, SNPs podem promover: *splicing* alternativo, alterações no padrão de expressão de genes (como no caso de alterações em sequências de promotores), geração ou supressão dos codons de terminação ou poliadenilação na molécula de RNA mensageiro e alteração nos codons de iniciação de tradução (GUIMARÃES; COSTA, 2002).

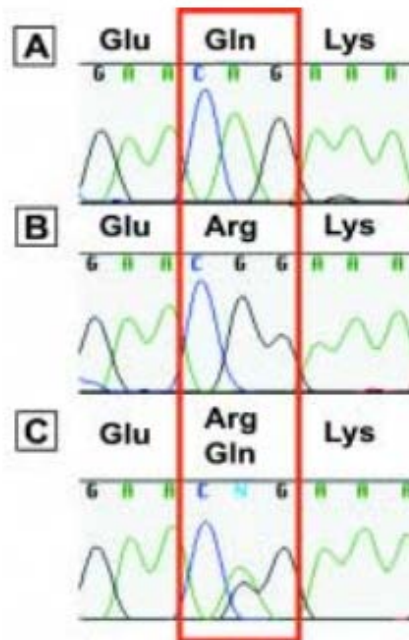


Figura 5. Alinhamento de sequências correspondentes ao mesmo segmento genômico de três indivíduos, mostrando um SNP não sinônimo A/G, que causa a substituição de glutamina para arginina na sequência proteica (GUIMARÃES; COSTA, 2002).

A Forma de SNP bi-alélico, isto é, quando há apenas uma forma alternativa para o polimorfismo em questão é geralmente a mais comum. No entanto, também podem existir formas tri e tetra-alélicas, menos frequentes, em razão da raridade de mutações (BROOKES, 1999).

As substituições mais frequentes que ocorrem no DNA são as que envolvem bases nitrogenadas de mesma característica estrutural, ou seja, trocas entre duas purinas (A/G ou G/A) ou duas pirimidinas (C/T ou T/C) e são denominadas transições. As transversões são substituições de uma purina por uma pirimidina ou vice-versa (GUIMARÃES; COSTA, 2002).

Para diploides e autopoliploides, o fator principal que influencia a eficiência da descoberta de SNPs é a discriminação contra as sequências parálogas, bem como aumento da amostragem para garantir a recuperação de todas as variantes alélicas. O processo é mais complicado para aloploides,

como a espécie *Coffea arabica*, pois existem diferentes variantes de nucleotídeo único.

A presença dos diversos tipos de variações gera nos aloploidos níveis ainda mais elevados de complexidade durante a comparação das sequências (Figura 6). O nível de dificuldade dessa comparação depende do grau de estudo dos genomas e informações disponíveis (KAUR; FRANCKI; FORSTER, 2011)

As Variantes Homólogas de sequência de nucleotídeo único (SNPs) surgem entre pares de cromossomos idênticos (KAUR; FRANCKI; FORSTER, 2011) de qualquer subgenoma em cada um dos progenitores diploides correspondentes. Variações da sequência Homoeólogas (HSVs) surgem nas coordenadas correspondentes de nucleotídeos de genes que possuem cópias iguais nos diferentes subgenomas da planta poliploide (SOMERS et al., 2003)

Variações da sequência Parálogas (PSVs) surgem entre genes duplicados que foram originados por uma sequência ancestral comum e coexistem atualmente no mesmo genoma (KAUR; FRANCKI; FORSTER, 2011). Quando ocorrem dentro de genomas ou subgenomas individuais são denominadas PSVs tipo I e quando ocorrem entre as sequências dos subgenomas diferentes são denominados de PSVs tipo II (KAUR; FRANCKI; FORSTER, 2011).

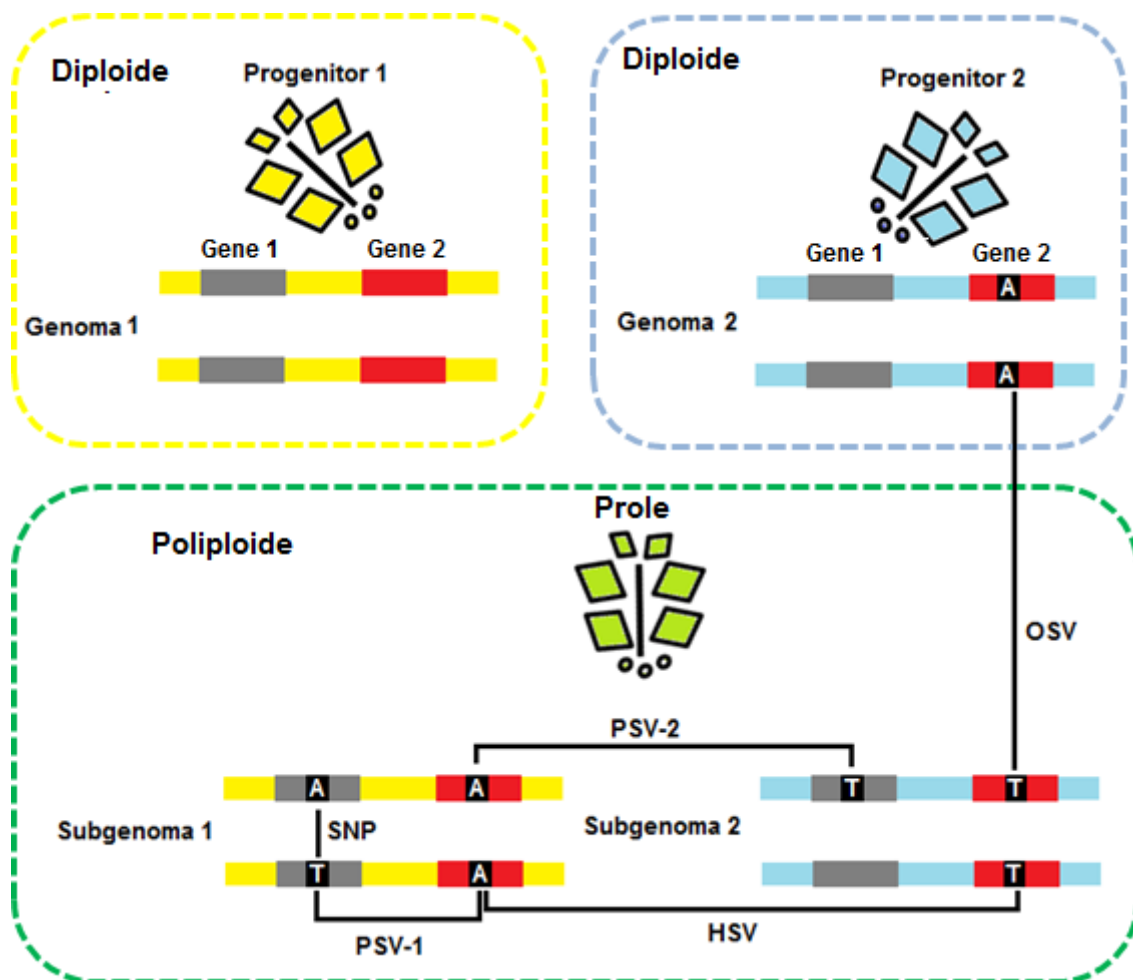


Figura 6. . Representação das categorias de variação de sequência dentro e entre subgenomas de uma planta alopoliploide e genomas diploides correspondentes (KAUR; FRANCKI; FORSTER, 2011).

A área de maior impacto dos marcadores moleculares no melhoramento vegetal é através do emprego da seleção assistida para identificação de genótipos superiores em populações segregantes (FEDERIZZI, 1998). Através dessa tecnologia, pode-se incrementar a eficiência no melhoramento de plantas, pois menor número de progênies por combinação é necessário, bem como menor número de gerações para a estabilização dos genótipos (BARBOSA-NETO, 1998).

Usando um mapa genético, genes putativos que afetam características de interesse podem então ser detectados testando associações estatísticas entre variações do marcador e qualquer característica de interesse.

Estas características podem ser de herança simples, muitos fatores que causam doenças em plantas são controlados por um ou poucos genes. Alternativamente, elas podem ser quantitativas complexas, Características desse tipo são aqueles cuja expressão fenotípica apresenta uma variação contínua, atribuída à segregação simultânea de muitos genes distribuídos em regiões definidas como QTLs (“Quantitative Trait Loci”).

A maioria das características economicamente importantes está na segunda categoria (CHOUDHARY; CHOUDHARY; SHEKHAWAT, 2008).

Uma das principais aplicações da análise genética é a construção de mapas de ligação. Foram reportados mapas genéticos completos de *C. canephora* (LASHERMES et al., 2001) e de um híbrido interespecífico (*C. pseudozanguebarie* x *C. liberica*) (KY et al., 2000). Pearl et al.(2003) desenvolveram um mapa genético para uma população pseudo-F2 gerada a partir do cruzamento de um híbrido de *C. arabica* com *C. canephora* e uma planta *C. arabica* verdadeira. Porém mapeamento com F2 verdadeiro ainda não existe.

Estes mapas genéticos fornecem informações importantes sobre o genoma do cafeeiro e a organização dos cromossomos. Particularmente, podem ser utilizados para mapear genes importantes (LASHERMES; CROS; COMBES, 1996), além de loci que codificam características quantitativas (QTLs), como compostos bioquímicos que

conferem ao café qualidade de bebida (BARRE et al., 1998) ou resistência á doenças (ALVARENGA et al., 2011).

Desenvolvimento de uma população de mapeamento normalmente levaria de 5 a 7 anos. O processo demorado e caro para desenvolvimento de uma população de mapeamento, juntamente com o baixo nível de diversidade genética dentro da espécie adiou construção mapa genético para esta planta (PEARL et al., 2004).

Outro fator limitante para o desenvolvimento de um mapa genético para *C.arabica* é o pouco conhecimento sobre genoma da espécie. Visando diminuir essa lacuna o projeto foi criado o Projeto Genoma Café (VIEIRA et al., 2006).

O Projeto Genoma-Café, desenvolvido por iniciativa do Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café, Embrapa Café, FAPESP e Embrapa Recursos Genéticos e Biotecnologia em parceria com diversas instituições de pesquisa (IAC, UNICAMP, USP, UNESP, IAPAR, UFLA, UFV, Epamig e Incaper). O Projeto resultou em um total de 200.000 ESTs (*Expressed Sequence Tags*) sequenciados que, após análises de bioinformática, foram agrupados em aproximadamente 35.000 genes distintos (VIEIRA et al., 2006).

A partir de dados do projeto do genoma café e da iniciativa de sequenciamento de EST de *C. canephora*, Vidal et al. (2010) encontraram uma frequência de 0,3934 SNPs a cada 100pb. A maioria dos polimorfismos encontrados corresponde a diferenças entre os subgenomas de *Coffea eugenioides* e *Coffea canephora*. Polimorfismos intra-especificos não foram encontrados nas sequencias das variedades analisadas. Porém, Yanagui

(2010) ao analisar sequências de genótipos da Etiópia, entre outras, encontrou 25 SNPs, representando 18% do total encontrando em *C.arabica*. No entanto, estes polimorfismos ainda precisam ser validados.

3- OBJETIVOS

O presente trabalho visa otimizar a técnica de SNaPshot para validação de SNPs em *Coffea arabica* em uma coleção Etíope da espécie.

4- MATERIAS E MÉTODOS

4.1. MATERIAL VEGETAL

Para o grupo teste, onde foi realizada a confirmação dos SNPs foram utilizados 11 genótipos de *Coffea arabica* (123^a,E017,E464,E238,516/69, 123b/121, 272/175, 007/87, Mundo Novo, Typica, Catuaib), 2 *Coffea canephora* (Canephora 621^a e Canephora 621^a) e 1 *Coffea eugenioides*. Os utilizados em Yanagui (2010).

Os SNPs confirmados foram selecionados para validação em uma Coleção e 143 genótipos da coleção da Etiópia.

4.2 ISOLAMENTO DO DNA GENÔMICO

O DNA genômico foi isolado a partir de folhas pelo método CTAB. Aproximadamente 5 g de folhas de café foram maceradas em nitrogênio líquido e acondicionadas em tubos falcon de 50 ml. Após a maceração foram

seguidos os passos descritos em Brasileiro e Carneiro (1988) com as modificações visando evitar a oxidação do material e purificação do DNA obtido, a seguir: adição de 0,4 ml de metabissulfito de sódio 10% ao tampão de extração e purificação com fenol-clorofórmio. Após a extração, verificação da qualidade foi realizada através de eletroforese em gel de agarose 0,8% e a quantificação foi feita por espectrofotometria.

4.3- Identificação de SNPs e desenho do oligos SNaPshot

A busca e identificação dos SNPs foram realizadas anteriormente por Yanagui, 2010 no instituto agrônomo do Paraná (IAPAR). Dos 144 SNPs encontrados foram selecionados 10 para validação, os quais representam diferenças intra-específicas em *Coffea arabica*. Os SNPs selecionados foram SNPs 1, 4, 5, 6, 15, 17, 18, 19, 22 e 25 (Tabela 1).

Tabela 1. Relação dos 25 SNPs encontrados em *C. arabica* Colunas 1 a 25 representam os SNPs. Linha representa o genótipo analisado. NA = informação não disponível (YANAGUI, 2010).

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>	NA	G/A	NA	A/T	T	T	NA	C/T	NA	NA	T	C/A	G/T	6 (T)/5 (T)	C	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	A
E-017	<i>C. arabica</i>	C	G/A	C/T	A/T	T/A	T/G	NA	C/T	NA	NA	T	NA	G/T	NA	NA	NA	C	C	A	NA	NA	NA	NA	NA	NA	A/C
E-123A	<i>C. arabica</i>	C	G/A	C/T	NA	T	T/G	NA	C/T	G/C	NA	T	C/A	G/T	NA	NA	NA	NA	C	A	NA	NA	NA	NA	NA	NA	A/C
E-123B	<i>C. arabica</i>	C	G/A	C	A/T	T/A	T	NA	C/T	G/C	NA	T	C/A	G/T	NA	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	NA	NA	A/T	T	T	NA	C/T	NA	NA	T	C/A	G/T	NA	NA	NA	NA	C	A	NA	NA	NA	NA	NA	NA	NA
E-238	<i>C. arabica</i>	C	NA	NA	T	T/A	T/G	NA	C/T	NA	NA	C	NA	G/T	NA	NA	NA	T/C	C	A	NA	G/A	G	A/G	A/T	A/C	
E-464	<i>C. arabica</i>	C	NA	NA	T	T	T/G	NA	C/T	NA	NA	T	NA	G/T	6 (T)/5 (T)	C	NA	T/C	C	A	NA	NA	NA	NA	NA	NA	A/C
E-516	<i>C. arabica</i>	A/C	NA	NA	T	T/A	T/G	NA	C/T	G/C	NA	NA	C/A	G/T	NA	NA	NA	NA	NA	NA	NA	G/A	G/A	A/G	A/T	A/C	
Bourbon	<i>C. arabica</i>	C	NA	C/T	A/T	T	T	NA	C/T	G/C	NA	T	NA	G/T	NA	NA	NA	NA	C	A	NA	NA	NA	NA	NA	NA	A/C
Typica	<i>C. arabica</i>	C	G/A	C/T	T	T	T	NA	C/T	G/C	T/C	T	NA	G/T	NA	NA	NA	T/C	C	A	NA	NA	NA	NA	NA	NA	NA
Catuai	<i>C. arabica</i>	C	G/A	C/T	A/T	T	T	NA	C/T	NA	NA	T	NA	G/T	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	A/C
Mundo Novo	<i>C. arabica</i>	C	G/A	NA	NA	T	T/G	T/A	C/T	G/C	NA	T	NA	G/T	NA	NA	NA	C	C	A	NA	G/A	G/A	A/G	A/T	NA	
C1007	<i>C. canephora</i>	NA	G	T	T	T	T	A	T	NA	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	NA	T	T	NA	T	6 (T)	C	T	C	C	A	G	NA	NA	G	A	NA	
C3001	<i>C. canephora</i>	C	G	T	T	T	T	A	T	NA	T	T	NA	T	NA	C	T	C	C	A	NA	NA	NA	G	A	A	
C4001	<i>C. canephora</i>	NA	G	NA	T	NA	T	A	T	G	NA	T	NA	T	6 (T)	C	T	C	C	A	NA	NA	NA	NA	NA	A	
G2020	<i>C. canephora</i>	C	G	T	T	T	T	A	T	NA	T	T	A	T	NA	C	T	C	C	A	G	G	A	G	A	NA	
G21	<i>C. canephora</i>	C	G	T	T	T	T	NA	T	NA	T	T	NA	T	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	A	
UW002	<i>C. canephora</i>	NA	NA	NA	T	T	T	A	T	G	T	T	NA	T	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	A	
UW099	<i>C. canephora</i>	C	G	T	T	T	T	A	T	NA	T	T	A	T	8 (T)	C	T	C	C	A	G	G	A	G	A	A	
Coffea eugenioides	<i>C. eugenioides</i>	C	G	NA	T	T	T	A	T	NA	T	T	NA	T	NA	C	T	C	C	A	G	G	NA	NA	NA	A	
Coffea racemosa	<i>C. racemosa</i>	C	G	T	T	T	T	NA	T	NA	T	T	NA	T	6 (T)	C	T	C	C	A	G	G/A	G	A	A/T	A	
Psilanthus bengalensis	<i>P. bengalensis</i>	C	G	NA	T	T/A	T	A	T	NA	T	T	NA	T	NA	C	T	C	C	A	G	NA	NA	NA	NA	A	

Foram selecionados genes relativos à qualidade do café e regiões do DNA de alta relevância para análises filogenéticas. Os genes da sacarose sintase (Susy_1), da invertase vacuolar (Inv,) estão envolvidos no metabolismo da sacarose, que é um metabolismo central para a planta e está relacionado ao sabor da bebida e a tolerância a seca.

Na maioria das espécies de plantas, o carbono assimilado e transportado em forma de sacarose, um dissacarídeo de glicose e frutose juntos por uma ligação O-glicosídicas, e quebram a sacarose em dois monossacarídeos. Plantas superiores apresentam distintas isoenzimas que podem ser distinguidas pela sua localização celular (parede celular, vacúolo e citosol) e solubilidade (FOTOPOULOS, 2005).

Por outro lado a sucrose sintase é uma glicosil-transferase, a qual, na presença de UDP, converte sacarose em UDP – glicose (STURM; TANG, 1999).

Também relacionado ao metabolismo energético, foram testados oligos iniciadores relativos aos genes da enzima Álcool desidrogenase (ADH), enzima de maior importância envolvida no processo de respiração anaeróbica das plantas. A enzima é induzida quando a planta cresce em estresse anóxico (Davies, 1980). No metabolismo fermentativo, o piruvato é descaboxilado para formar acetaldeído, catalisado pela enzima PDC. Por sua vez o acetaldeído é convertido em Etanol usando NADH (KE; MATAOS; KADER, 1993). De acordo com Zang et al. (1994), o acúmulo de acetaldeído pode acelerar a deterioração das sementes.

Também foram selecionados oligos relativos a região TrnL-TrnF. Esta região do DNA cloroplasmial é amplamente utilizada em estudos filogenéticos em todos os níveis taxonômicos. O espaçador intergênico trnL-trnF e o intron trnL têm sido usados desde níveis intra e interespecíficos até níveis de família e tribo. Esta é a região não codante do cpDNA mais frequentemente utilizada em estudos filogenéticos (Tai Sang

Os oligos SNaPshot foram desenhados através da análise das sequências vizinhas aos SNPs encontrados por Yanagui (2010); no *software Codon Code Aligner*. Preferencialmente foram sequências a montante adjacentes ao SNP de interesse. Nos casos em que não foi possível utilizá-la em razão da impossibilidade de determinação da sequência, os oligos foram desenhados a partir da fita complementar.

A avaliação da qualidade do oligo foi feita com o *software OligoCalc*, onde os principais fatores analisados foram, a formação de estruturas secundárias, porcentagem de G/C e temperatura de *melting* (T_m).

4.4- VALIDAÇÃO DOS SNPs

Para validação dos SNPs encontrados foi utilizado o método SNaPshot multiplex com o kit ABI PRISM® SNaPshot™ Multiplex. O método possui quatro passos: Preparação do molde, preparação da reação, tratamento termo cíclico de pós-extensão e leitura da amostra de acordo com Figura7(ABI,2010).

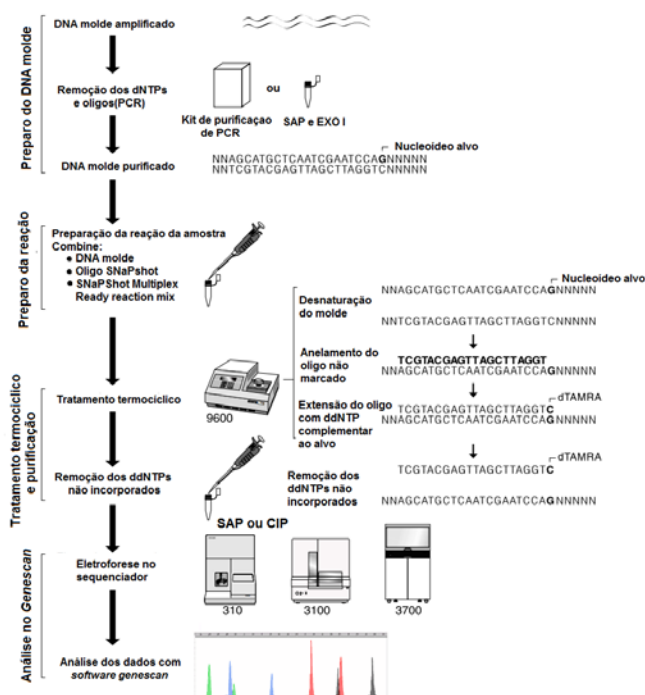


Figura 7. Fluxograma de trabalho para análise de SNPs usando kit SNaPshot™ Multiplex

Na fase de preparação do DNA molde foi realizada uma reação de PCR com oligos específicos das regiões onde estão localizados os SNPs de

interesse e logo em seguida o material foi purificado para remoção de ddNTPs e oligos não incorporados (ABI, 2010).

O DNA dos genótipos selecionados foi amplificado por reação de PCR combinando 2,0µL de tampão, 0,8µL de cloreto de magnésio (2 mM), 0,5µL de dNTP (2,5 mM), 1,0µL de oligonucleotídeo iniciador senso (10 mM), 1,0µL de oligo antisenso (10 mM), 1,0µL de Taq (1U), 25ng de DNA e 8,7µL de água ultrapura autoclavada para cada amostra. A eficiência da PCR foi confirmada em gel de agarose 1%, utilizando 5µL da amostra e 2µL de corante de frente (YANAGUI, 2010)

Para remoção de excesso de oligos, o produto da PCR foi submetido à purificação com Fosfatase Alcalina de Camarão (SAP) e Exonuclease 1 (EXO1) (ABI, 2010a). Foram adicionadas 5 (cinco) unidades de SAP e 2 (duas) unidades de EXO1, para 15µL de produto de PCR e homogeneizados. O produto então foi incubado a 37°C por 1 hora e em seguida a 75°C por 15 minutos para inativação das enzimas. O material foi mantido a -20°C (ABI, 2010).

O DNA molde purificado seguiu para etapa de preparação da reação SNaPshot, onde foi combinado com o oligo SNaPshot e o mix SNaPshotTM Multiplex Ready Reaction. O mix é composto por AmpliTaq[®] DNA Polymerase, marcador com fluorescência ddNTP e Tampão (ABI, 2010).

Os ddNTPs (Dideoxynucleotídeos trifosfato) são nucleotídeos que não possuem hidroxila ligada ao carbono 3' que foram marcados fluorescência específica (Tabela 2) (ABI, 2010).

Os oligos SNaPshot foram desenhados de maneira que fosse complementar a sequência imediatamente ao SNP de interesse no sentido 5'→3', tanto na fita senso como na antisenso. Para preparação da reação foram adicionados 5µL SNaPshot Multiplex Ready Reaction Mix, 3µL do produto da PCR, 1µL de oligo SNaPshot total. Cada oligo deve ser preparado para uma concentração de 0,2 µM. O volume final da solução deve ser igual à 10µL, sendo completado com água deionizada (ABI, 2010).

Após a preparação da reação a amostra seguiu para o para o termociclador, seguindo as seguintes especificações para uma repetição de 25 ciclos: ciclo de desnaturação (96°C, 10seg), anelamento (50°C, 5seg) e extensão (60°C, 30 seg) (ABI, 2010a). No início desta etapa ocorre a desnaturação do modelo, separando as moléculas de DNA. Em seguida o oligo SNaPshot anela-se imediatamente anterior ao SNP e o processo de extensão começa. É adicionado um ddNTP à extremidade 3' e a extensão cessa devido a falta de extremidade OH livre (Figura 8) (ABI, 2010).

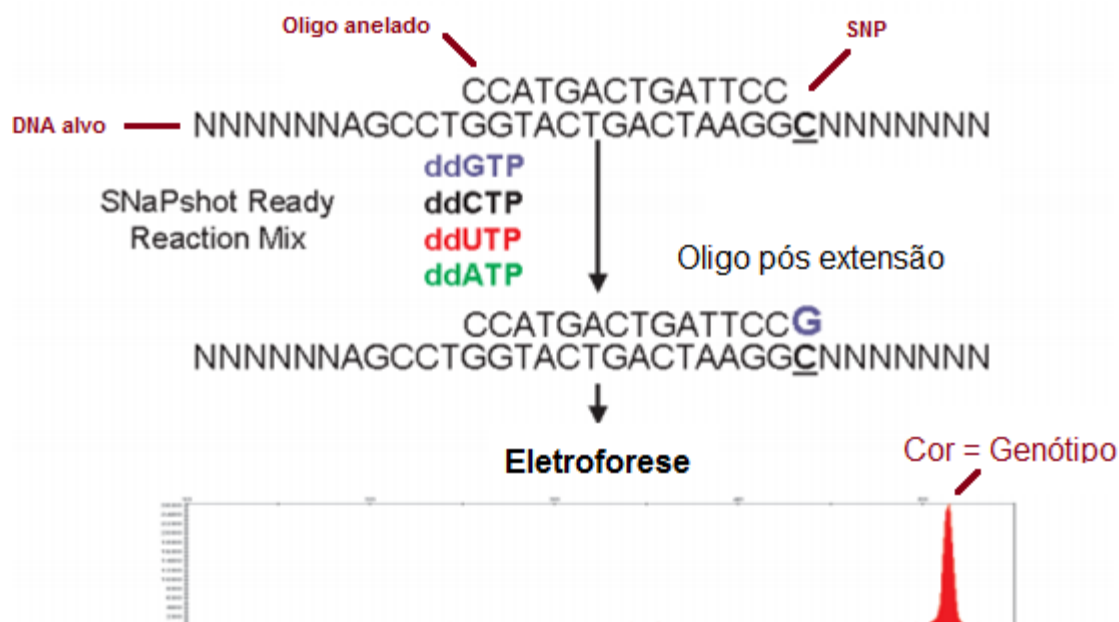


Figura 8. Método SNaPshot Multiplex. Introdução de dinucleotídeos fluorescentes baseados nos SNPs dos fragmentos de DNA amplificados.

Após o término da reação, iniciará a fase de pós-extensão ou de purificação, onde ocorre a remoção dos ddNTPs não incorporados através da adição da enzima Shrimp Alkaline Phosphatase (SAP) ao mix, que foi incubado a 37°C por 1 hora para sua atividade seguida a 75°C a 15 minutos para sua desativação (ABI, 2010). Depois do processo, as amostras foram armazenadas à -20°C, para posterior leitura no 3130xl *genetic analyzer*.

Para testes iniciais foi adotado o protocolo sugerido pelo Kit da *Applied Biosystems* (ABI, 2010) e Yanagui, 2010 descritos acima. A partir destes protocolos foram estabelecidas três modificações principais redução de reagentes da PCR; modificação dos parâmetros usados na reação SNaPshot e purificação das amostras. A metodologia descrita abaixo apresenta as modificações realizadas.

A reação de PCR foi realizada misturando 0,6 μ L de tampão, 0,24 μ L de cloreto de magnésio (2mM), 0,15 μ L de dNTP (2,5mM), 0,3 μ L de oligonucleotídeo iniciador (oligo) *senso* (10mM), 0,3 μ L de oligo antisenso (10mM), 0,3 μ L de Taq (1U), 1,5 μ L de DNA (25ng) e 2,61 μ L de água ultrapura autoclavada para cada amostra. Em seguida o material foi purificado através da adição de 0,8 unidades da enzima CIP e 2 Exonúclease 1 (EXO1). Para facilitar o processo as enzimas foram adicionadas aos tampões (1x) e misturados em um tubo de 2 μ L. A solução foi homogeneizada em um agitador de tubos tipo *vortex* e imediatamente aplicada na amostra. Em seguida a amostra seguiu para o termociclador onde foi incubada 37°C por uma hora para que ocorra atividade e em seguida por 15 minutos permaneceu em 80°C para inativação das enzimas purificadoras, EXO1 e CIP.

O material purificado seguiu para etapa de SNaPshot. Para preparação da reação são adicionados 5 μ L SNaPshot Multiplex Ready Reaction Mix, 3 μ L do produto da PCR, 2 μ L de oligo SNaPshot a 0,5 μ L. Após a preparação da reação a amostra seguiu para o termociclador onde ocorre a reação SNaPshot. Não houve alteração nesta fase, foi utilizado o procedimento padrão (ABI, 2010). Em seguida foi realizada a segunda etapa de purificação através da adição de 1U da enzima Calf Intestinal phosphatase (CIP) e tampão (1x). Terminada a purificação, 2 μ L das amostras contendo 1 dos 5 oligos são misturados para execução multiplex no sequenciador.

4.5- LEITURA E ANÁLISE DA AMOSTRA

Depois de passar pelos processos descritos acima, a amostra seguiu para a última etapa, a fase da análise da amostra, que foi realizada em duas etapas: leitura no seqüenciador *3130xl genetic analyzer* e posteriormente análise no programa computacional *GeneMapper*. No sequenciador as amostras de DNA são separadas em capilares contendo gel de eletroforese e são separadas de acordo com o tamanho. Um laser passa através do capilar e excita os corantes fluorescentes, que em resposta emitem luz em um comprimento de onda específico que é captado por uma câmera CCD, revelando a posição do fragmento e o nucleotídeo (ABI, 2005).

4.6 REAÇÕES CONTROLE

Para efeito de comparação e verificação da completa funcionalidade dos instrumentos, reações controle foram realizadas. Em cada kit ABI PRISM® SNaPshot™ Multiplex estão presentes oligos multiplex de controle e um tubo contendo DNA molde de com SNPs conhecidos (Tabela 3) (ABI, 2010)

Tabela 2. Oligos controle ABI PRISM® SNaPshot™ Multiplex

Mix controle Oligo multiplex	Tamanho do produto final (nt)	Cor do sinal	Heterozigossidade
20A oligo	21	Verde	Homozigoto
28G/A oligo	29	Azul e Verde	Heterozigoto
36G oligo	37	Azul	Homozigoto
44T oligo	45	Vermelho	Homozigoto
52C/T oligo	53	Preto e Vermelho	Heterozigoto
60C oligo	61	Preto	Homozigoto

Devido a influência do *dye* na mobilidade dos fragmentos de DNA, os tamanhos informados podem diferir algumas bases do tamanho real do fragmento. Em fragmentos menores a influência do *dye* tende a ser maior (Figura 9). Um controle negativo também deverá ser realizado a fim avaliar possíveis interferências externas (ABI, 2010).

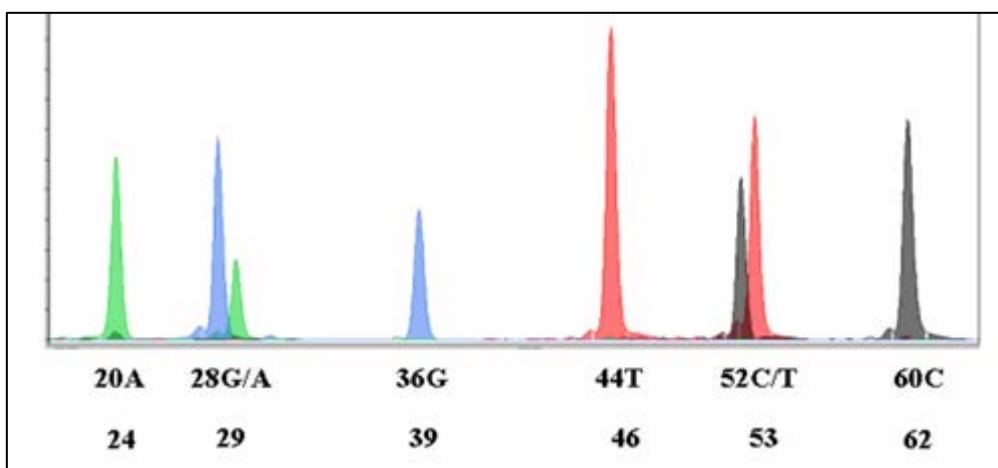


Figura 9. Determinação de padrão através do software GeneScan-120, mostrando tamanho do oligo seguido nucleotídeo complementar e abaixo tamanho relativo após

As reações foram preparadas conforme a Tabela 4. Durante todo procedimento as amostras controles foram mantidas em gelo. Após a preparação das reações, foi realizado o procedimento de tratamento termocíclico e de pós-

extensão, que foi o mesmo da amostra. Constatado o pleno funcionamento do aparelho pode dar início as reações das amostras selecionadas (ABI, 2010).

Tabela 3. Preparação das reações do controle positivo e negativo

Item	Controle positivo (μL)	Controle negativo (μL)
SNaPshot Multiplex Ready Reaction Mix	5	5
DNA molde multiplex controle	2	0
Oligo SNaPshot controle multiplex	1	1
Água deionizada	2	4
Total	10	10

4.7- ANÁLISE DOS DADOS COM *GeneMapper*

Após a análise da amostra pelo sequenciador *3130xl genetic analyzer* os dados ficam armazenados e são utilizados pelo programa *GeneMapper*. Os dados são apresentados no programa em forma de picos coloridos característicos de cada fluorescência. Os possíveis SNPs são identificados, a partir do tamanho do fragmento do oligo e da fluorescência relacionada ao SNP (Figura 10). Para análise dos dados fora do computador acoplado ao sequenciador foi utilizado o *software Peak scanner v0.1*.

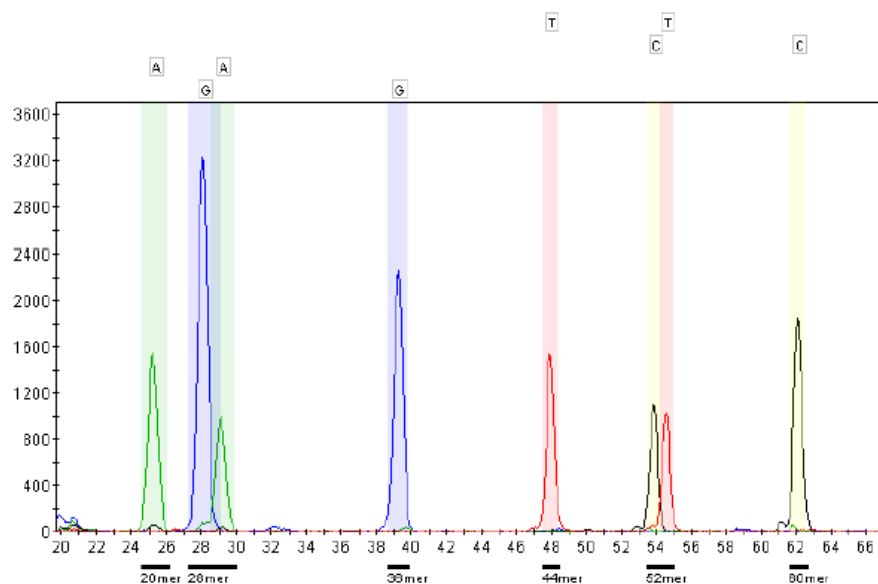


Figura 10. . Exemplo de um eletroferograma gerado pelo GeneMapper

5- RESULTADOS

5.1- OTIMIZAÇÃO DO PROTOCOLO

Para teste preliminares foram selecionados aleatoriamente 3 genótipos, 365\652, 037\302 e 237\141. Os testes seguiram o modelo padrão proposto no kit ABI PRISM[®] SNaPshot[™] Multiplex (ABI, 2010), onde foram feitas reações com os oligos em Multiplex, como descrito no tópico material em métodos e foi modificado gradualmente para solucionar problemas e melhorar a qualidade do processo. Como se pode observar na Figura 11, ocorreu a presença de grande quantidade de picos inespecíficos, que podem ser oriundos da purificação ineficiente ou da amplificação de fragmentos aleatórios, na região entre 10-90pb onde deveriam ocorrer apenas picos indicativos do

SNP que por sua vez não foram visualizados ou não foi possível diferenciá-los dos picos inespecíficos.

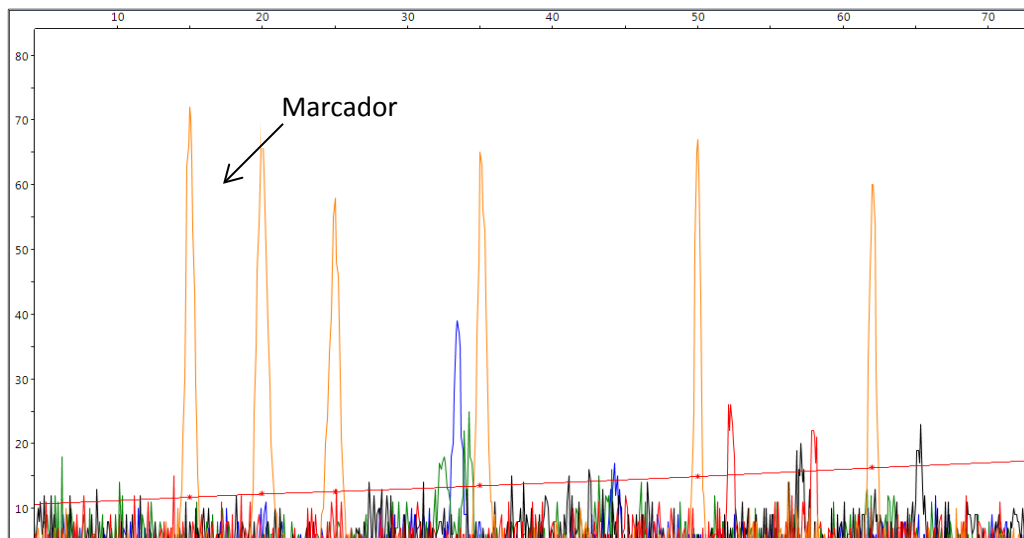


Figura 11. Eletroferograma do genótipo 365\652 analisado no software Peak Scanner Software v1.0 para 5 SNPs em multiplex.

A fim de diminuir a quantidade de picos inespecíficos e avaliar possíveis interações entre os oligos a fase da reação SNaPshot foi realizada separadamente para cada oligo, ao invés da reação em multiplex. Logo após 2,5 μ L de cada reação foi misturado em um poço de uma microplaca de 96 poços, totalizando 10 μ L. Este material foi purificado com 1 μ L de enzima CIP, seguindo corrida multiplex. A Figura 12 exibe um gráfico do procedimento após estas mudanças. Pode observar a melhoria na diferenciação dos picos, contudo ainda pouco definidos. Os picos indefinidos apesar de terem diminuído em quantidade ainda estão presentes.

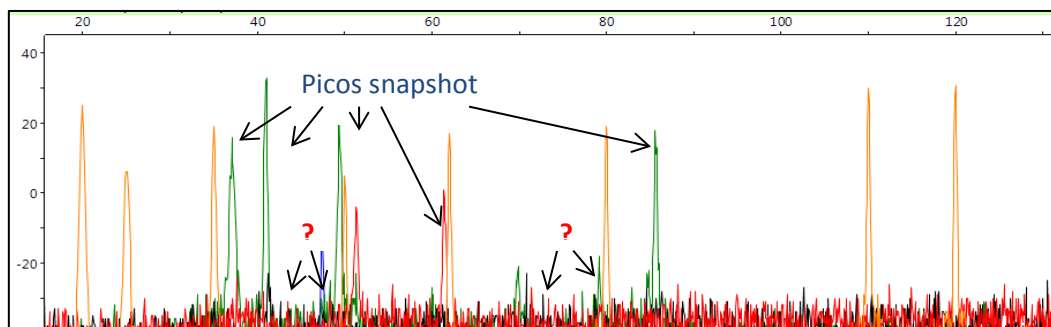


Figura 12. Eletroferograma do genótipo 365\652 analisado no software Peak Scanner Software v1. As reações foram feitas individualmente e a corrida em multiplex.

O próximo passo foi a mudança na concentração ($0,2\mu\text{M}$ e $0,5\mu\text{M}$) e volume ($1\mu\text{L}$ e $2\mu\text{L}$) dos oligos SNaPshot. O melhor resultado foi obtido no teste D com $0,5\mu\text{M}$ e $2\mu\text{L}$ (Figura13). Neste teste notou-se o aumento significativo na qualidade dos picos dos SNPs com uma diminuição dos inespecíficos. Entretanto o processo de purificação nesse teste foi realizado em amostras individualizadas, que pode ter contribuído para melhora dos resultados.

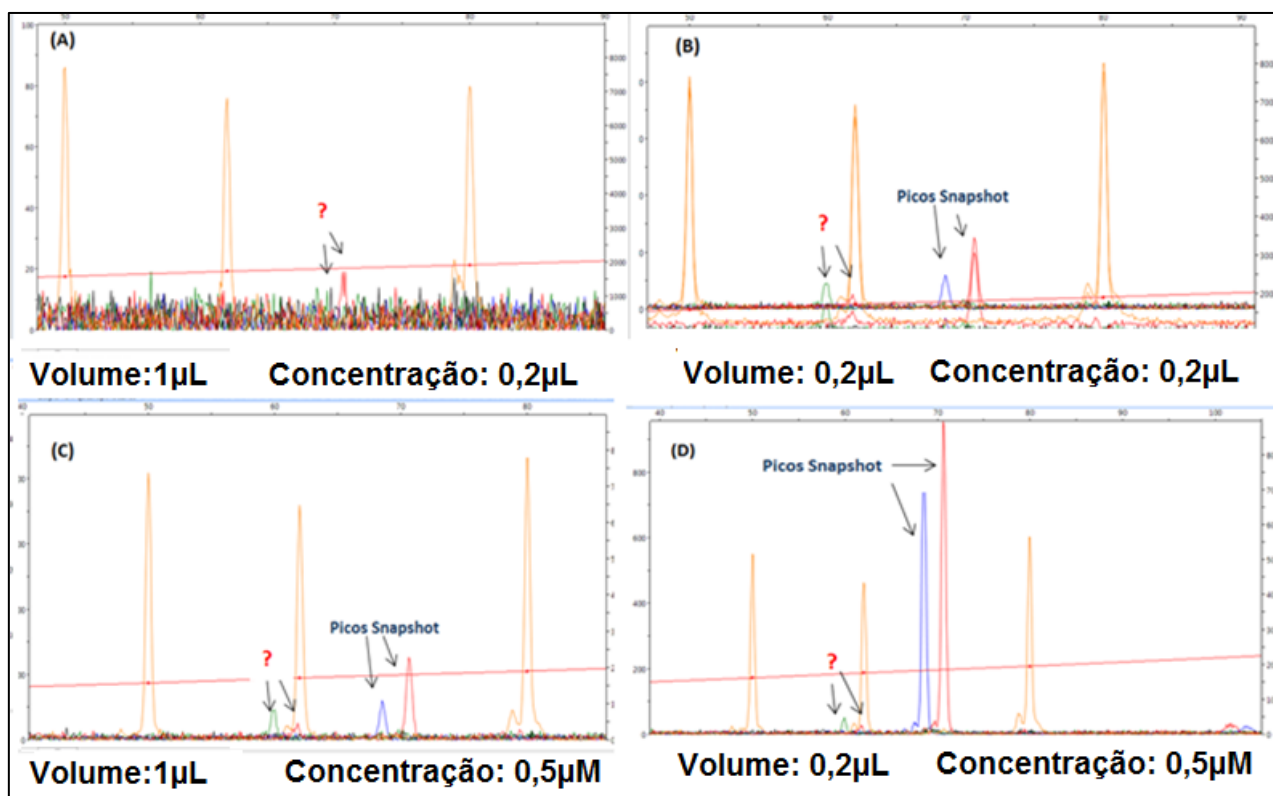


Figura 13. Resultado do teste de otimização do Volume e concentração dos Oligos SNaPshot e parâmetros modificados.

Para verificar o efeito da purificação das amostras após a reação do SNaPshot foi feito um teste com purificação individual e em conjunto. As reações que foram realizadas separadamente apresentaram picos com diferenciação maior e nenhum dos picos de SNP falhou em todas amostras (Amostra C; genótipo 237\141) (Figura 14). Nas amostras que foram purificadas em conjunto apenas genótipo 237\141 (Amostra A) apresentou os cinco picos de SNPs, os demais apresentaram ausência de alguns desses picos (Figura 14). Como é representado pela amostra B (Figura 14).

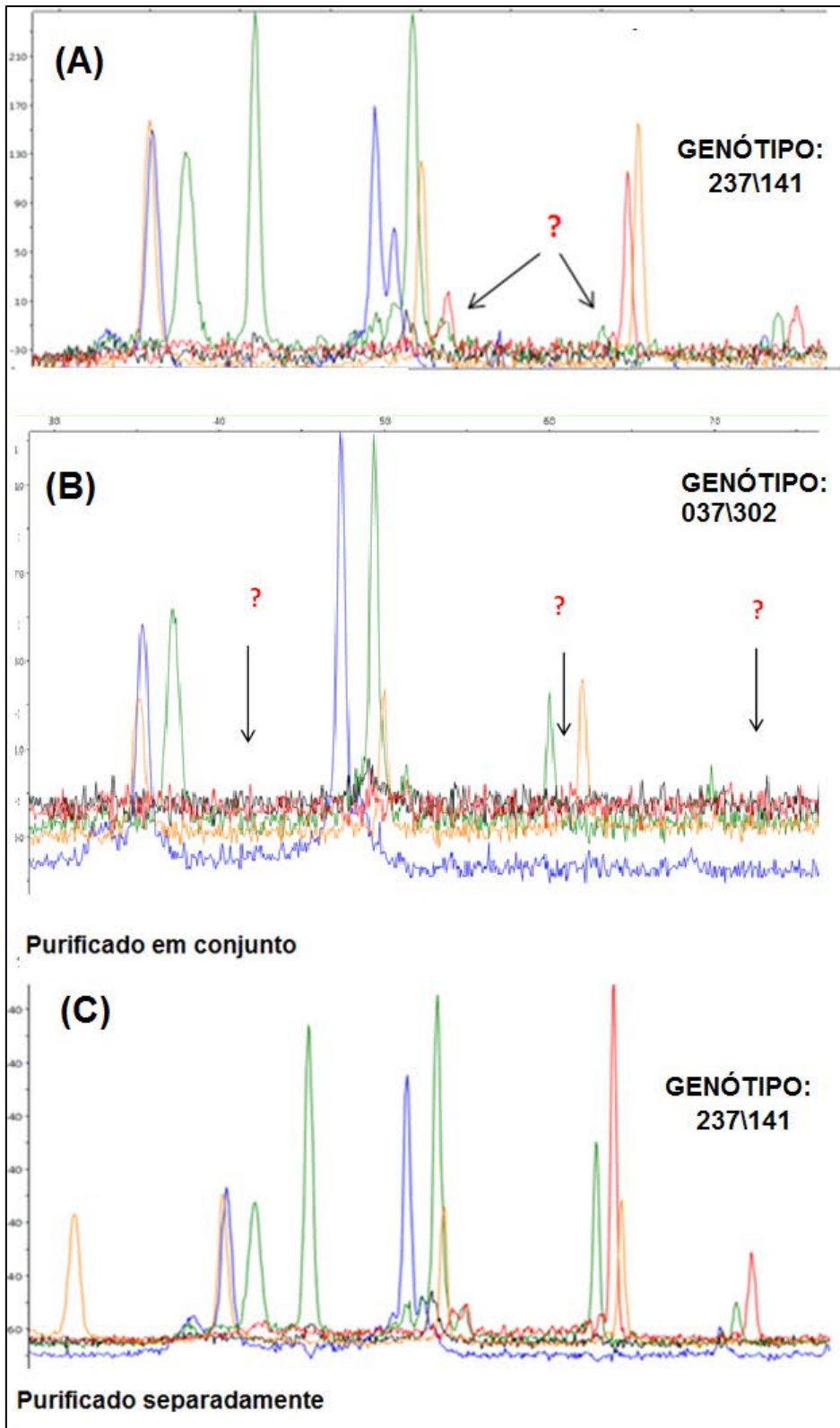


Figura 14. Eletroferograma da purificação após reação SNaPSHOT, em conjunto e individualmente.

Com as mudanças no processo houve o aumento no consumo de enzimas e o consumo. Para amenizar esse fator houve alterações no volume dos reagentes, já descritas em Materiais e Métodos. Nesta fase houve a redução de 75% dos reagentes em comparação ao protocolo original (YANAGUI, 2010). O volume final é suficiente para repetir a reação caso for necessário.

5.2- Oligo SNaPshot

No trabalho realizado por Yanagui, 2010 foram encontrados 25 SNPs. Dentre estes, apenas 10 possíveis SNPs intra-específicos de *Coffea arabica* foram selecionados para validação. A partir das sequências vizinhas aos SNPs selecionados foram desenhados 10 oligos SNaPshots. Estes formaram dois grupos, denominados de grupo A e grupo B, para realização de uma análise multiplex (Tabela 5). Uma cauda poli-T foi adicionada aos Oligos na extremidade 5' para que através da diferença de tamanho pudessem ser identificados na análise. Cada grupo multiplex possui 5 oligos de tamanhos diferentes. Os oligos do grupo A foram desenhados com tamanhos de 5 a 15pb de diferença entre eles. Em alguns casos a região de interesse da fita de DNA senso não foi utilizada devido a presença de GAPs ou SNPs. Nestes casos, os oligos foram desenhados a partir da fita complementar de DNA para contornar o problema (SNP22r, SNP17, SNP4 e SNP5).

No presente trabalho apenas os oligos do grupo (A) foram testados; e produziram resultados satisfatórios.

Tabela 4. Desenho dos Oligos SNaPShot

DNA molde	Oligo	Poli-T	Sequencia	CG%	pb
GRUPO A					
P14(Susy_c6_b)	SNP17	0	CGTTATATTGCTGACAAAAGGGG	43	23
CPS_SUS_1_4	SNP4	10	CCACACAGGTAGAATTGTTTTAC	38	34
P14(Susy_c6_b)	SNP15	20	CGTGGATCTGCAATTTGCGT	50	40
CPS_SUS_1_5	SNP5	30	CATCTTGACTTTCTTAGAACATTTT	28	55
CPS_SUS_1_8	SNP6	40	CTGATAATGATTGCCTTTTTTGT	30	63
GRUPO B					
ADH1.3	SNP1	0	GCAATGACGACATTCCTTACACT	43	23
P14(Susy_c6_b)	SNP19	9	GCTCACCATTCCGCGCTCTATTTGTTTGGGC	55	40
P14(Susy_c6_b)	SNP18	25	CCATTCCGCGCTCTATTTGTTTGGG	52	50
P25(Inv3_c1_c)	SNP22r	41	CCACGCACTCCCACATGCC	68	60
trnL trnF	SNP25	47	AAAGATGTTTCATTTATATGTATATC	20	73

Os picos gerados a partir dos oligos SNaPshot se distanciaram satisfatoriamente viabilizando a identificação dos SNPs em uma análise multiplex. As posições dos picos se mantiveram constantes em todas as amostras. A contaminação com picos estranhos na área de exposição do SNPs foi mínima na maioria das análises, contudo houve a presença desses picos em áreas distantes. A posição desses picos também são estáveis e facilmente inidentificáveis não influenciando na análise. Através da análise dos dados do grupo A foi observado que a posição dos picos variou de 5 a 14 pares de bases em relação ao marcador e os picos dos oligos menores distanciam-se menos em relação aos outros picos (Figura 15). Estes parâmetros foram utilizados para o desenho dos oligos do grupo B. Neste grupo os oligos foram desenhados de forma que o segundo menor oligo possui um diferença de 18pb em relação ao primeiro enquanto os demais uma distância de ± 10 pb.

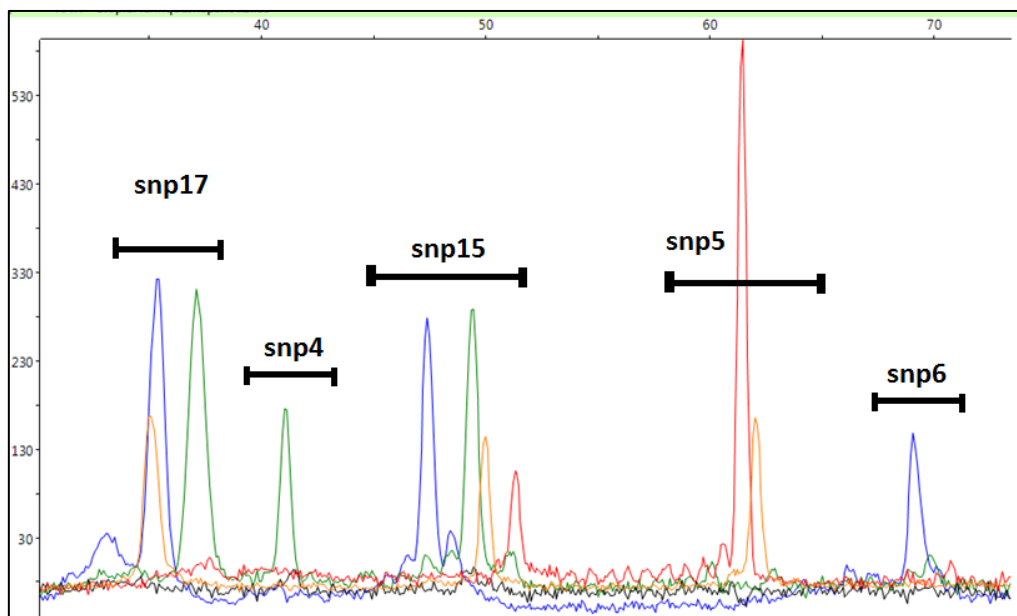


Figura 15. . Eletroferograma exibido no software Peak scanner v0.1. dos oligos do grupo A.

5.3- VALIDAÇÃO NO GRUPO TESTE

Foram validados 5 SNPs a partir de 15 genótipos. Destes, 13 pertencem a espécie *Coffea arábica*, 2 a *Coffea canephora* e 1 a *Coffea eugenioides*. Os SNP5 e SNP6 foram mais frequentes na população teste. SNP5 exibiu duas formas polimórficas, AT (54%) e T(46%) e SNP6 as formas GT (58%) e T (42%) nos genótipos analisados do grupo teste. No estudo do SNP 17 não foram observadas diferenças intraespecíficas, contudo os genótipos de *C. canephora* e o de *C. eugenioides* apresentaram o mesmo perfil alélico e ambos apresentaram perfis diferentes de *C. arabica*. Os SNPs 4 e 5 apresentaram poucas diferenças alélicas intraespecíficas. Este último exibindo um perfil majoritário trialélico.

Com exceção do SNP17 que não apresentou polimorfismos, os outros quatro SNPs serão validados nos 143 genótipos etíopes.

Tabela 5. Polimorfismos encontrados no grupo de genótipos teste.

	Genótipos	SNP17	SNP4	SNP15	SNP5	SNP6
E T I Ó P I A	123 ^a	C T	T G	C A T	A T	G T
	E017	C T	T G	C A T	A T	G T
	E464	C T	T G	C A T	A T	G T
	E238	C T	T G	C A T	A T	G T
	E237	C T	T	C A T	T	G
	516/69	C T	T G	C A T	T	T
	123B/121	C T	T G	C	A T	T
	272/175	C T	T G	C A T	A T	G T
	007/87	C T	T G	C A T	T	T
Mundo novo a	C T	T G	C A T	A T	G T	
Typica	C T				T	
Catuaib	C T	T G	C A T	T		
Canephora 621a	C	T G	C A T	T		
Canephora 621b	C		C A T		G T	
Eugeniodes	C	T G	C	T	T	

6.4- Validação do SNP5 na população da Etiópia

O SNP5 foi validado em 143 genótipos de *Coffea arabica* da Etiópia e exibiu 3 perfis alélicos (Figura 16). O perfil majoritário foi AT (76 genótipos), seguido por T (31 genótipos) e A (14 genótipos). Em 20 genótipos não foi possível a identificação dos SNPs (Figura 16).

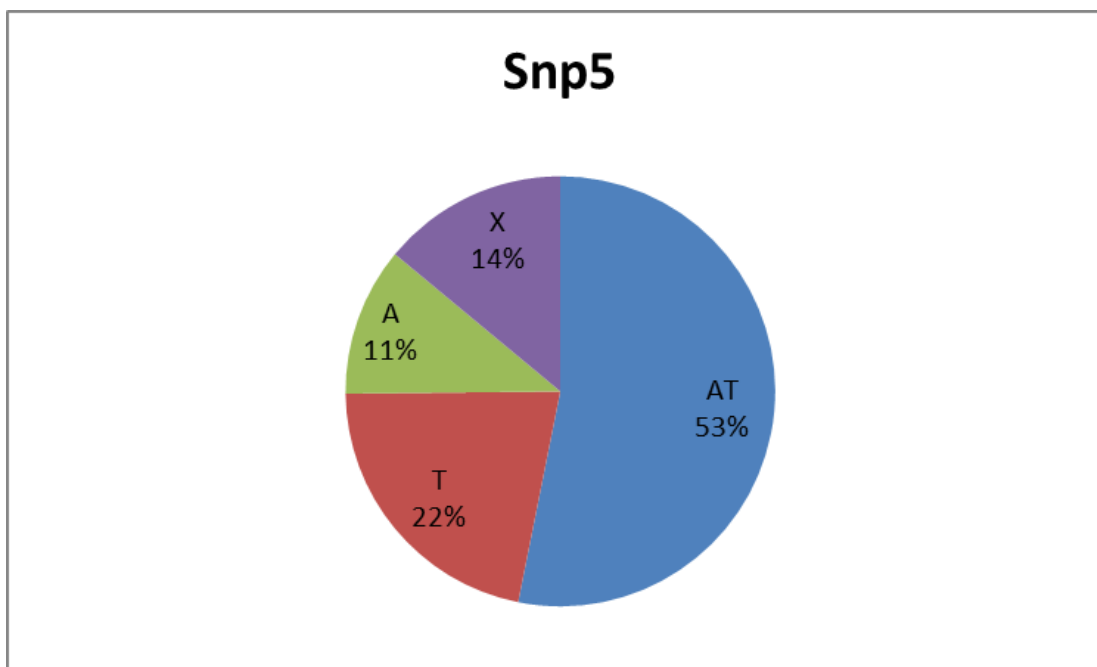


Figura 16. Perfil alélico da população de 143 genótipos da população da Etiópia para o SNP5. Não foi possível analisar os SNPs em 14% das amostras (X).

7- CONCLUSÃO

O presente trabalho visou otimizar a técnica de Snapshot em *Coffea arabica*. Ao final dos testes foi possível realizar com sucesso a reação snapshot em multiplex em um conjunto de 5 (cinco) SNPs.

Foi observado que apesar dos SNPs serem a o tipo de variação mais abundante no genoma, nos genótipos estudados a variabilidade foi baixa. No entanto, como visto nos testes realizados com SNP5, perfis alélicos diferenciados podem aparecer quando estudados em populações maiores.

8- REFERÊNCIAS

ABI, APPLIED BIOSYSTEMS . **Applied Biosystems 3730/3730xl DNA Analyzers and 3130/3130xl Genetic Analyzer**, 2005. Disponível em: <www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/general_documents/cms_041990.pdf>. Acesso em: 14 jun. 2011

ABI, APPLIED BIOSYSTEMS. **SNaPshot Multiplex Kit ®™**, 2010. Disponível em: <www3.appliedbiosystems.com/cms/groups/mcb_support/documents/general_documents/cms_041203.pdf>. Acesso em: 14 jun. 2011

ABIC, ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ **Estatísticas-Exportações**. Disponível em: <www.abic.com.br/estat_pagricola.html>. Acesso em: 14 jun. 2011.

ALVARENGA, S. M. et al. Marcadores moleculares derivados de sequências expressas do genoma café potencialmente envolvidas na resistência à ferrugem. **Pesquisa Agropecuária Brasileira**, v. 46, n. 8, p. 890-898, ago. 2011.

ANTHONY, F. et al. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. **TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik**, v. 104, n. 5, p. 894-900, abr. 2002.

ANTHONY, F.; DUSSERT, S.; DULLOO, E. Coffee genetic resource. In: ENGELMANN, M. E. (Ed.). **Conserving coffee genetic resources. Biodiversity International**. [S.l.] Biodiversity International, 2007. p. 12-22.

BARBOSA-NETO, J. F. Seleção assistida por marcadores moleculares. In: MILACH, S. C. . (Ed.). **ores moleculares em plantas**. Porto Alegre: UFRGS, 1998. p. 75-80.

BARRE, P. et al. Inheritance of caffeine and heteroside contents in an interspecific cross between a cultivated coffee species *Coffea liberica* var

dewevrei and a wild species caffeine-free *C. pseudozanguebariae*. **Theoretical and Applied Genetics**, v. 96, n. 2, p. 306-311, 1998.

BRASILEIRO, A.C.M. CARNEIRO, V. T. C. Manual de transformação genética de plantas. **EmbrapaSPI/Embrapa-Cenargen**, 1988.

BROOKES, A. J. The essence of SNPs. **Gene**, v. 234, n. 2, p. 177-186, 1999.

CHARRIER, A.; BERTHAUD, J. Botanical classification of coffee. In: CLIFFORD, M.; N, W. (Eds.). **Botany, biochemistry and production of beans and beverage**. Connecticut: The AVI Publishing Company, 1985. p. 13-47.

CHOUDHARY, K.; CHOUDHARY, O. P.; SHEKHAWAT, N. S. Marker Assisted Selection: A Novel Approach for Crop Improvement. **American-Eurasian Journal of Agronomy**, v. 1, n. 2, p. 26-30, 2008.

CONAB, C. N. DE A. **Acompanhamento da Safra Brasileira Café Safra 2010 segunda estimativa, maio/2010**. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/12_11_08_09_10_48_boletim_portugues_novembro_2012.pdf>. Acesso em: 12 nov. 2011.

CONAB, C. N. DE A. **Acompanhamento da Safra Brasileira Café Safra 2010 segunda estimativa, Janeiro/2014**. Disponível em: http://conab.gov.br/OlalaCMS/uploads/arquivos/14_01_09_09_18_57_boletim_cafe_-_original.pdf. Acesso em: 3 abril. 2014.

DAVIS, A. P. et al. A Global Assessment of Distribution, Diversity, Endemism, and Taxonomic Effort in the Rubiaceae. **Annals of the Missouri Botanical Garden**, v. 96, n. 1, p. 68-78, 2009.

DAVIES, David Denzil. Anaerobic metabolism and the production of organic acids. In: Davies, DD (ed) **The biochemistry of plants, Vol 2**. Academic, New York, p 581–611, 1980.

EMBRAPA, EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA
Hitórico do café. Disponível em:
<www.sapc.embrapa.br/index.php/portal/historico>. Acesso em: 14 jun. 2011.

FEDERIZZI, L. C. Estrutura de um programa de melhoramento de plantas e possíveis aplicações de marcadores moleculares: visão do melhorista. In: MILACH, S. C. K. (Ed.). **Marcadores moleculares em plantas**. Porto Alegre: UFRGS, 1998. p. 3-15.

FOTOPOULOS, Vasileios. Plant invertases: structure, function and regulation of a diverse enzyme family. **Journal of Biological Research**. V.4. p.127-137, 2005.

GUIMARÃES, P. E. M.; COSTA, M. C. R. SNPs: Sutis diferenças de um código. **Biotechnologia Ciência e Desenvolvimento**, v. 26, p. 24-27, 2002.

JEHAN, T.; LAKHANPAUL, S. Single nucleotide polymorphism (SNP)– Methods and applications in plant genetics: A review. **Indian Journal of Biotchnology**, v. 5, n. October, p. 435-459, 2006.

KAUR, S.; FRANCKI, M. G.; FORSTER, J. W. Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. **Plant Biotechnology Journal**, v. 10, n. 2, p. 1-14, 2011.

KE, Danyang; Mateos, Mila; Kader, Adel. A. Regulation of Fermentative Metabolism in fruits and Vegetables by Controlled Atmospheres. Proceedings from the sixth International Controlled Atemosphere Resarch Conference. NRAES-71.Cornell University, Ithaca, NY, pp 63-77. 1993.

LASHERMES, P. et al. Phylogenetic relationships. **Applied Genetics**, v. 94, p. 947-955, 1997.

LASHERMES, P. et al. Molecular characterisation and origin of the Coffea arabica L. genome. **Molecular & general genetics**, v. 261, n. 2, p. 259-66, mar. 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 National Research Council Canada Genome Conseil national de recherches Canada**, v. 44, n. 4, p. 589-596, 2001.

LASHERMES, P.; CROS, J.; COMBES, M. Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. **Theoretical and Applied Genetics**, v. 93, p. 626-632, 1996.

M.A, MINISTÉRIO DA AGRICULTURA. **Café no Brasil**. Disponível em: <www.agricultura.gov.br/vegetal/culturas/cafe/saiba-mais>. Acesso em: 27 nov. 2011.

MALUF, M. P.; OLIVEIRA, A. C. B. Genoma Café: Perspectivas para o melhoramento da cultura. **O agrônomo**, v. 59, n. 1, p. 68-70, 2007.

MORAIS, T. P. DE; MELO, B. DE. Biotecnologia aplicada ao melhoramento genético do cafeeiro. **Ciência Rural**, v. 41, n. 5, p. 753-760, 2011.

OROZCO-CASTLLO, C. et al. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. **Theoretical and Applied Genetics**, v. 87, n. 1545, p. 934-940, 1994.

PEARL, H. M. et al. Construction of a genetic map for arabica coffee. **Tag Theoretical And Applied Genetics Theoretische Und Angewandte Genetik**, v. 108, n. 5, p. 829-835, 2004.

PINTO-MAGLIO, C. A. F. Cytogenetics of coffee. **Brazilian Journal Of Plant Physiology**, v. 18, n. 1, p. 37-44, 2006.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, v. 5, n. 2, p. 94-100, 2002.

SCHIFINO-WITTMANN, M. T. POLIPLOIDIA E SEU IMPACTO NA ORIGEM E EVOLUÇÃO DAS PLANTAS. **Revista Brasileira Agrociência**, v. 10, n. 2, p. 151-157, 2004.

SANG, Tao.; Crawford, Daniel. J.; Stuessy, Tod. F. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). **American Journal of Botany**. 84:1120-1136. 1997.

SIQUEIRA, T. V. DE. A cultura do café: 1961-2005. **BNDES Setorial**, v. 22, p. 205-270, set. 2005.

STURM. Arnd; Tang, Guo-Quing. The sucrose-cleaving enzymes of plants are crucial for development., growth and carbon partitioning. **Elsevier Science Ltd**. Vol.4, n.10, p. 401 -407. Out. 1999.

UKERS, W. H. History of Coffe Propagation. In: UKERS, W. H. (Ed.). **All about Coffe**. New York: The Tea and Coffee Trade Journal Company, 1922. p. 35-39.

VIDAL, R. O. 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**, v. 154, n. 3, 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, mar. 2006.

YANAGUI, K. **Análise da diversidade nucleotídica intra análise da diversidade nucleotídica intra e interespecífica de coffea spp**. Londrina: Universidade Estadual de Londrina, 2010.

ZHANG, M.; MAEDA, Y.; FURIHATA, Y.; NAKAMAR, Y.; ESASHI, Y. A mechanism of seed deterioration in relation to the volatile compounds evolved by dry seeds themselves. **Seed Science Research**, Wallingford, v. 4, p. 49-56, Mar. 1994.