



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

BRUNA SILVESTRE RODRIGUES DA SILVA

**VALIDAÇÃO DE MARCADORES MICROSSATÉLITES E  
DIVERSIDADE GENÉTICA DE ACESSOS DE *Coffea arabica*  
PROVENIENTES DA ETIÓPIA**

---

Londrina  
2014

BRUNA SILVESTRE RODRIGUES DA SILVA

**VALIDAÇÃO DE MARCADORES MICROSSATÉLITES E  
DIVERSIDADE GENÉTICA DE ACESSOS DE *Coffea arabica*  
PROVENIENTES DA ETIÓPIA**

Exame de Defesa de Dissertação de mestrado apresentado ao Programa de Pós-Graduação em Biotecnologia Agroindustrial da Universidade Estadual de Londrina, como requisito para a obtenção do título de Mestre.

Orientador Dr: Luiz Filipe Protasio Pereira.  
Co- Orientador Dr: Douglas Silva Domingues.

Londrina  
2014

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

**Dados Internacionais de Catalogação-na-Publicação (CIP)**

S586v Silva, Bruna Silvestre Rodrigues da.  
Validação de marcadores microssatélites e diversidade genética de acessos de  
*Coffea arabica* provenientes da Etiópia / Bruna Silvestre Rodrigues da Silva. –  
Londrina, 2014.  
56 f. : il.

Orientador: Luiz Filipe Protasio Pereira.  
Coorientador: Douglas Silva Domingues.  
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. Café – Melhoramento genético – Teses.  
3. Café – Mapeamento cromossômico – Teses. 4. Café – Biotecnologia – Teses.  
5. Coffea arabica – Teses. I. Pereira, Luiz Filipe Protasio. II. Domingues, Douglas  
Silva. III. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa  
de Pós-Graduação em Biotecnologia. IV. Título.

CDU 663.1

BRUNA SILVESTRE RODRIGUES DA SILVA

**VALIDAÇÃO DE MARCADORES MICROSSATÉLITES E  
DIVERSIDADE GENÉTICA DE ACESSOS DE *Coffea arabica*  
PROVENIENTES DA ETIÓPIA**

Exame de Defesa de Dissertação de mestrado apresentado ao Programa de Pós-Graduação em Biotecnologia Agroindustrial da Universidade Estadual de Londrina, como requisito para a obtenção do título de Mestre.

**BANCA EXAMINADORA**

---

Orientador: Luiz Filipe Protasio Pereira  
Empresa Brasileira de Pesquisa Agropecuária  
– EMBRAPA – Café

---

Prof. Dr. André L.M. Oliveira  
Universidade Estadual de Londrina – UEL

---

Prof. Dra. Lilian Padilha  
Empresa Brasileira de Pesquisa Agropecuária  
– EMBRAPA – Café

Londrina, 25 de maio de 2017.

## **AGRADECIMENTOS**

Aos meus pais Maria Eudália Rodrigues da Silva e Sidnei da Silva por acreditarem sempre nos meus ideais, e por apoiarem nos momentos de fraqueza.

Aos meus irmãos Jaqueline Silvestre Rodrigues da Silva e Jeancarlos Rodrigues da Silva que sinto muita saudade pela distância, obrigado pela companhia e carinho dado a mim todos os 24 anos de convivência.

Aos meus segundos pais Lécio Silva e Oinei da Silva, meus dois exemplos de vida. Obrigada pelos ensinamentos, palavras de conforto, apoio financeiro, e pelo amor dedicado a mim desde meu nascimento. Não estaria aqui vencendo mais uma etapa se não fosse por eles. Muito obrigado por terem me apoiado nessa jornada que eu sei que não é fácil, mais me torna a cada dia uma pessoa mais realizada.

Ao meu orientador, Dr. Luiz Filipe Protasio Pereira, pela sua amizade, competência profissional, e pela oportunidade de realização do Mestrado no IAPAR.

Ao Dr. Douglas Silva Domingues pela ajuda na realização deste projeto e pelas contribuições recebidas.

Ao Consórcio de Pesquisa Café e CAPES pelo apoio financeiro neste trabalho.

Ao Consórcio Pesquisa Café pela concessão da bolsa de estudo.

A Universidade Estadual de Londrina, ao curso de Pós-graduação em Biotecnologia e os professores pela dedicação. Aos colegas de turma pela amizade e companheirismo.

Aos pesquisadores do laboratório de Biotecnologia do IAPAR pela amizade, companheirismo, pela ajuda nas análises e obtenção dos resultados deste trabalho.

SILVA, Bruna Silvestre Rodrigues. **Validação de marcadores microssatélites e diversidade genética de acessos de *Coffea arabica* provenientes da Etiópia**. 56 f. 2014. Dissertação (Mestrado em Biotecnologia Agroindustrial) – Universidade Estadual de Londrina. Londrina, 2014.

## RESUMO

O café é uma valiosa *commodity* agrícola, sendo sua bebida a mais popular do mundo. Em *Coffea arabica*, o desenvolvimento de novas cultivares é demorado podendo levar mais de 20 anos. Além disso, a estreita base genética de *C. arabica* dificulta a obtenção de cultivares resistentes a pragas e a doenças, ou tolerantes a estresses abióticos. Sabe-se que o uso de germoplasmas exóticos em programas de melhoramento pode aumentar a variabilidade genética das variedades cultivadas. Os acessos selvagens de *C. arabica* geneticamente distantes das variedades cultivadas proporcionam novos alelos para enriquecer sua variabilidade genética. Assim, a utilização de marcadores moleculares para análise da diversidade e seleção de genótipos representa uma importante ferramenta para auxiliar o melhoramento e tentar diminuir esses problemas. Visando aumentar o número de marcadores moleculares disponíveis, neste trabalho foi realizada uma análise *in silico* para a busca de marcadores microssatélites (SSR) a partir de dados de RNA-seq de Iapar 59 e de sequências de BACs de Híbrido de Timor 832/2. A análise *in silico* resultou em 23 motivos SSR que foram classificados quanto à natureza e frequência. Os motivos di-, tri- e tetranucleotídeos juntamente com as classes (AT)<sub>n</sub>, (TA)<sub>n</sub>, (GA)<sub>n</sub> e (TC)<sub>n</sub> foram identificados como os mais frequentes para ambos os genótipos. Junto aos 58 novos SSR, 77 SSR foram selecionados a partir de uma extensa revisão bibliográfica sobre SSR em *C. arabica* e/ou *C. canephora*. Foram construídos *primers* iniciadores para 135 SSR, dos quais 122 resultaram em produtos de amplificação e 31 foram polimórficos em um painel de 24 acessos do centro primário de origem da espécie. Desses, 20 SSR foram altamente informativos com valores de PIC acima de 0,7, e sete foram descritos pela primeira vez. Um total de 165 alelos foram obtidos com média de 5,3 alelos por loco. O estudo de diversidade e estrutura genética dos acessos da Etiópia separou os genótipos em três grupos, não havendo separação geográfica dos acessos entre as regiões Sudeste e Sudoeste do Vale do Rift.

**Palavras-chave:** Café. Marcadores. SSR. Análise *in silico*. Vetores BAC. RNA-seq.

SILVA, Bruna Silvestre Rodrigues. **Microsatellite markers validation and genetic diversity of access *Coffea arabica* from Ethiopia**. 56 p. 2014. Dissertation (Master's degree in Agroindustrial Biotechnology) – Universidade Estadual de Londrina. Londrina, 2014.

## ABSTRACT

The *Coffee* is a valuable agricultural *commodity*, being their most popular drink in the world. In *Coffea arabica*, the development of new cultivars is time consuming and can take over 20 years. Furthermore, the narrow genetic base of *C. arabica* difficult to obtain resistance to diseases and pests cultivars or tolerance to abiotics stresses. The use of exotic germplasm in breeding programs can increase the genetic variability present in cultivated varieties and increase your resistance to stress. The wild access of *C. arabica* genetically distant from cultivated varieties provide new alleles to enhance their genetic variability. Thus, the use of molecular markers for diversity analysis and genotype selection is an important tool to help improve and try to lessen these problems. To increase the number of molecular markers available in our laboratory, this work was performed *in silico* analysis to search for microsatellite markers (SSR) from RNA-seq data of lapar 59 and sequences of BACs Timor Hybrid 832/2. *In silico* analysis resulted in 23 motifs SSR that were classified according to the nature and frequency. SSR motifs di-, tri-and tetranucleotide with the classes (AT)<sub>n</sub> (TA)<sub>n</sub> (GA)<sub>n</sub> (CT)<sub>n</sub> were identified as the most common for both genotypes. Along with 58 new SSR, 77 SSR were selected from an extensive literature review on *C. arabica* and *C. canephora* SSR. Primers were designed for 135 SSR, 122 amplified, and 31 were polymorphic in a panel of 24 access from the primary center of origin of the species (Ethiopia). From those, 20 SSRs were highly informative with PIC values above 0.7, and seven were first described. A total of 165 alleles were obtained, with average of the 5.3 alleles per locus. The study of genetic diversity and structure of the Ethiopian access separated the genotypes into três groups, with no geographical separation of access between the Southeast and Southwest regions of the Rift Valley.

**Keywords:** Coffee. Markers. SSR. *In silico* analysis. BAC vectors. RNA-seq.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	8
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	10
2.1	O CAFÉ .....	10
2.1.1	Origem e Diversidade Genética de <i>C. arabica</i> .....	11
2.2	MARCADORES MOLECULARES .....	13
2.2.1	Marcadores Microsatélites (SSR) .....	15
2.2.2	Marcadores SSR em <i>Coffea</i> .....	17
2.2.3	Desenvolvimento e Busca de Marcadores SSR .....	17
<b>3</b>	<b>OBJETIVO GERAL</b> .....	20
3.1	OBJETIVOS ESPECÍFICOS .....	20
<b>4</b>	<b>MATERIAIS E MÉTODOS</b> .....	21
4.1	BUSCA DE MICROSSATÉLITES .....	21
4.2	DESENHO DOS <i>PRIMERS</i> PARA AMPLIFICAÇÃO DOS SSR .....	22
4.3	SELEÇÃO DOS MARCADORES .....	22
4.4	MATERIAL VEGETAL .....	22
4.5	EXTRAÇÃO DE DNA GENÔMICO .....	24
4.6	CONSTRUÇÃO DOS “ <i>BULKS</i> ” DE DNA .....	24
4.7	AMPLIFICAÇÃO DOS SSR VIA PCR .....	25
4.8	ANÁLISE DOS DADOS .....	26
4.9	DIVERSIDADE E ESTRUTURA GENÉTICA .....	26
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	28
5.1	BUSCA DOS SSR POR ANÁLISE <i>IN SILICO</i> .....	30
5.2	SELEÇÃO DOS <i>PRIMERS</i> POLIMÓRFICOS .....	32
5.3	ANÁLISE DA DIVERSIDADE GENÉTICA DOS 24 ACESSOS DA ETIÓPIA .....	39
5.4	ESTRUTURA GENÉTICA DOS ACESSOS .....	39
<b>6</b>	<b>CONCLUSÕES</b> .....	44

<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	45
<b>ANEXOS</b> .....	54
Anexo 1. Relação dos pares de primers SSR obtidos de revisão bibliográfica .....	54
Anexo 2. Relação dos pares primers SSR desenhados para <i>C. arabica</i> cv. Iapar 59 .....	55
Anexo 3. Relação dos pares de primers SSR desenhados para CaHT 832/2.....	56

## 1 INTRODUÇÃO

O café é uma das principais *commodities* agroindustriais de países em desenvolvimento e uma das mais populares bebidas não alcoólicas. É cultivado principalmente em regiões tropicais e subtropicais, sendo uma importante fonte de renda em países em desenvolvimento na África, Ásia e América Latina (INTERNATIONAL COFFEE ORGANIZATION, 2013).

O cultivo do café em sua maior parte é de *Coffea arabica*, representando 65% a 70% da produção mundial (YU et al., 2011), sendo considerado como um café de qualidade superior, aroma agradável e sabores diversificados. O restante é de *C. canephora*, sendo muito utilizado principalmente para a fabricação de cafés solúveis e em *blends* comerciais com *C. arabica* (MORAIS et al., 2009).

O sucesso da produção agrícola depende fortemente do uso de cultivares com qualidades diferenciadas e adaptadas a diferentes ambientes de cultivo (VEGRO & FERREIRA, 2000). O melhoramento de plantas tem um papel fundamental no desenvolvimento da agricultura ao gerar novas variedades em espécies de interesse agrônômico, tolerantes a doenças, pragas e a estresses abióticos que são fatores restritivos ao alcance do máximo potencial produtivo de uma cultivar (MILACH, 1998). Para tanto, a utilização de ferramentas biotecnológicas como os marcadores moleculares, permite a caracterização genética de espécies de interesse agrônômico e a escolha precoce de características com menor tempo e custo.

Dentre os marcadores moleculares atualmente disponíveis, *Simple Sequence Repeats* (SSR) ou microssatélites são sequências de DNA compostas por repetições curtas em tandem em um dado loco (TAUZ & RENZ, 1994). Essa classe de marcadores tem sido amplamente utilizada devido à simplicidade da técnica utilizada, rapidez, grande poder de resolução e altos níveis de polimorfismo. Além disso, são amplamente dispersos nos genomas de plantas possibilitando uma discriminação precisa mesmo entre indivíduos geneticamente relacionados. A presença de diferentes números de elementos repetitivos faz do SSR um marcador genético altamente variável, multialélico e de grande conteúdo informativo (CAIXETA et al., 2009), fazendo deles um dos mais versáteis marcadores de DNA usados em pesquisas genéticas de plantas (CSENCISICS et al., 2010).

Estudos sobre a caracterização do germoplasma de *Coffea spp.* incluindo análises com marcadores moleculares, relataram a baixa diversidade genética

de *C. arabica* (ANTHONY et al., 2002). As justificativas para a estreita base genética e diversidade molecular da espécie incluem a sua recente origem, sua reprodução preferencialmente autógama e ao seu histórico de dispersão limitado a poucos indivíduos. Hoje a maioria das cultivares resultam de apenas duas bases populacionais, *Typica* e *Bourbon*. Assim, mesmo com as vantagens dos marcadores SSR, poucos foram descritos até a data para o café, significando a necessidade de expandir em número esta classe de marcador.

As coleções de acessos de *C. arabica* da Etiópia (Centro de origem da espécie) possuem maior variabilidade fenotípica para tamanho de folhas, altura, concentração de cafeína, açúcares e diterpenos. Além disso, pelo estudo com marcadores moleculares nesta coleção foi observada variabilidade genética muito maior do que nas cultivares comerciais analisadas, demonstrando o potencial desses acessos para estudos de mapeamento genético associado a características de interesse (POT et al., 2010).

Devido à importância dos acessos do centro primário de origem da espécie, o Instituto Agrônomo do Paraná (IAPAR) recebeu do Instituto Agrônomo de Campinas (IAC) 132 acessos de *C. arabica* provenientes da Etiópia, sendo a maioria da região Oeste do Vale do Rift.

Este trabalho buscou marcadores SSR em dados de sequenciamento de genoma e de transcriptoma, e visando aumentar o número de marcadores SSR no estudo foi realizada uma extensa revisão bibliográfica. Foi realizada também a validação destes marcadores em um painel de genótipos representativos da coleção da Etiópia presente no IAPAR. A disponibilização dos marcadores SSR informativos no painel permitirá analisar a diversidade e estrutura genética dos acessos selvagens, sendo posteriormente uma importante informação para a criação de uma coleção núcleo, além de melhorar a conservação, acessibilidade e uso de recursos genéticos da coleção de acessos da Etiópia (HAMON et al., 1999).

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1 O CAFÉ

O Brasil é considerado o maior produtor mundial de café sendo responsável por 30% do mercado internacional (ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ, 2013), maior exportador e o segundo maior consumidor, atrás apenas dos Estados Unidos. O Estado de Minas Gerais representa 55,0% da produção nacional, seguido dos Estados de Espírito Santo, São Paulo e Bahia (CONAB, 2014).

A introdução de cafeeiros no Brasil iniciou-se em 1727 e aos poucos tornou um dos produtos mais tradicionais da agricultura brasileira. De acordo com o primeiro levantamento da Conab para a produção da safra cafeeira, em 2014 o país deverá colher entre 46,53 e 50,15 milhões de sacas (60 quilos) de café beneficiado. O resultado representa desde uma redução de 5,4% a um crescimento de 2,02%, quando comparado com a produção de 49,15 milhões de sacas obtidas no ciclo anterior (CONAB, 2014).

O gênero *Coffea* pertence à família Rubiaceae e possui um total de 124 espécies (Davis et al., 2011), sendo a maioria diplóide ( $2n = 22x$  cromossomos) e predominantemente autoincompatíveis. Dentre as espécies, apenas *C. arabica* e *C. canephora* são importantes comercialmente e cultivadas em larga escala (VAN DER VOSSSEN, 1980; LASHERMES et al., 2000). Análises de hibridação *in situ* no genoma de *C. arabica* revelaram que é uma espécie tetraplóide ( $2n = 4x = 44$ ) e é derivada da hibridação entre duas espécies ancestrais estreitamente relacionadas *C. eugenoides* e *C. canephora*, ambas diplóides com  $2n = 2x = 22$  (LASHERMES et al., 1999). Além disso, a sua origem recente é estimada em aproximadamente 665.000 de anos (YU et al., 2011).

Existem três cultivares amplamente cultivadas em todo mundo, e que foram selecionadas a partir de duas variedades botânicas distintas, *Typica* e *Bourbon*. A cultivar Caturra é um mutante do grupo *Bourbon*; Mundo Novo é um híbrido de *Bourbon* e *Typica*, enquanto Catuaí é um híbrido de Mundo Novo e Caturra. Estas cultivares possuem alto rendimento e produzem uma bebida de alta qualidade (ANTHONY et al., 2002).

Além das variedades comerciais, um híbrido de *C. arabica* chamado de Híbrido de Timor é utilizado para a introgressão de resistência a pragas e doenças incluindo a ferrugem da folha do café (*Hemileia vastatrix*), doença da baga do café

(*Colletotrichum kahawae*), doença bacteriana (*Pseudomonas syringae* pv. *garcae*), e nematóide das galhas (*Meloidogyne exigua*) (CHARRIER, 1997). Devido à sua compatibilidade com *C. arabica*, o Centro de Investigação das Ferrugens do Cafeeiro (CIFC) em Portugal utilizou como progenitores masculinos os clones 832/1 e 832/2 de CHT para a criação de híbridos, como os germoplasmas de Sarchimor (*C. arabica* cv. Villa Sarchi 971/10 x TH CIFC832/2) e Catimor (*C. arabica* cv. Caturra x TH CIFC832/1) (MENDES et al., 1998).

Já a cultivar *C. arabica* cv. IAPAR 59 é uma variedade derivada do germoplasma Sarchimor (Villa Sarchi 971/10 e Híbrido de Timor 832/2) e possui resistência à ferrugem e ao nematóide *Meloidogyne exigua*, altamente disseminado no Brasil (CARVALHO et al., 2008).

Vários estudos têm demonstrado que a diversidade genética de *C. arabica* é baixa quando comparada com *C. canephora* devido a sua biologia reprodutiva, origem e evolução (GELETA et al., 2012). Os esforços para o desenvolvimento de genótipos superiores de *C. arabica* apresentam uma série de limitações. Além da sua estreita base genética, o longo período de juvenilidade (três a quatro anos), a diferença em nível de ploidia, e a falta de ferramentas eficientes para o auxílio do melhoramento (mapas de ligação e marcadores para características agronômicas), constituem os principais fatores que dificultam a obtenção de novas cultivares (BARUAH et al., 2003).

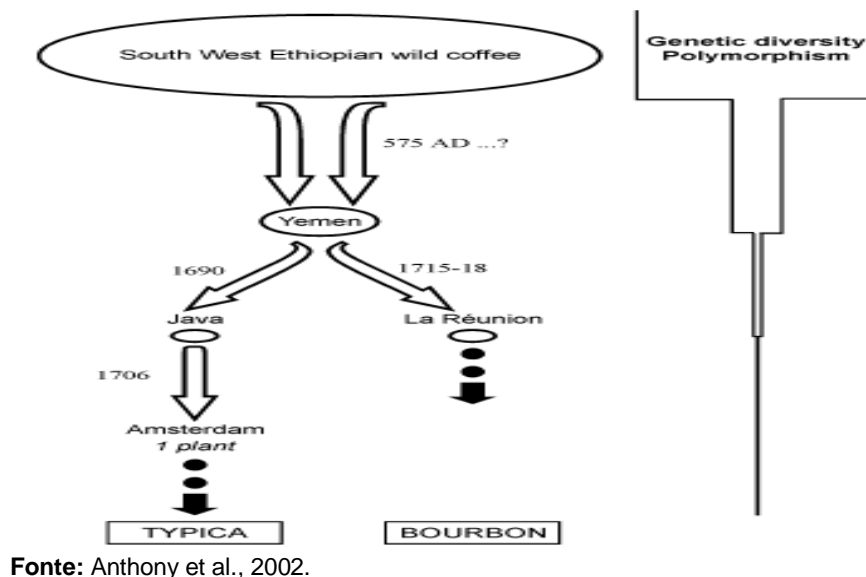
### 2.1.1 Origem e Diversidade Genética de *C. arabica*

A espécie *C. arabica* é uma planta perene e autógama, sendo nativa dos altiplanos do sudoeste da Etiópia (SYLVAIN, 1955), mas também há registros da espécie selvagem no Monte Marsabit no Quênia e Monte Imatong no Sudão (ANTHONY et al., 1987; BERTHAUD & CHARRIER, 1988).

O primeiro registro sobre o café foi de Razes, um médico árabe do século X (SMITH, 1985), entretanto o cultivo de *C. arabica* teve início na Arábia no século V, especificamente no Yemen (CARVALHO, 1945). O início do cultivo de *C. arabica* foi caracterizado por sucessivas reduções na diversidade das variedades comerciais, onde poucas plantas foram transferidas da Etiópia ao Yemen pelos árabes. Após um longo período no Yemen, uma ou poucas plantas foram espalhadas ao redor do mundo, resultando na estreita base genética da espécie (FERWERDA, 1976) (Figura 1).

A cultivar *Typica* foi originada de uma única planta da Indonésia (CHEVALIER; DAGRON, 1928), da qual no final do século XVII os europeus levaram sementes de Mocha (Yemen) para ilha de Java, e no início do século XVIII foram introduzidas nos jardins botânicos de Amsterdam. Já *Bourbon* foi originada de poucas plantas introduzidas a partir do Yemen para a Ilha Bourbon, (atualmente conhecida como ilha La Reunion) por volta de 1717 (FERWERDA, 1976), na qual apenas duas sobreviveram. As populações de Amsterdam e La Reunion originaram então as duas variedades distintas que são a base das cultivares modernas (ANTHONY et al., 2001) (Figura 1).

Figura 1: Representação esquemática das principais etapas da história do cultivo do café.



Fonte: Anthony et al., 2002.

Sabe-se que o centro de origem de uma espécie é a sua maior reserva natural de diversidade, representando assim um importante patrimônio e recurso para pesquisas biológicas (SEIFU et al., 2007). Uma alternativa para as limitações no melhoramento de *C. arabica* existe no centro de origem da espécie, localizado nos altiplanos do sudoeste da Etiópia (ANTHONY et al., 1999). Assim, populações naturais são possíveis fontes de novos genes de resistência necessários para lidar com estresses bióticos e abióticos. (AGA et al., 2003).

Devido a importante base genética dos acessos de *C. arabica* da Etiópia, e com o objetivo da conservação de seus recursos genéticos, foram organizadas prospecções para a coleta do germoplasma do cafeeiro em regiões de ocorrência natural da espécie (FAO, 1968). Esta iniciativa resultou em uma coleta de 621 amostras

de sementes de vários pontos de coleta na Etiópia que foram documentadas e enviadas a seis instituições internacionais na Índia, Tanzânia, Etiópia, Costa Rica, Peru e Portugal (FAO, 1968). Amostras de sementes foram posteriormente repassadas ao Instituto Agrônomo de Campinas (IAC), que por sua vez forneceu 132 acessos ao IAPAR.

Desde então esta coleção de acessos tem sido avaliada fenotipicamente, onde uma grande variabilidade foi observada para níveis de cafeína, diterpenos, açúcares e ácidos clorogênicos (SILVAROLLA et al., 2004; POT et al., 2008; PAGIATTO, 2013); resistência a pragas e doenças como: *Colletotrichum kahawae* (VAN DER VOSSSEN & WALYARO, 1980); *Hemileia vastatrix* (VAN DER GRAAFF, 1981); *Meloidogyne incognita* (ANZUETO et al., 2001); diferenças quanto à produtividade (BERTRAND et al., 2005); assim como tamanho de folhas e frutos (POT et al., 2008).

Estudos quanto à análise de diversidade molecular também foram realizados com diversos tipos de marcadores: RAPD (LASHERMES et al., 1996; ANTHONY et al., 2001; AGA et al., 2003; CHAPARRO et al., 2004); AFLP e SSR (ANTHONY et al., 2002); ISSR (AGA et al., 2005); e SSR (SILVESTRINI et al., 2007).

## 2.2 MARCADORES MOLECULARES

Até a década de 60, os estudos genéticos eram realizados utilizando marcadores morfológicos de fácil identificação e geralmente regidos por um único gene (KNAPP, 1991). Entretanto, esses marcadores são frequentemente afetados pela ação gênica de dominância, efeito ambiental, pleiotropia e epistasia (PATERSON et al., 1991). Para resolver estes problemas, foram desenvolvidos marcadores isoenzimáticos, que são baseados na análise do produto da expressão de genes (MOSS, 1982). Apesar das suas vantagens sobre os marcadores morfológicos, como a co-dominância e insensibilidade a pleiotropia e epistasia, o uso destes marcadores ainda se tratava de um modo indireto de estudar os genes (BORÉM, 1998).

Posteriormente, as técnicas de biologia molecular foram se aprimorando e o DNA passou a ser o foco no desenvolvimento de novos marcadores, sendo inicialmente utilizados no melhoramento de plantas na década de 80 (SOLLER & BECKMANN, 1983). Os marcadores moleculares são definidos como características de DNA que diferenciam dois ou mais indivíduos (PÍPOLO & GARCIA, 2006), e são segmentos de DNA que estão fisicamente ligados a locos que determinam características de interesse (ALZATE-MARIN et al., 2005). Ao contrário dos marcadores

fenotípicos, possuem a vantagem de serem ambientalmente independentes e podem ser detectados em qualquer tipo de tecido e fase de desenvolvimento da planta (FERREIRA & GRATTAPAGLIA, 1998).

O desenvolvimento de marcadores moleculares veio dar um impulso maior às pesquisas genéticas, propiciando a detecção de polimorfismos de DNA com comportamento mendeliano, passível de serem utilizados em diferentes áreas da genética e do melhoramento de plantas (FERREIRA & GRATTAPAGLIA, 1998).

Têm sido muito utilizados em análises genéticas com as mais diversas finalidades e entre as suas principais aplicações nos programas de melhoramento de plantas, estão o monitoramento e organização da variabilidade genética e a seleção assistida por marcadores (SAM). Para auxiliar no melhoramento, os marcadores podem ser utilizados para estimativas de diversidade, fluxo gênico, taxa de cruzamento, parentesco, e na construção de mapas genéticos (LEE, 1995). Além disso, os marcadores têm grande aplicação na proteção de cultivares através da identificação de clones, linhagens, cultivares, híbridos e paternidade (SALLES et al., 2003).

Atualmente os distintos tipos de marcadores moleculares diferenciam-se pela tecnologia utilizada para revelar variabilidade no DNA, pela expressão genética, distribuição no genoma, custo de implementação e operação (GRIFFITHS et al., 2000). De acordo com as técnicas de detecção, os marcadores moleculares podem ser classificados em três classes: baseados em hibridização; em PCR (Reação em Cadeia da Polimerase), e em sequenciamento (BORÉM; FRITSCHÉ-NETO, 2013).

Entre os marcadores moleculares identificados por hibridização estão os RFLP (*Restriction Fragment Length Polymorphisms*) (BOSTEIN et al., 1980) minissatélites ou VNTR (Variable Number of Tandem Repeats) (JEFFREYS et al., 1985), e DArT (BORÉM; FRITSCHÉ-NETO, 2013). Com o advento da técnica de PCR, concebida por Kary Mullis em meados da década de 80, acelerou-se o desenvolvimento de novas classes de marcadores. Assim, os marcadores baseados na síntese *in vitro* de cópias de DNA incluem os marcadores RAPD (*Random Amplified Polymorphic DNA*) (WILLIAMS, 1990); SCAR (*Sequence Characterized Amplified Regions*); STS (*Sequence Tagged Sites*) (PARAN & MICHELMORE, 1993); SSR (*Simple Sequence Repeats*) (LITT & LUTTY, 1989); AFLP (*Amplified Fragment Length Polymorphism*) (VOS et al., 1995), e ISSR (*Inter Single Sequence Repeats*) (BORÉM; FRITSCHÉ-NETO, 2013). Com a popularização dos dados

genômicos e transcriptômicos, os marcadores SNP (*Single Nucleotide Polymorphisms*) (BORÉM & FRITSCHÉ-NETO, 2013) vem ganhando notoriedade, pois são as formas mais frequentes de variação genética (90%) constituindo um número ilimitado de marcadores potenciais em estudos de diversidade.

### 2.2.1 Marcadores Microsatélites (SSR)

Os cientistas descobriram o DNA satélite em 1960, quando centrifugaram o DNA em gradiente de densidade e verificaram que ele apresentava-se distribuído em duas ou mais faixas de densidade: uma banda principal contendo genes, e bandas secundárias que foram chamadas de bandas satélites. Essas bandas satélites mostraram-se como sendo constituídas de sequências de DNA repetitivas (GRIFFITHS et al., 2000).

No fim da década de 80, os marcadores SSR ou STR (*Short Tandem Repeats*) foram isolados e descritos simultaneamente por três cientistas como *loci* polimórficos presentes no DNA que consistem de sequências de 1 a 6 bases repetidas em tandem (LITT & LUTY, 1989; TAUZ, 1989). Eles foram inicialmente utilizados em pesquisas relacionadas a doenças degenerativas e neurológicas em humanos demonstrando grande aplicabilidade para outras espécies (HOSHINO et al., 2012).

Os SSR são classificados de acordo com a pureza e complexidade dos motivos ou unidades de repetição, podendo ser perfeitos, quando possuem o mesmo motivo  $(AT)_n$ ; imperfeitos, quando os motivos são interrompidos por nucleotídeos que não fazem parte da unidade de repetição  $(AT)_{12}GC(AT)_8$ , e compostos, quando há dois ou mais motivos diferentes em tandem  $(AT)_7(GC)_6$ . As repetições compostas podem ser perfeitas ou imperfeitas (SELKOE & TOONEN, 2006). São também classificados como mono, di, tri, tetra, penta ou hexanucleotídeos (BORÉM & FRITSCHÉ-NETO, 2013).

Para a análise do conteúdo polimórfico, variações de tamanho do produto amplificado estão relacionadas com o número das unidades de repetição e indicam o nível de polimorfismo em um loco específico na população (WEISING et al., 2005). Com isso, sabe-se que um maior número de unidades de repetição está associado a uma maior variação genotípica e que motivos SSR mais curtos como os

dinucleotídeos possuem geralmente mais repetições do que motivos mais extensos como tetranucleotídeos (ZALAPA et al., 2012).

Vários modelos de dinâmica evolutiva dos microssatélites foram propostos, sendo que a maioria deriva do modelo de mutação definido como *stepwise* (ELLEGREN, 2004) que ocorre por *slippage* ou “escorregamento” da DNA polimerase. Taxas de mutação nestas sequências repetitivas também ocorrem devido ao *crossing over* desigual durante a meiose, e retrotransposição (BORÉM & FRITSHE-NETO, 2013). Essas alterações no DNA causam um pareamento incorreto dos microssatélites durante a replicação resultando em cópias herdáveis de DNA com tamanhos diferentes (LEVINSON & GUTMAN, 1987).

Esses arranjos repetitivos em tandem passam por uma taxa elevada de  $10^{-7}$  a  $10^{-3}$  mutações por loco por geração (BUSCHIAZZO & GEMMELL, 2006). Devido a esses erros que podem ocorrer durante a replicação do DNA, diferentes indivíduos de uma mesma espécie podem apresentar um número variado de repetições dentro de um mesmo loco, os quais constituem diferentes alelos (TAUZ, 1989), tornando marcadores úteis em análises genéticas.

Os SSR estão presentes em regiões codificadoras e não codificadoras do genoma de eucariotos e procariotos (MATULA & KYPR, 1999). Possuem grande vantagem no melhoramento, pois são informativos, codominantes, ambientalmente independentes e podem ser detectados em qualquer tipo de tecido e fase de desenvolvimento da planta (FERREIRA & GRATAPAGLIA, 1998). Uma das características mais importantes dos marcadores SSR é que eles apresentam múltiplos alelos por loco, e essa hipervariabilidade faz com que sejam mais informativos do que os marcadores SNP (Polimorfismo de um único nucleotídeo), que são geralmente bialélicos (HAMBLIN et al., 2007).

As sequências de DNA que flanqueiam os microssatélites são geralmente conservadas entre os indivíduos de uma mesma espécie, ou até mesmo entre espécies geneticamente relacionadas, o que permite selecionar oligonucleotídeos iniciadores (*primers*) específicos que amplificam via PCR (*Polimerase Chain Reaction*) fragmentos contendo o DNA repetitivo em todos os genótipos (BORÉM & CAIXETA, 2009). Assim, estes marcadores podem ser compartilhados entre laboratórios como sequências de *primers*, melhorando os esforços de cooperação em pesquisa e desenvolvimento, atuando como regentes

universais de mapeamento genético, transportáveis entre pedigrees, populações, e entre espécies do mesmo gênero (BRONDANI et al., 1998).

### 2.2.2 Marcadores SSR em *Coffea*

Os marcadores microssatélites têm demonstrado eficiência em avaliações de diversidade genética e estudos filogenéticos para o café (MONCADA & MCCOUCH, 2004; MALUF et al., 2005; SILVESTRINI et al., 2007). Entretanto, a disponibilidade de SSR polimórficos ainda é limitante para a espécie *C. arabica* (MISSIO et al., 2011).

Cento e cinquenta marcadores microssatélites foram descritos para as espécies do gênero *Coffea* até o ano de 2003 (BARUAH et al., 2003), e um total de 224 microssatélites foram descritos até 2008 (HENDRE et al., 2008). Neste mesmo ano, 37 SSR haviam sido mapeados para a espécie *C. canephora* (HENDRE et al., 2008) e nenhum para *C. arabica*. Até o ano de 2009, trezentos e cinquenta e três marcadores SSR foram relatados para o café, dos quais 165 (47%) foram obtidos para *C. canephora* e 188 (53%) para *C. arabica* (MISSIO et al., 2009).

Esses resultados demonstram a necessidade de ampliar o número destes marcadores altamente informativos para uma gestão eficiente e melhoria dos recursos de germoplasma do café (BARUAH et al., 2003).

### 2.2.3 Desenvolvimento e Busca de Marcadores SSR

O desenvolvimento dos marcadores SSR tornou-se acessível por volta de 1996, principalmente devido às estratégias de enriquecimento de bibliotecas de DNA (POWELL et al., 1996). A descoberta de *loci* SSR requeria previamente a construção de bibliotecas enriquecidas para sequências SSR seguido pelo sequenciamento dos clones e análise das sequências (EDWARDS et al., 2009). Significativos esforços, experiência e recursos financeiros eram necessários para o bom desenvolvimento de marcadores microssatélites por esta técnica, pois são necessários em grande número para diferentes estudos do gênero de interesse (ZALAPA et al., 2012).

Atualmente, as tecnologias NGS (*Next Generation sequencing*) tem proporcionado um novo método tanto para mapeamento genético quanto para quantificação de transcriptoma (WANG; GERSTEIN; SNYDER, 2009), permitindo o sequenciamento do genoma expresso (RNA-Seq) ou de todo o genoma (MIKHEYEV et al., 2010). Estas tecnologias de sequenciamento de alto rendimento não exigem a criação de bibliotecas, produzem rapidamente uma grande quantidade de sequências com custos reduzidos do que outros métodos, fornecendo confiabilidade e a rápida caracterização das cultivares (HENDRE et al., 2008). As duas principais tecnologias NGS utilizadas para a descoberta de *loci* SSR são os sequenciadores 454 da Roche (Pirosequenciamento) e *Genomic Analyser* da Illumina (Sequenciamento por síntese).

Uma das aplicações atuais e emergentes das tecnologias NGS é o rápido desenvolvimento e o baixo custo de marcadores moleculares (EKBLÖM & GALINDO, 2011). Estudos recentes demonstraram o uso eficiente de NGS em transcriptoma ou sequências de genoma para a descoberta em grande escala de *loci* SSR em plantas (ZALAPA et al., 2012), e a principal vantagem do desenvolvimento desses marcadores a partir do RNA-Seq é a possibilidade de encontrar associações com genes funcionais e, portanto com fenótipos (LI et al., 2002).

Os vetores BAC (*Bacterial Artificial Chromosome*) permitem a inserção de grandes fragmentos de DNA, com o tamanho típico de um fragmento de 80 a 200 Kb (PETERSON et al., 2000). Possuem estabilidade pela presença do fator F, o que impede que mais de um BAC habite simultaneamente uma mesma bactéria (YUKSEL & PATERSON, 2005), e são relativamente fáceis de manipular e de se propagarem. São os vetores mais utilizados para o sequenciamento de genomas seguido de rearranjos *in silico*, principalmente quando não há um genoma de referência disponível, pois fragmentos maiores facilitam a montagem do genoma (WICKER et al., 2006).

Com o advento da nova geração de sequenciadores no ano de 2005, houve conseqüentemente um aumento exponencial na quantidade de dados gerados e de sequências submetidas aos bancos públicos (VARSHNEY, 2009). A fim de analisar o grande volume de dados gerados, os computadores tornaram-se instrumentos fundamentais para pesquisas biológicas, culminando no surgimento da Bioinformática (LUSCOMBE et al., 2001).

Junto à NGS foram criados *softwares* de busca de marcadores SSR fornecendo uma identificação mais fácil. Essas ferramentas computacionais foram rotineiramente adotadas e se tornaram o método padrão na descoberta de polimorfismos genéticos (EDWARDS et al., 2009). Em conjunto, estas ferramentas vêm proporcionando um grande avanço nas pesquisas biológicas, sendo possível a obtenção e a utilização de marcadores moleculares em estudos de variabilidade genética de forma mais eficiente (VARSHNEY et al., 2009).

### 3 OBJETIVO GERAL

Identificar marcadores SSR e validá-los em um painel representativo do centro primário de origem de *C. arabica* visando estudos de diversidade genética.

#### 3.1 OBJETIVOS ESPECÍFICOS

- Realizar uma revisão de literatura para obter um maior número de *primers* SSR no estudo.
- Realizar análise *in silico* para a busca de marcadores SSR em dados de sequenciamento de transcriptoma (RNA-seq) em *C. arabica* cv. Iapar 59, e em dados de sequências genômicas de BACs de Híbrido de Timor 832/2.
- Identificar e caracterizar os motivos SSR encontrados.
- Validar os SSR em um painel representativo da coleção selvagem de *C. arabica* proveniente da Etiópia.
- Desenvolver marcadores SSR informativos para estudos de diversidade genética.
- Aumentar o conhecimento sobre a estrutura dos acessos selvagens de *C. arabica* presentes no IAPAR.

## 4 MATERIAIS E MÉTODOS

### 4.1 BUSCA DE MICROSSATÉLITES

Inicialmente, foi realizado um levantamento bibliográfico dos marcadores SSR mais informativos para *C. arabica* e/ou *C. canephora* que ainda não haviam sido utilizados em trabalhos anteriores no Laboratório de Biotecnologia do IAPAR.

Também foi realizada uma análise *in silico* de sequências repetitivas em dados originados a partir de sequenciamento NGS de transcriptoma (Clapar 59) e de genoma (CHT 832/2).

Em trabalhos anteriores no laboratório, foi realizado o sequenciamento de Híbrido de Timor 832/2 utilizando a tecnologia Illumina HISEQ 2.000 pelo método *paired end* no Centro de Ciências Genômicas (Universidade Carolina do Norte – EUA). Vinte BACs foram submetidos ao sequenciamento *shotgun* em uma linha do Illumina. Um BAC montado com 179 *contigs* (120 Kb) foi utilizado no estudo.

Para a cultivar Iapar 59 foi realizado RNA-seq utilizando o sequenciamento Illumina HISEQ 2.000, método *single end* pelo Centro de Ciências Genômicas (Universidade Carolina do Norte – EUA). Para esta finalidade, 12 amostras de diferentes tecidos de Iapar 59 foram sequenciadas em uma linha. Um total de 13.596 *contigs* foram montados e 77 *contigs* com um tamanho de 605 Kb foram analisados.

A identificação das sequências repetitivas ou motivos SSR foi realizada pela ferramenta *tool SSRIT (Simple Sequence Repeat Identification Tool)* desenvolvido por Cartingour e utilizado por Poncet et al. (2006) e Baruah et al. (2003). Os parâmetros do programa foram definidos para a detecção de motivos di-, tri-, tetra-, penta- e hexanucleotídeos com número mínimo de 3 repetições. Quanto ao tamanho da repetição, os motivos foram filtrados para anotação com a unidade mínima de repetição de 5 unidades de repetição para di-, 4 unidades de repetição para tri-, e 3 unidades de repetição para tetra-, penta e hexanucleotídeos.

Os SSR foram caracterizados de acordo com a unidade de repetição e a frequência com que os motivos ocorreram em ambos os genótipos.

#### 4.2 DESENHO DOS *PRIMERS* PARA AMPLIFICAÇÃO DOS SSR

Após a análise *in silico*, os *contigs* que continham sequências repetitivas foram utilizados para o desenho dos *primers* específicos complementares a cada região flanqueadora dos motivos SSR utilizando o programa Primer 3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) (ROZEN & SKALETSKY, 2000), como descrito por Varshney et al. (2002) e Baruah et al. (2003). Os critérios utilizados para o desenho dos *primers* foram temperatura média de anelamento ( $T_m$ ) entre 56° e 60°C, conteúdo de GC entre 30% e 60%, *amplicons* de 70 a 300 nucleotídeos, e *primers* com o tamanho de 18 a 22 pares de bases.

Pelo programa Oligo Calc: *Oligonucleotide Properties Calculator* (<http://www.basic.northwestern.edu/biotools/oligocalc.html>) foi realizada a escolha dos *primers* com base na não complementaridade entre eles.

#### 4.3 SELEÇÃO DOS MICROSSATÉLITES

Para a validação dos SSR, foram utilizados 58 marcadores SSR obtidos a partir da análise *in silico* de *C. arabica* cv. Iapar 59 e Híbrido de Timor 832/2 juntamente com 77 marcadores provenientes de revisão bibliográfica, onde foram anotados os marcadores SSR com maior conteúdo informativo em *C. arabica* e/ou *C. canephora* que ainda não haviam sido utilizados em trabalhos anteriores do no Laboratório de Biotecnologia do IAPAR. Os 135 *primers* SSR utilizados no estudo estão representados nas tabelas 1, 2 e 3 em anexo.

#### 4.4 MATERIAL VEGETAL

A validação foi realizada em um painel de 24 acessos representativos da diversidade existente na coleção de 132 acessos de *C. arabica* que se encontram disponíveis no IAPAR (Tabela 1).

Tabela 1. Painel com a ordem dos 24 acessos de *C. arabica* utilizados para seleção dos marcadores SSR e seus respectivos locais de coleta.

Origem	Kaffa	Illubador	Sidamo	Harar	Gojjam
Acessos	E044/122	E148/254	E022/163	E007/87	E565/10
	E516/069	E208/193	E021/011		
	E130/169	E370/196	E018/494.1		
	E131/018	E363/735	E238/22		
	E123A/231	E196/117	E237/171		
	E335/219	E454/107			
	E332/24	E087/194			
	E272/175	E464/113			
	E383/142				

Fonte: o próprio autor.

Os 24 acessos selecionados no estudo foram coletados em várias regiões da Etiópia, sendo nativos das províncias de Kaffa, Illubador, Gojjam do lado Oeste do Vale do Rift, e Sidamo e Harar do lado Leste do Vale do Rift (Figura 2).

Figura 2. Sítios de coleta dos acessos de *C. arabica*. Organização política da Etiópia de acordo com a FAO (1968), exceto pela separação da Eritreia.



Fonte: Silvestrini et al., 2007.

#### 4.5 EXTRAÇÃO DE DNA GENÔMICO

O DNA genômico foi isolado a partir de folhas de plantas adultas pelo método CTAB (DOYLE & DOYLE, 1990). Aproximadamente 5 g de folhas de café foram maceradas em nitrogênio líquido e acondicionadas em tubos falcon de 50 mL. Após a maceração seguiu-se os passos descritos por Brasileiro e Carneiro. (1998) com as modificações necessárias para evitar a oxidação do material e para purificar o DNA obtido, como adição de 0,4 mL de metabissulfito de sódio 10% ao tampão de extração e purificação com fenol-clorofórmio, respectivamente.

Após a extração a qualidade do DNA foi verificada em gel de agarose 0,8%, corado com brometo de etídeo e a quantificação estimada por espectrofotometria em NanoDrop com absorvância de 260 e 280 nm.

#### 4.6 CONSTRUÇÃO DE “*BULKS*” DE DNA

Para diminuir os custos e o tempo das análises, optou-se por utilizar a estratégia de *bulks* ou *pools* de DNA, uma adaptação da técnica de BSA (*Bulk Segregant Analysis*), e 15 amostras de DNA foram divididas em três *bulks*.

Os *bulks* de genótipos de *C. arabica* foram construídos com base nos dados de origem geográfica: um com oito plantas representando genótipos da região Oeste do Vale do Rift da Etiópia (***Bulk 1***), outro com cinco plantas representando genótipos da região Leste do Vale do Rift (***Bulk 2***), e um terceiro representado por um único acesso (E007/87) (***Bulk 3***), pois foi o que mais se destacou das demais em análises anteriores de genotipagem (FERREIRA, 2010) servindo como controle para esta estratégia (Tabela 2). Para cada marcador SSR foram realizadas três reações de amplificação, onde quantidades equimolares de DNA dos acessos de cada grupo foram combinados a uma concentração final de 100 ng/ $\mu$ L.

Tabela 2. Genótipos utilizados para formação dos *bulks* de *C. arabica*.

Genótipos B1	Sítio de Origem	Genótipos B2	Sítio de Origem	Genótipos B3	Sítio de Origem
E332/24	Kaffa	E018/494.1	Sidamo	E007/87	Harar
E123A/231	Kaffa	E022/163	Sidamo		
E516/069	Kaffa	E238/22	Sidamo		
E131/018	Kaffa	E021/011	Sidamo		
E272/175	Kaffa	E237/171	Sidamo		
E464/113	Illubador				
E454/107	Illubador				
E565/10	Gojjam				

Fonte: o próprio autor.

Os marcadores que apresentaram padrão polimórfico entre os grupos foram avaliados no painel de 24 genótipos descritos acima a fim de genotipar esta população. Os acessos individuais foram diluídos para uma concentração final de 5 ng/ $\mu$ L.

#### 4.7 AMPLIFICAÇÃO DOS SSR VIA PCR

As reações de amplificação dos *bulks* de DNA foram realizadas utilizando o Kit comercial *Go Taq Green Master Mix* da *Promega*, com um volume final de 10  $\mu$ L, contendo 5  $\mu$ L de Mix, 1  $\mu$ L de cada primer (1  $\mu$ M - forward e reverse), 2,0  $\mu$ L de DNA genômico e 1,0  $\mu$ L de água mili-Q autoclavada. Para o painel de 24 genótipos, as reações de amplificação foram conduzidas em um volume final de 10  $\mu$ L, contendo 5  $\mu$ L de DNA genômico, 2  $\mu$ L de tampão (5 X Green Go Taq® Flexi Buffer – *Promega*), 0,2  $\mu$ L de dNTPs (dNTP Mixture 10mM – *Amresco*), 0,6  $\mu$ L de  $MgCl_2$  (25 mM – *Promega*), 0,2  $\mu$ L de cada *primer* (10  $\mu$ M - forward e reverse), 0,1  $\mu$ L de Taq DNA polimerase a 1U (Go Taq® Flexi – *Promega*), e 1,7  $\mu$ L de água mili-Q autoclavada. As amplificações foram realizadas em termociclador GeneAmp® - PCR System 9700 (*Applied Biosystems*), com os seguintes parâmetros: 95° C por 2 minutos; seguido de 30 ciclos de 95° C por 50 segundos; 50° C por 1 minuto, 72° C por 1 minuto, e extensão final de 72°C por 5 minutos.

Os produtos de PCR foram submetidos à eletroforese em gel de poliacrilamida 10%, utilizando para a corrida TAE 1x (Tris base, ácido acético glacial e EDTA 500 mM, pH: 8,0), voltagem de 220 e coloração por brometo de etídio. Os géis foram visualizados sob luz ultravioleta e capturados pelo sistema digital Kodak

KS 120. O tamanho molecular dos produtos amplificados foi estimado utilizando uma escala de DNA de 50 pb como um padrão (*Ludwig biotec*).

#### 4.8 ANÁLISE DOS DADOS

A informação genética de cada loco SSR foi avaliada pelo número de alelos por loco, frequência alélica por loco e PIC (*Polymorphism Information Content* - PIC). O número médio de alelos por marcador foi obtido pela soma de todos os alelos amplificados por loco. A frequência alélica por loco foi calculada dividindo o número de cada alelo pelo número total de alelos amplificados em cada genótipo, e o PIC para cada marcador foi obtido pela fórmula:

$$PIC = 1 - \sum_{j=1}^n p_{ij}^2$$

onde  $p_{ij}$  representa a frequência do  $j$ -ésimo alelo para o  $i$ -ésimo primer (ANDERSON et al., 1993). Esta é uma estimativa da diversidade genética, indicando o poder de discriminação genética de cada marcador em um teste de identidade (WEIR, 1990).

#### 4.9 DIVERSIDADE E ESTRUTURA GENÉTICA

Os dados alélicos dos marcadores foram utilizados para determinar a diversidade genética e as relações filogenéticas entre os 24 acessos testados. Apesar dos microssatélites serem codominantes, a análise molecular de cada acesso foi baseada na presença/ausência de cada fragmento amplificado devido à natureza tetraplóide de *C. arabica*. Esta estratégia foi adotada pela dificuldade de distinguir entre dialelos duplex e simplex e entre os diferentes tipos de combinações de trialelos nos *loci* SSR (SILVESTRINI et al., 2007). No entanto, na avaliação dos géis cada banda ou alelo foi tratado como um caráter independente, com a presença da banda designada por 1 e ausência por 0 (zero) como realizado por AGA et al. (2003), AGGARWAL et al. (2007), e SILVESTRINI et al. (2007), sendo os dados

utilizados para a construção de uma matriz binária demonstrando o perfil dos SSR em cada genótipo.

Para realizar a análise de diversidade genética, uma matriz de dissimilaridade foi computada a partir dos alelos de cada marcador SSR. Esta matriz foi utilizada para a construção de um dendograma de diversidade global utilizando o método de agrupamento (UPGMA) (SNEATH & SOKAL, 1973), implementado no *software* DARwin 5 (PERRIER et al., 2003). A distância genética entre os acessos foi estimada com base no coeficiente de Dice (DICE, 1945). Um *bootstrap* de 1.000 foi utilizado para avaliar a confiabilidade da topologia da árvore.

A Análise de Coordenadas Principais (PCO) baseada nas distâncias genéticas de Nei. (1978) foi realizada pelo programa FAMD 1.31 (SCHLÜTER et al., 2006). Foi utilizada também uma abordagem de agrupamento *bayesiano* descrito por Pritchard et al. (2000), implementado no *software* STRUCTURE 2.3.4. O STRUCTURE foi executado utilizando o método estatístico  $\Delta K$  (EVANNO et al., 2005), inferindo  $K=5$ , Modelo *Admixture*, 20 corridas para cada valor de  $K$ , aplicando 10.000 ciclos burnin e 10.000 interações de Markov Chain Monte Carlo (MCMC) para a análise.

## 5.0 RESULTADOS E DISCUSSÃO

### 5.1 BUSCA DOS SSR POR ANÁLISE *IN SILICO*

Dos 13.596 *contigs* gerados pela montagem dos dados de RNA-seq para lapar 59, 77 *contigs* foram analisados e 21 pares de *primers* foram desenhados. Para o genótipo CHT 832/2, 179 *contigs* de um BAC foram analisados gerando 37 pares de *primers*.

Os motivos mais frequentes obtidos da análise *in silico* para ambos genótipos foram di-, tri- e tetranucleotídeos respectivamente. Para lapar 59, os dinucleotídeos foram mais frequentes com 82% do total dos motivos identificados, enquanto que para CHT 832/2, os motivos mais frequentes foram di e trinucleotídeos representando 63 e 22% (Tabela 3).

Pode-se observar que para o genótipo CHT832/2 obtivemos 5 vezes mais motivos SSR encontrados do que para lapar 59. Isso devido o genótipo Híbrido de Timor possuir maior variabilidade genética em relação ao lapar 59, que é uma variedade melhorada.

Tabela 3. Frequência dos motivos SSR encontrados em contigs de lapar 59 e CHT832/2.

	lapar 59		CHT 832/2	
	Nº	%	Nº	%
Di-SSR	9	82	37	63
Tri-SSR	2	18	13	22
Tetra-SSR	---	---	9	15
Total	11	100	59	100

Clapar 59- RNA-seq; CHT832/2- BAC genômico. **Fonte:** o próprio autor.

Os resultados da análise *in silico* reforçam os observados por Morgante & Olevieri. (1993), que descreveram que as repetições di e tri são amplamente distribuídas em plantas. Em *Coffea spp*, Aggarwal et al. (2007) e Missio et al. (2009) relataram que em geral os marcadores SSR desenvolvidos foram principalmente compostos por motivos di e trinucleotídeos. Marcadores dessa natureza são mais frequentes e polimórficos do que tetra-, penta- ou hexanucleotídeos (PONCET et al., 2004).

Para ambos os genótipos, os motivos mais frequentes quanto à natureza de repetição foram das classes  $(AT)_n$ ,  $(TA)_n$ ,  $(TC)_n$  e  $(GA)_n$ , e o tamanho dos motivos variaram de 3 a 13 unidades de repetição (Tabela 4).

Tabela 4. Caracterização dos motivos SSR quanto à natureza e tamanho da unidade de repetição encontrados para Iapar 59 e CHT 832/2.

Iapar 59				CHT 832/2											
Tipo de motivo SSR	Número de SSR	%	Tamanho da unidade de repetição	Tipo de motivo SSR	Número de SSR	%	Tamanho da unidade de repetição								
TA	2	18.2	5 5 -	TA	9	15.2	5	5	5	5	5	6	7	9	13
AT	2	18	5 8 -	AT	6	10.2	5	6	7	7	9	12	-	-	-
TC	3	27.2	5 6 6	AG	6	10.2	5	5	5	6	7	8	-	-	-
GA	2	18.2	5 8 -	TC	4	6.8	5	5	5	9	-	-	-	-	-
ATC	2	18.2	4 5 -	GA	4	6.8	5	8	8	8	-	-	-	-	-
				AAAT	4	6.8	3	3	4	5	-	-	-	-	-
				TTA	3	5.1	4	4	4	-	-	-	-	-	-
				ATTT	3	5.1	3	3	3	-	-	-	-	-	-
				CT	2	3.4	6	8	-	-	-	-	-	-	-
				GT	2	3.4	9	9	-	-	-	-	-	-	-
				CA	2	3.4	11	12	-	-	-	-	-	-	-
				AC	2	3.4	7	13	-	-	-	-	-	-	-
				GTG	2	3.4	4	6	-	-	-	-	-	-	-
				TGC	2	3.4	4	4	-	-	-	-	-	-	-
				ATT	2	3.4	5	5	-	-	-	-	-	-	-
				CCA	2	3.4	4	5	-	-	-	-	-	-	-
				GAT	2	3.4	4	5	-	-	-	-	-	-	-
				TAAA	2	3.4	3	3	-	-	-	-	-	-	-
Total	11			Total	59										

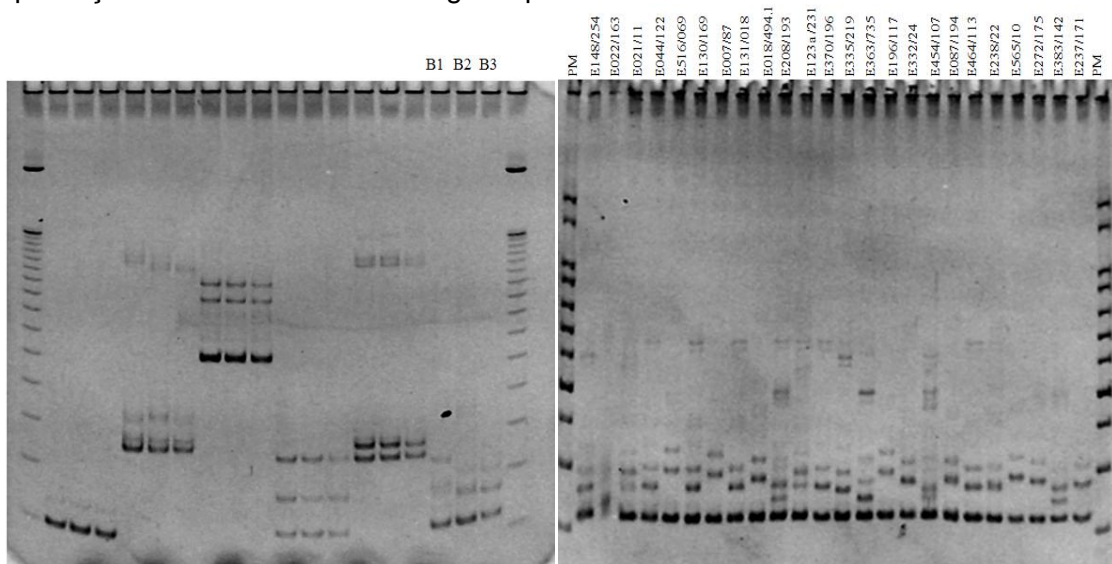
Clapar 59-RNA-seq; CHT 832/2-genômico. **Fonte:** o próprio autor.

Pereira et al. (2011), buscaram marcadores SSR a partir de banco de dados de EST-SSR de café com o objetivo de descobrir locos SSR associados à resistência ao bicho mineiro (*Leucoptera coffeella*). Do total dos motivos SSR obtidos da análise *in silico*, 35,83% dos SSR eram di- e 39,02% dos SSR resultaram em trinucleotídeos. As classes  $(AT)_n$ ,  $(TC)_n$ ,  $(GA)_n$ , e  $(TA)_n$  foram também mais frequentes.

## 5.2 SELEÇÃO DOS *PRIMERS* POLIMÓRFICOS

Os 58 pares de *primers* obtidos da análise *in silico* mais os 77 da revisão de literatura foram analisados em 15 acessos da coleção da Etiópia divididos em três *bulks*. Quando ocorria a visualização de alelos polimórficos entre os *bulks*, os SSR eram posteriormente avaliados no painel de 24 acessos selecionados no estudo (Figura 3).

Figura 3: Gel de Poliacrilamida revelando o polimorfismo dentro de *Coffea arabica*. A - Identificação de marcadores polimórficos pela estratégia de *bulks* de DNA (B1, B2 e B3 = grupos). A seta indica o polimorfismo revelado pelo marcador **Ca61**. B - Perfil de amplificação do marcador **Ca61** em genótipos individuais.



PM- peso molecular de 50 pb. **Fonte:** o próprio autor.

Dos 135 *primers* SSR desenhados, 31 *primers* foram polimórficos com padrões de bandas reprodutíveis. Destes 31 *primers*, 21 são compostos por motivos di-, 7 por motivos tri-, e 3 por tetranucleotídeos.

Todos os 77 *primers* SSR desenhados a partir da revisão bibliográfica foram testados pela estratégia de *bulks* de DNA. Destes, 47 SSR foram monomórficos e 24 foram polimórficos no painel de 24 acessos da Etiópia. Para os 37 *primers* SSR desenhados para CHT 832/2, todos foram testados em *bulks*, 22 *primers* foram monomórficos e 6 foram polimórficos no painel de 24 acessos da Etiópia. E dos 21 SSR obtidos e testados em *bulks* na cultivar Iapar 59, 9 dos SSR foram monomórficos, e apenas um *primer* (Clapar 14) demonstrou-se polimórfico no painel (Tabela 6).

Tabela 6. Número de *primers* SSR polimórficos e alelos nos acessos da Etiópia.

	Nº	<i>Primers</i> analisados em <i>bulks</i>	<i>Primers</i> monomórficos	<i>Primers</i> não amplificados	<i>Primers</i> analisados em painel	Polimorfismos no painel	%	Número de alelos
<i>Primers</i> revisão bibliográfica	77	77	47	6	24	24	31	141
<i>Primers</i> análise <i>in</i> <i>silico</i> CHT 832/2	37	37	22	5	10	6	16	22
<i>Primers</i> análise <i>in</i> <i>silico</i> Iapar 59	21	21	9	3	9	1	5	2
Total	135	135	25	13	67	31	100	165

Nº- número total de *primers* utilizados no estudo. **Fonte:** o próprio autor.

O sucesso desta estratégia é observado em outras culturas como a soja, onde Rabel et al. (2010) demonstraram que as análises em “*bulks*” reduziram em cinco vezes o número das reações necessárias para a realização da genotipagem com uma conseqüente diminuição de custos, tempo e reagentes. Em estudos anteriores em nosso laboratório, de 292 *primers* utilizados, 48 marcadores microssatélites polimórficos foram obtidos por esta estratégia. Desses, quatorze SSR apresentaram 2 a 5 alelos em genótipos que representam parte da diversidade da coleção da Etiópia (FERREIRA, 2010).

Pode-se observar com a utilização de *bulks* de DNA que houve uma redução de custos e tempo, onde em uma única placa de 96 poços foi possível realizar PCR de 32 microssatélites, em 15 acessos, dentre os três *pools* de DNA. Em relação aos géis, a análise também foi facilitada, pois em um único gel de poli-acrilamida foi possível avaliar a presença de oito microssatélites, permitindo a análise de 16 marcadores por dia.

### 5.3 ANÁLISE DA DIVERSIDADE GENÉTICA DOS 24 ACESSOS DA ETIÓPIA

Após a análise dos *bulks*, os 31 SSR polimórficos foram analisados no painel de 24 genótipos da coleção da Etiópia.

Foi observado um total de 165 alelos com 2 a 11 alelos por loco detectado, e uma média de 5,3 alelos SSR. (Tabela 7). Missio et al. (2009) sequenciaram bibliotecas enriquecidas e desenharam 96 *primers* SSR, que posteriormente foram testados em cultivares de *C. arabica*. Vinte e um SSR foram polimórficos com um total de 1 a 4 alelos. O marcador SSR Ca81 apresentou um alelo no estudo, onde em nossos resultados foram obtidos sete alelos. Este resultado demonstra a baixa variabilidade genética das variedades comerciais em relação aos materiais selvagens da Etiópia.

Tabela 7. Número de alelos, frequências alélicas e conteúdo polimórfico dos 31 marcadores microssatélites obtidos a partir dos perfis genéticos de 24 acessos da Etiópia.

Marcador	Nº de alelos	Alelos	f.a	PIC
CM2	6	210	0,10	0,8
		214	0,14	
		220	0,31	
		240	0,33	
		244	0,02	
		260	0,10	
CFG792b2	4	170	0,27	0,7
		180	0,27	
		200	0,23	
		280	0,23	
CFCA2812	5	150	0,31	0,7
		152	0,33	
		168	0,01	
		170	0,29	
		172	0,06	
CFCA331 <sup>2</sup>	6	130	0,48	0,7
		136	0,02	
		138	0,23	
		140	0,06	
		150	0,18	
		154	0,02	
CFCA360 <sup>2</sup>	4	160	0,31	0,7
		168	0,23	
		176	0,30	
		198	0,16	
CaM03 <sup>b</sup>	3	170	0,48	0,5
		174	0,02	
		200	0,50	
M24	11	140	0,02	0,9
		150	0,02	
		160	0,13	
		164	0,11	
		168	0,20	
		170	0,14	
		172	0,09	
		174	0,04	
		196	0,06	
		198	0,05	
		200	0,14	

Tabela 7, Continuação...

Marcador	Nº de alelos	Alelos	f.a	PIC
M47	6	130	0,46	0,7
		140	0,22	
		148	0,03	
		150	0,21	
		160	0,06	
		170	0,01	
SSRCa 002	7	252	0,09	0,7
		258	0,11	
		260	0,09	
		262	0,04	
		264	0,53	
		270	0,04	
SSRCa052	2	135	0,56	0,5
		150	0,44	
SSRCa 085	4	80	0,56	0,6
		84	0,02	
		90	0,20	
		100	0,22	
SSRCa 091	6	70	0,01	0,7
		80	0,01	
		90	0,26	
		100	0,31	
		110	0,31	
		120	0,10	
SSRCa 081	7	142	0,02	0,8
		150	0,23	
		160	0,30	
		164	0,30	
		170	0,04	
		180	0,04	
		200	0,06	
LEG11	4	150	0,38	0,6
		156	0,01	
		171	0,42	
		180	0,19	

Tabela 7, Continuação...

Marcador	Nº de alelos	Alelos	f.a	PIC
LEG12	8	140	0,31	0,8
		160	0,05	
		178	0,18	
		180	0,10	
		182	0,01	
		200	0,16	
		210	0,04	
		240	0,16	
LEG13	5	150	0,33	0,7
		160	0,33	
		170	0,01	
		180	0,26	
		186	0,06	
LEG32	7	185	0,02	0,8
		191	0,31	
		200	0,10	
		203	0,31	
		206	0,13	
		209	0,04	
		230	0,08	
M32	9	110	0,32	0,8
		120	0,05	
		130	0,25	
		135	0,03	
		145	0,12	
		147	0,06	
		155	0,09	
		160	0,05	
		170	0,03	
Sem nome	7	170	0,47	0,7
		228	0,01	
		230	0,12	
		233	0,13	
		236	0,16	
		239	0,02	
		251	0,08	

Tabela 7, Continuação...

Marcador	Nº de alelos	Alelos	f.a	PIC
AJ250253	9	260	0,11	0,8
		264	0,12	
		266	0,11	
		270	0,02	
		290	0,07	
		298	0,27	
		300	0,12	
		310	0,13	
		312	0,04	
AJ250256	4	146	0,30	0,7
		148	0,26	
		150	0,19	
		158	0,25	
DCM01	5	260	0,38	0,7
		262	0,17	
		264	0,38	
		270	0,04	
		380	0,04	
CaM41	6	170	0,72	0,5
		178	0,05	
		180	0,09	
		200	0,06	
		228	0,03	
		230	0,04	
CFGA249	5	244	0,02	0,6
		250	0,02	
		248	0,04	
		350	0,42	
		352	0,42	
		400	0,08	
Clapar14	2	270	0,44	0,5
		276	0,56	
CHT03	3	170	0,45	0,6
		200	0,32	
		210	0,23	
CHT04	3	350	0,36	0,7
		402	0,36	
		441	0,28	

Tabela 7, Continuação...

Marcador	Nº de alelos	Alelos	f.a	PIC
CHT12	4	110	0,47	0,6
		114	0,07	
		140	0,39	
		144	0,06	
CHT25	2	250	0,79	0,3
		262	0,21	
CHT34	2	192	0,54	0,5
		194	0,46	
CHT35	8	142	0,23	0,7
		146	0,40	
		234	0,02	
		240	0,08	
		250	0,18	
		254	0,02	
		260	0,07	
		262	0,01	
<b>Média</b>	<b>5,3</b>	<b>165</b>	<b>0,2</b>	<b>0,7</b>

f.a= frequência alélica; PIC= *Polymorphism Information Content*.

Fonte: o próprio autor.

As frequências alélicas variaram de 0,01 a 0,8 com 78 dos alelos (47,3%) com frequências entre 0,2 a 0,8, e 87 (52,7%) dos alelos considerados raros pela baixa frequência (0,2). Do total dos alelos identificados, um alelo obteve alta frequência (0,8). Silvestrini et al. (2007), identificaram 54 alelos em 73 acessos de *C. arabica* da Etiópia, e desses, 31 foram considerados raros, 16 com frequências entre 0,2 e 0,8, e sete dos alelos foram encontrados em alta frequência.

A alta quantidade de alelos raros encontrados em nosso estudo deve se dar em função da variabilidade genética da população utilizada, pois é uma população selvagem, com maior diversidade genética e, portanto, com maior probabilidade de serem detectados novos alelos, aumentando a informatividade genética de cada marcador.

O valor de PIC é de fundamental importância para determinar o nível de polimorfismo dentro de uma população, ou seja, quanto maior o valor de PIC maior será a informação polimórfica obtida. Ele varia de zero para o loco monomórfico, a um para o loco com muitos alelos com frequências similares. Os valores de PIC variaram de 0,3 (CHT25) a 0,8 (M24), com média de 0,7. Dos 31 *primers*, 20 apresentaram valores de PIC superior a 0,7 demonstrando ser

informativos para estudos de diversidade genética. Essas análises permitem identificar entre os marcadores testados, aqueles geneticamente mais informativos para testes de identidade genética. Apesar de alguns marcadores terem demonstrado valores de 0,3 a 0,5 de PIC (CaM03<sup>b</sup>, SSRCa052, CaM41, Clapar14, CHT25), é importante observar que as análises realizadas em gel de poliacrilamida possuem menor resolução comparada à eletroforese capilar, o que pode ter interferido na observação dos alelos e conseqüentemente na obtenção do PIC.

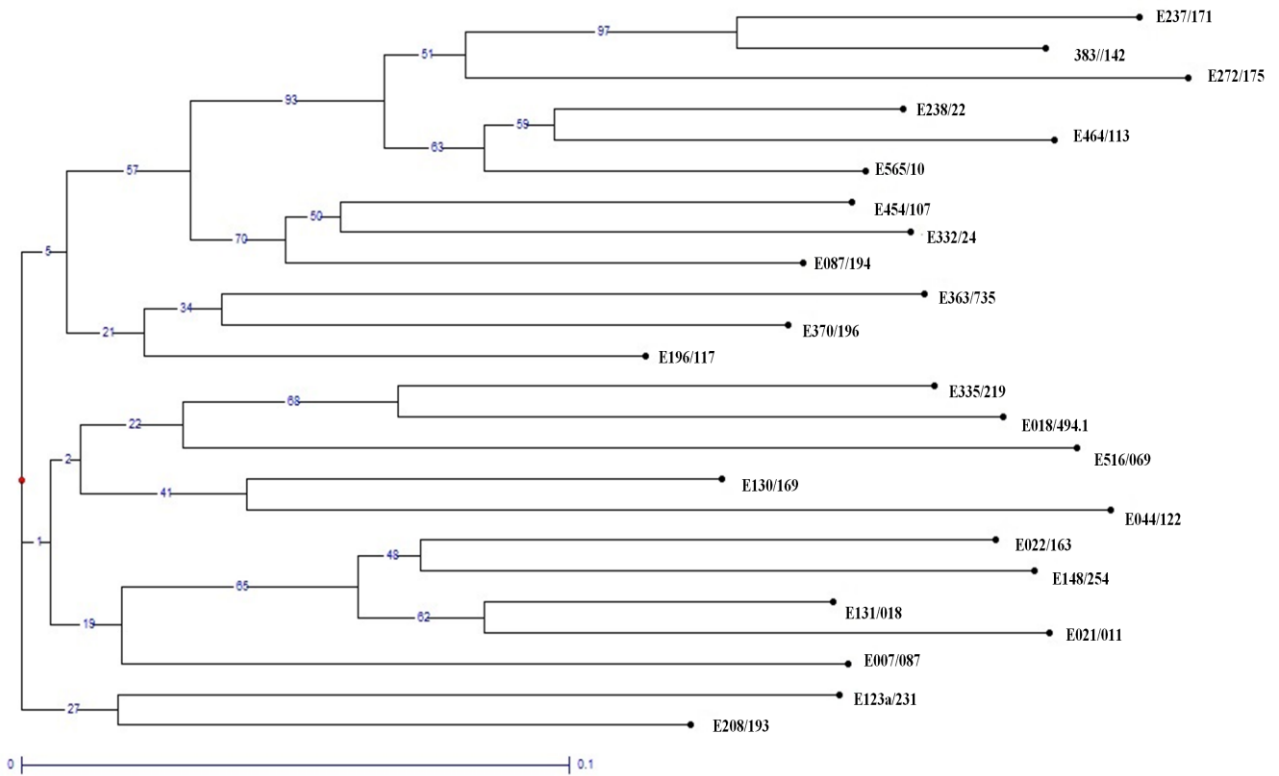
Um ponto importante para os marcadores polimórficos obtidos da análise *in silico* foi a origem dos dados, onde os SSR desenvolvidos para o genótipo CHT 832/2 foram de sequenciamento de BACs genômicos, enquanto que para lapar 59 foram analisados dados de transcriptoma (RNA-seq). Assim, neste estudo seis marcadores SSR de CHT 832/2 foram validados no painel e apenas um marcador SSR obtido de lapar 59 foi polimórfico. Portanto um nível inferior de polimorfismo observado nos SSR obtidos de dados de transcriptoma pode ocorrer devido a alta conservação de sequências de DNA em regiões transcritas no genoma, o que não ocorre com os SSR genômicos, que são amplamente distribuídos (VARSHNEY et al., 2005), principalmente em regiões intergênicas.

Missio et al. (2011) utilizaram marcadores SSR genômicos e EST-SSR já descritos na literatura para avaliar a diversidade genética de 24 acessos de café. Maiores porcentagens de polimorfismos entre os marcadores obtidos de dados genômicos em comparação com os EST-SSR foram observados, demonstrando a sua eficiência em diferenciar todos os acessos de *C. arabica* estreitamente relacionados.

#### 5.4 ESTRUTURA GENÉTICA DOS ACESSOS

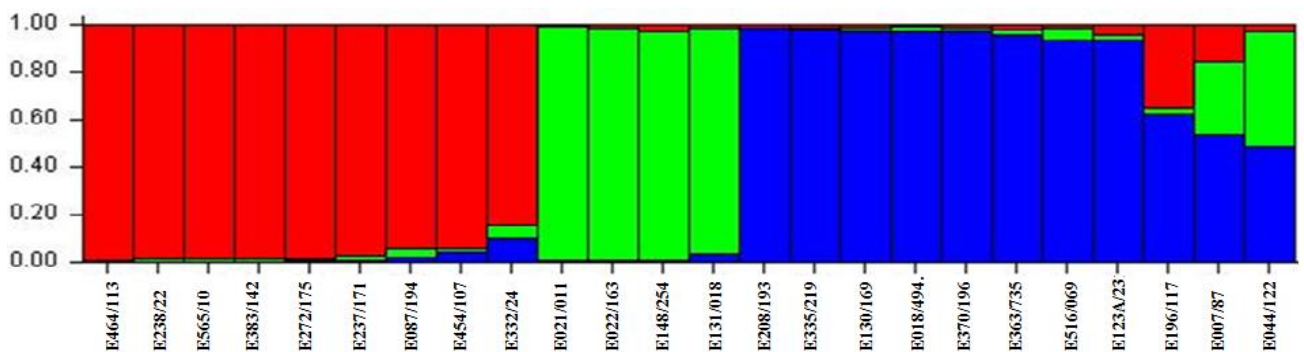
A ancestrabilidade entre os acessos observados através do dendograma de diversidade global (Figura 3) gerou três principais grupos que se distinguiram entre os acessos tetraplóides, fato corroborado pela análise da estrutura populacional ( $K=5$ ) (Figura 4) e pela PCO (Figura 5).

Figura 3. Dendograma de diversidade global entre os 24 acessos de *C. arabica* com base no coeficiente de Dice utilizando o método UPGMA.



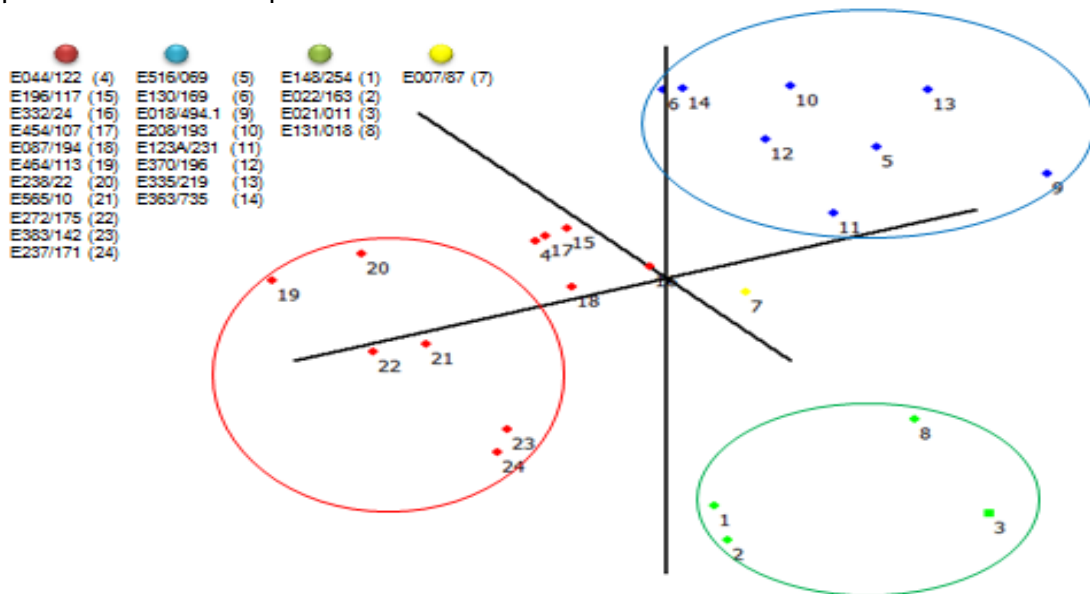
Os números sobre os ramos são valores de bootstrap (%) obtidos a partir de 1.000 análises replicadas. **Fonte:** o próprio autor.

Figura 4. Gráfico de estrutura genética a partir dos dados dos 31 marcadores SSR.



Programa STRUCTURE inferindo  $K=5$ . Cada cor representa um grupo genético. **Fonte:** O próprio autor.

Figura 5. Análise da coordenada principal entre 24 acessos de *C. arabica* provenientes da Etiópia.



As coordenadas X, Y e Z correspondem a 15,85%, 15,66% e 10,38%, respectivamente. **Fonte:** o próprio autor.

O dendograma de diversidade global gerado a partir dos acessos de *C. arabica* do centro de origem reforça a sua considerável diversidade genética, demonstrando que alguns dos genótipos da coleção da Etiópia não constituem um grupo geneticamente homogêneo (ANTHONY et al., 2002, SILVESTRINI et al., 2007; LÓPEZ-CARTNER et al., 2009, e FERREIRA, 2013). Em trabalhos anteriores (ANTHONY et al., 2001, ANTHONY et al., 2002; LÓPEZ-GARTNER et al., 2009), outras espécies de café e acessos comerciais de *C. arabica* foram utilizados para o estudo de diversidade genética, e reforçaram que há uma diversidade genética maior dos acessos de *C. arabica* do centro de origem em relação aos comerciais. Silvestrini et al. (2007), em um estudo de diversidade genética em acessos comerciais e selvagens de *C. arabica*, demonstraram que a diversidade dos acessos da Etiópia foi maior do que as cultivares demonstrando a sua importância como fonte de variabilidade genética para os programas de melhoramento do café.

Estudos relatam que a distinção entre os cafeeiros das regiões Sudoeste e Sudeste da Etiópia é resultante da presença da falha tectônica, o "Grande Vale do Rift", que atravessa a Etiópia do Nordeste ao Sudoeste (LASHERMES et al, 1996b; CROS et al, 1998; ANTHONY et al., 2001). Nesse estudo, a maioria dos acessos do lado Leste da Etiópia ficaram no grupo 2 (E018/494.1, E022/163, E021/11, E007/87), porém não no mesmo subgrupo e o

genótipo E238/22 e 237/171 dentro do grupo 3, não havendo uma partição clara entre genótipos de ambos os lados do Vale do Rift. Já com base nos marcadores AFLP, Anthony et al. (2002) relataram a partição dos acessos em dois grupos, separando o acesso E238 dos demais acessos do lado Sudoeste da Etiópia.

Na identificação dos 3 grupos pelo dendograma, o primeiro grupo incluiu um genótipo de Kaffa e um de Illubador. Uma justificativa seria que ambas as províncias estão localizadas lado a lado, e por causa disso, é esperado que o processo de adaptação tenha sido similar e portanto, se agrupam com base em suas proximidades geográficas. O segundo grupo incluiu principalmente genótipos de Kaffa e Sidamo. E o terceiro grupo foi formado pelos genótipos de Kaffa, Illubador e Gojjam. Pode-se observar também, que os genótipos da província de Sidamo ficaram agrupados entre os acessos de ambos os lados do Vale do Rift.

Similarmente, Aga et al. (2003) relataram que as amostras coletadas em Jima (Oeste do Vale do Rift) foram encontradas agrupadas com os acessos coletados em Bale (Leste do Vale do Rift). Tal agrupamento não pode explicar em termos de proximidade de origem, pois os pontos de coleta possuem uma distância estimada em mais de 300 Km. Uma justificativa para nossos resultados seria que provavelmente, os acessos tiveram uma origem comum sendo possível a transferência das melhores plantas entre diferentes populações, e não seleção local de variedades (CHAPARRO et al., 2004). Silvestrini et al. (2007) agruparam 73 acessos da Etiópia de acordo com suas distribuições geográficas, porém não foi verificada uma estruturação consistente entre os grupos, principalmente no agrupamento do lado Leste.

Lópes-Gartner et al. (2009), utilizando 47 SSR em 68 acessos de *C. arabica* da Etiópia, obtiveram também 3 agrupamentos. Houve como em nossos resultados, o agrupamento de genótipos coletados dos dois lados do Vale do Rift, e também de genótipos nativos de províncias próximas como Kaffa e Illubador. O fato dos genótipos de Illubador serem encontrados agrupados em outros grupos pode ser devido ao fato de Illubador estar em uma região com ambiente contrastante, podendo explicar a sua distribuição em diferentes grupos. Embora Anthony et al. (2001) conseguiram separar alguns acessos de ambos os lados do Vale do Rift, os agrupamentos formados continham acessos coletados em províncias próximas (Gojjam, Illubador e Shoa), assim como alta similaridade de acessos das regiões Sul

e Sudeste (Sidamo e Harar) com os da região Sudoeste da Etiópia (Shoa, Illubador, Kaffa).

Ferreira, 2010, em estudo da diversidade de 24 acessos da Etiópia com 292 locos SSR em gel de poliacrilamida, obteve 48 marcadores polimórficos com 2 a 5 alelos por loco detectado. No dendograma obtido os acessos de ambos os lados Sudeste e Sudoeste da Etiópia foram agrupados entre si, não havendo correlação entre os sítios de coleta e seus resultados. Já a mesma autora estudou a diversidade dos 132 acessos da coleção da Etiópia por sistema semi-automatizado com os 48 SSR obtidos no estudo anterior. Desses, 38 SSR foram informativos com um total de 2 a 13 alelos por loco detectado. A análise de agrupamento formou 2 grupos, porém não houve uma separação significativa dos acessos de acordo com suas origens geográficas (FERREIRA, 2013).

Pela análise da estrutura dos acessos em nosso estudo é possível sugerir que os acessos das províncias de Gojjam, Illubador e Kaffa são geneticamente relacionados. Do total dos acessos, 20 possuem ancestrabilidade semelhante, porém os acessos E007/087, E196/117, E044/122 e E332/24 demonstraram assim como na PCO possuir mistura de genótipos ancestrais de ambos os três agrupamentos. Já o acesso E007/087, o único da província de Harar, demonstrou ser o mais diverso entre os estudados e o que mais compartilha similaridade genética entre os três grupos em todas as análises, sendo um genótipo alvo para posteriores estudos de melhoramento da espécie.

Nossos resultados corroboram com os trabalhos de ANTHONY et al., 2001, CHAPARRO et al., 2004 e LOPES-GARTNER et al., 2009, no qual os acessos originários de províncias próximas foram mais estreitamente relacionados do que os coletados na mesma província, não existindo uma relação muito clara entre os sítios de origem geográfica dos acessos da Etiópia e os agrupamentos resultantes.

Sabe-se que em espécies selvagens, o nível de fluxo gênico é determinado principalmente por dois mecanismos, movimento mediado pelo humano ao longo das principais rotas de comércio e colonização natural resultante da polinização cruzada. Assim, podemos sugerir que as plantas da mesma origem foram geneticamente separadas pelo processo de domesticação (SILVESTRINI et al., 2007). Outra hipótese seria de que as plantas de café etíope do Sul e Sudeste poderiam ter sido introduzidas a partir do Sudoeste por seres humanos, ou

recolhidas nas florestas do Sul e Sudeste antes dessas serem destruídas, não existindo diferença na flora de cada lado da falha tectônica (WHITE, 1973).

Desde que estudos filogenéticos sugeriram a recente origem do gênero *Coffea* (LASHERMES et al, 1996b; CROS et al, 1998), a presença da falha tectônica ainda não explicou totalmente a distinção entre as populações de *C. arabica* das regiões Sudoeste e Sudeste da Etiópia. Não se pode descartar também possíveis erros na coleta e identificação deste material, como também na necessidade de se ampliar mais marcadores SSR informativos e indivíduos no estudo para o melhor entendimento da estrutura desta população.

## 6. CONCLUSÕES

- A análise *in silico* foi eficiente para a busca e caracterização dos marcadores SSR em *C. arabica* com os motivos di-, tri- e tetranucleotídeos das classes  $(AT)_n$ ,  $(TA)_n$ ,  $(GA)_n$  e  $(TC)_n$  os mais frequentes em ambos os genótipos.
- Dos 135 loci SSR testados, validou-se 31 *primers* polimórficos pela estratégia de *bulks*, e destes, sete foram descritos pela primeira vez na literatura.
- Com a estratégia de *bulks* de DNA houve uma grande economia de materiais e tempo otimizando as análises dos resultados.
- A análise da estrutura genética separou os acessos em três grupos, não havendo correlação entre os sítios de coleta na Etiópia e os agrupamentos resultantes.

## REFERÊNCIAS BIBLIOGRÁFICAS

ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ. **Histórico**. Disponível em: <<http://www.abic.com.br>>. Acesso em: 20 fever. 2014.

AGA, E.; BRYNGELSSON, T.; BEKELE, E.; SALOMON, B. Genetic diversity of forest arabica coffee (*Coffea arabica* L.) in Ethiopia as revealed by random amplified polymorphic DNA (RAPD) analysis. **Hereditas**, 138:36–46, 2003.

AGA, E.; BEKELE, E.; BRYNGELSSON, T. Inter-simple sequence repeat (ISSR) variation in forest coffee trees (*Coffea arabica* L.) populations from Ethiopia. **Genetica** 124:213–221, 2005.

AGGARWAL, R.K.; HENDRE, P.S.; VARSHNEY, R.K.; BHAT, P.R.; KRISHNAKUMAR, V.; SINGH, L. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. **Theor Appl Genet**, 114:359–372, 2007.

ANDERSON, J.A.; CHURCHILL, G.A.; AUTRIQUE, J.E.; TANKSLEY, S.D. Optimizing parental selection for genetic linkage maps. **Genome**, 36:181-186, 1993.

ALZATE-MARIN, A.L.; CERVIGNI, G.D.L.; MOREIRA, M.A & BARROS, E.G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia Brasileira**, 30:333-342, 2005.

ANTHONY, F.; BERTHAUD, J.; GUILLAUMET, J.L.; LOURD, M. Collecting wild *Coffea* species in Kenya and Tanzania. **Plant Genet Resour News** 69:23–29, 1987.

ANTHONY, F.; C. ASTORGA & J. BERTHAUD. **Los recursos genéticos: las bases de una solución genética a los problemas de La caficultura latinoamericana**. In: B. Bertrand & B. Rapidel (Eds.), *Desafíos de la caficultura centroamericana*, p. 369–406. IICA, San José, Costa Rica, 1999.

ANTHONY, F.; BERTRAND, B.; QUIROS, O.; WILCHES, A.; LASHERMES, P.; BERTHAUD, J.; CHARRIER, A. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. **Euphytica**, v.118, p.53-65, 2001.

ANTHONY, F.; COMBS, C.; ASTORGA, C.; BERTRAND, B.; GRAZIOSI, G.; LASHERMES, P. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. **Theor Appl Genet**, 104: 894–900, 2002.

ANZUETO, F.; BERTRAND, B.; SARAH, J.L.; ESKES, A.B.; DECAZY, B. Resistance to *Meloidogyne incognita* in Ethiopian *Coffea arabica* accessions. **Euphyta**, 118: 1–8, 2001.

BARUAH, A.; NAIK, V.; HENDRE, P.S.; RAJKUMAR, R.; RAJENDRAKUMAR, P and AGGARWAL, R.K. Isolation and characterization of nine microsatellite markers from *Coffea Arabica* L., showing wide cross-species amplifications. **Molecular Ecology**, Notes 3: 647–650, 2003.

BERTHAUD J.; CHARRIER A. Genetics resources of Coffea. In: Clarke RJ, Macrae R (eds) Coffee, vol 4. Agronomy. **Elsevier Applied Science**, London, pp 1–42, 1988.

BERTRAND, B.; ETIENNE, H.; CILAS, C.; CHARRIER, A.; BARADAT, P. *Coffea arabica* hybrid performance for yield, fertility and bean weight. **Euphytica**, 141(3): 255-262, 2005.

BORÉM, A. **Melhoramento de plantas**. 2º edição, Viçosa: Editora UFV, p.453, 1998.

BORÉM, A.; CAIXETA, E.T. **Marcadores moleculares**. 2ª ed. Viçosa: UFV, 532p, 2009.

BOSTEIN, D.; WHITE, R.L.; SKOLNICK, M.; DAVIS, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal Human Genetics**, Chicago, v.32, n.3, p.314-331, 1980.

BORÉM, A.; FRITSHE-NETO, R. **Biotecnologia aplicada ao melhoramento de plantas**. Viçosa: UFV, 2013.

BRASILEIRO, A.C.M.; CARNEIRO, V.T.C. **Manual de transformação genética de plantas**. Brasília: EmbrapaSPI/Embrapa-Cenargen, pag 309, 1998.

BRONDANI, R.P.V. et al. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **TAG Theoretical and Applied Genetics**, v. 97, n. 5-6, p. 816-827, 28 out. 1998.

BUSCHIAZZO, E & GEMMELL, N.J. The rise, fall and renaissance of microsatellites in eukaryotic genomes. **BioEssays**, 28 : 1040 – 1050, 2006.

CAIXETA, E.T.; OLIVEIRA, A.C.B.; BRITO, G.G.; SAKIYAMA, N.S. Tipos de marcadores moleculares. In: BORÉM, A.; CAIXETA, E.T (2ed). **Marcadores moleculares**. Viçosa: Folha de Viçosa, p.11-78, 2009.

CARVALHO, A. Distribuição geográfica e classificação botânica do gênero *Coffea* com referência especial à espécie *Arabica*. Boletim da Superintendência dos Serviços do Café, 21:174–180, 1945.

CARVALHO, C.H.S. et al. Cultivares de Café Arábica de Porte Baixo. In: CARVALHO, C.H.S. **Cultivares de café: origem, características e recomendações**. Brasília, DF: Embrapa Café, p.27- 334. 2008.

CHAPARRO, A. P.; CRISTANCHO, M. A.; CORTINA, H. A AND GAITAN, A. L. Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. **Genetic Resources and Crop Evolution**, 51:291–297, 2004.

CHARRIER, A. **A coffee** In: Charrier A, Jacquot S, Hamon D, Nicolas C (eds) Cirad, Repères Tropical Plant Breeding, 128–152, 1997.

CHEVALIER, A.; DAGRON, M. **Recherches historiques sur lês débuts de la culture du caféier en Amérique**. Commun Actes Acad Sci Colonial Paris, 1928.

COMBES, M.C.; ANDRZEJEWSKI, S.; ANTHONY, F.; BERTRAND, B.; ROVELLI, P.; GRAZIOSI, G.; LASHERMES, P. Characterization of microsatellites loci in *Coffea arabica* and related coffee species. **Molecular Ecology**, v.9, p.1171-1193, 2000.

Companhia Nacional de Abastecimento (Brasil). **Acompanhamento da safra brasileira de café**. Levantamentos de safra – 1º Levantamento de Café safra 2014. Brasília: CONAB, 2012. Disponível em <[http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14\\_01\\_09\\_09\\_18\\_57\\_boletim\\_cafe\\_\\_original.pdf](http://www.conab.gov.br/OlalaCMS/uploads/arquivos/14_01_09_09_18_57_boletim_cafe__original.pdf)>.1-20. Acesso em 11 março. 2014.

CRISTANCHO, M.A.; GAITÁN, A.L. Isolation, characterization and amplification of simple sequence repeat loci in coffee. **Crop Breeding and Applied Biotechnology**, 8: 321-329, 2008.

CROS, J.; M.C. COMBES.; P. TROUSLOT.; F. ANTHONY.; S. HAMON.; A. CHARRIER & P. LASHERMES. Phylogenetic relationships of *Coffea* species: new evidence based on the chloroplast DNA variation analysis. **Mol Phylogenet Evol** 9: 109–117, 1998.

CSENCICS.; D.S.B RODBECK AND HOLDEREGGER, R. Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. **Journal of Heredity**, 101 : 789 – 793, 2010.

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

DICE, L.R. Measures of the amount of ecologic association between species. **Ecology**, 26:297–302, 1945.

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

EDWARDS, D.; STAJICH, J.; HASEN, D. **Bioinformatics: Tools and Applications**. New York: Springer, 2009.

EKBLOM, R AND J. GALINDO. Applications of next generation sequencing in molecular ecology of non-model organisms. **Heredity**, 107 : 1 – 15, 2011.

ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nat Rev Genet**, 5:435-45, 2004.

EVANNO, G., REGNAUT, S., GOUDET, J. Detecting the number of clusters of individuals using the programa STRUCTURE: a simulation study. **Molecular Ecology**. 14, p. 2611-2620, 2005.

FAO. **FAO Coffee Mission to Ethiopia**, 1964–65. FAO, Rome, Italy, 1968.

FERREIRA, M.E AND GRATTAPAGLIA, D. **Introdução ao Uso de Marcadores Moleculares em Análises Genéticas**. 2nd edition. Brasília: Embrapa-Cnargen, 1998.

FERREIRA, R.V. **Seleção de marcadores microssatélites polimórficos em *Coffea arabica***. [Trabalho de Conclusão de Curso]. Londrina: Universidade Estadual de Londrina, Centro de Ciências Biológicas, 2010.

FERREIRA, R.V. **Genotyping of a collection of *Coffea arabica* using microsatellite markers**. 61 p. 2013. Mestrado em Genética e Biologia Molecular – Universidade Estadual de Londrina. Londrina, 2013.

FERWERDA, F. P. **Coffee**. in N.W. Simpsons, ed. Evolution of crop plants. Longman, London. p. 257-260, 1976.

GELETA, M.; HERRERA, I.; MONZON, A.; BRYNGELSSON, T. **Genetic Diversity of Arabica Coffee (*Coffea arabica* L.) in Nicaragua as Estimated by Simple Sequence Repeat Markers**. The Scientific World Journal, Volume 2012, Article ID 939820, 2012.

GRFFITHS, W.C.; GREGORY, M.P. MILLER, J.H.;SUZUKI, D,T.; LEWONTIN, R.C.; GELBART, W.M. **An introduction to Genetic Analysis**. 7 ed. New York: W.H Freeman & Co. 2000.

HAMBLIN, M. T.; M. L. WARBURTON AND E. S. BUCKLER. Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. **PLoS ONES** 2: e1367, 2007.

HAMON, P.; SEGUIN, M.; PERRIER, X.; GLASZMANN, C. **Diversité génétique des plantes tropicales cultivées**. CIRAD, Paris, France, 1999.

HENDRE, P.S. et al. Development of new genomic microsatellite markers from robusta coffee (*Coffea canephora* Pierre ex A. Froehner) showing broad cross-species transferability and utility in genetic studies. **BMC plant biology**, v. 8, p. 51, jan, 2008.

HOSHINO, A.A.; BRAVO, J.P.; NOBILE, P.M.; MORELLI, K.A. Microsatellites as tools for genetic diversity analysis. In: Mahmut Caliskan. (Org.). Genetic Diversity in Microorganisms. **Genetic Diversity in Microorganisms**. Rijeka: InTech, 149-170, 2012.

INTERNATIONAL COFFEE ORGANIZATION. **Trade statistics**. Disponível em: <[http://www.ico.org/trade\\_statistics.asp](http://www.ico.org/trade_statistics.asp)> Acesso em: 25 fevereiro 2014.

JEFFREYS, A.J.; WILSON, V.; THEIN, S.L. Hypervariable ‘minisatelite’ regions in human DNA. **Nature**, 316:76-79, 1985.

KNAPP, P.E. Studies of glial lineage and proliferation in vitro using an early marker of committed oligodendrocytes. *J. Neurosci Res*, 30(2):336-45, 1991.

LASHERMES, P.; TROUSLOT, P.; ANTHONY, F.; COMBES, M.C.; CHARRIER, A. Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. **Euphytica**, 87:59–64, 1996.

LASHERMES, P.; COMBES, M.C.; ROBERT, J.; TROUSLOT, P.; D'HONT, A.; ANTHONY, F.; CHARRIER, A. Molecular characterisation and origin of the *Coffea arabica* L. genome. **Mol Gen Genet**, 261: 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.

LEVINSON, G AND GUTMAN, G.A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. **Molecular Biology and Evolution**, 4: 203-221, 1987.

LI, Y. C.; A. B. KOROL.; T. FAHIMA.; A. BEILES AND E. NEVO. Microsatellites: Genomic distribution, putative functions, and mutational mechanisms: A review. **Molecular Ecology**, 11: 2453 –2465, 2002.

LITT, M. & LUTY, J. A. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. **Am. J. Hum. Genet**, 44, 397-401, 1989.

LÓPEZ-GARTNER, G.; CORTINA, H.; MC COUCH, SUSAN.R.; MONCADA, M.D.P. Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. **Tree Genetics and Genomes**, v.5, p.435-446, 2009.

LUSCOMBE, N. M.; GREENBAUM, D.; GERSTEIN, M. **What is bioinformatics? An introduction and overview**, Yearbook of Medical Informatics, 2001.

MALUF, M.P.; SILVESTRINI, M.; RUGGIERO, L.M.C.; GUERREIRO, FILHO, O.; COLOMBO, C. Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. **Sci Agric**, 62(4):366–373, 2005.

MATULA, M AND KYPR, J. Nucleotide sequences flanking dinucleotide microsatellites in the human, mouse and *Drosophila* genomes. **Journal of Biomolecular Structure and Dynamics**, 17: 275-280, 1999.

MENDES, A.N.G AND GUIMARÃES, R.J. **Genética e Melhoramento do Caffeiro**. Lavras: UFLA/FAEPE, 1998.

MIKHEYEV, A.S.; VO, T.; WEE, B.; SINGER, M.C.; PARMESAN, C. Rapid microsatellite isolation from a butterfly by de novo transcriptome sequencing: Performance and a comparison with AFLP-derived distances. **PLoS ONES**, e1212, 2010.

MILACH, S.C.K. Marcadores de DNA – Aplicações no melhoramento e plantas. *Biotecnologia Ciência e Desenvolvimento*, ano 1, n.5. Disponível em:

<http://www.biotechnologia.com.br/revista/bio05/marcadoresdna.pdf> - Acessado em:22/04/2010, 1998.

MISSIO, R.F. et al. Development and validation of SSR markers for *Coffea arabica* L. **Crop Breeding and Applied Biotechnology**, 9: 361-371, 2009.

MISSIO, R.F.; CAIXETA, E.T.; ZAMBOLIM, E.M.; PENA, G.F.; ZAMBOLIM, L.; DIAS, L.A.S AND SAKIYAMA, N.S. Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers. **Genet. Mol. Res**, 10 (4): 2366-2381, 2011.

MONCADA, P.; MCCOUCH, S. Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome**, 47: 501-509, 2004.

MORAIS, S. A. L.; AQUINO, F. J. T.; NASCIMENTO, P. M.; NASCIMENTO, E. A.; CHANG, R. Compostos bioativos e atividade antioxidante do café conilon submetido a diferentes graus de torra. **Química Nova**, São Paulo, v. 32, n. 2, p. 327-331, 2009.

MOSS, D.W. **Izoenzimas**. Capman & Hall, London & New York, 1982.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from small number of individuals. **Genetics** v. 89: p.583-590. 1978.

PAGIATTO, Natália Ferrarezi. **Análise de diterpenos e cafeína em uma coleção da etiópia de *Coffea arabica***. 77 f. Dissertação (Mestrado em Biotecnologia) – Universidade Estadual de Londrina, Londrina. 2013.

PARAN, I.; MICHELMORE, R.W. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. **Theor. Appl. Genet**, 85:985-993, 1993.

PATERSON, A.H.; TANKSLAY, S.D.; SORRELS, M.E. DNA markers in plant improvement. **Advances in Agronomy**, 46:39-90, 1991.

PEREIRA, G.S.; PADILHA, L.; PINHO, E.V.R.; TEIXEIRA, R.K.S.; CARVALHO, C.H.S.; MIRIAN PERES MALUF, M.P.; CARVALHO, B.L. Microsatellite markers in analysis of resistance to coffee leaf miner in Arabica coffee. **Pesq. agropec. Bras**, Brasília, v.46, n.12, p.1650-1656, dez. 2011.

PERRIER, X.; FLORI, A.; BONNOT, F. **Data analys methods. In Genetic diversity of cultivated tropical plants**. Science Publishers, Inc., Enfield, N. H. p. 43-76, 2003

PETERSON, D. G.; TOMKINS, J. P.; FRISCH, D. A.; WING, R. A.; PATERSON, A. H. Construction of plant artificial chromosome (BAC) libraries: An illustrated guide. **Journal of Agricultural Genomics**, Wallingford, v. 5, p. 1-3, 2000.

PÍPOLO, V.C.; GARCIA, J.C. **Biotecnologia aplicada ao melhoramento genético**. In: Biotecnologia na agricultura: aplicações e biossegurança. Cascavel: COODETEC, 2006.

PONCET, V.; HAMON, P.; MINIER, J.; CARASCO-LACOMBE, C.; HAMON, S.; NOIROT, M. SSR cross-amplification and variation within coffee trees (*Coffea* spp.). **Genome**, 47:1071–108, 2004.

POT, D.; FERREIRA, L.P.; DIAS, R.C.E.; DURAND, N.; GUYOT, B.; RAMOS, J.; SANDRIN-GARCIA, P.; BENASSI, M.T.; MARRACCINI, P.; PEREIRA, L.F.P.; LEROY, T.; VIEIRA, L.G.E. Genetic and molecular determinism of diterpene metabolism in *Coffea* spp. In: 22nd Internacional CONFERENCE ON COFFEE SCIENCE, Campinas. 22nd Internacional Conference on Coffee Science, p. 174-174, 2008.

POT, D.; SCHOLZ, M. B. S.; LANNES, S. D.; DEL GROSSI, L.; PEREIRA, L. F. P.; VIEIRA, L. G.; SERA, T. Phenotypic analysis of *Coffea arabica* accessions from Ethiopia: contribution to the understanding of *Coffea arabica* diversity. In: 22nd INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, Campinas. **Anais...** Campinas, p.165, 2010.

POWELL, W.; G.C. MACHRAY & J. PROVAN. Polymorphism revealed by simple sequence repeats. **Trends Plant Sci**, 1: 215–222, 1996.

PRITCHARD, J. K.; STEPHANS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v. 155, p. 945–959, 2000.

RABEL, M. et al. Marcadores Moleculares Microssatélites na Avaliação de Sementes de Soja com Variação na Coloração do Hilo. **Revista Brasileira de Sementes**, 32, 2, 019-025, 2010.

ROZEN, S AND SKALETSKY, H.J. Primer 3 for general users and for biologist programmers. In: Krawetz S and Misener S (eds) **Bioinformatics methods and protocols: methods in molecular biology**. Humana Press, Totowa, p. 365-386, 2000.

SALLES, G. et al. Marcadores Microssatélites em Espécies Vegetais, 2003.

SCHLÜTER, P. M.; HARRIS, S. A. Analysis of multilocus fingerprinting data sets containing missing data. **Mol. Ecol.** 6, 569-572, 2006.

SELKOE, K.A & TOONEN, R.J. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. **Ecology Letters**, Vol.9, No.5, (May 2006), pp. 615–629, ISSN 1461-0248, 2006.

SEIFU, S.; BELLACHEW, B. Genetic erosion - a threat to the conservation and sustainable utilization of Ethiopian coffee (*Coffea arabica* L.) germplam. 21ST INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, Montpellier, France, 11-15 September, 2006, pp. 960-967, 2007.

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

SILVESTRINI, M.; JUNQUEIRA, M.G.; FAVARIN, A.C.; GUERREIRO-FILHO, O.; MALUF, M.P.; SILVAROLLA, M.B AND COLOMBO, C.A. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. **Genetic Resources and Crop Evolution**, 54: 1367-1379, 2007.

SNEATH, P. AND SOKAL, R. Numerical Taxonomy. W.H. Freeman Guillaumet J.L. and Hallé F. 1978. **Echantillonnage du materiel and Company**, San Francisco, 1973.

SOLLER, M; BECKMANN, J.S. Restriction fragment length polymorphisms and genetic improvement. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, Madrid, vol. 6, pp. 396-404, 1983.

SYLVAIN, P.G. Some observations on *Coffea arabica* L. in Ethiopia. Turrialba 5:37–53, 1955.

TAUTZ, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucl Acids Res**, 17:6463–6471, 1989.

TAUTZ, D AND RENZ, M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. **Nucleic Acids Research**, 12: 4127-4138, 1994 .

VAN DER GRAAFF, N. Selection of Arabica coffee types resistant to coffee berry disease in Ethiopia. PhD Thesis Wageningen University, 1981.

VAN DER VOSSSEN, HAM, WALYARO D.J. Breeding for resistance to coffee berry disease in *Coffea arabica* L. **Euphytica**, 29(3): 777–791, 1980.

VARSHNEY, R.K.; THIEL, T.; STEIN, N.; LANGRIDGE, P.; GRANER, A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. **Cell Mol Biol Lett**, 7:537–546, 2002.

VARSHNEY, R.K; GRANER, A AND SORRELLS M.E. Genic microsatellite markers in plants: features and applications **Trends in Biotechnology**, 23: 48-55, 2005.

VARSHNEY, R.K.; NAYAK, S.N.; MAY, G.D.; JACKSON, S.A. Next-generation sequencing technologies and their implications for crop breeding. **Trends in Biotechnology**, v.27, n.9, p.522-530, 2009.

VEGRO, C. L. R; FERREIRA, C. R. R. P. T. **Evolução das despesas com defensivos agrícolas e fertilizantes para a safra de café 2000/01 nos estados de São Paulo e Paraná**. Informações Econômicas, v.30, p.53-59, 2000.

VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; VAN DE LEE, T.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.; ZABEAU, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Research**, v.23, p.4407-4414, 1995.

WANG, Z.; GERSTEIN, M; SNYDER, M. RNA-Seq: A revolutionary tool for transcriptomics. **Nature Reviews Genetics**, London, v.10, n.1, p.57-63, 2009.

WICKER, T. et al. 454 sequencing put to the test using the complex genome of barley. **BMC Genomics**, 7, 275, 2006.

WHITE, E. **The vegetation of Africa**. UNESCO, Paris (FRA), 356, 1973.

WEIR, B.S. Genetic data analysis methods for discrete genetic data Sinauer Association. **Sunderland: Massachusetts**, p. 445, 1990.

WEISING, K.; NYBOM, H.; WOLFF, K. and KAHL, G. **DNA Fingerprinting in Plants. Principles, Methods and Applications**. 2<sup>nd</sup> Edition. Taylor & Francis Group, 2005, 444 p. ISBN 0-8493-1488-7, 2005.

WILLIAMS, J.G.K. Polimorphisms amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Research**, Oxford, v.18, p.6531-6535, 1990.

YU, Q.; GUYOT, R.; KOCHKO, A.; DE BYERS, A.; NAVAJAS-PÉREZ, R.; LANGSTON B, J.; DUBREUIL-TRANCHANT, C. D.; PATERSON, A. H.; PONCET, V.; NAGAI, C.; MING, R. **Micro-collinearity and genome evolution in the vicinity of an ethylene receptor gene of cultivated diploid and allotetraploid coffee species (*Coffea*)**. *Plant Journal*, v. 67, p. 305–317, 2011.

YUKSEL, B.; PATERSON, A. H. Construction and characterization of a peanut *Hind*III BAC library. **Theoretical and Applied Genetics**, Oxford, v. 111, p. 630-639, 2005.

ZALAPA, J. E. et al. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. **American journal of botany**, v. 99, n. 2, p. 193-208, fev. 2012.

## ANEXOS

Anexo 1. Relação dos pares de primers obtidos de revisão bibliográfica.

Nome do primer	Forward	Reverse	Motivo de repetição	Natureza da repetição	Referência
CM2	TGTGATGCCATTAGCCTAGC	TCCAACATGTGTGGTGATT	(AC) <sub>10</sub> (AT) <sub>9</sub>	Di	Baruah et al., 2003
CM8	GCCAATTGTGCAAAAGTGCT	ATTCATGGGGCCTTTGTCTT	(GA) <sub>7</sub> (GT) <sub>12</sub>	Di	Baruah et al., 2003
CM16	TGGGAAAAGAAAGGATATAGACAAGAG	GAGGGGGGCTAAGGGAATAACATA	(GA) <sub>14</sub>	Di	Baruah et al., 2003
CFGA502	AAGCCACCCAGAAAACAGCACATC	ATTTGCTTCTCATGTTCCCTTTCA	(AG) <sub>27</sub>	Di	Moncada and Couch, 2004
M25	CCCTCCCTGCCAGAAAGAAC	AACCACCGTCCTTTCCCTCG	(GT) <sub>2</sub> CT(GT) <sub>2</sub> (GT) <sub>12</sub>	Di	Combes et al., 2000
CFGA627 <sup>2</sup>	GGGAAGGAATCTTTCAACTCT	CTTGAAATACCATGCAACC	(AG) <sub>16</sub>	Di	Cristancho et al., 2008
CFGA792b <sup>2</sup>	GATCAGAACTTTGAGTCAAGCA	AATGTGGCAGCTAGAAAGTG	(AG) <sub>12</sub>	Di	Cristancho et al., 2008
CFCA281 <sup>2</sup>	GCGTCCACGTGTTAAGTCTT	TCAAGTGGCAGACATGTCCAC	(AC) <sub>13</sub>	Di	Cristancho et al., 2008
CFCA331 <sup>2</sup>	TGATGGACAGGAGTTGATGG	CACCTATTTTGCCAACTTACC	(CT) <sub>17</sub> (AC) <sub>18</sub>	Di	Cristancho et al., 2008
CFCA360 <sup>2</sup>	TTAAGACATCGGTGCAATCA	TGTGTACTGGGTTTTTGTGATG	(AC) <sub>15</sub>	Di	Cristancho et al., 2008
CaM03 <sup>9</sup>	CGCGTTGCTCCCTCTGTCTCT	TGGGGAGGGGCGGTGTT	AC	Di	Geleta et al., 2012
CM5 <sup>5</sup>	GTACCACACCTCTCTCTGC	TGGAGGTAACGGAAGCTCTG	CCT	Tri	Geleta et al., 2012
Sat207 <sup>4</sup>	GAAGCCGTTTCAAGCC	CAATCTCTTTCCGATGCTCT	nd	nd	Geleta et al., 2012
SSR09 <sup>6</sup>	TTGGCTTTTGTCCCTCCTCCTCTG	AGCCCATTTTCCCTCTCATCTTCAAG	GT	Di	Geleta et al., 2012
M24	GGCTCGAGATATCTGTTTATG	TTTAAATGGGCTAGGGTCC	(CA) <sub>15</sub> (CG) <sub>4</sub> CA	Di	Combes et al., 2000
AJ250257	GACCCATTACATTTACACACAC	GCATTTTGTTCACACTGTA	(CTCAC) <sub>4</sub> (CA) <sub>9</sub>	Hexa/Di	Mahuf et al., 2005
M47	TGATGGACAGGAGGTGATGG	TGCCAATCTACCTACCCCTT	(CT) <sub>9</sub> (CA) <sub>9</sub> (CT) <sub>4</sub> (CA) <sub>5</sub>	Di	Combes et al., 2000
SSRCa 002	CTGTCCCAACCAACCAAAA	CTTCAACCCCAACACAC	(TTCC) <sub>3</sub> .....(GT) <sub>17</sub>	Tetra/Di	Missio et al., 2009
SSRCa 006	CTTGCTCAGTGAACCATCC	TGCCCTTATGCCACTACTAAA	(CT) <sub>6</sub>	Di	Missio et al., 2009
SSRCa 018	CTTCGGTTTCAAGCTCTCTC	ATTTTGGCAGGGTATGTTT	(GT) <sub>18</sub> (GA) <sub>10</sub>	Di	Missio et al., 2009
SSRCa 021	GCTGAGAGTTTGGAGGGAAA	CCGACGTAGTTGATGATTGA	(GGA) <sub>3</sub> N <sub>4</sub> (AAG) <sub>2</sub>	Tri	Missio et al., 2009
SSRCa 033	GTTTTTACGGCAGCAGATTA	TTCAAAAAGTCAACTATTCTCC	(AAT) <sub>3</sub> ...(GA) <sub>4</sub> N <sub>5</sub> (CA) <sub>2</sub>	Tri/Di	Missio et al., 2009
SSRCa 040	AGGGATGTAGAACCAGCAAAA	CCAATAGCTCACAACAAAAGG	(GAG) <sub>3</sub> A(AAG) <sub>3</sub>	Tri/Di	Missio et al., 2009
SSRCa 045	GACTTGTGCAATCCCTCA	GCGCATGTGAAGAGAAAAGT	(TTTAC) <sub>3</sub> ...(AC) <sub>3</sub> T(CA) <sub>3</sub>	Tetra/Di	Missio et al., 2009
SSRCa 052	GATGGAAAACCCAGAAAAGTTG	TAGAAGGGCTTTGACTGGAC	(TTG) <sub>7</sub>	Tri	Missio et al., 2009
SSRCa 054	CGGAACCCAACTAACATCTC	GCAGGCTTCCATTGTCTGT	(AAA) <sub>3</sub>	Tetra	Missio et al., 2009
SSRCa 055	AAGGAAAACAACACCCAAAGA	CGAGACAAGAGAGGGGAAA	(ATC) <sub>3</sub> ...(AAGG) <sub>2</sub> N <sub>2</sub>	Tri/Tetra	Missio et al., 2009
SSRCa 061	GCAGGTGCAAGTGATAAAAG	CGTCTTGTGATGTTAGGG	(CCA) <sub>2</sub> (CT) <sub>5</sub>	Tetra/Di	Missio et al., 2009
SSRCa 065	ATCTAACAAAATCCCGTCA	ATCGGTCCCGCTTCTAAT	(AG) <sub>2</sub> (AAG) <sub>3</sub>	Di/Tri	Missio et al., 2009
SSRCa 085	ATGTGAAAATGGGAAGGATG	CACAGGAAAGTGACACGAAAG	(TC) <sub>24</sub>	Di	Missio et al., 2009
SSRCa 087	TCATCTCGCAGACACACTAC	GCAGAGATGATCACAAGTCC	(TC) <sub>22</sub>	Di	Missio et al., 2009
SSRCa 091	CGTCTCGTATCAGCTCTC	TGTTCTCGTCTCTCTCTC	(GT) <sub>8</sub> (GA) <sub>10</sub>	Di	Missio et al., 2009
SSRCa 092	ATAGCCTGAGCCGTAACCA	GGGTAATTATGACGAGGGACA	(CCA) <sub>7</sub> CT(TCCACC) <sub>5</sub>	Tri/Di/Hexa	Missio et al., 2009
EST-SSR 007	AGTGGCTGGGAACAAAGAGA	TTCTCCTCCGCAACACAGAG	(GTC) <sub>5</sub>	Tri	Missio et al., 2011
EST-SSR 010	CTTCTTCAATCCAACAACACG	TGCCATTCACACTGTGCTACT	(TG) <sub>7</sub> (TA) <sub>4</sub> (GT) <sub>9</sub>	Di	Missio et al., 2011
EST-SSR 027	ATGGAAGTGTCTTGTGCGTG	ATGTGCGGTGGTCCGGTCAAA	(AT) <sub>3</sub> (AC) <sub>4</sub> (AT) <sub>9</sub>	Di	Missio et al., 2011
EST-SSR 029	TTAACCTCTGCCACACA	CGCCAAATAAATCCCTCCA	(AT) <sub>10</sub>	Di	Missio et al., 2011
EST-SSR 047	GGCTCAATTAAGCCTCATCATC	CAGCCGTTTGAAGTAATC	(CCA) <sub>5</sub> N <sub>3</sub> (CTG) <sub>6</sub>	Tri	Missio et al., 2011
EST-SSR 073	GAGGCTTCCCAACCAACA	GGATACGAGAGTCCCTTCCA	(GGC) <sub>2</sub> (GGC) <sub>5</sub>	Tri	Missio et al., 2011
SSRCa 081	ACCGTTGTGATATCTTTG	GGTTGAACCTAGACCTTATTT	(CT) <sub>38</sub>	Di	Missio et al., 2009
LEG9	AGGTTTCCAAAGGAGATGAGC	GAAGACAAGTCCATCGTCCAA	GAT	Tri	Pereira et al., 2011
LEG11	CACCTGAAGGCCTGGAAGAAT	AGCATCTGCAGCCTCCATAG	TGG	Tri	Pereira et al., 2011
LEG12	CACCATAGCAACTTCAAACACG	CACATCCAGGAACCTTGTCT	TC	Di	Pereira et al., 2011
LEG13	GAAAGGGAAGAGGGGCAAG	GTGGTGGAGGAAGGGATTC	GAA	Tri	Pereira et al., 2011
LEG22	TTTCACTCCCTGCAACAC	TCTTCTAAGGCCAGCAAGGA	TC	Di	Pereira et al., 2011
LEG26	TGAAGTGCCTCCTGTTTCT	CGTCAGCTCAAGACTGTGC	ATA	Tri	Pereira et al., 2011
LEG28	TGTTACAGCTAAACCCAAACC	TTGACGGTGCAGATGTTGAT	AGT	Tri	Pereira et al., 2011
LEG32	GGGTGATGGAAAAGCAAAATG	CCAGCATCAGCAAAGTAAAGG	AGA	Tri	Pereira et al., 2011
AJ308738	AAAAAGTGTCCATGTCAA	GGGGCGTTCAGTTATAAACA	(TG) <sub>8</sub>	Di	Silvestrini et al., 2007
AJ308746	AGGCCTCATCTCAAAAACC	AGCGTTACTTGTAGGCCAAA	(TC) <sub>14</sub> (CA) <sub>11</sub> (CA) <sub>16</sub>	Di	Silvestrini et al., 2007
AJ308762	CTGGGTTGGTCTGATTTTG	GGTCCCAGAGATTTCTCC	(TG) <sub>16</sub>	Di	Silvestrini et al., 2007
AJ308767	TGCTTAGGCATGATATAGGA	CACGTGCAAGTACACATTTTA	(CA+TA) <sub>38</sub>	Di	Silvestrini et al., 2007
AJ308785	TGGCCGTGATAATAACACGC	ATGTGGCAATCTAAAGCCAA	(TG) <sub>21</sub>	Di	Silvestrini et al., 2007
M32	AACCTCTCCATCCCGCATTC	CTGGGTTTTCTGTGTTCTGC	(CA) <sub>3</sub> (CA) <sub>3</sub> (CA) <sub>18</sub>	Di	Combes et al., 2000
Sem nome	CCTCCCTCAGTCAATTTCA	CTTGGTCTCCCTCTTTTTC	(ATC) <sub>14</sub>	Di	Silvestrini et al., 2007
AJ308754	AAGGGGAGTGGATAAGAAAG	GGCTGGATTTGTGCTTTAAG	(CA) <sub>12</sub>	Di	Silvestrini et al., 2007
AJ308764	CACCTGGCATTAGAAAGCACCC	GGCAAAGTCAATGATGACTC	(CA) <sub>14</sub>	Di	Silvestrini et al., 2007
AJ250251	ATTCTCTCCCTCTCTCG	TGTGTGGCGTTTTCTTG	(CA) <sub>6</sub> (CA) <sub>4</sub> (CA) <sub>3</sub> (CA) <sub>18</sub>	Di	Silvestrini et al., 2007
AJ250252	ACCCGAAAGAAAGAACCAAG	CCACACAACCTCTCTCATTC	(GT) <sub>4</sub> (GA) <sub>4</sub> (GT) <sub>4</sub> (GT) <sub>6</sub>	Di	Silvestrini et al., 2007
AJ250253	CTTGTGAGTCTGTGCGCTG	TTTCCCTCCCAATGTCTGTA	(GA) <sub>2</sub> (GT) <sub>3</sub> TT(GT) <sub>4</sub> TT(GT) <sub>7</sub> /(GA) <sub>11</sub> /(TC) <sub>7</sub> (CT) <sub>7</sub> GT	Di	Silvestrini et al., 2007
AJ250256	AGGAGGGAGGTGTGGGTGAAG	AGGGGAGTGGATAAGAAAG	(GT) <sub>11</sub>	Di	Mahuf et al., 2005
CFGA189	CATCCATCCGAAAACCTTGTAAAG	CAGCACTGGCAAAATAGCAACTCTT	(AG) <sub>10</sub>	Di	Moncada and Couch, 2004
CaM16	AAGGCAGTGAAGCGGGACAAA	TGGGGAGAGCTGCAGTTGGAGG	(TC) <sub>11</sub>	Di	Hendre et al., 2008
CFGA465	ACCCCTTACTACTTATTTACTCTC	ACATCCCTTGGCAATTTCTTC	(AG) <sub>15</sub>	Di	Moncada and Couch, 2004
CFGA249	TAAGAAGCCACGTGACAAGTAAGG	TATGGCCCTTCTCGCTTAGTT	(AG) <sub>13</sub>	Di	Moncada and Couch, 2004
CorEST-SSR03	GGACAAAACACCCGCCCAAAAATA	AGGGAGACAGAGGAAGGGAATATT	(TC) <sub>13</sub>	Di	Bhat et al., 2005
CaM45	CGCGCCAGTGAATTCGAGCTC	TCGCCATTTGGAGCTGTGATTTCA	(GT) <sub>8</sub> (GA) <sub>5</sub>	Di	Hendre et al., 2008
CaM33	CGCATTAGCGGTGGGAGAA	CAGAGGTTGTGCGGTGAGTGGAGAA	(A) <sub>13</sub> -5pb-(AG) <sub>18</sub>	Mono/Di	Hendre et al., 2008
CaM03	CGCGCTTGTCTCTGTCTCT	TGGGGAGGGGCGGTGTT	(AC) <sub>11</sub>	Di	Hendre et al., 2008
DCM01	TTTTTGGGAAATGAAGGTGC	TGCACTTCAAGATCCCTTT	(AG) <sub>15</sub>	Di	Ramesh et al., 2007
CFGA202	AGGGAAAGTGTGCAATTTACTACA	TGTGTAGCTTTTCTTTTGTGATTT	(AG) <sub>11</sub>	Di	Moncada and Couch, 2004
CaM41	CATCGTCTCCATCGTTGCTCTATC	CCCTCCCTCTTTCTATCTAAT	(TAAA) <sub>5</sub>	Tetra	Hendre et al., 2008
CaM36	TGGTTTTAGTTTGTATTATTGATGTGAT	CGAGCCCTCCCTTGTCA	(TTA) <sub>7</sub>	Tri	Hendre et al., 2008
CFGA311	GGACGGACAAAATGACAAAAATG	TACTCCCTCCCGACATTTG	(AG) <sub>10</sub>	Di	Moncada and Couch, 2004
CFGA547	AAGGCATGCGGCGGGAGTAT	TCGTCAAGGACAATCTAAAGC	(AG) <sub>18</sub>	Di	Moncada and Couch, 2004
CFGA276	GCTACAGAAACCAGAAAATGAGAA	CTCTTTCAAGTTCTTATGGGCAG	(AG) <sub>21</sub>	Di	Moncada and Couch, 2004
CFGA285	CAGCCCTGGCAACTATTTTT	TCCTTGTCAAAATGGGTAGAAACT	(AG) <sub>18</sub>	Di	Moncada and Couch, 2004

nd: não possui dados na referência original. **Fonte:** o próprio autor.

Anexo 2. Relação dos pares de *primers* SSR desenhados para CaHT 832/2.

<i>Primers Contigs</i>	Sequência 5'- 3' senso	Sequência 5'- 3' anti-senso	Motivo de repetição	Natureza da repetição	Nome do contig
CHT01	GACCGTAACTGCCACAAATC	GTAGCGTGCTGGAATTCAC	(TGA) <sub>4</sub>	Tri	scaffold6 size10802
CHT02	GAGGACTTCTGGGACCATT	TGGTACTCTGCCTTTGGAAC	(AT) <sub>7</sub>	Di	scaffold8 size27678
CHT03	GTCTCTCCGCTTTTCTTCC	CTTGGTTGCCTGTTTCTTAA	(CT) <sub>8</sub>	Di	scaffold8 size27678
CHT04	TCAAATTTCCATGACCGTCT	TTTTTCCGCAATAGCATCTT	(TTA) <sub>4</sub>	Tri	scaffold8 size27678
CHT05	CCATTCATCGTTTGTCCATT	CCTCGAAGGCTACCAGTACA	(GTG) <sub>4</sub>	Tri	scaffold8 size27678
CHT06	TCAAACCTCGAAACCCACATT	AGACGGTCATGGAAATTTGA	(TGC) <sub>4</sub>	Tri	scaffold8 size27678
CHT07	AGCAGGTGTTGTTTTCATCC	ATATCAGTTGCGAAGGCATC	(CCA) <sub>4</sub>	Tri	scaffold8 size27678
CHT08	AGGCCCTTCTCTCCTGTATG	ACCACCGCTTCTCTCTTCTT	(AAAAAT) <sub>3</sub>	Penta	scaffold8 size27678
CHT09	GAATTTCTCAAGCGCAATGT	GATCCACCATATGCCTCAAG	(TAAA) <sub>3</sub>	Tetra	scaffold14 size15334
CHT10	CGTATTTGTGACCTCGAAT	TTTACCGGTATTTTTCATTTG	(TATT) <sub>3</sub>	Tetra	scaffold14 size15334
CHT11	AGGCCTAATCCCAACAATCT	ATCCATGGCTTGCAGAAATA	(AT) <sub>7</sub>	Di	scaffold17 size8074
CHT12	CCGAGCATTGTGACTCGTAT	CAGGAAAAACAGAGACGAA	(AT) <sub>9</sub>	Di	scaffold19 size38401
CHT13	AAGGGAGAAATTGGAAGTCG	ACACCTCAACATTGGCATCT	(TGC) <sub>4</sub>	Tri	scaffold19 size38401
CHT14	AGTTTTGGGAGCTTTCTGGT	ACACCTCAACATTGGCATCT	(GTG) <sub>6</sub>	Tri	scaffold19 size38401
CHT15	AAGGGAGAAATTGGAAGTCG	ACACCTCAACATTGGCATCT	(GAT) <sub>4</sub>	Tri	scaffold21 size6818
CHT16	TACCCGCAGAATAGACATCC	AGCATGGACACACTTCAAATC	(ATTT) <sub>3</sub>	Tetra	scaffold30 size8115
CHT17	GCAGGAAAAACAAACACACA	CATGACAAACGAAAAAGCACA	(AT) <sub>6</sub>	Di	scaffold40 size33925
CHT18	TCTGCACTAGAACACCACGA	GCTCCATGGAAGAAAGAAACA	(TTA) <sub>4</sub>	Tri	scaffold40 size33925
CHT19	GCAGTTCTTCTCGTTTGA	AAATGTTGGAGGGGATTAGG	(CTG) <sub>4</sub>	Tri	scaffold45 size8881
CHT20	GGAGTTAGGCTCGGATCAAT	CGTCCTTTTCATTGGACAAA	(ACA) <sub>3</sub>	Tetra	scaffold45 size8881
CHT21	TGGACCACACAAAATGATGA	TTCTGGAATCGAGAGGCTAA	(AAAT) <sub>3</sub>	Tetra	scaffold46 size4171
CHT22	TTTTGAGCCAAACGCTTTT	GGTGTTAATCGACCACTGCT	(TTTG) <sub>3</sub>	Tetra	scaffold47 size12462
CHT23	TGGTGTGAAGGGGATATTTG	ACTGGGATTCTGTCCTGTGA	(AGC) <sub>4</sub>	Tri	scaffold57 size16806
CHT24	TCATTCATCCAACGCTGATA	GGAAACCGGTAGGTTATCCA	(ATTT) <sub>3</sub>	Tetra	scaffold57 size16806
CHT25	CCTGTCTTGGCTCTACCTGA	TCTGTTGATCCGTTGTTGATG	(CTAT) <sub>3</sub>	Tetra	scaffold59 size4646
CHT26	AAAATAAAAGGGCGGTGAAG	AGATCGGGGAGTTGAAGAAC	(TAAA) <sub>3</sub>	Tetra	scaffold61 size3184
CHT27	TGCATCCATCATTTCAAGTGT	ATTGAGCCACCTTGAGTCAG	(AGGA) <sub>3</sub>	Tetra	scaffold67 size3824
CHT28	GAACCTCGACCCTTCACTT	CCAACACGCTTCAAATTTT	(AATA) <sub>3</sub>	Tetra	scaffold91 size5659
CHT29	TAACAGCAATCGCTGAACAA	CGAGGGGCCATACCTATAAA	(AGAAGC) <sub>3</sub>	Hexa	scaffold97 size3334
CHT30	CAAAATAATGCAAGGGCAAT	GAATTCGCTCAAAAAGCAGA	(AT) <sub>5</sub>	Di	scaffold101 size1410
CHT31	GGCTGGTTGTATGAGGAATG	AGGGCACACTATGGACAGAA	(GAGT) <sub>3</sub>	Tetra	scaffold103 size2138
CHT32	AGGGCTAAGAGGCAATAACC	ACAAGCAAGATGGATGGCTA	(AG) <sub>7</sub>	Di	scaffold115 size831
CHT33	AGGGCTAAGAGGCAATAACC	ACAAGCAAGATGGATGGCTA	(GA) <sub>5</sub>	Di	scaffold115 size831
CHT34	CCGACGGGTCTCTTTTAT	TCTTTACGGGATTGCTCTG	(AG) <sub>5</sub>	Di	scaffold121 size2154
CHT35	AAACCCAACCTGGCTTTTT	CATCGCTCTCTTTCTCATC	(TTCC) <sub>3</sub>	Tetra	scaffold121 size2154
CHT36	ATGTAGGTGGTGTGGGCTA	ACTTTGATTTGCGATGGTTG	(AAT) <sub>4</sub>	Tri	scaffold131 size858
CHT37	CAGCCTCTCTTTGAACCA	TGGATGAATTGGGTATGGAC	(AATG) <sub>3</sub>	Tetra	scaffold137 size526

Fonte: o próprio autor.

Anexo 3. Relação dos pares *primers* SSR desenhados para *C.arabica* cv. Iapar 59.

<i>Primers</i> <i>Contigs I59</i>	Sequência 5'-3' senso	Sequência 5'-3' anti-senso	Motivo de repetição	Natureza da repetição	Nome do contig
CIapar01	GGTTGGTGGCTTGTCTAATG	ACGTACTCGAATGCTTCTGG	(AT)5	Di	37301
CIapar02	TCACTTACTCCAAACCGTCTTT	ACGAGACTGGGGAAGAGAAT	(TC)6	Di	37084
CIapar03	GTGGCTGAAATGCTTGTTTT	TACTACACCCTCACCCAACC	(TC)6	Di	37360
CIapar04	TTACACCAACCAACACAACG	AGGAGGAGCCTTTGATCCTA	(CA)5	Di	34901
CIapar05	TTGATTTTGTGAGGGTCGAT	GCTGCTTCTCTATGGCTACG	(GAT)5	Tri	34901
CIapar06	AAAAATGACAGGCAAAGCAT	CCTTCTGGACAGCACAAATTC	(GA)8	Di	37436
CIapar07	GAAGGCAAAGAACAATGCAC	CTGGTGGGATACCTGACAAC	(AAT)4	Tri	36526
CIapar08	CAAAAACCTCCGAGTGAGGA	GAAGAAGAAGAGGCAAGAGGA	(TC)5	Di	37443
CIapar09	TTCTCATCCAAAGCCATCAT	CAAATTCTGAAGCCAGTGAAA	(TG)5	Di	26884
CIapar10	ACGCTACAGCCTCAAAGATG	CACCCTGGTCAAGAAGAAGA	(ATC)4	Tri	37403
CIapar11	AATCCATCTCCCCATCTCTC	GTGCCATGTTGAGCAATGTA	(CTG)4	Tri	37221
CIapar12	TTGTTGAGGCAGGAAATGAT	AACCAAGCAGCAAAGTTCAC	(ATC)5	Tri	37159
CIapar13	TGTGGACAGAGACAGGACAA	CATCATCAGCGTGATTGTGT	(GA)5	Di	37248
CIapar14	GCGGATCTAACCAAGTAGCC	ATGATGCCGGTGATGTTTAT	(TTC)4	Tri	37248
CIapar15	AAGGTCATCCCCTCATTAGC	CCTATCCCCAGCCAAGTATT	(TCA)4	Tri	37466
CIapar16	CCACCTAACTGCTGCTGAAT	GTAAGCCTCGTGAATTGGAA	(CTT)4	Tri	37466
CIapar17	GGTTGCAAGATGAGTTGGAT	GGGTTTTGTTGGATTTGTG	(AC)5	Di	36710
CIapar18	GGTTGCAAGATGAGTTGGAT	GGGTTTTGTTGGATTTGTG	(TA)5	Di	36710
CIapar19	AGTTGAACCTGTGTGTGCAA	GGACAAGAAGGACGAAACAA	(AT)8	Di	37348
CIapar20	AGTTGAACCTGTGTGTGCAA	GGACAAGAAGGACGAAACAA	(TA)8	Di	37348
CIapar21	CGGAGATTGATATGCGTCTT	TGAAGATGAACCATGTGCAG	(CT)5	Di	36399

Fonte: o próprio autor.