

LUÍS FELIPE VENTORIM FERRÃO

MARCADORES MICROSSATÉLITES EM ESTUDOS DE DIVERSIDADE,  
MAPEAMENTO GENÉTICO E ANÁLISES DE QTLs EM *Coffea canephora*

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2013

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

F373m  
2013

Ferrão, Luís Felipe Ventrím, 1988-  
Marcadores microssatélites em estudo de diversidade,  
mapeamento genético e análises de QTLs em  
*Coffea canephora* / Luís Felipe Ventrím Ferrão. – Viçosa,  
MG, 2013.  
x, 134f. : il. ; (algumas color.) ; 29cm.

Orientador: Eveline Teixeira Caixeta.  
Dissertação (mestrado) - Universidade Federal de Viçosa.  
Inclui bibliografia.

1. *Coffea canephora* – Melhoramento genético. 2. Genômica.  
3. Marcadores genéticos. 4. *Hemileia vastatrix*. I. Universidade  
Federal de Viçosa. Programa de Pós-Graduação em Genética e  
Melhoramento. II. Título.

CDD 22. ed. 633.732

LUÍS FELIPE VENTORIM FERRÃO

MARCADORES MICROSSATÉLITES EM ESTUDOS DE DIVERSIDADE,  
MAPEAMENTO GENÉTICO E ANÁLISES DE QTLs EM *Coffea canephora*

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

APROVADA: 05 de fevereiro de 2013

---

Laércio Zambolim

---

Cosme Damião Cruz  
(Coorientador)

---

Eveline Teixeira Caixeta  
(Orientadora)

Aos meus pais Romário Gava Ferrão e Liliâm Maria Ventorim Ferrão, e aos meus irmãos Guilherme Ventorim Ferrão e Arthur Ventorim Ferrão pela força, torcida, dedicação e amor oferecidos durante toda a minha caminhada acadêmica.

OFEREÇO

À todos os meus familiares, amigos, orientadores e professores pela confiança depositada e auxílio na conclusão de mais uma etapa importante da minha vida.

DEDICO

## AGRADECIMENTOS

À Deus pela vida, saúde e pelas pessoas maravilhosas que colocou no meu caminho.

À Universidade Federal de Viçosa (UFV) e ao Curso de Pós-Graduação em Genética pelo ensino de excelência que me proporcionaram.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudo.

À Dra. Eveline Teixeira Caixeta, pela orientação segura, pelos conhecimentos transmitidos, pelas oportunidades e, principalmente, por ser uma pessoa de diálogo fácil o que contribuiu muito para o meu crescimento.

Aos meus orientadores Prof. Cosme Damião Cruz, Pesquisadora Maria Amélia Gava Ferrão e Prof. Ney Sussumu Sakiyama pela disposição em compartilhar seus conhecimentos e pelas sugestões valiosas na fase de finalização deste trabalho. À Dra. Eunize Maciel Zambolim, pelo auxílio nos trabalhos de laboratório e pelo espírito de colaboração. Ao Prof. Laércio Zambolim pelos ensinamentos e sugestões na área de fitopatologia.

Aos Professores Paulo Roberto Cecon, Fabyano Fonseca, Everaldo Gonçalves de Barros, Pedro Crescêncio Souza Carneiro e Leonardo Bhering pelos valiosos ensinamentos transmitidos durante o mestrado, seja nas disciplinas ou nas conversas “informais”.

Aos pesquisadores Romário Gava Ferrão, por compartilhar seu vasto conhecimento sobre a cultura do cafeeiro; e Flávio França de Souza, pela paciência e amizade demonstrada no início dos meus trabalhos.

À todos os colegas que passaram pelo Laboratório BioCafé, que não vou arriscar-me a nomeá-los para que não cometa nenhuma injustiça, por proporcionarem momentos de descontração e aprendizado.

Ao Incaper, em nome dos pesquisadores Romário, Maria Amélia, Aymbiré, Paulo Volpi, Abraão Verdin, José Lani e Aires Ventura, pelos conhecimentos transmitidos e disponibilização das populações e dos dados de campo utilizados neste estudo.

Aos meus avós, tios e primos da Família Ventorim e Ferrão pela torcida constante.

À “família Cecon”: Paulo, Neusa, Camila, Roberta e Paulinha; pela amizade sincera e momentos alegres que me proporcionaram aqui em Viçosa.

Aos amigos da Rep, por transformarem a minha estadia em Viçosa um momento prazeroso, divertido e inesquecível.

A todos os meus colegas de curso, em especial ao David, Iara e Carlos (Cacá), pela amizade sincera, momentos de descontração e pelo companheirismo apresentado em todo o mestrado.

A minha namorada Juliana e seus familiares, pela cumplicidade, amizade e amor transmitidos.

Por fim, mas não menos importante, aos meus pais Romário e Liliâm, e aos meus irmãos Guilherme e Arthur, pela torcida constante e amor incondicional mesmo à 400 km de casa.

## **BIOGRAFIA**

LUÍS FELIPE VENTORIM FERRÃO, filho de Romário Gava Ferrão e Liliâm Maria Ventorim Ferrão, nasceu no dia 17 de maio de 1988, em Vitória, Estado do Espírito Santo.

Concluiu o ensino fundamental em 2002, na Escola Normal Nossa Senhora do Carmo em Viçosa-MG e o ensino médio em 2005, no Centro Educacional Charles Darwin em Vitória-ES.

Em janeiro de 2011 diplomou-se em Ciências Biológicas na Universidade Federal de Viçosa (UFV), na modalidade bacharelado. Durante a graduação foi bolsista de iniciação científica no Departamento de Estatística, sob orientação do Professor Paulo Roberto Cecon; e no Departamento de Fitotecnia, sob orientação do Professor Ney Sussumu Sakiyama.

Em março de 2011, iniciou o curso de mestrado no Programa de Pós Graduação em Genética e Melhoramento, área de concentração Genética Vegetal, na Universidade Federal de Viçosa (UFV), sob orientação da Dra. Eveline Teixeira Caixeta, pesquisadora da Embrapa Café.

Atualmente está aprovado no curso de doutorado em Genética e Melhoramento de Plantas, na Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP) em Piracicaba- SP, onde iniciará seus trabalhos na área de Genética Estatística, sob orientação do Prof. Antônio Augusto Franco Garcia.

## ÍNDICE

<b>RESUMO .....</b>	<b>VII</b>
<b>ABSTRACT .....</b>	<b>IX</b>
<b>1. INTRODUÇÃO GERAL.....</b>	<b>1</b>
<b>2. REVISÃO DE LITERATURA .....</b>	<b>3</b>
2.1) A ESPÉCIE <i>COFFEA CANEPHORA</i> .....	3
2.2. MELHORAMENTO GENÉTICO DE <i>C. CANEPHORA</i> .....	4
2.3. FERRUGEM ALARANJADA DO CAFÉ ( <i>HEMILEIA VASTATRIX</i> ).....	8
2.4. MARCADORES MOLECULARES .....	10
2.4.1. Base genética dos marcadores <i>Microssatélites</i> .....	12
2.4.2 <i>Marcadores microssatélites em café</i> .....	13
2.5 BANCOS DE GERMOPLASMA E AVALIAÇÃO DOS RECURSOS GENÉTICOS.....	14
2.5.1. <i>Diversidade genética em C. canephora</i> .....	16
2.6. MAPEAMENTO GENÉTICO E AS PARTICULARIDADES EM ESTUDOS DE <i>C. CANEPHORA</i> . .....	18
2.7. ANÁLISE DE QTLs.....	22
2.7.1. <i>Regressão de Haseman e Elston</i> .....	27
2.7.2. <i>Regressão de Fulker e Cardon</i> .....	28
2.7.3. <i>Importância do Mapeamento de QTL e sua aplicação em estudo com café</i> .30	
<b>3. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>32</b>
<b>4. CAPÍTULO 1: EFEITO DA CODIFICAÇÃO DE DADOS MOLECULARES EM ESTUDOS GENÉTICOS INTRA-POPULACIONAIS .....</b>	<b>39</b>
<b>5. CAPÍTULO 2: ANÁLISES GENÔMICAS E COMPONENTES DE RESISTÊNCIA DE <i>Coffea canephora</i> À <i>Hemileia vastatrix</i>. .....</b>	<b>67</b>
<b>6. CAPÍTULO 3: DESENVOLVIMENTO E VALIDAÇÃO DE EST-SSR PARA <i>Coffea arabica</i>: TRANSFERABILIDADE E APLICAÇÃO EM ESTUDOS DE CARACTERIZAÇÃO MOLECULAR E MAPEAMENTO GENÉTICO. ....</b>	<b>102</b>
<b>7. CONCLUSÕES GERAIS.....</b>	<b>134</b>

## RESUMO

FERRÃO, Luís Felipe Ventrím. M. Sc., Universidade Federal de Viçosa, fevereiro de 2013. **Marcadores microssatélites em estudos de diversidade, mapeamento genético e análises de QTLs em *Coffea canephora***. Orientadora: Eveline Teixeira Caixeta, Coorientadores: Cosme Damião Cruz, Maria Amélia Gava Ferrão e Ney Sussumu Sakiyama.

O uso de marcadores moleculares em estudos genético apresenta-se como uma técnica potencialmente capaz de auxiliar no melhoramento genético de *Coffea canephora*. Neste contexto, os marcadores microssatélites (SSR) tem sido preferencialmente escolhidos, pois agregam qualidades importantes, dentre as quais se destacam a reprodutibilidade, altos níveis de polimorfismos, expressão codominante e multialelismos. Em programas de melhoramento genético essa tecnologia pode ser utilizada para diferentes finalidades com destaque para os estudos de diversidade genética, que auxiliam na avaliação dos recursos genéticos disponíveis e seleção de potenciais genitores; e nas análises de QTLs, que permitem a elucidação do caráter quantitativo e utilização dessas informações em programas de seleção assistida por marcadores (SAM) e clonagem posicional de genes. Neste sentido o presente trabalho teve como objetivo geral demonstrar a aplicação dos marcadores microssatélites em estudos de diversidade, mapeamento genético e análises de QTLs em *C. canephora*. Nos estudos de diversidade genética, apresentados no capítulo 1, o resultado da codificação de dados nas avaliações de germoplasma foi mensurado e observou-se uma perda significativa de informações genéticas, o que influencia no manejo e caracterização dos recursos genéticos armazenados nas coleções. Para os estudos de mapeamento e análises de QTLs, apresentados no capítulo 2, os marcadores SSR foram utilizados para a construção do mapa de ligação genético e identificação de fatores genéticos associados à resistência à ferrugem (*Hemileia vaxtatrix*), que atualmente é a principal doença foliar do cafeeiro. Neste estudo, progênies de irmãos completos de uma população de melhoramento do Incaper (Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural) foram genotipadas e fenotipadas, considerando componentes de resistência avaliados em campo e em laboratório. O resultado do mapeamento de QTL foi a identificação, no grupo de ligação 1, de QTLs significativos para todas as características consideradas no estudo, o que tornou essa região genômica potencialmente útil em estudos de clonagem posicional ou SAM, em programas de melhoramento genético. Por fim, visto a importância dos marcadores moleculares nos

estudos com cafeeiros, no capítulo 3, foi descrito um conjunto de 101 novos *primers* SSRs, minerados de sequencias expressas (EST) do Projeto Brasileiro do Genoma Café. O nível de polimorfismos, a taxa de transferabilidade e sua aplicação em estudos de caracterização molecular e mapeamento genético dentro do gênero *Coffea* foram considerados e mostrou que os EST-SSR descritos são ferramentas de grande valia para estudos genéticos com cafeeiros. Dessa forma, os resultados obtidos nesse estudo são importantes para os estudos genéticos dentro da espécie, uma vez que foram disponibilizados conhecimentos importantes que tangem o manejo de germoplasmas, seleção de genótipos resistentes à ferrugem, diversidade genética e seleção assistida por marcadores moleculares.

## ABSTRACT

FERRÃO, Luís Felipe Ventorim. M. Sc., Universidade Federal de Viçosa, February, 2013. **Microsatellite markers in diversity studies, genetic mapping and QTL analysis in *Coffea canephora***. Adviser: Eveline Teixeira Caixeta, Co-advisers: Cosme Damião Cruz, Maria Amélia Gava Ferrão and Ney Sussumu Sakiyama.

The use of molecular markers in genetic studies is an important technique that is able to assist the breeding programs of *Coffea canephora*. In this context, the microsatellite (SSR) has been preferentially chosen because it has reproducibility, high levels of polymorphism, codominant expression and multiallelism. In breeding programs, this technology can be used for different purposes. In genetic diversity studies, it is important for the evaluation of genetic resources and selection of potential parents for crossings. In analysis of QTL, it allows the elucidation of quantitative trait and this information can be used in programs that aim the marker assisted selection (MAS) or positional cloning of genes. In this sense, the present study aimed to demonstrate the application of microsatellite markers in diversity studies, genetic mapping and QTL analysis in *C. canephora*. In genetic diversity studies, presented in Chapter 1, the result of the transformation of data was measured in germplasm studies. We observed a significant loss of genetic information, which influences the management and characterization of genetic resources stored in these collections. In the QTL and mapping studies, presented in Chapter 2, the SSR markers were used to construct the genetic linkage map and identification of genetic factors associated with resistance to coffee leaf rust (*Hemileia vastatrix*), which is the major foliar disease of coffee. In this study, we used a population of full-sib progenies, that is maintained in the Incaper (Institute, Technical Assistance and Rural Extension of the Espírito Santo State). This population was phenotyped, considering the components of resistance evaluated in field and in laboratory. We identified, in linkage group 1, a significant QTL for all traits considered in this study and this genomic region is potentially useful in studies of positional cloning or MAS in breeding programs. Finally, because the importance of molecular markers in studies with coffee, in chapter 3, we described a set of 101 new SSR primers, mined from expressed sequences (EST) Coffee Genome Brazilian Project. The level of polymorphisms, the rate of transferability and its application in studies of molecular characterization and genetic mapping within the genus *Coffea* were considered. The EST-SSR showed that they are powerful tools for genetic studies with coffee. Thus, the results obtained in this study are important for

genetic studies within species, since they were made available important knowledge that can be used in germplasm management, selection of genotypes resistant to rust, genetic diversity and molecular marker-assisted selection.

## 1. INTRODUÇÃO GERAL

Das diferentes atividades ligadas ao negócio agrícola, o café está entre as de maior importância econômica e social. Com papel de destaque no cenário mundial, o Brasil lidera a produção e exportação de grãos e é o segundo maior consumidor da bebida no mundo. Pertencente ao gênero *Coffea*, que possui centenas de espécies, a bebida amplamente conhecida como café é produzida a partir dos grãos de apenas duas delas: *Coffea arabica*, que contribui com o aroma e sabor adocicado; e *Coffea canephora*, que apresenta maiores quantidades de cafeína, sólidos solúveis e oferece corpo à bebida. Na década de 50, com a ascensão do café solúvel, a espécie *C. canephora*, considerada naquela época como uma espécie de qualidade inferior para bebida, começou a ser explorada comercialmente em misturas (*blends*) com *C. arabica* nas indústrias de torrado e moído em função de sua menor acidez e maior quantidade de sólidos solúveis. Além de contrabalançar a acidez e conferir corpo ao produto industrializado, os *blends* obtiveram boa eficiência industrial o que resultou em preços baixo e, conseqüentemente, maior competitividade no mercado. Este fato impulsionou a produção mundial de *C. canephora*.

Popularmente conhecido como café Robusta, atualmente, o Brasil destaca-se como o segundo maior produtor de *C. canephora* do mundo, ficando atrás apenas do Vietnã. Originário de regiões tropicais, quentes, úmidas e de baixa altitude da África; em território nacional, a cultura ganhou destaque naquelas regiões de temperatura média entre 22 e 26<sup>0</sup>C e déficit hídrico inferior a 220 mm/ano. Dessa forma, a produção cafeeira concentrou-se nos estados do Espírito Santo, Rondônia e, em menor escala, na Bahia, Mato Grosso, Pará e Minas Gerais. Além da importância na indústria, a espécie destaca-se também pela alta produtividade de grãos, menor susceptibilidade a estresses abióticos e resistência à ferrugem, doença causada pelo fungo biotrófico (*Hemileia vastratix*). Todas essas características fizeram do café Robusta alvo de programas de melhoramento genético, que priorizaram a melhoria das características agrônomicas associadas à produção e qualidade final do produto, visando a redução de custo de produção e maior sustentabilidade da atividade. Dentre essas melhorias, a atenção para com as doenças é um ponto importante, pois é um fator limitante da produção de café.

Dentre as doenças da cultura do cafeeiro a ferrugem, é considerada a principal podendo levar a perdas na ordem de 35-50% na produção de *C. arabica*. Em *C. canephora* o número de trabalhos com a doença ainda são modestos e poucos conclusivos, todavia, já foi observado que a amplitude fenotípica é elevada, com materiais genéticos variando de altamente resistentes a total susceptibilidade.

Atualmente a medida de manejo mais utilizado para o controle da ferrugem é o uso de defensivos agrícolas, em especial, os fungicidas protetores e sistêmicos. Contudo, sabe-se que o uso desses produtos proporciona o aumento no custo de produção e eleva a probabilidade de contaminação do meio ambiente. Somado a isso, o uso intensivo de fungicidas sistêmicos pode favorecer o aumento da frequência de isolados de *H. vastatrix* resistentes, diminuindo assim a efetividade das aplicações e, conseqüentemente, a sustentabilidade da atividade. Neste contexto, o desenvolvimento de variedades geneticamente resistentes é a melhor alternativa para o controle da doença. Para isso os marcadores moleculares são considerados uma ferramenta valiosa, visto que podem auxiliar não apenas na seleção de genótipos resistentes, mas também no melhor entendimento da herança e na piramidação de genes de resistência.

Os estudos utilizando dados moleculares, quando comparados com os fenotípicos, apresentam a vantagem de serem mais polimórficos, neutros fenotipicamente e aplicados em qualquer estágio de desenvolvimento da planta. Dentre as diferentes classes de marcadores moleculares reportadas na literatura, os microssatélites ou SSR destacam-se por apresentarem distribuição frequente e aleatória no genoma, alta reprodutibilidade, simplicidade, altos níveis de polimorfismos e expressão codominante. Essas características fazem com que esses marcadores sejam os mais utilizados em trabalhos com plantas, sobretudo, nos estudos de caracterização molecular, mapeamento genético e análise de QTLs. Desse modo, na concepção do presente trabalho, propôs-se utilizar os marcadores microssatélites para estudos genéticos em *C. canephora*, principalmente no que tange os estudos de diversidade e manejo de bancos de germoplasmas; no direcionamento dos programas de melhoramento genético, sobretudo, na seleção de genótipos resistentes para ferrugem e na análise de caracteres de herança complexas; e, finalmente, na validação e disponibilização de novos marcadores EST-SSRs com grande aplicabilidade em estudos de caracterização molecular, mapeamento e diversidade genética no gênero *Coffea*. Para tanto, foram propostos os seguintes objetivos específicos: i) Avaliação de marcadores

moleculares dominantes e codominantes em estudos de diversidade genética intra-populacional; ii) Plasticidade fenotípica, componentes de resistência e mapeamento de QTLs, aplicados ao estudo da ferrugem (*Hemileia vastatrix*); (iii) Desenvolvimento e validação de novos EST-SSRs para o gênero *Coffea*.

## 2. REVISÃO DE LITERATURA

### 2.1) A espécie *Coffea canephora*

Com origem na África, o cafeeiro pertence à família Rubiaceae, que contém aproximadamente 500 gêneros e mais de seis mil espécies. Dentre todos estes gêneros, particular atenção tem sido dada ao gênero *Coffea*, ao qual se inclui duas espécies cultivadas de grande importância econômica: *Coffea arabica* L. e *Coffea canephora* Pierre ex Froehner. Outras espécies, apesar de não apresentarem valor comercial, representam importantes fontes de variabilidade genética para diferentes características agrônomicas de interesse, que são de grande valia em programas de hibridação e de melhoramento genético (Ferrão et al., 2007).

Praticamente todo o café produzido e consumido no mundo é originado de variedades de *C. arabica* e *C. canephora*, que são espécies bem distintas. *C. arabica* é a espécie mais cultivada e comercializada no mundo, é alotetraplóide ( $2n=4x=44$ ), autógama, adaptada aos locais de altitudes elevadas e clima ameno. Já os cafeeiros da espécie *C. canephora*, também chamados de café Conilon ou Robusta, são diploides ( $2n=2x=22$ ), com autoincompatibilidade genética do tipo gametofítica, reproduzem-se por fecundação cruzada e são adaptados a regiões de altitudes mais baixas e temperaturas mais elevadas (Berthaud and Charrier, 1988; Ferrão, 2004; Leroy et al., 2005).

Tradicionalmente, a produção de café no Brasil e no mundo concentrava-se apenas na espécie *C. arabica* (café arábica). Entretanto, a partir do fim do século XIX, devido a um grande surto de ferrugem (*Hemileia vastatrix*), que afetou os cafezais do sul e leste da Ásia, o café Robusta, que se mostrava resistente à doença, passou a ser alvo de estudos científicos visando a sua exploração econômica (Charrier and Berthaud, 1988).

No Brasil, a espécie *C. canephora* foi introduzida por volta de 1920 no Estado do Espírito Santo, segundo consta, pelas mãos do ex-governador Jerônimo Monteiro. As primeiras sementes foram plantadas no município de Cachoeiro do Itapemirim e, posteriormente, levadas para a região norte do Estado. O objetivo inicial do cultivo foi ocupar as áreas consideradas marginais para o arábica (Ferrão et al., 2007). Na década de 50, com a ascensão do café solúvel, o café Robusta começou a ser explorado comercialmente em misturas (*blends*) com café arábica, nas indústrias de torrado e moído em função de sua menor acidez e maior quantidade de sólidos solúveis. Além de contrabalançar a acidez e conferir corpo ao produto industrializado, os *blends* obtiveram boa eficiência industrial o que resultou em preços baixo e conseqüentemente, maior competitividade no mercado. Este fato impulsionou a produção mundial de *C. canephora*, liderado principalmente por países como Vietnã, Brasil e Indonésia. Originário de regiões tropicais, quentes, úmidas e de baixa altitude da África; em território nacional a cultura ganhou maior destaque naquelas regiões de temperatura média entre 22 e 26 °C e déficit hídrico inferior a 220 mm/ano. Dessa forma, a produção cafeeira da espécie concentrou-se nos estados do Espírito Santo e Rondônia e em menor escala na Bahia, Mato Grosso, Pará e Minas Gerais.

Segundo dados da Conab (2012), atualmente 12,5 milhões de sacas de café Robusta foram produzidos no Brasil, o que representa 24,6% da produção nacional. Em razão disso, o agronegócio da cultura atua de forma importante na geração de empregos, tributos e formação de receita cambial para o Brasil, desempenhando assim, papel fundamental no desenvolvimento social e econômico do país.

## **2.2. Melhoramento genético de *C. canephora***

O resultado imediato do incremento da produção de *C. canephora*, a partir da década de 50, foi o aumento em pesquisas relacionadas com a espécie. Nas últimas décadas a pouca ênfase dada ao café Robusta em detrimento da cafeicultura do arábica deve-se, principalmente, ao fato do primeiro ser considerado café de qualidade inferior. Tal situação foi sendo revertida com o desenvolvimento de programas de melhoramento genético, que objetivam o aumento da produtividade, estabilidade de produção, resistência a pragas e doenças e a melhoria da qualidade de bebida (Ferrão et al., 2007)

A presença de variabilidade genética na população é condição básica para obtenção de sucesso em um programa de melhoramento genético. Associado a isso, é de fundamental importância a existência de médias altas das características de interesse, pois permite a seleção de genótipos superiores e possibilita o incremento da frequência de genes favoráveis nos materiais genéticos estudados. Concomitantemente, estratégias que permitam identificar genótipos realmente superiores são imprescindíveis. Entretanto, para definir estratégias de melhoramento e seleção de plantas, é requerido um aprimorado conhecimento sobre a biologia floral da espécie; forma de reprodução e propagação; números de cromossomos; o germoplasma existente, sua localização e facilidade de utilização; e informações sobre a base genética, herança, herdabilidade dos caracteres e correlação entre eles. Somado a isso, faz parte do planejamento a definição do(s) local (is) de execução da pesquisa, como também, a escolha das condições e estratégias mais adequadas para realização dos experimentos. Sempre considerando os objetivos da pesquisa, a estrutura física, os recursos financeiros e humanos necessários, e o tempo disponível para a realização do trabalho. (Ferrão et al., 2007).

Para realização de melhoramento do cafeeiro, são necessários procedimentos especiais, devido a particularidades importantes da espécie, quais sejam: período normalmente mais longo para obter as flores para a realização dos cruzamentos e, assim, produzir sementes; custo alto das avaliações de campo, devido à necessidade de áreas maiores e de tempo longo; necessidade de avaliação da precocidade e da longevidade produtiva; e necessidade de avaliação da oscilação anual de produção, florescimento e produção anual sobre a mesma planta (Ferrão, 2004; Ferrão et al., 2007).

Outro aspecto importante da espécie *C. canephora* é a alogamia natural. Tal característica reprodutiva limita a fixação de caracteres quando propagado via sexuada, fazendo com que populações naturais sejam altamente heterozigotas e com ampla variabilidade genética (Fonseca, 1999). Neste sentido, segundo Ferrão et al. (2007), o melhoramento genético de *C. canephora* apresenta a possibilidade de se trabalhar com a propagação vegetativa, o que em termos genéticos significa obter ganhos maiores e mais rápidos, quando comparado com outras culturas perenes que não apresentam essa característica. Assim sendo, diferentes estratégias de melhoramento vêm sendo utilizadas para espécie, sendo que em quase todas utiliza-se da reprodução assexuada para fixação do genótipo. Dentre os métodos de melhoramento comumente utilizados,

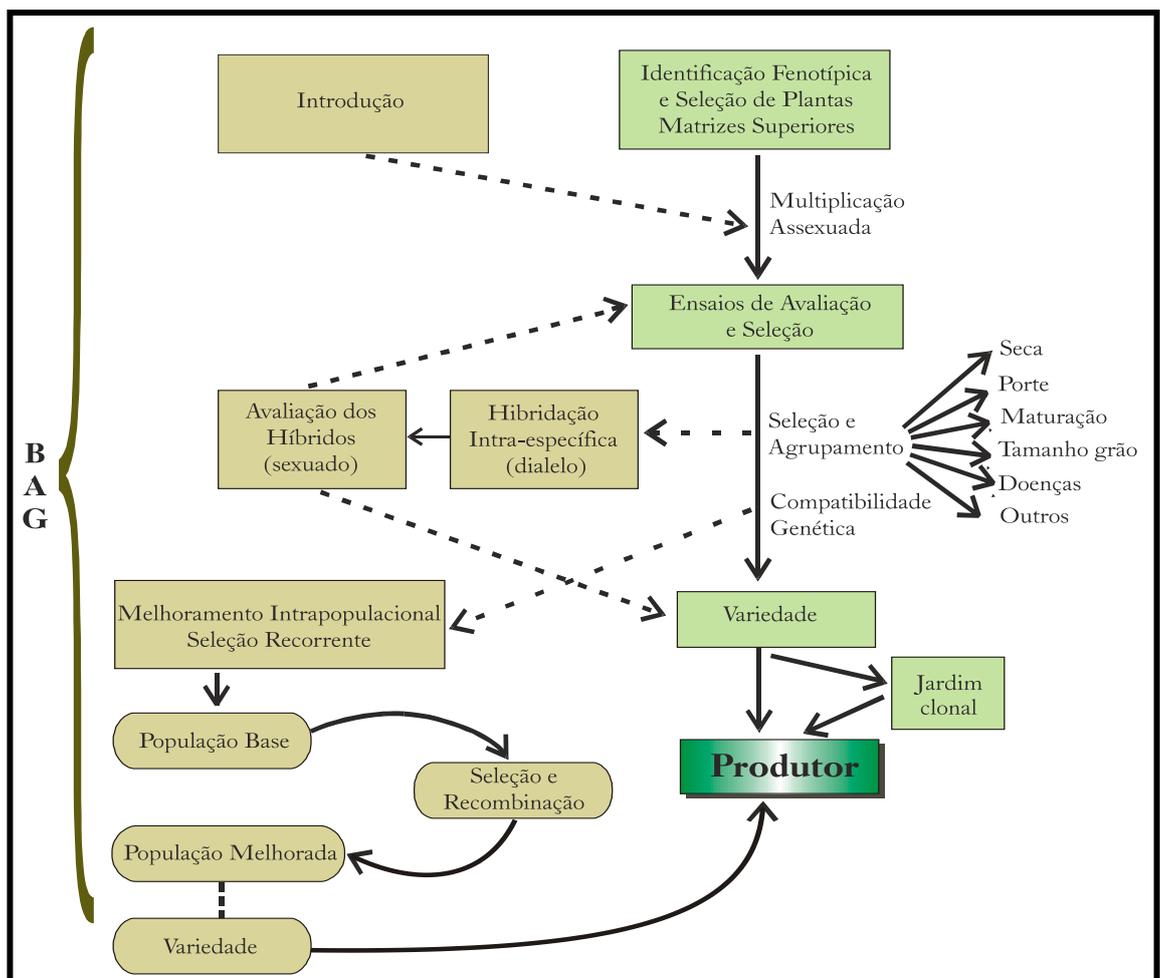
destacam-se: introdução de germoplasma, seleção clonal, hibridação, seleção recorrente e haploidiploidização. Dentre essas, um dos que se adéqua muito bem ao café Robusta é a seleção recorrente recíproca interpoblacional, especialmente, por se tratar de uma espécie alógama, de fácil propagação vegetativa e que apresenta germoplasma geneticamente estruturado em populações polimórficas (Ferrão et al., 2007; Souza, 2011). Essas características permitem que populações sejam melhoradas de forma independente, compondo assim grupos heteróticos bem definido e com indivíduos altamente heterozigotos, que podem ser utilizados para exploração da heterose.

Uma importante recomendação em programas de melhoramento genético de *C. canephora* é a condução paralela de reproduções assexuada e sexuada. A primeira apresenta a vantagem de fixar um genótipo a qualquer tempo, sem que haja a necessidade de avançar gerações para tal finalidade. Entretanto, apresenta como desvantagem o estreitamento da base genética da população, dificultando o estabelecimento de estratégias de melhoramento. Para minimizar essa questão recomenda-se no melhoramento genético usar em paralelo as estratégias que envolvam as duas formas de reproduções, pois a sexuada, permite a recombinação genética recuperando assim a variabilidade genética e minimizando os possíveis riscos de estreitamento da base genética da cultura (Charrier and Berthaud, 1988; Ferrão, 2004; Ferrão et al., 2007; Fonseca, 1999). Essa estratégia tem sido utilizada com sucesso na Costa do Marfim, África e no Estado do Espírito Santo, Brasil.

No Brasil, um dos principais programas de melhoramento genético da espécie é desenvolvido no Estado do Espírito Santo, pelo Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural – Incaper. Iniciado em 1985, o programa do Incaper desenvolveu, liberou e recomendou para plantio seis variedades de Conilon, sendo cinco deles clonais e uma propagada por sementes. A preferência pelo grupo varietal Conilon se deve a sua ampla adaptação, tolerância à seca, facilidade de manejo e ao contexto histórico-cultural que foi criado no Estado. Atualmente, o Estado do Espírito Santo destaca-se como o maior produtor de *C. canephora* do país, sendo responsável por 77,8% de todo o café Conilon produzido no Brasil (Conab, 2012). No Estado, as variedades de café do Incaper ocupam cerca de 50% da área cultivada e estima-se que de 1993, data do lançamento da primeira variedade, até os dias atuais a produção de café no Espírito Santo aumentou 277% (Merlo, 2012). Muito deste sucesso

é devido a forte parceria estabelecida entre o Instituto, fornecedor das tecnologias, e os produtores, que seguem as recomendações técnicas estabelecidas (Ferrão et al., 2007).

O fluxograma do programa de melhoramento do Incaper está resumido na Figura 1. O que se observa é um programa de melhoramento dinâmico e baseado em diferentes estratégias, mas que passa quase sempre pela multiplicação assexuada dos seus materiais. Como já citado, a propagação vegetativa permite que o melhorista aproveite integralmente a variância genética, o que possibilita a fixação de caracteres de interesse. Além da vantagem genética, que em termos práticos significa economia de tempo e de áreas experimentais, o desenvolvimento de variedades clonais é de grande interesse também dos produtores, visto que o agrupamento de genótipos com diferentes épocas de maturação permite o escalonamento da colheita (Ferrão, 2004; Ferrão et al., 2007; Fonseca, 1999).



**Figura 1.** Fluxograma do programa de melhoramento genético de *Coffea canephora* do Incaper, segundo Ferrão et al. (2007).

Outros programas de melhoramento genético de *C. canephora* têm sido conduzidos no Brasil pelo Instituto Agrônomo de Campinas (IAC), São Paulo; e pela Embrapa, em Rondônia e pela Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG) em associação com a Universidade Federal de Viçosa (UFV). No continente asiático merecem destaque os estudos realizados na Índia, Indonésia e Vietnã; sendo o último o maior produtor de café Robusta do mundo. Na África, centro de diversidade da espécie, estudos voltados para o melhoramento de café são realizados pela Costa do Marfim, Madagascar, Camarões e Uganda (Souza, 2011).

### **2.3. Ferrugem alaranjada do café (*Hemileia vastatrix*)**

A ferrugem alaranjada do cafeeiro é causada pelo fungo biotrófico *Hemileia vastatrix*, Basidiomicota, Ordem Uredinales e Família Pucciniaceae. A literatura registra a sua primeira ocorrência em 1861, sendo que em 1868, no Ceilão, atualmente Sri Lanka, foram relatados os efeitos devastadores da doença na cafeicultura da ilha. No Brasil, o primeiro relato foi feito na década de 70 na Bahia, e em menos de uma década atingiu toda a América Latina, tornando-se principal doença do cafeeiro. Apesar de não haver dados conclusivos, estima-se que no Brasil a perda de produção no café arábica seja na ordem de 35-50%, dependendo das condições ambientais e do manejo adotado na lavoura (Zambolim et al., 1999). Os sintomas da doença são fáceis de identificar e caracterizam-se pelo aparecimento de manchas amareladas translúcidas na face inferior das folhas até a formação de pústulas circulares, pulverulentas, de cor amarela e alaranjada, cobertas pelos uredósporos do fungo, que dão um aspecto de um “pó amarelado”.

A resistência do cafeeiro as raças fisiológicas de *H. vastatrix* tem sido atribuída à nove genes dominantes (SH1 a SH9), simples ou associados, que foram identificadas nas espécies *C. arabica*, *C. canephora* e *C. liberica* (Bettencourt et al., 1992; Bettencourt and Rodrigues, 1988; Prakash et al., 2005b; Wagner and Bettencourt, 1965). Esses fatores conferem resistência dos genótipos e permitem classificá-los em grupos de acordo com a interação com as raças fisiológicas do patógeno. Contudo, a obtenção de variedades com resistência durável tem sido dificultada pela grande variabilidade genética das populações naturais das raças. Por exemplo, no Brasil já

foram relatados no gênero *Coffea*, a ocorrência de 15 raças fisiológicas e a presença de genótipos com resistência vertical e horizontal, indicando que os níveis de resistência podem variar de acordo com a raça e os genótipos analisados. No país, as pesquisas têm mostrados que a raça II é a mais frequente em estudos com *C. arabica* e *C. canephora* (Zambolim et al.,2005)

Atualmente a medida de manejo mais utilizado para o controle da ferrugem é o uso de defensivos agrícolas, em especial, os fungicidas protetores e sistêmicos. Contudo, sabe-se que o uso desses produtos não é a forma mais adequada de controle a doença, visto que aumenta o custo de produção e eleva a probabilidade de contaminação do meio ambiente. Somado a isso, o uso intensivo de fungicidas pode favorecer o aumento da frequência de isolados de *H. vastatrix* resistentes, diminuindo assim a efetividade das aplicações e, conseqüentemente, a sustentabilidade da atividade. Nesse contexto, o desenvolvimento de variedades geneticamente resistentes é a melhor alternativa para o controle da doença. No Brasil, trabalhos de melhoramento com esses fins são realizados pelo Instituto Agrônômico de Campinas (IAC), Universidade Federal de Viçosa (UFV), Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig), Instituto Agrônômico do Paraná (Iapar) e Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper).

Além do desenvolvimento de variedades resistentes, uma segunda estratégia para conferir maior resistência às plantas no campo é denominado de “multilinhas” e pode ser implementado no cultivo de *C. canephora*. Como na cultura, é normal o plantio dos genótipos em linhas, recomenda-se a utilização de variedades resistentes entre os materiais, de modo a servirem como barreira física e genética para o patógeno, reduzindo assim o aumento da frequência de isolados do patógeno capazes de suplantar a resistência dessa variedade.

A ferrugem do cafeeiro foi mais bem estudada na espécie *C. arabica*, enquanto que em *C. canephora* as informações são modestas, sobretudo, pelo entendimento de que a espécie era altamente resistente à doença. No entanto, observa-se que os níveis de resistência são alterados de genótipo para genótipo em uma escala de plantas que podem variar de resistentes, passando por níveis moderados de resistência, e chegando até a completa susceptibilidade. Um importante estudo sob a epidemiologia e resistência do cafeeiro à ferrugem foi relatado por Capucho (2012), que dentre as conclusões mencionada está o fato de que (i) a biologia de *H. vastatrix*, avaliada para germinação e

infectividade, na espécie *C. canephora* se assemelha à biologia do fungo em *C. arabica*; (ii) as características das lesões apresentadas na variedade Catuaí Vermelho IAC 144 (*C. arabica*, susceptível) foram muito semelhantes aquelas observadas em alguns Conilons (grupo varietal de *C. canephora*), rejeitando a ideia de que uma espécie é resistente e outra é susceptível; e (iii) diferentes níveis de resistência foram observados com plantas de Conilon variando de alta resistência a susceptibilidade.

Essas e outras informações sobre a doença reforçam a importância da inclusão da ferrugem como mais uma variável de interesse agrônomo nos programas de melhoramento genético de *C. canephora*. Assim, faz-se necessário o uso de métodos adequados em programas de melhoramento que permitam a identificação e seleção de plantas com maiores níveis de resistência. Neste contexto, os marcadores moleculares são uma ferramenta valiosa, visto que podem auxiliar não apenas na seleção de genótipos resistentes, mas também no melhor entendimento da resistência a ferrugem e na piramidação de genes de resistência.

#### **2.4. Marcadores moleculares**

O termo marcador tem sido utilizado para designar fatores morfológicos, fisiológicos, bioquímicos ou genéticos que por co-segregação com os genes de interesse, permitem o estudo comparativo de genótipos e de suas progênies. Entre esses, os marcadores de DNA merecem destaque pela grande aplicabilidade em estudos genéticos, para diferentes fins, em plantas cultivadas.

Nos estudos genéticos dos anos 60, os marcadores utilizados eram controlados por genes associados a caracteres morfológicos (características fenotípicas de fácil visualização), sendo esses as principais fontes de seleção utilizadas pelos melhoristas de plantas. No entanto, o desenvolvimento de marcadores mais moderno, fez com que esses ficassem em desvantagem por se apresentarem em número limitado nas espécies, o que resultava na redução da possibilidade de associação entre esses marcadores e características de interesse econômico (Ferreira and Grattapaglia, 1998).

Segundo esses mesmos autores, esse quadro começou a mudar com o advento das modernas técnicas de biologia molecular, que permitiram a detecção de polimorfismo diretamente ao nível de DNA. Nesse contexto, surgiram os marcadores

moleculares, que quando comparados com marcadores fenotípicos apresentam como vantagem: a) maior nível de polimorfismo por loco estudado; b) neutralidade fenotípica, com mínimo ou nulo efeito epistático ou pleiotrópico; c) em geral, codominantes, contendo maior quantidade de informação genética; e d) podem ser identificados em qualquer estágio de desenvolvimento da planta, inclusive em amostras de células ou tecidos.

Em artigo de revisão, Agarwal et al. (2008) relataram que na literatura são listados mais de vinte diferentes tipos de marcadores moleculares. De acordo com os mesmos, a maioria desses pode ser dividida em duas classes de acordo com a metodologia utilizada: (1) baseados em técnicas de hibridização ou (2) baseado em técnicas de PCR (*Polymerase Chain Reaction*).

Outro importante ponto a ser considerado é a adequação do tipo de marcador molecular ao objetivo da pesquisa, uma vez que, cada marcador é baseado em diferentes princípios, o que os fazem diferir quanto à consistência e tempo para obtenção de resultados; nível de polimorfismos; custo e facilidade de implementação; e reprodutibilidade (Ferrão et al., 2013). Em estudos de diversidade genética de *C. canephora* diferentes trabalhos são relatados na literatura utilizando, principalmente, os seguintes tipos de marcadores baseados em PCR: RAPD (*Random Amplified Polymorphic DNA*), AFLP (*Amplified Fragment Length Polymorphism*) e SSR (*Simple Sequence Repeats*) (Aggarwal et al., 2007; Ferrão et al., 2013; Lashermes et al., 2000; Lashermes et al., 1997; Musoli et al., 2009; Souza, 2011)

Em estudos recentes, Ferrão et al. (2013) avaliaram a eficiência dos marcadores RAPD, AFLP e SSR em estudos genéticos com *C. canephora*. Os microsatélites (SSRs), por se tratar de um marcador codominante e multialélico, foram considerados o mais informativo dos três e apresentou os melhores resultados nos estudos de diversidade genética. Frequente e distribuído ao acaso pelo genoma, os SSRs combinaram informatividade, ampla cobertura do genoma, robustez e reprodutibilidade o que torna o marcador adequado para estudos com a espécie. Assim sendo, a seguir será detalhada a base genética do marcador SSR e sua importância nos estudos dentro do gênero *Coffea*.

#### 2.4.1. Base genética dos marcadores *Microssatélites*

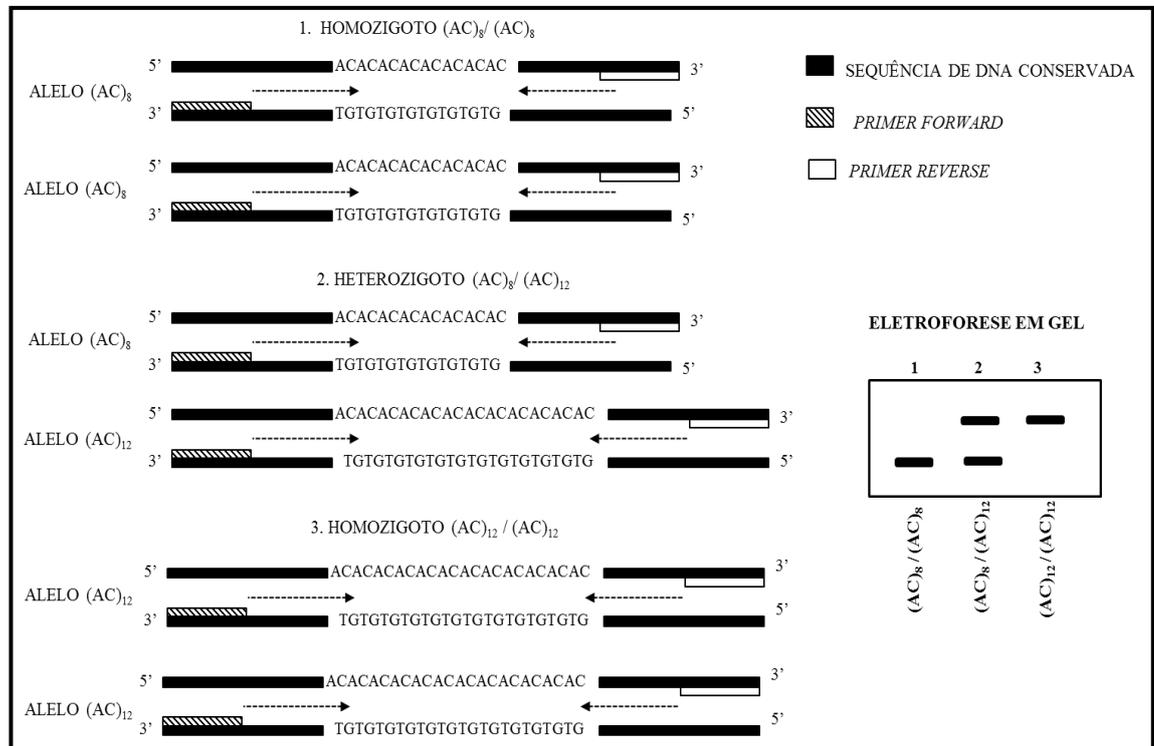
Experimentos realizados nos anos 80 demonstraram que os genomas de eucariotos são densamente povoados por diferentes classes de sequências repetidas. Essas regiões foram denominadas microssatélites ou SSR e apresentam de 1 a 6 nucleotídeos de comprimento, repetidos em tandem (Agarwal et al., 2008).

Enquanto esse número de repetições em tandem de um microssatélite geralmente varia, as sequências de DNA que os flanqueiam geralmente são conservadas entre indivíduos de uma mesma espécie. Isso permite o desenho de *primers* específico (20 a 30 bases de comprimento) para as sequências adjacentes a um dado microssatélite de forma que, através de uma reação de PCR seja possível amplificar esse loco em diferentes genótipos (Figura 2).

Os produtos dessa amplificação são observados em gel de poliacrilamida e as diferenças no número de repetições são responsáveis pelo polimorfismo de tamanho entre os fragmentos de PCR. A explicação exata para a essas variação no número de unidades repetidas ainda não é totalmente conhecida, entretanto acredita-se que seja causada por eventos mutagênicos dinâmicos e complexos, tais como: crossing over desigual, retrotransposição e, sobretudo, deslizamento da DNA polimerase (*slippage*) (Ferreira & Grattapaglia, 1998). Essa alta taxa de variação, faz das regiões microssatélites um loco genético altamente polimórfico, multialélico e de grande conteúdo informacional. Essas características fazem dos SSR um dos preferidos para estudos genéticos em plantas.

Segundo Caixeta et al. (2009), os marcadores SSR tem sido desenvolvidos em várias espécies de plantas cultivadas, substituindo rapidamente outros marcadores em vários tipos de estudos genético. Isso se deve, sobretudo, à distribuição frequente e aleatória, que permite uma cobertura completa do genoma; alta reprodutibilidade; simplicidade; baixo custo e altos níveis de polimorfismos. Além disso, outra vantagem desses marcadores sobre os demais é a expressão codominante, ou seja, ambos os alelos de um indivíduo heterozigoto são visualizados. Por outro lado, a grande limitação dessa classe de marcadores é a necessidade de serem isolados e desenvolvidos especificamente para cada espécie, não sendo possível utilizar *primers* anônimos, como

no RAPD. Somado a isso, o desenvolvimento desses marcadores envolve um processo demorado, trabalhoso e de alto custo.



**Figura 2.** Base genética dos marcadores microssatélites (Caixeta et al., 2012)

#### 2.4.2 Marcadores microssatélites em café

Atualmente, devido às características citadas anteriormente, os marcadores microssatélites têm sido o preferido para estudos de mapeamento genético, diversidade, análise de QTLs, seleção assistida e manutenção de germoplasma em café. Apesar dessas vantagens o número de marcadores SSRs para o gênero *Coffea* é ainda considerado limitado (Hendre et al., 2008). Dentre os motivos para o número modesto de SSRs, destacam-se as dificuldades de desenvolvimento de *primers* específicos.

A identificação de locos microssatélites, normalmente, ocorre por duas abordagens: via bibliotecas genômicas de DNA ou por sequenciamento de cDNA. O uso de bibliotecas genômicas de DNA, apesar de relativamente simples apresenta baixa eficiência, custo elevado e é laborioso; o que limita o uso na maioria das espécies vegetais (Guichoux et al., 2011). Por outro lado, nas últimas décadas, o avanço nas técnicas de sequenciamento permitiu que milhares de sequências expressas de DNA

(ESTs) fossem sequenciadas e depositadas em bancos públicos e privados de DNA. Sequências ESTs são fragmentos sequenciados de DNA complementar às sequências de RNA mensageiro e representa parte de uma região transcrita do genoma em certas condições fisiológicas. Esse panorama abriu a oportunidade para o desenvolvimento de marcadores SSRs com maior facilidade e menores custos. Assim, ESTs-SSRs (SSRs gênicos) passaram a ser desenvolvidos com a vantagem de estarem relacionados com porções funcionais do genoma e apresentarem altos níveis de transferabilidade entre espécies relacionadas (Guichoux et al., 2011; Kalia et al., 2010).

O panorama atual do número de marcadores SSRs para o gênero *Coffea*, mostra que é necessário aumentar os esforços para o desenvolvimento e validação de novos marcadores, visto a sua importância nos estudos genéticos. Na literatura estão listados para a espécie *C. canephora* e *C. arabica* cerca de 400 SSRgenômicos (Baruah et al., 2003; Combes et al., 2000; Coulibaly et al., 2003; Cristancho and Gaitán, 2008; Hendre et al., 2008; Leroy et al., 2005; Missio et al., 2010; Moncada and McCouch, 2004; Poncet et al., 2007; Poncet et al., 2004), e 68 EST-SSRs (Aggarwal et al., 2007; Bhat et al., 2005; Geromel et al., 2006; Poncet et al., 2006). Trata-se de um número modesto quando comparamos com outras importantes espécies cultivadas.

A seguir serão detalhadas as aplicações dessa tecnologia em estudos de diversidade, mapeamento genético e análise de QTLs; em *C. canephora*. Ressalta-se que as técnicas moleculares foram responsáveis por grande parte dos avanços obtidos nessas áreas, pois permitiram a obtenção de resultados mais acurados e, também, uma melhor compreensão dos fenômenos complexos, que são difíceis de serem elucidados utilizando apenas dados fenotípicos.

## **2.5 Bancos de germoplasma e avaliação dos recursos genéticos**

Em um programa de melhoramento, o estudo da diversidade genética e a avaliação dos recursos genéticos disponíveis são de primordial importância, principalmente no início do programa, onde ocorre o planejamento e as definições das estratégias de trabalhos.

Neste contexto, os bancos de germoplasmas são de grande valia, pois são estruturas físicas onde estão centralizadas todas as atividades relativas ao manejo de

recursos genéticos. Dentre essas atividades destacam-se: aquisição de material, caracterização, avaliação, distribuição, intercâmbio e conservação dos materiais ali depositados. O termo germoplasma tem sido definido como todo o material hereditário de uma espécie, ou ainda, como todo o patrimônio genético de uma espécie que é transmitida de uma geração para outra (Borém, 1998). Esse patrimônio pode estar na forma de pólen, anteras, plantas, sementes, tecidos (meristemas ou calos), células ou estruturas simples. A conservação desses materiais, denominados de acessos, podem seguir dois métodos básicos, denominados *ex situ* e *in situ*. Esses por sua vez, dependem da natureza do material, do objetivo e do alcance da conservação. Na conservação *in situ*, a biodiversidade é conservada dentro do ecossistema mantendo a dinâmica evolutiva do habitat original ou do ambiente natural. Por outro lado, a conservação *ex situ* é feita de forma artificial em bancos de germoplasmas, fora do habitat natural da espécie (Nass, 2001).

No Brasil, existem cerca de 180 bancos de germoplasmas, totalizando mais de 200 mil acessos de espécies vegetais, como grãos, fruteiras, florestais, hortaliças, entre outras, abrangendo 84 produtos (Valois et al., 2001). Dentre as centenas de coleções espalhadas pelo país, na literatura encontra-se registrado um pequeno número de bancos referentes à espécie *C. canephora*. As principais coleções desta espécie estão localizadas em São Paulo, no IAC; no Espírito Santo, no Incaper; em Minas Gerais, na Universidade Federal de Viçosa (UFV); e em Rondônia, na Embrapa Rondônia.

Para que os bancos de germoplasma possam ser usados com eficiência nos programas de melhoramento de plantas, é importante que haja o máximo de conhecimento da variabilidade ali presente. Neste sentido, a diversidade genética dentro dessas coleções pode ser mensurada seguindo, basicamente, duas abordagens: (i) cálculos de parâmetros genéticos da população, em particular a diversidade e sua partição em diferentes níveis; e (ii) análises da relação genética entre amostras. A primeira abordagem envolve o conhecimento da frequência alélica e o cálculo de diferentes parâmetros com o objetivo de mensurar a estrutura e a variação genética existente.

A maioria das pesquisas em bancos de germoplasma tem preferido o uso da segunda abordagem, que se refere aos enfoques comparativos que podem ser realizados entre e dentro de cada população ou grupos de indivíduos em estudo. Nestes tipos de estudos, a diversidade genética tem sido comumente estimada utilizando duas etapas

sequenciais: escolha de uma medida de similaridade (ou dissimilaridade); e a adoção de uma técnica de agrupamento para a formação dos grupos. Na primeira etapa, uma matriz de dissimilaridade genética (ou similaridade) é construída, a partir da análise de todas as possíveis combinações par-a-par, e é usada para a caracterização da estrutura genética, de modo a quantificarem e informarem sobre o grau de semelhança ou de diferença entre dois genótipos, populações ou indivíduos (Karp et al., 1997; Kosman and Leonard, 2005). Posteriormente, é necessário a utilização de métodos multivariados de modo a auxiliar na classificação, ordenamento e resumo do grande número de informações que é gerado. Esse tipo de estudo pode ser realizado utilizando diferentes metodologias estatísticas, dependendo do conjunto de dados e da finalidade da pesquisa. Revisão detalhada sobre os métodos são apresentados por Mohammadi and Prasanna (2003).

Mais recentemente, uma abordagem bayesiana tem sido utilizada nos estudos de diversidade genética e estrutura de população. Implementada no *software* STRUCUTURE (<http://pritch.bsd.uchicago.edu/software.html>) o algoritmo foi desenvolvido por Pritchard et al. (2000), para inferência de estrutura populacional e classificação de indivíduos quanto a sua população a partir de dados genotípicos. Para isso, assume-se um modelo no qual existem K populações (onde K pode ser desconhecido) que são caracterizadas a partir da frequência do conjunto de alelos de cada loco. Os indivíduos amostrados são então classificados (probabilisticamente) em populações, ou separados em duas ou mais populações se for assumido o modelo de misturas (*admixed*). Atualmente a abordagem, implementada no STRUCTURE, é uma das mais utilizadas em estudos de diversidade para diferentes fins. Além dos resultados robustos, soma-se o fato do *software* permitir fazer inferências sobre identificação de migrantes, estimação de frequências alélicas e estudos de zonas híbridas; utilizando classe de marcadores moleculares dominantes e codominantes. Extensão do método original para usos com marcadores dominantes, poliploides e outras particularidades foram apresentados por Falush et al. (2003), Falush et al. (2007) e Hubisz et al. (2009).

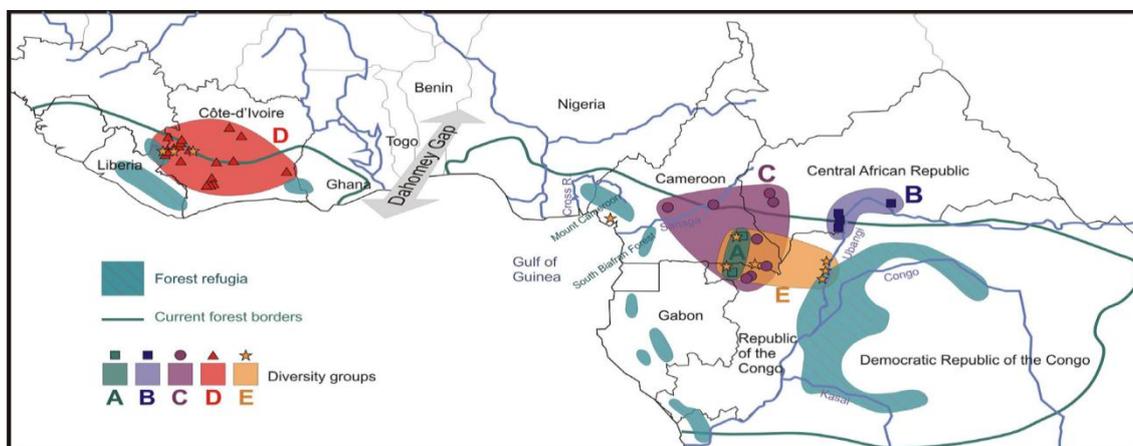
### 2.5.1. Diversidade genética em *C. canephora*

Avaliações fenotípicas, bioquímicas e moleculares têm sido empregadas para estudo da diversidade genética em *C. canephora*, tanto em populações naturais quanto nas coleções de germoplasma. Esses estudos indicam que a espécie é dividida em dois grandes grupos de acordo com seus respectivos centros de diversidade e está fortemente relacionada com o isolamento geográfico, remetente das últimas glaciações ocorridas há 18 mil anos (Cubry et al., 2012; Ferrão et al., 2007; Gomez et al., 2009). O primeiro grupo, chamado de Guineano, compreende os genótipos da região oeste africana e apresentam folhas menores, menor vigor e porte, frutos pequenos, bebida de qualidade inferior, tolerância à seca e são susceptíveis a ferrugem (*H. vastatrix*). Estudos recentes indicam que este grupo é estruturado em subpopulações (Cubry et al., 2012). O segundo grupo é denominado de Congolês e o centro de diversidade é a região central da África. Este grupo é estruturado em cinco subgrupos, denominados de: SG1, SG2, B, C e, mais recentemente, um grupo de acessos selvagens foi chamado de UW (Musoli et al., 2009). Na Figura 3 são representados os centros de diversidades dos grupos citados.

O subgrupo SG1 reúne os genótipos chamados de Kouillous (Conilons, no Brasil) e apresentam características adaptativas semelhantes, em parte, àquelas do grupo Guineano. Trata-se do “subgrupo” de melhor adaptação no Brasil e que está presente nas principais variedades desenvolvidas no país. Os subgrupos SG2, B e C compreendem os genótipos do tipo Robusta, que são plantas mais altas, vigorosas, de folhas e frutos maiores, resistentes à ferrugem e com maior sensibilidade à seca. No processo de domesticação da espécie, o grupo Guineano ficou restrito ao seu local de origem, enquanto que o grupo Congolês foi disseminado para outros países e progressivamente melhorado nos países produtores (Ferrão et al., 2007; Souza, 2011).

Na literatura são relatados diferentes estudos de diversidade genética com *C. canephora* (Cubry et al., 2012; Gomez et al., 2009; Maurin et al., 2007; Musoli et al., 2009; Prakash et al., 2005a). Independente se o método de avaliação são marcadores moleculares, sequências de DNA ou caracteres fenotípicos; observa-se dentro da espécie uma ampla base genética o que resulta em altos níveis de polimorfismos e de diversidade. Dentre os motivos para isso, destacam-se a reprodução por alogamia, a ampla distribuição geográfica (maior do gênero *Coffea*) e o processo de domesticação, uma vez que o cultivo não ficou restrito à apenas uma pequena região. Dessa forma, a grande variabilidade observada entre e dentro das populações naturais de *C. canephora* faz desta espécie uma magnífica fonte de alelos para o desenvolvimento de novas

variedades, bem como, para a utilização em programas de melhoramento genético de *C. arabica* (Souza, 2011).



**Figura 3.** Posição geográfica dos grupos de *Coffea canephora*: Guineano (letra D), no Oeste Africano; Congolense e os seus subgrupos (letra A,B,C e E), na região central do continente. Relatado por Gomez et al. (2009).

## 2.6. Mapeamento genético e as particularidades em estudos de *C. canephora*.

O desenvolvimento de mapas genéticos tem sido um dos objetivos dos geneticistas desde a redescoberta dos trabalhos de Mendel, há cerca de 100 anos, visto que é uma ferramenta importante nos estudos de herança e manipulação de germoplasmas. Mapas genéticos são formados por grupos de ligação, nos quais está representada a ordem e a distância entre marcadores genéticos, o que permite o melhor entendimento da segregação e da relação física dos alelos. Embora não proporcionem informação em nível molecular dos genes, os mapas destacam-se por auxiliar na fundamentação de estudos comparativos e de evolução; no entendimento dos processos biológicos; e de organização dos cromossomos. Se portadores de uma elevada quantidade de informação, os mapas proporcionam simplificação na manipulação genética (Ferreira and Grattapaglia, 1998). Entre as aplicações práticas no melhoramento de plantas, destacam-se as oportunidades para a escolha de genitores, seleção assistida por marcadores (MAS) e a clonagem de genes de interesse.

A construção dos mapas foi viabilizada graças aos trabalhos de Thomas Morgan, que estudando caracteres genéticos de moscas (*Drosophila melanogaster*) observou distorções na Segunda Lei de Mendel. De acordo com Morgan, alguns genes presentes

nos mesmos cromossomos poderiam trocar segmentos, fato conhecido com *crossing over* ou permutação. O resultado dessa permutação é chamado de recombinação e permite a produção de gametas “não parentais”, ou seja, gametas que sofreram permuta e se tornaram diferentes daqueles gametas “parentais”. A partir dessas descobertas Morgan observou que a proporção de progênies recombinantes variava muito, o que acabava refletindo de algum modo na distância entre os genes nos cromossomos. Tal observação foi confirmada em 1913, pelo pesquisador A.H. Sturtevant que sugeriu o uso da porcentagem de recombinação como indicador quantitativo da distância linear entre dois genes na construção de mapas genéticos (Carneiro and Viera, 2002).

Bastante útil em estudos básicos de genética, na primeira metade do século XX, o uso de mapas em programas de melhoramento de plantas foi limitado devido ao baixo número de genes marcados que estavam disponíveis para a maioria das espécies de interesse. Estudos desse tipo ficaram restritos as espécies de amplo conhecimento citológico como: milho, ervilha, trigo, tomate e aveia (Carneiro and Viera, 2002). Diante dessas limitações, na década de 60 surgiram os marcadores bioquímicos (isoenzimas) que permitiu a construção de mapas genéticos de algumas culturas. Na década de 80, com o advento dos marcadores moleculares, métodos genético-estatísticos e ferramentas computacionais; a análise de ligação genética tornou-se uma das mais básicas e indispensáveis metodologias da genética (Ferreira and Grattapaglia, 1998).

Neste contexto, em que biologia molecular e a estatística passaram a andar atrelados, a construção de mapas genéticos em plantas passou a envolver, basicamente, quatro etapas. A primeira delas, esta relacionada com a triagem de marcadores genéticos, sendo que para isso, fatores como conteúdo informativo e acessibilidade devem ser levados em consideração. A segunda etapa envolve o desenvolvimento de uma população segregante com locos em desequilíbrio de ligação, para isso, é fundamental o uso de genitores contrastantes de modo a maximizar os níveis de polimorfismos. Posteriormente, realiza-se a análise de herança dos marcadores genéticos, de forma a selecionar adequadamente aqueles que apresentam as frequências esperadas para a segregação de um loco único em uma população. A ocorrência de distorções de segregação indica que o modelo genético adotado pode ser inapropriado, ou que os dados são de baixa qualidade, ou que o processo de amostragem não foi aleatório. Finalmente, na quarta etapa, estabelece-se a distância e ordem dos marcadores

genéticos com base na porcentagem de recombinação. Para isso é necessário definir a frequência máxima de recombinação e o LOD mínimo para inferir se dois locos estão ligados (Liu, 1998; Schuster and Cruz, 2004). Para tanto, diversas metodologias foram desenvolvidas e implementadas em programas computacionais como MapMaker (Lander et al., 1987); GMENDEL (Liu and Knapp, 1990), JoinMap (JW, 2011), “GQMOL” (<http://www.ufv.br/dbg/gqmol/gqmol.htm>) e OneMap (Margarido et al., 2007).

Ao se observar as propriedades biológicas dos materiais em estudo, para os trabalhos de mapeamento em plantas, dois tipos básicos de populações podem ser utilizadas: as oriundas de cruzamentos controlados (endogâmicas) e as populações naturais (exôgâmicas). O primeiro tipo engloba as populações  $F_2$ , retrocruzamentos, duplo-haplóides e linhagens endogâmicas recombinantes (RILs); enquanto que o segundo inclui as famílias de irmãos completos, meios irmãos e populações oriundas de intercruzamentos. A escolha de uma população em detrimento da outra deve ser feita levando em consideração o número de indivíduos amostrados, as características de interesse dos genitores e, sobretudo, o sistema reprodutivo da espécie em estudo (Liu, 1998).

Grande parte dos mapas genéticos já desenvolvidos é originada de populações de cruzamentos controlados, derivadas de  $F_1$ , onde os genitores são linhas puras. Para este caso, o desequilíbrio de ligação gerado pode ser estimado nas populações segregantes, permitindo assim, a análise de ligação (Lin et al., 2003). Somado a isso, para o caso de espécies diploides e com dois alelos segregando, as fases de ligação são conhecidas o que facilita às análises. No entanto, em culturas como maracujá, eucalipto, cana de açúcar, citros, café (*Coffea canephora*) e outras espécies florestais; não é possível obter populações derivadas de linhagens endogâmicas, devido a fatores biológicos como autoincompatibilidade, depressão por endogamia ou longo período juvenil. Neste caso as análises genéticas-estatísticas para a construção de mapas baseiam-se em populações desenvolvidas pelo cruzamento biparental entre indivíduos não homozigóticos (*outcrossing*), gerando  $F_1$  segregantes formada por indivíduos que são irmãos completos (*full sib*) ou meios irmãos (*half sib*).

Para esse tipo de população o estabelecimento de mapas genéticos apresenta algumas particularidades que não são observadas no mapeamento de populações controladas. Segundo Lu et al. (2004), há duas razões para isso. A primeira esta

relacionada com número de alelos do marcador e os diferentes padrões de segregação. Revisão detalhada do assunto é discutido por Wu et al. (2002), e está apresentado simplificado na Tabela 1, onde considerando organismos diploides existem 18 tipos distintos de cruzamento (configurações genéticas) para cada loco do marcador, caracterizando, portanto, quatro possíveis tipos de segregações (A, B, C e D). O resultado disso é a existência de diferentes funções de verossimilhança para as possíveis configurações genéticas, o que é um complicador nas análises genética-estatística. A segunda particularidade, está relacionada com as fases de ligação entre os diferentes marcadores, uma vez que estas não são conhecidas *a priori* para os pais, o que faz necessário o desenvolvimento de algoritmos para caracterização da fase de ligação mais provável. Ressalta-se que o conhecimento da fase de ligação é requerido para a detecção de eventos de recombinação, uma vez que esta define a configuração dos alelos de um par de locos heterozigotos em cromossomos homólogos de um parental.

**Tabela 1** - Configuração genética dos locos marcadores, segundo o tipo de cruzamento, e os respectivos padrões de bandas observadas (alelos) nos genitores e nas progênie oriundas de cruzamentos entre genitores não-endogâmicos.

Tipo de Cruzamento	Genitores		Progênie					
	Cruzamentos	Bandas Observadas	Bandas observadas	Observação	Bandas observadas	Segregação	Nº de fenótipos	
A	1	$ab \times cd$	$ab \times cd$	Assimétrico	$ac, ad, bc, bd$	1:1:1:1	4	
	2	$ab \times ac$	$ab \times ac$	Assimétrico	$a, ac, ba, bc$	1:1:1:1	4	
	3	$ab \times co$	$ab \times c$	Assimétrico	$ac, a, bc, b$	1:1:1:1	4	
	4	$ao \times bo$	$a \times b$	Assimétrico	$ab, a, b, o$	1:1:1:1	4	
B	B1	$ab \times ao$	$ab \times a$	Assimétrico	$ab, 2a, b$	1:2:1	3	
	B2	$ao \times ab$	$a \times ab$	Assimétrico	$ab, 2a, b$	1:2:1	3	
	B3	$ab \times ab$	$ab \times ab$	Simétrico	$a, 2ab, b$	1:2:1	3	
C	8	$ao \times ao$	$a \times a$	Simétrico	$3a, o$	3:1	2	
D	D1	9	$ab \times cc$	$ab \times cc$	Assimétrico	$ac, bc$	1:1	2
		10	$ab \times aa$	$ab \times aa$	Assimétrico	$a, ab$	1:1	2
		11	$ab \times oo$	$ab \times o$	Assimétrico	$a, b$	1:1	2
		12	$bo \times aa$	$b \times aa$	Assimétrico	$ab, a$	1:1	2
		13	$ao \times oo$	$a \times o$	Assimétrico	$a, o$	1:1	2
	D2	14	$cc \times ab$	$cc \times ab$	Assimétrico	$ac, bc$	1:1	2
		15	$aa \times ab$	$aa \times ab$	Assimétrico	$a, ab$	1:1	2
		16	$oo \times ab$	$o \times ab$	Assimétrico	$a, b$	1:1	2
		17	$aa \times bo$	$aa \times b$	Assimétrico	$ab, a$	1:1	2
		18	$oo \times ao$	$o \times a$	Assimétrico	$a, o$	1:1	2

Adaptado de Wu *et al.*, (2002)

Por outro lado, a possibilidade de construção de um único mapa genético integrando todos os marcadores moleculares apresenta como principal vantagem a possibilidade de aumentar a saturação do mapa de ligação e estender a caracterização da variação polimórfica em todo o genoma (Garcia et al., 2006). Além disso, a localização de locos controladores de características quantitativas (QTLs) pode ser facilitada nas integrações de mapa (Maliepaard et al., 1998). Para isso, Ritter et al. (1990) e Ritter and Salamini (1996) propuseram os primeiros métodos estatísticos para estimação da frequência de recombinação entre marcadores com diferentes padrões de segregação em plantas. Discussão sobre problemas e particularidades dessa metodologia é relatado por Maliepaard et al. (1998), enquanto que Wu et al. (2002), Lu et al. (2004) e Jansen (2005) apresentaram metodologias alternativas para estimação simultânea da fração de recombinação e fase de ligação. Muitas dessas ideias motivaram a criação do pacote *Onemap* em ambiente R e trabalhos semelhantes também foram desenvolvidos por Bhering (2008) e implementados no aplicativo computacional *GQMOL software* brasileiro utilizado para mapeamento genético e análise de dados moleculares.

Para o melhoramento genético do cafeeiro, a importância desses estudos é a possibilidade do uso dessas informações em trabalhos que objetivam a identificação e posicionamento de regiões do genoma que controlam características de interesse. O estudo que se presta a atender esses objetivos é denominado de Mapeamento de QTLs, o qual está fundamentada no estabelecimento de associação entre o caráter quantitativo (fenótipo) e os marcadores moleculares (genótipo), alocados em mapas de ligação genética.

## **2.7. Análise de QTLs**

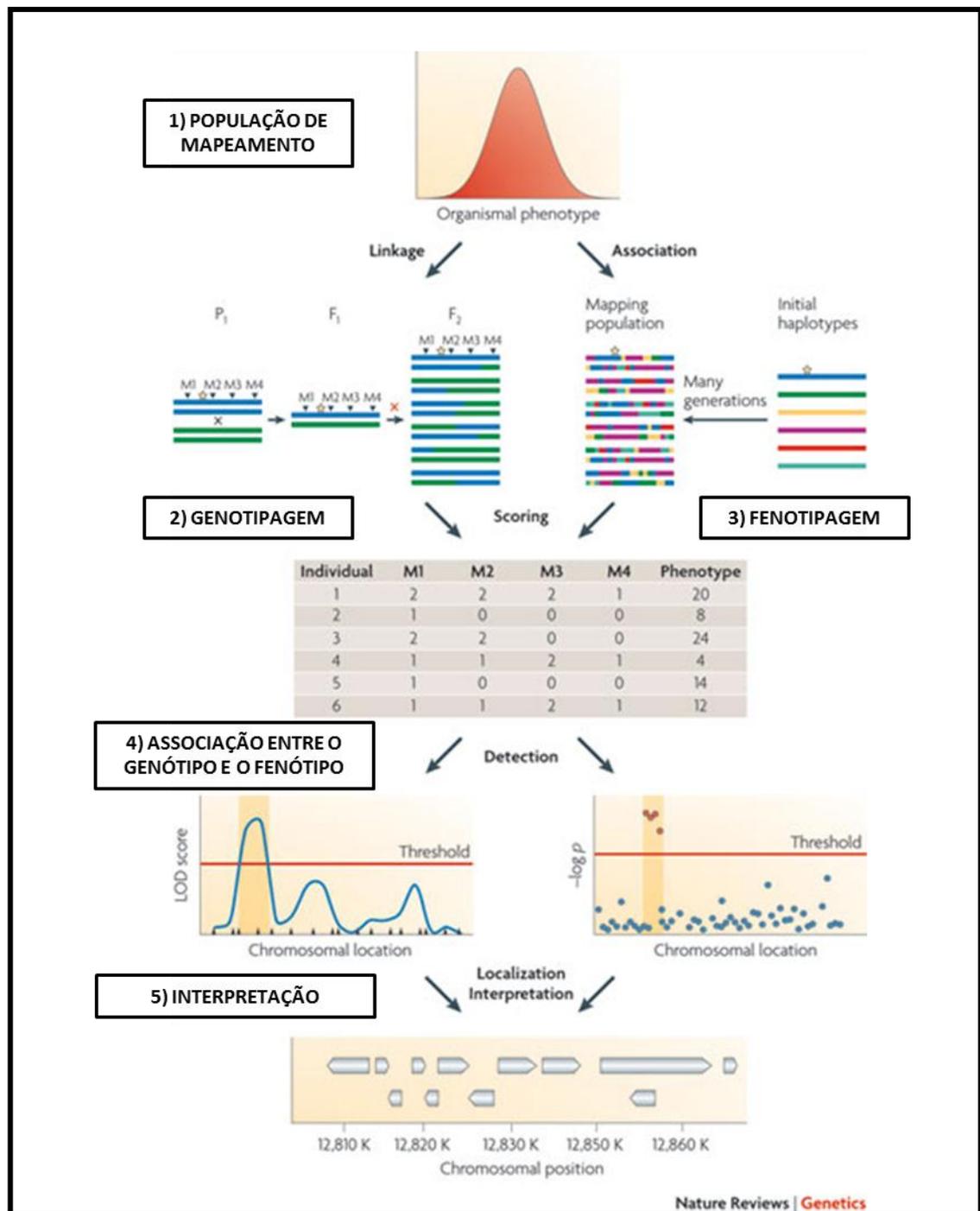
O termo QTL foi utilizado pela primeira vez por Gerldermann em 1975 e se refere aos locos controladores de características quantitativas. Ao contrário dos caracteres simples estudados por Mendel e determinados por um único loco e dois alelos, as características consideradas complexas são determinadas por um grande número de genes e são influenciadas por fatores ambientais em diferentes graus. A grande maioria das características de importância econômica é poligênica. Os fenótipos resultantes da expressão desses genes apresentam uma variação contínua na população

ao invés de classes discretas, como observado nos caracteres de herança mendeliana simples. A variação contínua inviabiliza o agrupamento dos genótipos em classes, tornando o indivíduo *per si* uma unidade de estudo de pouco valor. Assim, para os estudos de características complexas faz-se necessário utilizar modelos quantitativos que contemplam um tipo de herança que é determinada pela segregação de vários locos, sendo que cada um deles apresenta um pequeno efeito no genótipo. Para isso, esses modelos se baseiam, na grande maioria das vezes, nos princípios de genética de populações, com ênfase nas estimativas de médias, variâncias e covariâncias (Tanksley, 1993).

A melhor compreensão dos caracteres complexos ocorreu com o advento dos marcadores moleculares e percepção de suas vantagens. O uso da biologia molecular possibilitou o estudo desses caracteres sob outro aspecto, sobretudo, por possibilitar a associação direta entre o fenótipo e o genótipo. Assim, os estudos que focam a identificação, o posicionamento e as estimativas dos efeitos e das interações (epistasia) dos caracteres quantitativos foram denominados de mapeamento de QTLs. O fundamento teórico desses estudos baseia-se na elucidação de um oligogene ou poligene (gene controlador de um caráter de herança complexa) através de um monogene (gene controlador de um caráter de herança simples) que estiverem fisicamente ligados. Em outras palavras, a detecção e mapeamento de QTLs são feita com base na análise da segregação conjunta dos locos marcadores (monogenes) e dos valores fenotípicos (oligogenes ou poligenes) observados. Contudo, a premissa que fundamenta essas análises é a existência de desequilíbrio de ligação entre os marcadores e os QTLs. Sem esse pré-requisito, os alelos do marcador e do QTL ocorrerão em combinações equivalentes à distribuição independente, fazendo assim com que não haja nenhuma associação entre os dois locos (Mackay et al., 2009).

A exemplo dos experimentos visando a construção de mapas genéticos de ligação, a maioria das análises de QTL em plantas envolve o uso de populações experimentais derivadas de linhagem endogâmica. Dentre os motivos para isso, destaca-se a maior simplicidade para o desenvolvimento de modelos genético-estatísticos, uma vez que há apenas dois alelos segregando em cada loco e as frequências alélicas e as fases de ligação entre marcador e os QTLs são conhecidas. A análise dos dados obtidos pode ser realizada por meio de diferentes metodologias, nos quais se incluem desde as técnicas mais simples, como as análises de marcas individuais, até os modelos mais

complexos que envolvem a análise de múltiplas regiões do genoma simultaneamente. Um resumo da análise de QTL é representado na Figura 3.



**Figura 3.** Resumo das cinco etapas necessárias para o mapeamento de QTLs. 1) Obtenção de uma população de mapeamento, que pode ser uma população experimental (esquerda) ou não (direita); 2) Genotipagem da populações; 3) Medição fenotípica do caráter de interesse; 4) Detecção dos QTLs pela associação do genótipo ao fenótipo utilizando métodos estatísticos próprios; 5) Interpretação dos resultados com estimação da posição, efeito e interação (epistasia) dos QTLs. Adaptado de Mackay et al. (2009)

O mapeamento de QTL utilizando marcas individuais é chamado de análise de marca simples e baseia-se na comparação entre as médias das classes genotípicas dos marcadores. Para isso, utilizam-se testes estatísticos como o teste *t*, análise de variância (ANOVA), regressão linear simples e testes de razão de verossimilhança. Estes procedimentos têm em comum o fato de testarem a hipótese nula de que a média do caráter em questão é independente do genótipo do marcador. A rejeição da hipótese nula leva a crer a possibilidade da existência de pelo menos um QTL ligado à marca. Dentre as vantagens dessa metodologia, destaca-se a simplicidade, rapidez de execução e a não necessidade de mapas de ligação. Por outro lado, a desvantagem é o confundimento que é gerada entre o efeito e a posição do QTL, de modo que não é possível a distinção entre um QTL de pequeno efeito situado próximo ao marcador, de um QTL de grande efeito, mas situado distante do marcador. Dessa forma, as análises de marca simples caracterizam-se por ser um passo preliminar nas análises exploratórias, fornecendo indícios importantes da presença ou não de QTLs.

Um pouco mais complexo, porém de grande poder estatístico, surgiram os modelos de mapeamento por intervalo. Proposto por Lander and Botstein (1989), a metodologia utiliza o princípio da máxima verossimilhança para verificar a possível presença de um QTL entre um par de marcas adjacentes em um grupo de ligação. Para tanto, é necessário à presença de mapas de ligação. Assim, em um intervalo, várias posições são testadas para a presença de um QTL. Uma característica desse método é a busca contínua de QTLs ao longo do genoma, através da obtenção de curvas para os resultados dos testes estatísticos, comumente expressos na forma de LOD score. Esse teste estatístico compara a função de verossimilhança da hipótese de nulidade (ausência de QTL) com a hipótese alternativa (QTL na posição testada), resultando na detecção de regiões significativas dentro de grupos de ligação. Dentre as vantagens do mapeamento por intervalo, destacamos a possibilidade de inferência da provável posição dos QTLs ao longo do genoma e o não confundimento entre efeito e posição dos QTLs. No entanto, como ponto negativo dessa metodologia está o surgimento dos chamados “QTLs fantasmas”, que na verdade são artefatos causados por QTLs que são mapeados próximos uns aos outros, o que acaba gerando falsos positivos nas análises e o menor poder estatístico. Este último, causado pelo fato de serem considerados apenas dois marcadores por vez nas análises, ao invés, de todos os marcadores simultaneamente.

Isso faz com que as informações adicionais fornecidas por outros marcadores sejam desprezadas, diminuindo assim o poder de detecção do método de intervalo simples.

Visando melhorar as estimativas de QTLs e, conseqüentemente, corrigir os problemas citados do método anterior, Zeng (1994) propôs uma nova metodologia para análises de QTLs. Conhecido intervalo composto (*Composite Interval Mapping, CIM*), o método combina o modelo de intervalo proposto por Lander and Botstein (1989), e um modelo de regressão múltipla, utilizado para controlar a variação fora do intervalo mapeado. Para isso, são incluídos nas análises marcadores localizados em outras regiões do genoma (cofatores), com o intuito de diminuir o resíduo do modelo, aumentando assim o poder de detecção e a precisão das estimativas de efeito e posição. No entanto, a escolha dos marcadores a serem utilizados como cofatores é um ponto crítico dessa metodologia e deve ser feita de maneira criteriosa. A inclusão errônea pode ter efeito contrário no método, causando assim, perda da precisão em função da superparametrização do modelo. Outros trabalhos estenderam os conceitos de mapeamento por intervalos sob vários enfoques, com destaque para os trabalhos de Jiang and Zeng (1995) e Kao et al. (1999).

Os métodos apresentados até o momento são de uso específico para populações experimentais, como  $F_2$ , retrocruzamentos ou RIL's. Contudo, já foi citado que para algumas espécies cultivadas, inclusive *C. canephora*, a obtenção de linhagens endogâmicas é impossibilitada por diferentes fatores biológicos, tais como autocompatibilidade, depressão endogâmica ou complexidade do genoma. Assim, nessas espécies o mapeamento de QTL deve ser realizado a partir de cruzamentos entre indivíduos não aparentados em populações de irmãos completos ou meios irmãos. Dentre as particularidades a ser consideradas nessas análises, destacam-se os marcadores com diversos padrões de segregação, o número de QTLs variando de um a quatro (em diploides), a fase de ligação desconhecida entre os marcadores e entre os QTL e os marcadores (Lin et al., 2003).

O mapeamento de QTLs nessa situação normalmente é conduzido utilizando outras abordagens estatísticas, tais quais: análise de marcas simples, *duplo pseudo-testcross* (Grattapaglia and Sederoff, 1994), mapeamento baseados em mapas integrados (Gazaffi, 2009; Lin et al., 2003) e o uso de regressão linear baseado em IBD. Neste trabalho, será discutido o último método com destaque para o trabalho original de Haseman and Elston (1972) e sua derivação, apresentado por Fulker and Cardon (1994).

### 2.7.1. Regressão de Haseman e Elston

O método proposto por Haseman and Elston (1972) foi originalmente proposto para o mapeamento de QTLs em humanos. Derivado dos métodos clássicos de estimativas de componentes de variância e de análises de regressão, a metodologia de Haseman e Elston é baseada na simples intuição de que indivíduos parecidos fenotipicamente compartilham um maior número de alelos do QTL do que indivíduos fenotipicamente discordantes (Almasy and Blangero, 2009). Com isso, é montada uma regressão, na qual o quadrado da diferença fenotípica entre pares de irmãos é regredida em função no número de alelos que eles compartilham. Em outras palavras, a variável dependente  $Y_j$  do modelo de regressão é representado pelo quadrado da diferença fenotípica entre os indivíduos do  $j$ -ésimo par de irmãos, enquanto que a variável independente  $\pi_j$  do modelo é dado pela quantidade de alelos idênticos por descendência (*IBD-identical by descent*) para os pares de irmãos considerados. Com isso, é adotado o seguinte modelo de regressão, no qual o quadrado da diferença fenotípica é regredido em relação ao IBD dos descendentes:

$$Y_j = \beta_0 + \beta_1 \pi_j + \varepsilon_j, \text{ onde:}$$

$Y_j$  = quadrado da diferença fenotípica entre indivíduos do  $j$ -ésimo par de irmãos, dado por

$\pi_j$  = valor de alelos idênticos por descendência (IBD) do QTL;

$\beta_0$  e  $\beta_1$  = são os parâmetros dos modelos de regressão.

O valor de  $Y_j$  é dado por:

$$Y_j = (X_{1j} - X_{2j})^2, \text{ onde:}$$

$X_{1j}$  e  $X_{2j}$  são os valores observados da característica para o primeiro e o segundo irmão do  $j$ -ésimo para.

O valor assumido pela variável independente  $\pi_j$  é uma estimativa bayesiana dada pela fórmula:

$$\pi_j = 1/2 f_{j1} + f_{j2}, \text{ onde:}$$

$f_{j1}$  = probabilidade de que o  $j$ -ésimo par de irmão tenha 1 alelo IBD;

$f_{j2}$  = probabilidade de que o  $j$ -ésimo par de irmão tenha 2 alelo IBD.

Quando os genótipos dos indivíduos e o dos parentais são conhecidos, o valor de  $\pi_j$  pode ser facilmente calculado para cada par de irmãos, considerando o tipo de cruzamento dos quais eles são provenientes. Dentro do conceito de mapeamento de

QTLs via regressão de Haseman e Elston, a ideia do IBD é de grande importância, pois quantifica a semelhança genética entre o par de indivíduos avaliados. Sob hipótese de ligação, um par de indivíduos aparentados e, conseqüentemente, mais semelhantes fenotipicamente irão compartilhar uma maior proporção de alelos IBD do que sob a hipótese nula (sem ligação), no qual os fenótipos dos pares serão pouco relacionados e independentes do compartilhamento de alelos no gene marcador.

Ao considerar ausência de dominância ( $d=0$ ), assumindo que existe ligação completa entre o marcador e o QTL, reunindo as informações quadráticas sob o aspecto quantitativo e levando em consideração os valores de  $\pi_j$ , chega-se ao seguinte modelo de regressão linear:

$$E(Y_j/\pi_j) = \alpha + \beta\pi_j, \text{ onde:}$$

$$\alpha = \sigma_e^2 + 2\sigma_g^2$$

$$\beta = -2\sigma_g^2$$

Ainda, se o QTL se posiciona a uma distância do loco marcador, com uma frequência de recombinação  $c$ , e considerando ausência de dominância e completa informação dos parentais, então  $\beta$  pode ser calculado por:  $\beta = -2(1-2c)^2\sigma_g^2$ ; e a hipótese de não ligação ( $\beta=0$ ) genética entre o marcador e o QTL é possível de ser testada pelo teste  $t$  unilateral. Com isso, a regressão de Hanseman e Elston é considerada uma metodologia robusta para a detecção de ligação e estimativas da variância do QTL, que já foi utilizada em trabalhos com humanos e plantas (Almasy and Blangero, 2009; Gray-McGuire et al., 2010; Larkin et al., 2008; Tomaz, 2008). No entanto, como em todos os testes baseados em marca simples, esse método apresenta o inconveniente de realizar apenas a detecção do QTL, sem diferenciar o efeito do QTL e a distância do marcador.

### 2.7.2. Regressão de Fulker e Cardon

Descrito por Fulker and Cardon (1994), o método foi derivado originalmente da ideia de regressão por IBD proposta por Hanseman e Elston. Contudo, já foi comentada que o grande entrave dessa regressão é a impossibilidade de distinção entre o efeito do QTL e a distância do marcador, o que gera confusão entre a variância genotípica e a frequência de recombinação. Com isso, ao observar um valor significativo de  $\beta$  nas

análises de regressão, fica indefinido se é um QTL de efeito maior ou um QTL de baixo efeito, porém muito próximo ao marcador (Fulker and Cardon, 1994).

Uma solução para resolver o problema e, conseqüentemente, aumentar o poder estatístico das análises, foi o desenvolvimento da metodologia de intervalo simples por meio da regressão de pares de irmãos. Chamado de regressão de Fulker e Cardon, o método foi desenvolvido com o propósito de isolar a variância do QTL da taxa de recombinação, bem como localizar o QTL no cromossomo. Para isso, tal procedimento utiliza dois marcadores flanqueando o QTL para estimar separadamente a posição e o efeito do QTL sobre o caráter. Admite-se que exista uma frequência de recombinação  $\theta_1$  entre o primeiro marcador e o loco do QTL, e que existe uma frequência de recombinação  $\theta_2$  entre o segundo marcador e o loco QTL.

Com isso, se for considerado a análise separada para cada loco do marcador pela metodologia de Haseman and Elston (1972), teríamos os seguintes resultados:

a) Para o marcador M1:  $\beta_1 = -2(1 - 2\theta_1)^2 \sigma_g^2$

b) Para o marcador M2:  $\beta_2 = -2(1 - 2\theta_2)^2 \sigma_g^2$

As frequências de recombinação do marcador 1 e 2, podem ser expressas através de uma função de mapeamento apropriada e por isso as equações são reescritas da seguinte maneira:

a)  $\beta_1 = \beta_q(1 - 2\theta_1)^2$

b)  $\beta_2 = \beta_q(1 - 2\theta_2)^2$

onde  $\beta_q$  é o valor de  $\beta$  que é obtido quando conhecemos a localização do QTL e aplicamos a abordagem convencional de Hanseman e Elston.

Para o cálculo do IBD é proposto a utilização de  $\pi_1$  e  $\pi_2$  para a estimativa de um valor médio de numero de alelos compartilhados para os marcadores M1 e M2. Denominado de  $\pi_q$ , os autores apresentaram uma expressão para o cômputo da média condicional de alelos IBD do QTL com função linear dos  $\pi_m$  nos dois marcadores de um intervalo. Essa expressão é dada por:

$$\hat{\pi}_q = \alpha + \beta_{\pi_1}\pi_1 + \beta_{\pi_2}\pi_2$$

Os valores de  $\alpha$  e  $\beta$  são obtido pela resolução matricial do sistema de equações normais, cujos estimadores são apresentados a seguir:

$$\hat{\alpha} = (1 - \hat{\beta}_{\pi_1} - \hat{\beta}_{\pi_2})/2$$

$$\hat{\beta}_{\pi_1} = [(1 - 2\theta_1)^2 - (1 - 2\theta_2)^2 - (1 - 2\theta_{12})^2] / [1 - (1 - 2\theta_{12})^4]$$

$$\hat{\beta}_{\pi_2} = [(1 - 2\theta_2)^2 - (1 - 2\theta_1)^2 - (1 - 2\theta_{12})^2] / [1 - (1 - 2\theta_{12})^4]$$

De posse das estimativas de  $\alpha$  e  $\beta$ , obtêm-se a proporção IBD ( $\pi_q$ ) para o QTL. Essa proporção depende essencialmente da fração de recombinação entre os locos. As frequências de recombinação podem ser computadas a partir da frequência gamética de cada genitor ou a partir da frequência genotípica da progênie. Informações multilocos entre pares de marcas adjacentes são usadas na estimação (Resende et al., 2012).

Obtido então a estimativa do valor de  $\pi_q$  para uma determinada posição do QTL, este valor é utilizado como variável independente no modelo de regressão  $Y_j = \beta_0 + \beta_q \pi_{qj}$ , sendo que para cada par de marcadores haverá um valor de IBD diferente. Assim, admite-se que a melhor regressão dentro de um intervalo é aquela que apresenta a menor soma de quadrados residuais. Nessa posição é possível então determinar a posição do QTL e o valor da contribuição dada pelo QTL para a variância genética do caráter (Tomaz, 2008). A herdabilidade também pode ser estimada pela fórmula:

$$h^2 = -\hat{\beta}_q / 2\sigma_f^2$$

O poder estatístico do método e sua aplicabilidade em estudos genéticos já foram relatados em diferentes trabalhos (Cardon and Fulker, 1994; Rosado et al., 2010; Silva et al., 2004; Tomaz, 2008).

### 2.7.3. Importância do Mapeamento de QTL e sua aplicação em estudo com café

A genética quantitativa, que fornece as bases teóricas para o melhoramento genético de plantas, permite a estimação de diversos parâmetros genéticos a partir de medidas fenotípicas que refletem a expressão gênica. O conhecimento desses parâmetros é de grande importância, pois direcionam um programa de melhoramento na medida em que auxiliam os melhoristas na prática da seleção e informam o progresso genético obtido no decorrer das gerações. Além disso, pode auxiliar na elucidação de questões complexas, como por exemplo, a interação genótipo x ambiente, número de genes envolvidos e interações alélicas.

Com o advento dos marcadores moleculares, os estudos de caracteres complexos deu um grande salto, sobretudo, pela possibilidade de mapear QTLs individualmente, o que é de fundamental importância para o entendimento da arquitetura genética dos caracteres de interesse. Estimados a posição, os efeitos e os possíveis tipos de interações dos QTLs, este é o primeiro passo para validação dessas regiões genômicas como possíveis genes candidatos. Após a validação, essas informações podem ser utilizadas em trabalhos práticos de seleção assistida por marcadores moleculares (MAS) ou clonagem posicional. Exemplos de QTLs que apresentam moderados e grandes efeitos são relatados com certa abundância na literatura (Price, 2006). Em tomate, por exemplo, são relatados QTLs de grande efeito (30%-67%) responsáveis pelo controle de tamanho, peso e formato do fruto (Alpert et al., 1995; Ku et al., 1999).

Para a espécie *C. canephora*, o número de QTLs mapeados ainda é pequeno. Dentro aqueles que são relatados na literatura, destaque é dado para os estudos com o loco S da incompatibilidade (Lashermes et al., 1996); restauração da viabilidade dos polens (Coulibaly et al., 2003); características morfológicas (Amidou et al., 2007); capacidade de embriogênese somática (Priyono et al., 2010); e, mais recentemente, para qualidade de bebida (Leroy et al., 2011) e para 15 características fenotípicas de interesse agrônomo (Souza, 2011). No trabalho desenvolvido por Souza (2011), ressalta-se a utilização da metodologia de mapeamento associativo em acessos de um banco de germoplasma de *C. canephora*. Dentre os resultados, destaque especial é dado para a identificação de marcadores SSRs associados significativamente com a produtividade, números de inflorescência por ramo, tamanho do entrenó, número de ramo por plantas e tamanho das plantas.

Com relação aos estudos de resistência à ferrugem (*H. vastatrix*), não são relatados na literatura estudos de QTLs em *C. canephora*. Para outras espécies do gênero *Coffea*, em especial para *C. arabica*, existe alguns trabalhos que buscam o melhor entendimento dos fatores genéticos relacionados com a resistência a doença (de Brito et al., 2010; Mahe et al., 2008; Pestana, 2010; Romero et al., 2010). No entanto, visto os danos causados na cultura e a susceptibilidade de algumas variedades de *C. canephora* à doença, é imperativo que sejam realizados estudos no sentido de compreender melhor os mecanismos de resistência, de modo a auxiliar no lançamento de variedades resistentes.

### 3. REFERÊNCIAS BIBLIOGRÁFICAS

- Agarwal, M., Shrivastava, N., and Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep* **27**, 617-631.
- Aggarwal, R. K., Hendre, P. S., Varshney, R. K., Bhat, P. R., Krishnakumar, V., and Singh, L. (2007). Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor Appl Genet* **114**, 359-72.
- Aitken, K. S., Jackson, P. A., and McIntyre, C. L. (2006). Quantitative trait loci identified for sugar related traits in a sugarcane (*Saccharum* spp.) cultivar x *Saccharum officinarum* population. *Theoretical and Applied Genetics* **112**, 1306-1317.
- Almasy, L., and Blangero, J. (2009). Human QTL linkage mapping. *Genetica* **136**, 333-340.
- Alpert, K. B., Grandillo, S., and Tanksley, S. D. (1995). Fw-2.2 - a Major Qtl Controlling Fruit Weight Is Common to Both Red-Fruited and Green-Fruited Tomato Species. *Theoretical and Applied Genetics* **91**, 994-1000.
- Amidou, N., Michel, N., Serge, H., and Valerie, P. (2007). Genetic basis of species differentiation between *Coffea liberica* Hiern and *C-canephora* Pierre: Analysis of an interspecific cross. *Genetic Resources and Crop Evolution* **54**, 1011-1021.
- Baruah, A., Naik, V., Hendre, P. S., Rajkumar, R., Rajendrakumar, P., and Aggarwal, R. K. (2003). Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. *Molecular Ecology Notes* **3**, 647-650.
- Berthaud, J., and Charrier, A. (1988). Genetics resources of *Coffea*. In "Coffee" (R. J. Clarke and R. Macrae, eds.), Vol. 4, pp. 1-42. Elsevier Applied Science, London,.
- Bettencourt, A. J., Lopes, J., and Palma, S. (1992). Factores genéticos que condicionam a resistência às raças de *Hemileia vastatrix* Berk. et Br. dos clones-tipo dos grupos 1, 2 e 3 de derivados de Híbrido de Timor. *Broteria Genética* **XIII**, 185-194.
- Bettencourt, A. J., and Rodrigues, J. C. J. (1988). "Principles and practice of coffee breeding for resistance to rust and other disease," Agronomy, Elsevier Applied Science, London.
- Bhat, P. R., Krishnakumar, V., Hendre, P. S., Rajendrakumar, P., Varshney, R. K., and Aggarwal, R. K. (2005). Identification and characterization of expressed sequence tags-derived simple sequence repeats markers from robusta coffee variety 'CxR' (an interspecific hybrid of *Coffea canephora* x *Coffea congensis*). *Molecular Ecology Notes* **5**, 80-83.
- Bhering, L. L. (2008). Mapeamento genético em famílias simuladas de irmãos completos, Universidade Federal de Viçosa, Viçosa, MG.
- Borém, A. (1998). "Melhoramento de Plantas," Universidade Federal de Viçosa, Viçosa, MG.
- Bundock, P. C., Potts, B. M., and Vaillancourt, R. E. (2008). Detection and stability of quantitative trait loci (QTL) in *Eucalyptus globulus*. *Tree Genetics & Genomes* **4**, 85-95.

- Cabral, P. G. C., Zambolim, E. M., Zambolim, L., Lelis, T. P., Capucho, A. S., and Caixeta, E. T. (2009). Identification of a new race of *Hemileia vastatrix* in Brazil. *Australasian Plant Disease Notes* **4**, 129-130.
- Caixeta, E. T., Ferrão, L. F. V., and Maciel-Zambolim, E. (2012). Marcadores Moleculares. In "Biotecnologia aplicada ao melhoramento de plantas" (A. Borem and R. Fristsche-Neto, eds.). Universidade Federal de Viçosa, Viçosa, MG.
- Caixeta, E. T., Oliveira, A. C. B. D., Brito, G. G. D., and Sakiyama, N. S. (2009). Tipos de marcadores moleculares. In "Marcadores Moleculares" (A. Borém and E. T. Caixeta, eds.), Vol. 2, pp. 9-78. Universidade Federal de Viçosa, Viçosa.
- Capucho, A. S. (2012). Epidemiologia e resistência do cafeeiro à ferrugem, Universidade Federal de Viçosa, Viçosa, MG.
- Cardon, L. R., and Fulker, D. W. (1994). The Power of Interval Mapping of Quantitative Trait Loci, Using Selected Sib Pairs. *Am J Hum Genet* **55**, 825-833.
- Carneiro, M. S., Camargo, L. E. A., Coelho, A. S. G., Vencovsky, R., Leite, R. P., Stenzel, N. M. C., and Vieira, M. L. C. (2002). RAPD-based genetic linkage maps of yellow passion fruit (*Passiflora edulis* Sims. f. *flavicarpa* Deg.). *Genome* **45**, 670-678.
- Carneiro, M. S., and Viera, M. L. C. (2002). Mapas genéticos em plantas. *Bragantia* **61**, 89-100.
- Charrier, A., and Berthaud, J. (1988). Principles and methods in Coffea plant breeding: Coffea canephora Pierre. In "Coffea agronomy" (R. J. Clarke and R. Macrae, eds.), Vol. 6, pp. 167-195. Elsevier Applied Science, London.
- Combes, M. C., Andrzejewski, S., Anthony, F., Bertrand, B., Rovelli, P., Graziosi, G., and Lashermes, P. (2000). Characterization of microsatellite loci in Coffea arabica and related coffee species. *Mol Ecol* **9**, 1178-80.
- Conab (2012). Cafés do Brasil: safra 2011/2012. Vol. 2012. <http://www.conab.gov.br/>, Brasília: MAPA/CONAB.
- Coulibaly, I., Revol, B., Noirot, M., Poncet, V., Lorieux, M., Carasco-Lacombe, C., Minier, J., Dufour, M., and Hamon, P. (2003). AFLP and SSR polymorphism in a Coffea interspecific backcross progeny [(C-heterocalyx x C-canephora) x C-canephora]. *Theoretical and Applied Genetics* **107**, 1148-1155.
- Cristancho, M., and Gaitán, A. L. (2008). Isolation, characterization and amplification of simple sequence repeat loci in coffee. *Crop Breeding and Applied Biotechnology* **8**, 321-329.
- Cubry, P., De Bellis, F., Pot, D., Musoli, P., and Leroy, T. (2012). Global analysis of Coffea canephora Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. *Genetic Resources and Crop Evolution*.
- de Brito, G. G., Caixeta, E. T., Gallina, A. P., Zambolim, E. M., Zambolim, L., Diola, V., and Loureiro, M. E. (2010). Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. *Euphytica* **173**, 255-264.
- Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.
- Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574-578.

- Ferrão, L. F. V., Caixeta, E. T., Souza, F. d., Zambolim, E. M., Cruz, C. D., Zambolim, L., and Sakiyama, N. S. (2013). Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. *Plant Systematics and Evolution* **299**, 225-238.
- Ferrão, R. G. (2004). Biometria aplicada ao melhoramento genético de café Conilon, Universidade Federal de Viçosa, Viçosa, MG.
- Ferrão, R. G., Ferrão, M. G. A., Fonseca, A. F. A., and Pacova, B. E. V. (2007). Melhoramento genético de *Coffea canephora*. In "Café Conilon" (R. G. Ferrão, A. F. A. Fonseca, S. M. Bragança, M. G. A. Ferrão and L. H. De Muner, eds.), Vol. 1, pp. 123-173. Incaper, Vitória, Brazil.
- Ferreira, M. E., and Grattapaglia, D. (1998). "Introdução ao uso de marcadores moleculares em análise genética," Embrapa recursos genéticos e Biotecnologia, Brasília, DF.
- Fonseca, A. F. A. d. (1999). Análises biométricas em café conilon (*Coffea canephora* Pierre), Universidade Federal de Viçosa, Viçosa, MG.
- Fulker, D. W., and Cardon, L. R. (1994). A sib-pair approach to interval mapping of quantitative trait loci. *Am J Hum Genet* **54**, 1092-103.
- Garcia, A. A. F., Kido, E. A., Meza, A. N., Souza, H. M. B., Pinto, L. R., Pastina, M. M., Leite, C. S., da Silva, J. A. G., Ulian, E. C., Figueira, A., and Souza, A. P. (2006). Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theoretical and Applied Genetics* **112**, 298-314.
- Gazaffi, R. (2009). Desenvolvimento de modelo genético-estatístico para mapeamento de QTLs em progênies de irmãos completos, com aplicação em cana-de-açúcar, ESALQ, Piracicaba, SP.
- Geromel, C., Ferreira, L. P., Guerreiro, S. M., Cavalari, A. A., Pot, D., Pereira, L. F., Leroy, T., Vieira, L. G., Mazzafera, P., and Marraccini, P. (2006). Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. *J Exp Bot* **57**, 3243-58.
- Gomez, C., Dussert, S., Hamon, P., Hamon, S., de Kochko, A., and Poncet, V. (2009). Current genetic differentiation of *Coffea canephora* Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evol Biol* **9**.
- Grattapaglia, D., and Sederoff, R. (1994). Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* **137**, 1121-37.
- Gray-McGuire, C., Guda, K., Adrianto, I., Lin, C. P., Natale, L., Potter, J. D., Newcomb, P., Poole, E. M., Ulrich, C. M., Lindor, N., Goode, E. L., Fridley, B. L., Jenkins, R., Le Marchand, L., Casey, G., Haile, R., Hopper, J., Jenkins, M., Young, J., Buchanan, D., Gallinger, S., Adams, M., Lewis, S., Willis, J., Elston, R., Markowitz, S. D., and Wiesner, G. L. (2010). Confirmation of linkage to and localization of familial colon cancer risk haplotype on chromosome 9q22. *Cancer Res* **70**, 5409-18.
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Leger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., and Petit, R. J. (2011). Current trends in microsatellite genotyping. *Mol Ecol Resour* **11**, 591-611.
- Haseman, J. K., and Elston, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* **2**, 3-19.

- Hemmat, M., Weeden, N. F., Manganaris, A. G., and Lawson, D. M. (1994). Molecular marker linkage map for apple. *J Hered* **85**, 4-11.
- Hendre, P. S., Phanindranath, R., Annapurna, V., Lalremruata, A., and Aggarwal, R. K. (2008). 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 Biol* **8**, 51.
- Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* **9**, 1322-1332.
- Jansen, J. (2005). Construction of linkage maps in full-sib families of diploid outbreeding species by minimizing the number of recombinations in hidden inheritance vectors. *Genetics* **170**, 2013-25.
- Jiang, C., and Zeng, Z. B. (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**, 1111-27.
- JW, V. A. N. O. (2011). Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet Res (Camb)* **93**, 343-9.
- Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., and Dhawan, A. K. (2010). Microsatellite markers: an overview of the recent progress in plants. *Euphytica* **177**, 309-334.
- Kao, C.-H., Zeng, Z.-B., and Teasdale, R. D. (1999). Multiple Interval Mapping for Quantitative Trait Loci. *Genetics* **152**, 1203-1216.
- Karp, A., Kresovich, S., Bhat, K., Ayad, W., and Hodgkin, T. (1997). "Molecular tools in plant genetic resources conservation: a guide to the technologie." International Plant Genetic Resources Institute, Rome, Italy.
- Kosman, E., and Leonard, K. J. (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Mol Ecol* **14**, 415-24.
- Ku, H. M., Doganlar, S., Chen, K. Y., and Tanksley, S. D. (1999). The genetic basis of pear-shaped tomato fruit. *Theoretical and Applied Genetics* **99**, 844-850.
- Lander, E. S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-99.
- Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., and Newberg, L. A. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174-81.
- Larkin, E. K., Patel, S. R., Elston, R. C., Gray-McGuire, C., Zhu, X., and Redline, S. (2008). Using Linkage Analysis to Identify Quantitative Trait Loci for Sleep Apnea in Relationship to Body Mass Index. *Ann Hum Genet* **72**, 762-773.
- Lashermes, P., Andrzejewski, S., Bertrand, B., Combes, M. C., Dussert, S., Graziosi, G., Trouslot, P., and Anthony, F. (2000). Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). *Theoretical and Applied Genetics* **100**, 139-146.
- Lashermes, P., Combes, M. C., Trouslot, P., and Charrier, A. (1997). Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical and Applied Genetics* **94**, 947-955.
- Lashermes, P., Couturon, E., Moreau, N., Paillard, M., and Louarn, J. (1996). Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. *Theoretical and Applied Genetics* **93**, 458-462.
- Leroy, T., De Bellis, F., Legnate, H., Kananura, E., Gonzales, G., Pereira, L. F., Andrade, A. C., Charmetant, P., Montagnon, C., Cubry, P., Marraccini, P., Pot,

- D., and de Kochko, A. (2011). Improving the quality of African robustas: QTLs for yield- and quality-related traits in *Coffea canephora*. *Tree Genetics & Genomes* **7**, 781-798.
- Leroy, T., Marraccini, P., Dufour, M., Montagnon, C., Lashermes, P., Sabau, X., Ferreira, L. P., Jourdan, I., Pot, D., Andrade, A. C., Glaszmann, J. C., Vieira, L. G. E., and Piffanelli, P. (2005). Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theoretical and Applied Genetics* **111**, 1032-1041.
- Lin, M., Lou, X. Y., Chang, M., and Wu, R. (2003). A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. *Genetics* **165**, 901-13.
- Liu, B. H. (1998). "Statistical genomics: linkage, mapping, and QTL analysis," CRC Press, Boca Raton, Florida, USA.
- Liu, B. H., and Knapp, S. J. (1990). Gmendel - a Program for Mendelian Segregation and Linkage Analysis of Individual or Multiple Progeny Populations Using Log-Likelihood Ratios. *Journal of Heredity* **81**, 407-407.
- Lu, Q., Cui, Y., and Wu, R. (2004). A multilocus likelihood approach to joint modeling of linkage, parental diplotype and gene order in a full-sib family. *BMC Genet* **5**, 20.
- Mackay, T. F., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**, 565-77.
- Mahe, L., Combes, M. C., Varzea, V. M. P., Guilhaumon, C., and Lashermes, P. (2008). Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.). *Molecular Breeding* **21**, 105-113.
- Maliepaard, C., Alston, F. H., van Arkel, G., Brown, L. M., Chevreau, E., Dunemann, F., Evans, K. M., Gardiner, S., Guilford, P., van Heusden, A. W., Janse, J., Laurens, F., Lynn, J. R., Manganaris, A. G., den Nijs, A. P. M., Periam, N., Rikkerink, E., Roche, P., Ryder, C., Sansavini, S., Schmidt, H., Tartarini, S., Verhaegh, J. J., Vrieling-van Ginkel, M., and King, G. J. (1998). Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics* **97**, 60-73.
- Margarido, G. R., Souza, A. P., and Garcia, A. A. (2007). OneMap: software for genetic mapping in outcrossing species. *Hereditas* **144**, 78-9.
- Maurin, O., Davis, A. P., Chester, M., Mvungi, E. F., Jaufeerally-Fakim, Y., and Fay, M. F. (2007). Towards a Phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences. *Ann Bot* **100**, 1565-83.
- Merlo, P. M. S. (2012). "Conilon Capixaba: 100 anos de desafios, crescimento e inovação.," Bumerangue Produção de comunicação, Vitória, ES.
- Missio, R. F., Caixeta, E. T., Zambolim, E. M., Zambolim, L., Cruz, C. D., and Sakiyama, N. S. (2010). Polymorphic information content of SSR markers for *Coffea* spp. *Crop Breeding and Applied Biotechnology* **10**, 89-94.
- Mohammadi, S. A., and Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants - Salient statistical tools and considerations. *Crop Science* **43**, 1235-1248.
- Moncada, P., and McCouch, S. (2004). Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome* **47**, 501-9.
- Musoli, P., Cubry, P., Aluka, P., Billot, C., Dufour, M., De Bellis, F., Pot, D., Bieysse, D., Charrier, A., and Leroy, T. (2009). Genetic differentiation of wild and

- cultivated populations: diversity of *Coffea canephora* Pierre in Uganda. *Genome* **52**, 634-46.
- Nass, L. L. (2001). Utilização de recursos genéticos vegetais no melhoramento. In "Recursos Genéticos & Melhoramento- Plantas" (L. L. Nass, A. C. C. Valois, I. S. D. Melo and M. C. Valares-Ingliš, eds.), Vol. 1, pp. 29-56. Fundação MT, Rondonópolis, MT.
- Pestana, K. N. (2010). Caracterização fenotípica e molecular da resistência do cafeeiro Hibrido de Timor a *Hemileia vastatrix*, Universidade Federal de Viçosa, Viçosa, MG, Brazil.
- Poncet, V., Dufour, M., Hamon, P., Hamon, S., de Kochko, A., and Leroy, T. (2007). Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. *Genome* **50**, 1156-1161.
- Poncet, V., Hamon, P., Minier, J., Carasco, C., Hamon, S., and Noirot, M. (2004). SSR cross-amplification and variation within coffee trees (*Coffea* spp.). *Genome* **47**, 1071-81.
- Poncet, V., Rondeau, M., Tranchant, C., Cayrel, A., Hamon, S., de Kochko, A., and Hamon, P. (2006). SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Mol Genet Genomics* **276**, 436-49.
- Prakash, N., Combes, M. C., Dussert, S., Naveen, S., and Lashermes, P. (2005a). Analysis of genetic diversity in Indian robusta coffee gene pool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLPs. *Genetic Resources and Crop Evolution* **52**, 333-343.
- Prakash, N. S., Ganesh, D., and Bhat, S. S. (2005b). Population dynamics of coffee leaf rust (*Hemileia vastatrix*) and recent advances in India. In "Durable resistance to coffee leaf rust" (L. Zambolim, E. M. Zambolim and V. M. P. Várzea, eds.), pp. 411-442. Universidade Federal de Viçosa, Viçosa, Brazil.
- Price, A. H. (2006). Believe it or not, QTLs are accurate! *Trends Plant Sci* **11**, 213-216.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-959.
- Priyono, Florin, B., Rigoreau, M., Ducos, J. P., Sumirat, U., Mawardi, S., Lambot, C., Broun, P., Petiard, V., Wahyudi, T., and Crouzillat, D. (2010). Somatic embryogenesis and vegetative cutting capacity are under distinct genetic control in *Coffea canephora* Pierre. *Plant Cell Rep* **29**, 343-357.
- Resende, M. D. V. d., Silva, F. F. e., Lopes, P. S., and Azevedo, C. F. (2012). "Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Múltivariada e Estatística Espacial," Universidade Federal de Viçosa, Viçosa.
- Ritter, E., Gebhardt, C., and Salamini, F. (1990). Estimation of Recombination Frequencies and Construction of Rflp Linkage Maps in Plants from Crosses between Heterozygous Parents. *Genetics* **125**, 645-654.
- Ritter, E., and Salamini, F. (1996). The calculation of recombination frequencies in crosses of allogamous plant species with applications to linkage mapping. *Genet Res* **67**, 55-65.
- Romero, G., Alvarado A, G., Cortina G, H., Ligarreto M, G., Galeano, N., and Herrera P, J. (2010). Partial resistance to leaf rust (*Hemileia vastatrix*) in coffee (*Coffea arabica* L.): genetic analysis and molecular characterization of putative candidate genes. *Molecular Breeding* **25**, 685-697.
- Rosado, T. B., Tomaz, R. S., Junior, M. F. R., Guimarães, L. M. d. S., Araújo, E. F., Alfnas, A. C., and Cruz, C. D. (2010). Detection of QTL associated with rust

- resistance using IBD-based methodologies in exogamic *Eucalyptus* spp. populations. *Crop Breed. Appl. Biotechnol.* **10**, 321-328.
- Schuster, I., and Cruz, C. D. (2004). "Estatística genômica aplicada a populações derivadas de cruzamentos controlados," Universidade Federal de Viçosa, Viçosa, MG.
- Silva, M. V. G. B., Martinez, M. L., Torres, R. A., Lopes, P. S., Euclides, R. F., Machado, M. A., and Arbex, W. (2004). Mapeamento de QTL em famílias de irmãos completos por meio de modelos aleatórios. *Arq. Bras. Med. Vet. Zootec.* **56**, 232-241.
- Souza, F. F. d. (2011). Estudos sobre a diversidade, estrutura populacional, desequilíbrio de ligação e mapeamento associativo em *Coffea canephora* Pierre ex Froehner, Universidade Federal de Viçosa, Viçosa, MG.
- Tanksley, S. D. (1993). Mapping polygenes. *Annu Rev Genet* **27**, 205-33.
- Tomaz, R. S. (2008). Mapeamento de loco de resistência à *Puccinia psidii* em *Eucalyptus* sp. por meio de marcadores moleculares microssatélites, UFV, Viçosa, MG.
- Valois, A. C. C., Nass, L. L., and Goés, M. (2001). Conservação ex situ de recursos genéticos vegetais. In "Recursos Genéticos & Melhoramento- Plantas" (L. L. Nass, A. C. C. Valois, I. S. Melo and M. C. Valadares-Ingles, eds.), Vol. 1, pp. 123-149. Fundação MT, Rondonópolis - MT.
- Wagner, M., and Bettencourt, A. J. (1965). Inheritance of reaction to *Hemileia vastatrix* Berk & Br. in *Coffea arabica* L. *Coffee Rusts Research Center*.
- Wu, R., Ma, C. X., Painter, I., and Zeng, Z. B. (2002). Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theor Popul Biol* **61**, 349-63.
- Zambolim, L., Vale, F. X. R., Pereira, A. A., and Chaves, G. M. (1999). Manejo integrado das doenças do cafeeiro. In "Produção de café com qualidade" (L. Zambolim, ed.), Vol. 1. Universidade Federal de Viçosa, Viçosa.
- Zambolim, L., Maciel-Zambolim, E., Vale, F. X. R., Pereira, A. A., Sakiyama, N. S., and Caixeta, E. T. (2005). Physiological races of *Hemileia vastatrix* Berk. et Br. in Brazil: Physiological variability, current situation and future prospects. In "Durable Resistance to Coffee Leaf Rust" (L. Zambolim, E. M. Maciel-Zambolim and V. M. P. Várzea, eds.), Vol. 1. Universidade Federal de Viçosa, Viçosa, Brasil.
- Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457-68.

#### **4. CAPÍTULO 1: EFEITO DA CODIFICAÇÃO DE DADOS MOLECULARES EM ESTUDOS GENÉTICOS INTRA-POPULACIONAIS**

##### **RESUMO**

O uso de marcadores moleculares para estudos de diversidade genética representou um avanço nessa área, sobretudo, pelo acréscimo no nível de polimorfismo acessado e a neutralidade fenotípica das informações. Marcadores codominantes, como os microssatélites (SSR), são aqueles que possuem a sensibilidade para discriminar os heterozigotos nos estudos genéticos. Apesar dessa vantagem, o que se tem observado é a preferência de alguns estudos em trabalhar com dados codificados, sobretudo, pela maior facilidade de avaliação, existência de poliploidias e necessidade de análise conjunta de diferentes tipos de marcadores. Com isso, o presente trabalho se propõe a estudar as consequências dessas codificações em dados simulados e em dados reais de café; além de sugerir uma alternativa para as análises conjunta de resultados de diferentes marcadores moleculares. Para os dados simulados foram propostos dois cenários, sendo o primeiro simulando o uso de marcadores SNPs e o segundo para marcadores SSR. Para os dados reais, foram utilizados dados de genotipagem de SSR em acessos de *Coffea canephora* do Banco de Germoplasma da Embrapa. Para esse tipo de dados, a diversidade genética foi estudada utilizando índices de dissimilaridade e abordagem bayesiana implementada no *software* STRUCTURE. Para os dados simulados, em ambos os cenários observou-se perda de informações genéticas ao realizar codificações dos dados. O mesmo resultado foi observado nos estudos com café, no entanto, essa perda de informação foi considerada dentro do contexto de um programa de melhoramento de plantas e as suas consequências foram ponderadas para avaliações de germoplasma e seleção de genitores para hibridizações. Para os estudos envolvendo diferentes classes de marcadores moleculares, foi discutida uma alternativa para a análise conjunta, no qual a informatividade, a cobertura, e a qualidade dos dados moleculares fossem ponderadas nos estudos de diversidade genética.

**Palavras chaves:** Simulação de dados; marcadores dominantes; marcadores codominantes; SSR; *Coffea canephora*; STRUCTURE.

## INTRODUÇÃO

A determinação da dissimilaridade genética entre indivíduos é um ponto decisivo em estudos de diversidade genética, pois diferentes metodologias podem ser utilizadas para este propósito. Nos anos 60, a dissimilaridade entre e dentro populações era quantificada por meio de avaliações visuais, resultantes da expressão de marcadores fenotípicos. A partir dos anos 80, com o advento da biologia molecular e da percepção de suas vantagens, esses mesmos estudos passaram a ser realizados com o auxílio de marcadores moleculares que permitiram a detecção de polimorfismos em nível de DNA. A grande vantagem do uso dessa tecnologia foi o acréscimo no nível de polimorfismo acessado e a neutralidade fenotípica das informações, ou seja, os resultados obtidos encontram-se livre da influência do ambiente. Essas características fizeram das análises moleculares uma ferramenta poderosa nos estudos de diversidade, sobretudo, pela maior acurácia e confiabilidade dos resultados obtidos (Souframanien and Gopalakrishna, 2004).

Os avanços rápidos nas tecnologias moleculares abriram novas perspectivas para a avaliação da diversidade e os estudos passaram a envolver, diferentes abordagens estatísticas. As abordagens que utilizam a frequência alélica se baseiam no cálculo de diferentes parâmetros com o objetivo de mensurar a estruturação e a variação genética existente dentro de uma população ou conjunto de genótipos. Outra abordagem refere-se aos enfoques comparativos da constituição genotípica que podem ser realizados entre e dentro de cada população ou grupos de indivíduos em estudo. Neste caso, uma matriz de dissimilaridade (ou similaridade) genética é construída, a partir da análise de todas as possíveis combinações genotípicas par-a-par, e é usada no estabelecimento de padrões de similaridades e dissimilaridades no material de estudo (Karp et al., 1997; Kosman and Leonard, 2005). O uso de métodos multivariados nessas análises é de grande importância, uma vez que permitem a classificação, ordenamento e resumo do grande número de informações que são geradas. Metodologias estatísticas com essa finalidade foram apresentadas por Mohammadi and Prasanna (2003). Mais recentemente, uma abordagem bayesiana tem sido utilizada nos estudos de diversidade genética e estrutura de populações. Implementada no *software* STRUCTURE (Pritchard et al., 2000), dados genotípicos são utilizados para classificação probabilística dos genótipos considerando K populações (onde K pode ser desconhecido). Além dos resultados robustos, soma-se o fato do *software* permitir fazer inferências sobre migrantes, frequências alélicas e

estudos de zonas híbridas; utilizando classe de marcadores moleculares dominantes e codominantes.

A maioria das pesquisas em programas de melhoramento genético de plantas tem preferido a utilização das abordagens genótípicas, sobretudo, na avaliação e caracterização dos bancos de germoplasmas, que auxiliam curadores e pesquisadores na obtenção de cultivares mais produtivos, ou de melhor qualidade, ou resistente a fatores bióticos e abióticos (Laurentin, 2009). Nestes tipos de estudos, a diversidade genética é calculada utilizando índices de dissimilaridade (ou similaridades), que permite a diferenciação em nível de acesso. A utilização do índice mais apropriado é um ponto chave para determinação da variabilidade genética. A escolha desse índice deve ser feita levando em consideração o tipo de marcador, a ploidia do organismo e o objetivo do estudo (Kosman and Leonard, 2005). Considerando os tipos de marcadores existentes, uma divisão de acordo com a sensibilidade discriminatória de cada um deles pode ser feita. Os marcadores dominantes são aqueles que não possuem a capacidade de diferenciar os genótipos homocigotos dominantes dos heterocigotos, pertencem a essa classe os marcadores RAPD (*Random Amplified Polymorphic DNA*), AFLP (*Amplified Fragment Length Polymorphis*), ISSR (*Inter-Simple Sequence Repeat*) e os DaRTs (*Diversity Arrays Technology*). Os marcadores denominados de codominantes são aqueles capazes de discriminar os genótipos heterocigotos e pertencem a esse grupo os RFLP (*Restriction Fragment Length Polymorphism*), SSR (*Simple Sequence Repeats*) e SNP (*Single Nucleotide Polymorphism*).

Os marcadores codominantes apresentam como vantagem a possibilidade de diferenciação dos genótipos heterocigotos, o que teoricamente enriquece as análises de diversidade (Ferrão et al., 2013). Contudo, em alguns trabalhos, o que se têm observado é a opção pela codificação dos dados, devido à facilidade de avaliação, existência de poliploidias e necessidade de análise conjunta de marcadores dominantes e codominantes. A marcação dos alelos na forma binária, com anotação apenas de presença ou ausência de bandas, o que simplifica o *score* e as análises estatísticas (Kosman and Leonard, 2005). Outro fator que dificulta a avaliação de marcadores codominantes é a análise de indivíduos poliploides. Nessas avaliações não é possível, pela análise visual, identificar quantos números de cópias do alelo está presente em um determinado heterocigoto, assim, uma maneira simples e rápida de avaliação é codificar os dados de marcadores codominantes (Bruvo et al., 2004; De Silva et al., 2005). Além

disso, com a disponibilidade de diferentes técnicas moleculares é comum que muitos grupos de pesquisas utilizem mais de um tipo de marcador para estudos de diversidade (Belaj et al., 2003; Ferrão et al., 2013; Gallego et al., 2005; Lamia et al., 2010). No entanto cada técnica difere quanto à informatividade, cobertura no genoma e qualidade dos dados gerados. Assim, uma forma de agregar informações provenientes de diferentes metodologias é avaliando todos os marcadores com o mesmo índice de dissimilaridade (ou similaridade) e, para isso, é necessária a transformação dos dados. Os três fatores listados representam as principais causas de transformação de dados. Assim, o presente trabalho objetiva responder os seguintes questionamentos: i) utilizando dados simulados em diploides, qual seria a perda de informação ao avaliar marcadores codominantes como dominantes? ii) Considerando dados reais, como essas diferenças de avaliação comprometem o manejo dos recursos genéticos em programas de melhoramento genético de plantas? ; iii) Qual seria a melhor alternativa para análise conjunta de dados provenientes de diferentes marcadores moleculares?

## **MATERIAL E MÉTODOS**

### *Dados simulados*

Para realização do estudo foram propostos dois cenários distintos. No cenário 1, uma população de 200 indivíduos e constituída de 500 locos bialélicos foi simulada. Neste cenário considerou-se o uso de marcadores SNP's para o estudo da diversidade. No cenário 2, foi simulada uma população de mesmo tamanho e mesmo número de locos, contudo o marcador utilizado era multialélico (1- 9 alelos) tal como os SSR's. O processo de simulação consistiu em estabelecer um número de amostras de simulação, chamada de réplicas ( $r$ ). Cada réplica ( $r$ ) foi formada por um número inicial ( $m$ ) de marcadores a ser avaliado. Também foi estabelecido um incremento ( $\Delta$ ) a ser dado aos sucessivos números iniciais ( $m$ ), de forma a ser possível estabelecer réplicas que variam de um tamanho inicial ( $m$ ) até um tamanho final ( $m'$ ), numa progressão aritmética de razão ( $\Delta$ ). Para ambos os cenários o valor de  $m$  e de  $\Delta$  foram 50, enquanto que o valor de  $m'$  foi igual a 500. Assim, 10 réplicas ( $r$ ) para cada cenário foram considerados.

A fim de estabelecer comparações entre as formas de avaliação, primeiramente, todas as réplicas, variando de  $m$  a  $m'$ , foram avaliadas como codominantes e realizou-se a análise de diversidade genética. Cada alelo recebeu um rótulo de acordo com o seu tamanho molecular. Posteriormente, as mesmas réplicas foram consideradas, contudo,

foram codificadas, anotando a presença e ausência de banda (dados binários). No cenário 2, além dessas duas codificações, uma outra anotação foi incluída para comparação. O alelo mais frequente da população foi codificado como 1, enquanto que os outros receberam 0. Essa última transformação é comum em trabalhos que utilizam marcadores multialélicos.

Como parâmetro de comparação, foi considerado que a verdadeira dissimilaridade genética foi obtida pela análise de 500 locos avaliados como codominantes. Essa pressuposição foi feita considerando a vantagem dessa classe de marcador em discriminar os genótipos homozigotos e heterozigotos. Para quantificar a perda de informação obtida pela codificação dos dados, cada réplica foi avaliada utilizando um índice apropriado. Para os dados codificados, a dissimilaridade genética foi calculada pelo Complemento do Coeficiente de Jaccard (1908), comumente utilizado nas análises para dados moleculares. Os dados codominantes foram avaliados considerando o Complemento do Índice Ponderado (Cruz et al., 2011), citado por Ferrão et al. (2013). A comparação das dissimilaridades genéticas foi realizada pela correlação de matrizes. A normalização da estatística Z de Mantel foi usada para determinar a associação entre duas matrizes e 1000 permutas aleatórias foram utilizadas para se testar a significância das correlações matriciais.

As simulações e as análises de diversidade genética foram feitas utilizando o programa computacional GENES (Cruz, 2006). Cinco simulações para cada cenário foram realizadas. Os resultados apresentados são uma média aritmética dos valores obtidos.

#### *Análise molecular de acesso de Coffea canephora*

Os estudos de diversidade genética com dados reais foram realizados em cafeeiros utilizando a espécie *C. canephora* ( $2x=2n=22$ ). Oitenta e sete acessos do Banco de Germoplasma da Embrapa Rondônia foram utilizados nessas análises (Tabela 1). A caracterização e o estudo dessa coleção foi realizada anteriormente por Souza (2011) e Ferrão et al. (2013). Os acessos mantidos nessa coleção pertencem a dois grupos varietais distintos: Conilon e Robusta. A existência de possíveis híbridos naturais deve ser considerada, uma vez que a espécie *C. canephora* é alógama.

**Tabela 1.** Acessos de *C. canephora* mantidos no Banco de Germoplasma da Embrapa Rondônia\*

Acessos	Código	Acessos	Código	Acessos	Código
Conilon Incaper 03	ES03	Cpafro183	RO183	Cpafro147	RO147
Conilon Incaper 110	ES110A	Cpafro015	RO015	Cpafro164	RO164
Conilon Incaper 28	ES028	Cpafro140	RO140	Cpafro189	RO189
Conilon Incaper 16	ES16	Cpafro001	RO001	Cpafro190	RO190
Conilon Incaper 45	ES45	Cpafro016	RO016	Cpafro036	RO036
Kouillou IAC 661	K661	Cpafro044	RO044	Cpafro089	RO089
Conilon Incaper V.1	ESV1	Cpafro101	RO101	Cpafro045B	RO045B
Conilon Incaper V.2	ESV2	Cpafro119	RO119	Cpafro077	RO077
Conilon Incaper V.3	ESV4	Cpafro155	RO155	Cpafro138	RO138
Robusta IAC 1675	R1675	Cpafro004	RO004	Cpafro142	RO142
Robusta IAC 2259	R2259	Cpafro042	RO042	Cpafro196	RO196
Robusta IAC 2257.1	R22571	Cpafro098	RO098	Cpafro193	RO193
Robusta IAC 2257.2	R22572	Cpafro160	RO160	Cpafro049	RO049
Robusta IAC 640.1	R6401	Cpafro184	RO184	Cpafro030	RO030
Robusta IAC 640.2	R6402	Cpafro018	RO018	Cpafro032	RO032
Robusta IAC 2258.1	R22581	Cpafro045	RO045	Cpafro076	RO076
Robusta IAC 2258.2	R22582	Cpafro146	RO146	Cpafro161	RO161
Robusta IAC 2258.3	R22583	Cpafro194	RO194	Cpafro035	RO035
Cpafro 006	RO006	Cpafro017	RO017	Cpafro038	RO038
Cpafro 047	RO047	Cpafro043	RO043	Cpafro073	RO073
Cpafro 199	RO199	Cpafro120	RO120	Cpafro139	RO139
Cpafro052	RO052	Cpafro010	RO010	Cpafro141	RO141
Cpafro151	RO151	Cpafro064	RO064	Cpafro171	RO171
Cpafro172	RO172	Cpafro086	RO086	Cpafro197	RO197
Cpafro003	RO003	Cpafro103	RO103	Cpafro025	RO025
Cpafro058	RO058	Cpafro203	RO203	Cpafro026	RO026
Cpafro059	RO059	Cpafro022	RO022	Cpafro072	RO072
Cpafro088	RO088	Cpafro024	RO024	Cpafro075	RO075
Cpafro096	RO096	Cpafro127	RO127	Cpafro115	RO115

\*Adaptado de Souza (2011) e Ferrão et al. (2013).

Para as análises em laboratório, folhas jovens e completamente desenvolvidas de cada cafeeiro foram coletadas, congeladas a -80 °C, liofilizadas, trituradas e conservadas a -20 °C. O DNA genômico foi extraído conforme protocolo descrito por Diniz et al. (2005). As análises moleculares foram realizadas utilizando o marcador codominante e multialélico SSR, no laboratório de Biotecnologia do Cafeeiro (BioCafé/Bioagro). Quarenta e sete *primers* SSRs foram utilizados na genotipagem (Tabela 2) e a amplificação dos microssatélites foi realizada conforme relatado por Missio et al. (2010). A marcação dos alelos foi realizada da mesma maneira que no cenário 2 em dados simulados, no qual os marcadores foram avaliados como codominantes (*Cod*), na forma binária (*Bin*) e dominantes considerando o alelo mais frequente (*Dom*).

**Tabela 2.** *Primers* microssatélites utilizados nas análises com dados reais de *Coffea canephora*.

Primer	Forward primer (5'>3')	Reverse primer (5'>3')
SSR-07 <sup>1</sup>	TGACATAGGGGGCTAAATTG	TTAATGGTGACGCTTTGATG
SSR-08 <sup>1</sup>	CACTGGCATTAGAAAGCACC	GGCAAAGTCAATGATGACTC
SSR-13 <sup>1</sup>	TGGCCGTGATAATAAACAGC	ATGTGGCAATCTAAAGCCAA
SSR-16 <sup>1</sup>	ACCCGAAAGAAAGAACCAAG	CCACACAACCTCCTCATT
SSR-21 <sup>1</sup>	GACCATTACATTTACACAC	GCATTTTGTTCACACTGTA
SSR-29 <sup>2</sup>	GGCTTCTTGGGTGTCTGTGT	CCATTGGCTTTGTATTTCTGG
SSR-30 <sup>3</sup>	ATGGGGCCAACCTGAATATG	CAGGGCATCTATCTACTTCTTT
SSR-34 <sup>3</sup>	GGAGACGCAGGTGGTAGAAG	TCGAGAAGCTTGGGGTGT
SSR-35 <sup>3</sup>	CTGGCATTAGAAAGCACCTTG	GCTTGGCTCACTGTAGGACTG
SSR-37 <sup>3</sup>	CAACACTATCTCTTGATTTTTCACT	CGTGCAAGTCACATACTTTACTAC
SSR-39 <sup>2</sup>	TCCCCATCTTTTTCTTTCC	GGGAGTGTTTTGTGTGGCTT
SSR-40 <sup>3</sup>	AAAGGAAAATTGTTGGCTTGA	TCCACATACATTTCCACAGTA
SSR-42 <sup>2</sup>	TTGCTTGCTTGTCTGTAT	TGACACGAGAGTTAGAAATGA
SSR-43 <sup>3</sup>	TTTTCTGGGTTTTCTGTGTCTC	TAACCTCCATTCCCGCATT
SSR-46 <sup>3</sup>	AATGAAGAAGAGGGTGGTG	CGAGGGTATTGTTTTCCAG
SSR-48 <sup>2</sup>	AGCAAGTGGAGCAGAAGAAG	CGGTGAATAAGTCGCAGTC
SSR-49 <sup>3</sup>	TGGAGAAGGCTGTTGAAACC	GGCGTGAAGCAAAAAGGTAT
SSR-52 <sup>3</sup>	GACCAAATGTCAGCTCATTG	GCCGACTGCTTTTTAGTGT
SSR-55 <sup>2</sup>	GCAGGTATTTGAAGGATGAACC	GTGTAGGTGGTGGCAGTGTGT
SSR-56 <sup>3</sup>	AGCTATCTTTATCTCACACACACA	GTTAGTGTTCGATTTGGTACTG
SSR-57 <sup>2</sup>	CTCGCTTTCACGCTCTCTCT	CGGTATGTTCTCGTTCCTC
SSR-59 <sup>3</sup>	CCAGCTCTCCTCACTCTTTTCA	GGTGGTGGAGGGTAATAGG
SSR-64 <sup>3</sup>	GTTAGTGTGCGACCGTGTGT	TTATGCCCTCCCCATATCT
SSR-65 <sup>3</sup>	CTCACACACACACTGTCACG	CGAATGAGGCTCCATCAC
SSR-70 <sup>4</sup>	GTAACCACCACCTCCTCTG	TGGAGGTAACGGAAGCTCTG
SSR-71 <sup>4</sup>	GCTAAGTTCAATTGCCCTGT	GGGTTAATTTGATTGCGTGA
SSR-74 <sup>4</sup>	TGGGGAAAAGAAGGATATAGACAAGAG	GAGGGGGGCTAAGGGAATAACATA
SSR-76 <sup>4</sup>	GGTCCCACTCTCAAGCTGAA	GGCAATTGATTCTGGAACCT
SSR-77 <sup>5</sup>	TCTCCTCTTCTTGCTGAGCC	AGATTCACCCTTCAAGTTTCTC
SSR-81 <sup>5</sup>	AGTAATGAACCTGCCGCTCTTT	TGTCATCTTGTGTTTTCCATCC
SSR-82 <sup>5</sup>	CAAAATGAAGGAGAAAAGTGGACA	TGGCTTCATCTCAACCTTCTC
SSR-84 <sup>5</sup>	AAGTAGATTGGTGAAAGGGAAGC	TCCTTCATTTTCTCCTCTGGTT
SSR-87 <sup>5</sup>	ATTCGACGACTCCAAAGCATA	CCTGTGGCCCTTCCTT
SSR-93 <sup>5</sup>	ATACAGCAATTTTGAGAGGAGGAA	TTCTGTGCCTTCCCAGTCA
SSR-100 <sup>5</sup>	ACCCTTACTACTTATTTACTCTC	ACATCCCCTTGCCATTTCTTC
SSR-102 <sup>5</sup>	CTCAGAGCTGCGGTGGTGTA	CCGGACGATCTTTCTTTCTTCA
SSR-106 <sup>1</sup>	CCCTCCCTCTTCTCCTCTC	TCTGGGTTTTCTGTGTTCTCG
SSR-114 <sup>6</sup>	TAACAGAAGCACAAAACC	TCTAAACCCACCTCACAAC
SSR-119 <sup>3</sup>	TTGCCATCATCGTTCATTCT	GCATAGTGTGCGTTGTGTTGTT
SSR-121 <sup>3</sup>	CGACACTTTCTTTGGCACTC	AGACACCCACCCATCCAC
SSR-122 <sup>3</sup>	CGTCTCGTTTACGCTCTCT	GATCTGCATGTACTGGTGCTTC
SSR-138 <sup>3</sup>	GCACTTCCCTCTCAACCAAC	ACTAGGACAGCAAATAGCATAACCC
SSR-151 <sup>3</sup>	TGGTTCAAGGTAATGTGGAAA	GGCCGTGATAATAAACAGCTA

**Referências:** <sup>1</sup> Combes et al. (2000); <sup>2</sup>Coulibaly et al. (2003); <sup>3</sup>Poncet et al. (2004); <sup>5</sup>Moncada and McCouch (2004); <sup>6</sup>Leroy et al. (2005).

Para a análise dos dados codificados (*Bin e Dom*), a dissimilaridade genética foi calculada pelo Complemento do Coeficiente de Jaccard (1908). Os dados codominantes foram avaliados considerando o complemento do Índice Ponderado (Cruz et al., 2011). Em todas as análises, o agrupamento para construção dos dendrogramas foi realizado pela metodologia *Neighbor Joining* (NJ). Todos os procedimentos estatísticos foram

feitos utilizando os *softwares* GENES (Cruz, 2006) e FigTree v1.3.1 (Rambaut, 2006) . As análises de agrupamento bayesiano foram realizadas pelo *software* STRUCTURE (Pritchard et al., 2000) o qual foi considerado valores de K variando de 1 a 5 com modelo de mistura genética e cinco repetições. Cada corrida foi implementada com um período de 10.000 *burn-in* seguido por 100.000 MCMC. O verdadeiro número de grupos genéticos foi estimado pelo cálculo de  $\Delta K$ , como descrito por Evanno et al. (2005), utilizando o *software* STRUCTURE HARVESTER (Earl and vonHoldt, 2012).

## RESULTADOS E DISCUSSÃO

### *Dados Simulados*

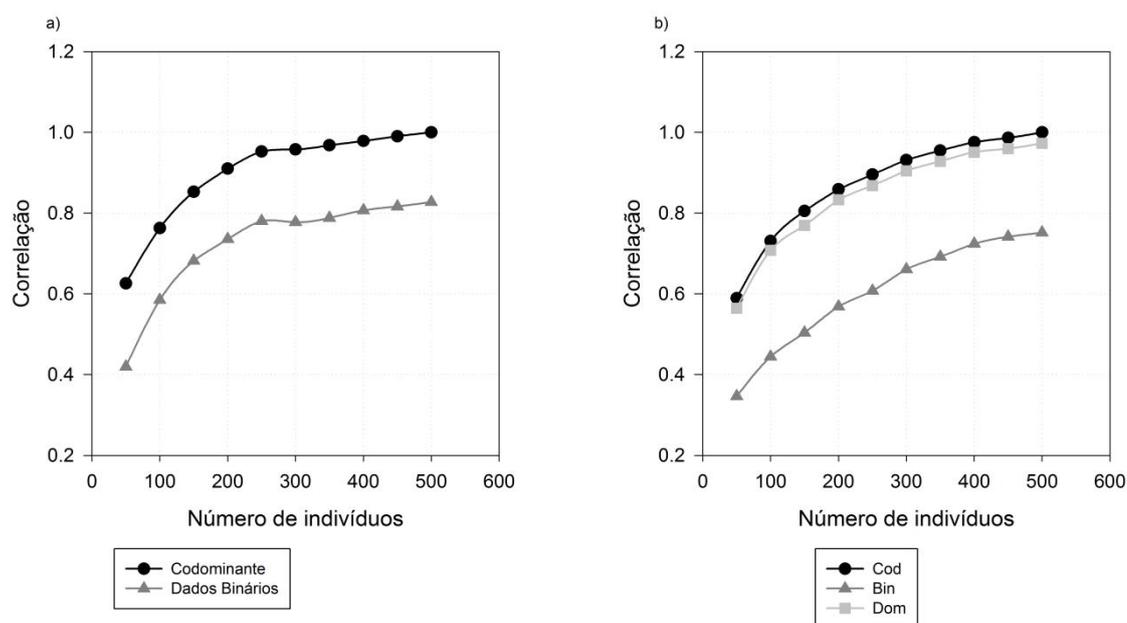
Em dados simulados, dois cenários foram propostos. O primeiro deles simulou o uso de marcadores SNPs, que caracterizam-se por serem codominantes e bialélicos. Trata-se de um marcador amplamente utilizado em estudos genéticos, pois possibilita a geração de um grande número de informações por ensaio molecular. Nessas análises observou-se uma diferença de 18% no valor de correlação entre os dados codificados e não codificados, o que representou perda de informações genéticas (Figura 1a).

No segundo cenário foi simulado o uso dos SSRs, que são marcadores codominantes e multialélicos, o que aumenta o nível de informatividade das análises genéticas. Neste cenário, dois tipos de codificações foram considerados. Na primeira, chamada de *Bin*, observou-se uma perda de informação na ordem de 3%. Nestas análises o marcador codominante multialélico foi avaliado de forma binária com anotação da presença (1) ou ausência da banda (0). Para a segunda codificação, chamada de *Dom*, a perda de informação foi na ordem de 25%. Neste caso, os genótipos que apresentavam o alelo mais frequente foram codificados como 1 (um) e os que possuíam os demais alelos foram codificados como 0 (zero) (Figura 1b). Para todas as codificações realizadas a distância genética entre pares de genótipos foi mensurada utilizando o Índice de Jaccard (1908), que é uma metodologia própria para estudos de marcadores dominantes. Os dados simulados mostraram que o uso de codificações associada a metodologia de análise de marcadores dominantes, resultou em perda de informações genética.

Ao utilizar dados genotípicos de marcadores codominantes é conveniente que às distâncias genéticas entre pares de indivíduos não sejam calculadas simplesmente em termos de coincidência de bandas que são compartilhados entre dois indivíduos, como é

feita nas avaliações com dados binários. Ao utilizar marcadores codominantes é recomendável que a avaliação seja ajustada para representar os padrões alélicos dos indivíduos em todos os locos estudados, representando assim o número total de locos e o número de alelos compartilhados entre eles (Kosman and Leonard, 2005). Estudos comparativos realizados por Ramos et al. (2011) indicam que o Índice Ponderado (Cruz et al., 2011) é um algoritmo eficiente para determinação da diversidade entre pares de indivíduos, pois pondera todos os parâmetros aqui citados.

Considerando dados reais foi possível visualizar como essa perda de informação influencia no manejo de recursos genéticos em programas de melhoramento de plantas. Para isso, marcadores microssatélites foram utilizados na espécie *C. canephora* e a dissimilaridade genética foi calculada considerando as formas de avaliação *Cod*, *Bin* e *Dom*.



**Figura 1.** a) Cenário 1. Simulação do uso de marcadores SNPs avaliados como codominantes e na forma binária; b) Cenário 2. Simulação do uso de marcadores SSRs avaliados como codominantes (*Cod*), na forma binária (*Bin*) e considerando o alelo mais frequente da população (*Dom*).

#### *Análise molecular de acesso de Coffea canephora*

Objetivando quantificar o resultado das codificações em estudos de diversidade genética em *C. canephora* foram utilizados dados reais de genotipagem de acessos do Banco de Germoplasma da Embrapa Rondônia. Com base nas análises genotípicas de

87 cafeeiros e 47 locos SSR foi realizada a análise de agrupamento baseados em coeficientes de dissimilaridade genética e a análise bayesiana no *software* STRUCTURE (Pritchard et al., 2000). Essa última análise trata-se de uma abordagem que permite a classificação probabilística dos genótipos em populações de acordo com o seu coeficiente de ancestralidade (Q).

Nas análises de agrupamentos baseados em coeficientes de dissimilaridade, os dados de marcadores SSR foram codificados como *Cod*, *Bin* e *Dom*, da mesma maneira que nos dados simulados do Cenário 2. Em todas as avaliações (*Cod*, *Dom* e *DomFreq*), para as duas abordagens, foi possível classificar os acessos em relação aos seus grupos varietais.

Popularmente conhecido como “Café Robusta”, a espécie *C. canephora* pode ser dividida em dois grupos varietais de acordo com o seu centro de diversidade e suas características adaptativas. O primeiro grupo, chamado de Conilon é composto por genótipos de folhas menores, menor vigor, menor porte, frutos pequenos, bebida de qualidade inferior e tolerantes à seca; o segundo grupo, chamado de Robusta é composto por genótipos mais altos, vigorosos, de folhas e frutos maiores, com melhor qualidade de bebida e maior sensibilidade à seca (Ferrão et al., 2007). Apesar das diferenças marcantes que separam os dois grupos varietais, a classificação dos acessos nas coleções de germoplasma não é uma tarefa fácil visto que, em razão da forma natural de reprodução estritamente alógama, as populações dessa espécie são formadas por plantas com elevada amplitude fenotípica e heterozigose. Além disso, deve-se considerar a existência de possíveis cruzamentos naturais entre os dois grupos varietais o que pode complicar a discriminação. Dessa forma, a avaliação correta da diversidade genética é de grande importância em um programa de melhoramento, de modo a fornecer aos melhoristas informações fidedignas para que a avaliação e seleção de genótipos sejam feitas de forma acurada. Assim, apesar de todas as análises serem eficientes na separação dos Conilons e Robustas a eficiência dessa discriminação variou para cada metodologia (Figura 2).

Assim como nos dados simulados os resultados das avaliações *Cod* e *Bin* foram similares (Figura 2a e 2b, respectivamente). Os dois dendrogramas apresentaram estruturação semelhante com a alocação de genótipos nos grupos dos Conilons, Robustas e híbridos naturais. A grande diferença nos dois resultados ficou restrita dentro dos grupos, especialmente ao grupo dos Conilons, representado na Figura 2 pelo

clado em azul. Nas avaliações *Cod* obteve-se uma estruturação em três subgrupos (Figura 2a). Dois deles se referem aos genótipos codificados como RO, que são provenientes de coletas no Estado de Rondônia, Brasil; e um terceiro subgrupo, denominado de ES, que é constituído de genótipos de Conilon coletados no Estado do Espírito Santo, Brasil. Os resultados apresentados no dendrograma mostram que a maioria dos genótipos puderam ser agrupados de acordo com a sua localidade de origem. Para a localidade de Rondônia foi observado ainda uma estruturação adicional em dois subgrupos, que foram denominados de RO.1 e RO.2. Segundo Souza (2011), o germoplasma da localidade de Rondônia começou a ser formado nas últimas quatro décadas a partir da introdução de sementes e estacas provenientes dos Estados de São Paulo e Espírito Santo, Brasil. Devido a maior semelhança de RO.2 com o grupo ES, admite-se que os genótipos desse subgrupo sejam originados do Espírito Santo, enquanto que o subgrupo RO.1 é formado por genótipos de origem de São Paulo, uma vez que apresentam características próprias que os distinguem do grupo ES.

Essa estruturação dos Conilons em três subgrupos obtida com os dados *Cod* não foi observada para as avaliações *Bin*. Os acessos mantidos nesse banco de germoplasma representam o material cultivado e conservado em diferentes instituições brasileiras de pesquisas (Ferrão et al., 2013; Souza, 2011). Assim, é esperado que os genótipos de mesma localidade apresentassem perfis moleculares semelhantes, pois compartilham características adaptativas. Essa tendência foi melhor observada na avaliação *Cod*, no qual a maioria dos acessos ES formaram um subgrupo independente.

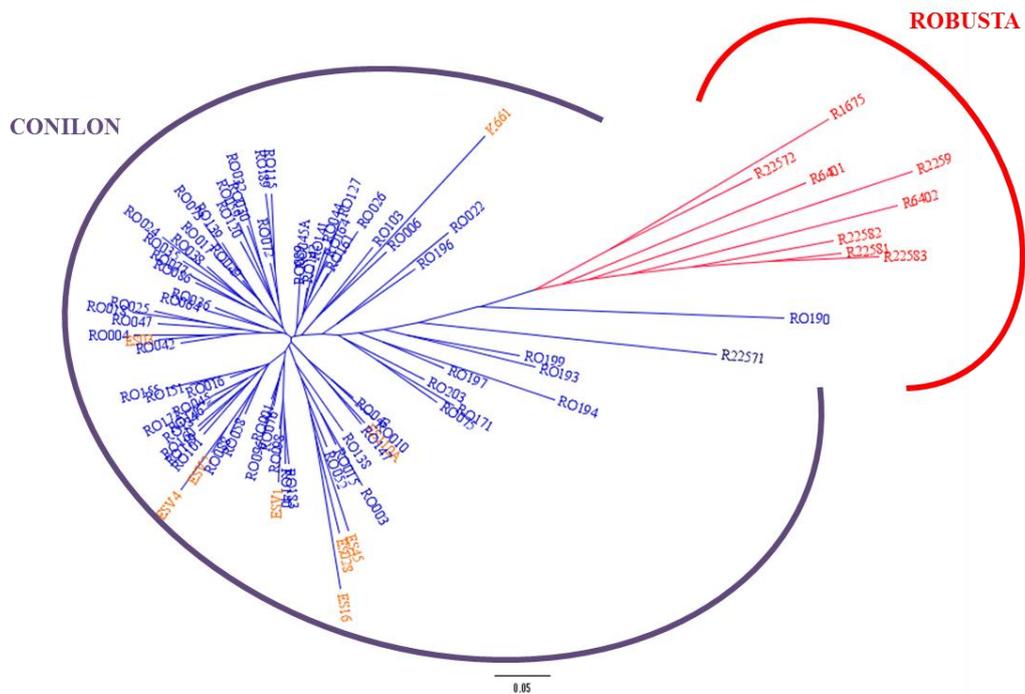
Esse resultado sugere que ao considerar materiais muito divergentes, como acessos de Robusta e Conilon, ambas as avaliações são eficientes na discriminação varietal. Contudo, para o caso de materiais similares e pertencentes a um mesmo grupo, como os acessos ES, a avaliação *Cod* fornece uma discriminação mais detalhada, sendo capaz de separar materiais que compartilham características adaptativas. Essa qualidade é resultante da forma de avaliação dos marcadores codominantes, cujos índices de similaridade ponderam o número de alelos compartilhados e o número de locos em estudo. Outra particularidade das avaliações *Cod* é a possibilidade de se trabalhar diretamente com as frequências alélicas (Karp et al., 1997), o que permite fazer inferências do nível de estruturação genética utilizando, sobretudo, as estatísticas F de Wright (1965, 1978) e as estatísticas G de Nei (1973).

Para o caso das avaliações, nas quais não são feitas distinção dos genótipos, como no caso de marcadores dominantes, as inferências de frequências alélicas podem ser realizadas, se algumas pressuposições sob o conjunto de dados forem realizadas (Bonin et al., 2007). Dentre essas, inclui-se a existência de Equilíbrio de Hardy Weinberg (EHW) na população, genes não ligados e equilíbrio de ligação. No entanto, ao considerar os acessos de coleções de germoplasma, na maioria das vezes, eles não constituem uma população no qual as pressuposições de EHW podem ser atendidas, o que dificulta a utilização dessas abordagens de inferência alélica com dados dominantes.

Se por um lado as avaliações *Cod* e *Bin* apresentaram dendrogramas similares, por outro, as avaliações *Dom* (Figura 2c) apresentaram resultados pouco robustos. Neste tipo de análise, não foi possível separar os híbridos corretamente e a estruturação dos grupos Conilon e Robusta foram diferentes daquelas observadas nas outras análises. O grupo Robusta foi formado com menor número de acessos e, dentro do grupo Conilon, o agrupamento dos materiais ES não apresentaram nenhuma lógica adaptativa. Além disso, foram observados valores nulos de dissimilaridade o que impossibilitou a discriminação de alguns acessos. Com isso, apesar da facilidade, este tipo de avaliação foi considerado a menos eficiente, não sendo recomendada para estudos de diversidade genética. No melhoramento genético de plantas o uso dessa metodologia pode resultar em perdas nos ganhos genéticos, sobretudo em programas que buscam a identificação de grupos heteróticos e genitores contrastantes para exploração do vigor híbrido.

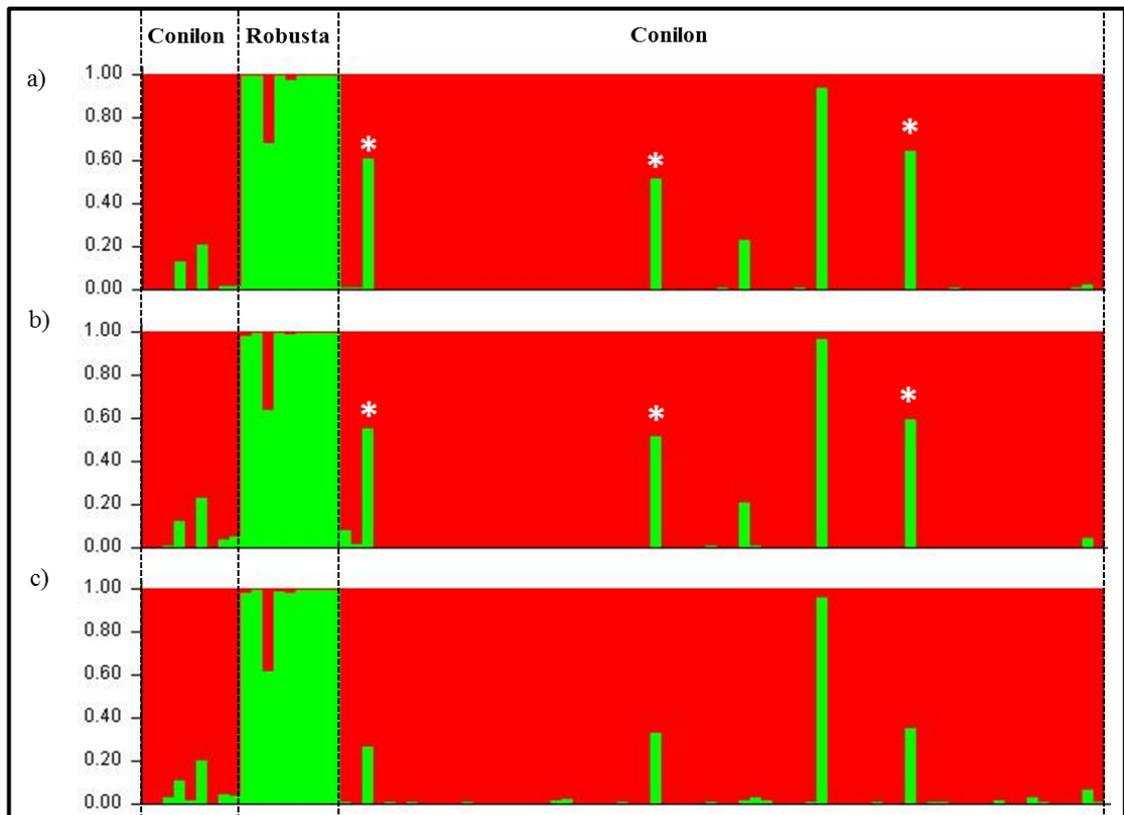
A avaliação correta da diversidade é um fator chave para escolha de genitores. A expectativa de que pais divergentes proporcionem bons híbridos decorre do fato de que, se dois pais são próximos geneticamente, entre si, a tendência é de que compartilhem muitos genes ou alelos em comum e no cruzamento destes haverá pouca complementaridade e baixo vigor em razão do baixo nível de heterozigozidade alélica no cruzamento. No manejo de recursos genéticos, a análise de diversidade genética pode indicar a existência de falsas duplicatas na coleção de germoplasma, resultando no descarte errôneo de materiais promissores. Além disso, a avaliação correta dos recursos genéticos armazenados nessas coleções é de grande valia para os programas de melhoramento, principalmente no início, quando ocorre o planejamento e as definições de estratégias de trabalho. Com isso, a caracterização e avaliação dos acessos depositados devem ser o mais acurado possível.





**Figura 2.** Dendrogramas *Neighbor Joining* (NJ) dos acessos de *C.canephora* do Banco de Germoplasma da Embrapa Rondônia. Clados em azul, vermelho e verde representam os grupos varietais: Conilon, Robusta e híbridos naturais, respectivamente. Códigos de acessos em laranja são provenientes da mesma localidade e merecem atenção especial. Três tipos de avaliações foram considerados: a) *Cod*; b) *Bin*; c) *Dom*.

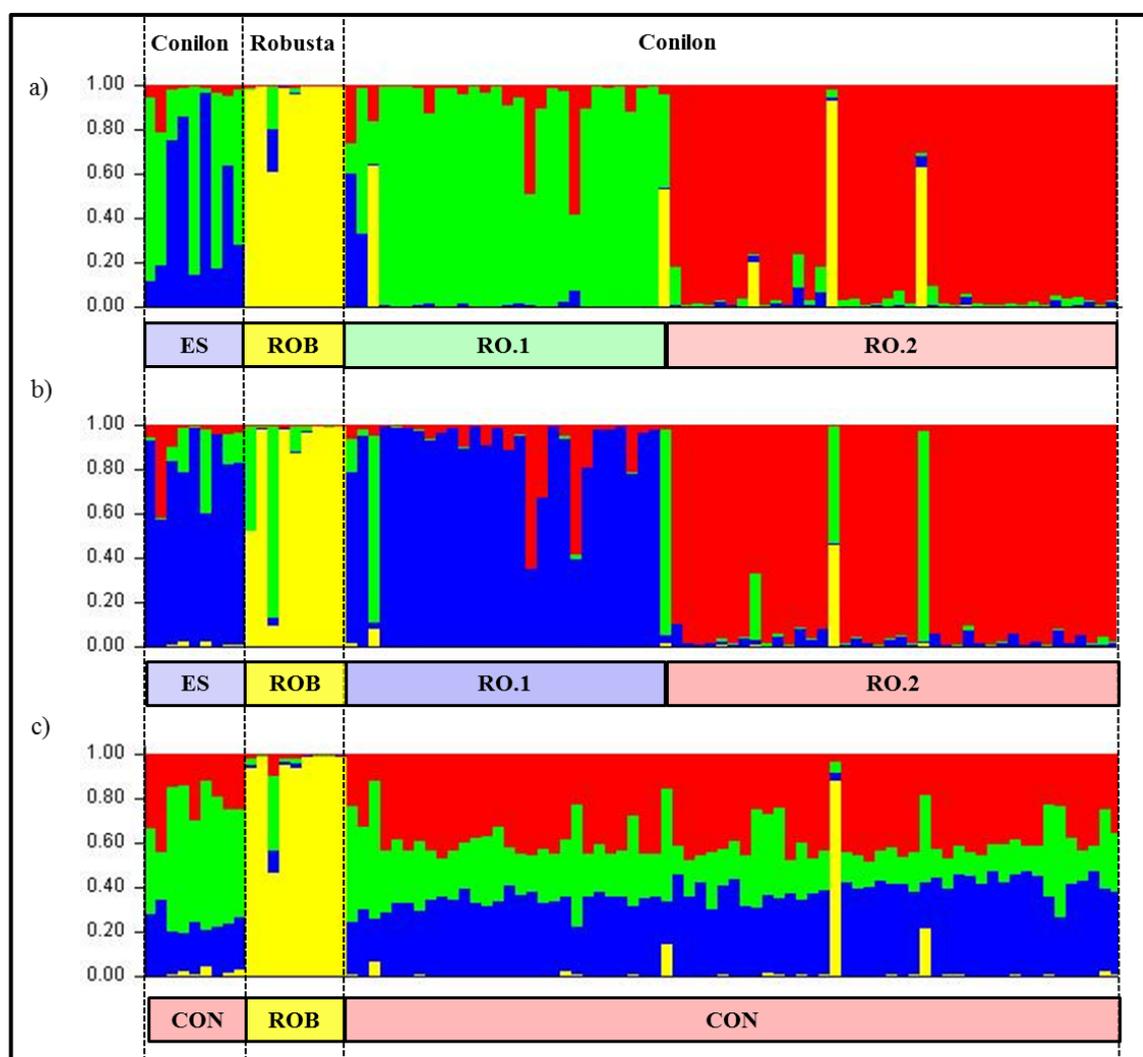
Utilizando a abordagem bayesiana os resultados encontrados foram semelhantes aos já apresentados. Segundo a metodologia proposta por Evanno et al. (2005), os maiores valores de  $\Delta K$  para todas as três avaliação foi obtido para  $K=2$ , mostrando que os acessos podem ser separado em dois grupos de materiais (Figura 3). Esses resultados mostraram que o maior nível de hierarquia dos materiais avaliados está relacionado com o grupo varietal. Assim, para as três avaliações, foi possível discriminar os Conilons dos Robustas. Todavia, assim como nas análises de agrupamento, apenas em *Cod* (Figura 3a) e em *Bin* (Figura 3b) foi possível discriminar com exatidão os híbridos naturais.



**Figura 3.** Gráfico de barras do *software* STRUCTURE utilizado para estudo de diversidade na espécie *C. canephora*. Os oitenta e setes genótipos estão representados na mesma ordem da Tabela 1 e foram divididos em dois grupos (K=2) de acordo com o grupo varietal. a) Avaliações *Cod*; b) avaliações *Bin*; e c) avaliações *Dom*.

Informações mais detalhadas sobre os subgrupos dentro do clado dos Conilons foram obtido considerando valor de K=4 e, novamente, os resultados foram semelhantes àqueles das análises de agrupamento (Figura 4). Ao considerar este valor de K, em *Cod* (Figura 4a) foi observado a semelhança dos perfis moleculares de genótipos da mesma localidade, além da discriminação entre genótipos dos subgrupos ES, RO.1, RO.2 e dos Robustas (ROB). Além disso, confirma-se a hipótese de que ES e RO.1 compartilham características adaptativas. Todavia, em *Bin* (Figura 4b) os genótipos dos subgrupos ES e RO.1 foram agrupados juntamente, admitindo que essa codificação não foi eficiente como na discriminação por localidade. Em estudos de mapeamento associativo esse tipo de informação é fundamental para a eficácia do estudo, visto que a não detecção de estruturas populacionais dentro do conjunto de dados pode ocasionar perda na acurácia das associações e, assim, invalidar os testes padrões (Ewens and Spielman, 1995). Finalmente, a avaliação *Dom* (Figura 4c) permitiu a separação apenas dos Robusta e Conilon não fornecendo nenhuma informação adicional de possíveis subgrupos, além

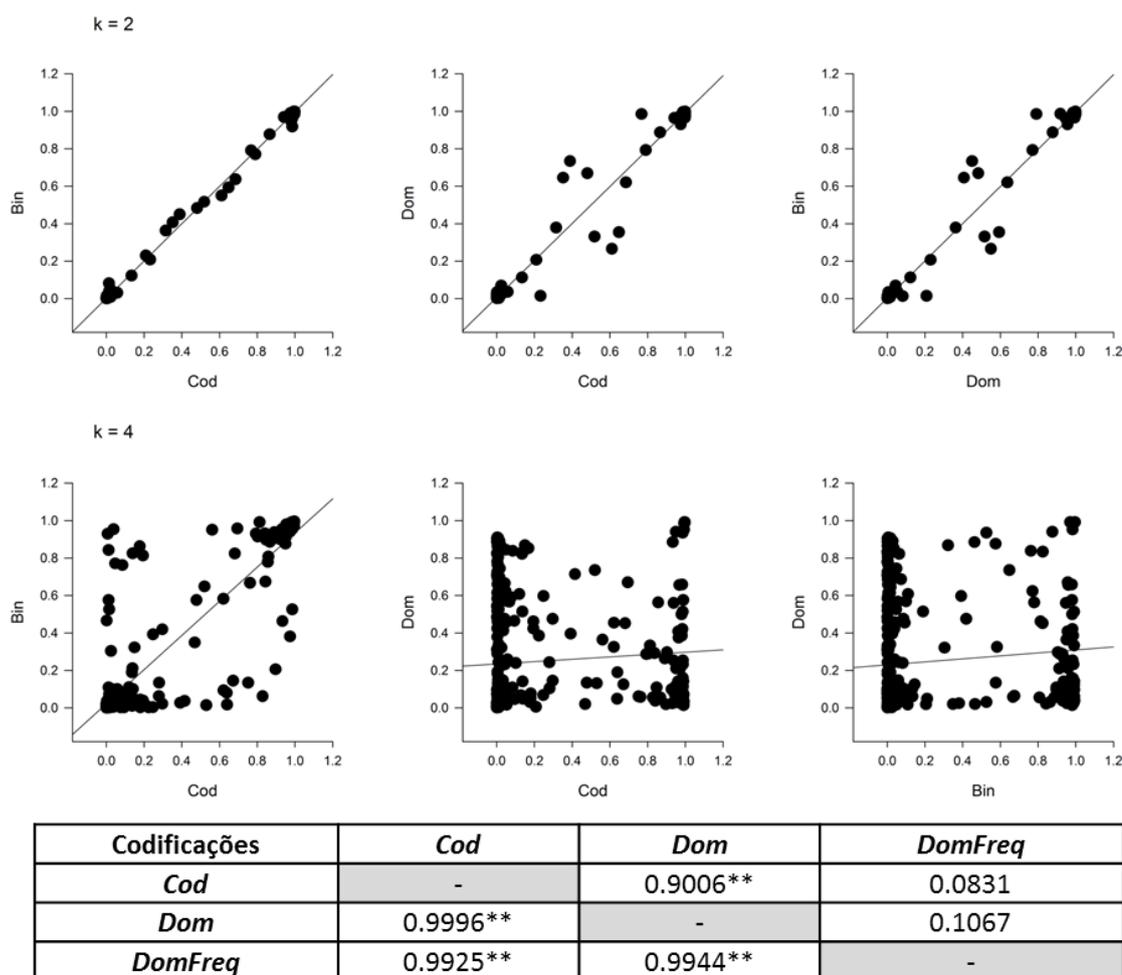
de, apresentar muitos perfis genotípicos idênticos o que é um falso indicativo da existência de duplicatas na coleção de germoplasma.



**Figura 4.** Gráfico de barras do *software* STRUCTURE utilizado para estudo de diversidade na espécie *C. canephora*. Os oitenta e setes genótipos estão representados na mesma ordem da Tabela 1 e foram divididos em dois grupos de acordo com o grupo varietal Conilon ou Robusta. O valor de  $K=4$  permitiu a estruturação dos genótipos em subgrupos de acordo com a localidade, esses foram denominados ES (localidade Espírito Santo), RO.1 e RO.2 (subdivisões dentro da localidade de Rondônia); as cores dos subgrupos indicam se os mesmos puderam ou não ser diferenciadas nas avaliações *Cod* (a); *Bin* (b); e *Dom* (c). A diferenciação clara dos três subgrupos e os Robustas (ROB) só foi observada nas avaliações *Cod*.

A similaridade dos resultados obtidos pela abordagem bayesiana para  $K=2$  e  $K=4$  foi quantificada, levando em consideração a correlação dos valores de coeficiente de ancestralidade ( $Q$ ) obtidos para *Cod*, *Bin* e *Dom* (Anexo 1). Ao considerar uma

estruturação mais simples, classificando os genótipos apenas quanto ao seu grupo varietal ( $K=2$ ), as três formas de avaliação foram altamente correlacionadas. Contudo, ao considerar a eficiência de *Cod*, *Bin* e *Dom* em acessar informações mais detalhadas, considerando um maior nível de estruturação ( $K=4$ ), observa-se valores menores de correlação (Figura 5).



\* Significativos pelo teste  $t$  ( $P=0,05$ ); \*\* significativo pelo teste  $t$  ( $P=0,01$ ).

**Figura 5.** Correlação de Pearson entre os valores de coeficientes de ancestralidade ( $Q$ ) para os diferentes tipos de codificações, considerando  $k=2$  e  $k=4$ . Na tabela os valores para  $k=2$  estão representados abaixo da diagonal e os de  $k=4$  acima da diagonal.

O algoritmo básico do *software* STRUCTURE foi desenvolvido, *a priori*, para análise de dados codominantes, assumindo locos não ligados, equilíbrio de ligação e equilíbrio de HW. Posteriormente, Falush et al. (2007) estenderam o método original para uso com marcadores dominantes e em espécies poliploides. Os elevados valores de

correlação observados para  $K=2$  mostraram que as informações acessadas pelas metodologias são semelhantes. Esses resultados sugerem que a extensão do método original foi eficiente. Todavia, na medida em que aumentamos a exigência informacional das análises, com valores maiores de  $K$ , observou-se um decréscimo dos valores de correlação entre as formas de avaliação.

A diminuição da acurácia ao codificar os dados já havia sido relatada por Falush et al. (2007) e segundo os autores, esse decréscimo é ocasionado pela ambiguidade genotípica gerado pela existência de alelos recessivos.

#### *Análise conjunta de marcadores dominantes e codominantes*

Confirmado a perda de informação e as suas consequências no manejo de recursos genéticos, o uso de dados reais demonstrou que, sempre que possível, a avaliação utilizando metodologias próprias para marcadores codominantes (*Cod*) devem ser preferidas. No entanto, com o avanço da biologia molecular, muitos grupos de pesquisas utilizam mais de uma classe de marcadores para estudos de diversidade de forma a tornar as avaliações mais precisas e confiáveis (Belaj et al., 2003; Gallego et al., 2005; Lamia et al., 2010). Um ponto importante a ser considerado nesse tipo de estudo é que cada tipo de marcador apresenta características próprias que devem ser consideradas nas análises conjuntas. Um exemplo prático disso é o caso dos marcadores AFLP e SSR. Se por um lado os AFLP destacam-se pelo grande número de polimorfismo por ensaios moleculares, por outro os SSR apresentam elevada informatividade e reprodutibilidade dos dados gerados. Dessa forma, é importante que ao analisar conjuntamente os dados, todas essas características sejam levadas em consideração.

Assim, para que as características intrínsecas de cada marcador sejam levadas em consideração propõem-se que para cada metodologia seja calculada uma matriz de (dis)similaridade utilizando os coeficientes mais apropriado. Dentre aquelas mais utilizadas para marcadores dominantes inclui-se o Coeficiente de Jaccard (1908), de Dice (1945) e o de coincidência simples (Sokal and Michener, 1958); para marcadores codominantes destacam-se o Índice Ponderado e Não Ponderado (Cruz et al., 2011); distância genética de Smouse and Peakall (1999) e, mais recentemente, o de Kosman and Leonard (2005). Posteriormente, essas matrizes são multiplicadas por um índice de ponderação obtido pela fórmula:  $IP = L/N \cdot i \cdot QND$ ; onde  $IP$  é o índice de ponderação da

matriz; L é o número de locos acessado para cada marcador individualmente; N é o número total de locos acessado por todos os marcadores; *i* é uma constante de índice de informatividade, e QND é a qualidade da natureza dos dados.

Diante da perda de informação observado nas codificações de dados reais e simulados, o índice de ponderação (IP) é uma metodologia mais adequada para análises conjuntas de marcadores, do que a simples codificação dos dados. A multiplicação das matrizes por este índice permite acrescentar dados importantes de cobertura do genoma, informatividade e qualidade dos dados. Assim, os marcadores mais completos serão ponderados com as maiores notas e terão maior representatividade nos estudos de diversidade genética.

A maior dificuldade para o cálculo de IP é definir os valores das constantes informatividade (*i*) e qualidade dos dados (QND). Considerando as particularidades de cada marcador, propõe-se alguns valores de *i* para marcadores dominantes e codominantes e, dentro de codominante, para bialélicos ou multialélicos (Tabela 3). Os resultados deste e de outros trabalhos (Gallego et al., 2005; Lamia et al., 2010; Powell et al., 1996; Russell et al., 1997; Varshney et al., 2007) comprovam que marcadores codominantes e multialélicos são mais informativos, por esse motivo são ponderados com um valor maior.

**Tabela 3.** Valores de informatividade (*i*) considerando diferentes tipos de marcadores moleculares.

Classe	Avaliação	Marcadores Moleculares	Informatividade ( <i>i</i> )
Codominante	Multialélico	SSR	1,00
	Bialélico	SNP e DaRTs	0,75
Dominante	Binário	AFLP, RAPD e ISSR	0,60

O conceito de QND foi apresentado por Varshney et al. (2007) e é calculado pela fórmula:  $QND = DC \times QM \times PR$ ; onde DC é a capacidade de documentação, QM é a qualidade do marcador e RP é a reprodutibilidade. Os valores de QM variam de acordo com a combinação de *primer* usada para cada tipo de marcador, dessa forma, para o cálculo final de QND utiliza-se o valor médio de QM. Na Tabela 4 é apresentada a escala de valores para este parâmetro.

Os valores de DC e PR são apresentados na Tabela 5 e foram sugeridos levando em consideração as características de cada metodologia. No caso do DC, valores

inferiores foram dados para os marcadores de amostragem ampla, como o AFLP, devido a grande quantidade de bandas/picos que dificultam a interpretação mesmo em sistemas automatizados de genotipagem. Por outro lado, marcadores baseados em métodos de hibridização utilizando plataformas sólidas de DNA, como os SNPs e os DaRTs, receberam valores maiores pois são documentados automaticamente em “*digital fashion*” o que é conveniente para o armazenamento em bancos de dados (Jaccoud et al., 2001; Varshney et al., 2007). Marcadores locos específicos, como o SSR são facilmente avaliados, pois em diploides apresentam no máximo duas bandas/picos. O valor intermediário de DC é resultado dos artefatos que são gerados durante os ensaios, o que dificulta a análise (Guichoux et al., 2011).

O valor de PR é uma medida de confiabilidade de cada técnica. Na caracterização de germoplasmas esse parâmetro é de grande valia, pois indica o quanto os resultados são reprodutíveis e podem ser compartilhados entre diferentes laboratórios. Neste sentido, os marcadores SNPs e DarTs são considerados os mais robustos e o RAPD o menos confiável. Para aqueles ensaios em que são realizadas repetições e existe um controle rigoroso dos dados, pode-se optar por desconsiderar o parâmetro QND do calculo do índice ponderado (IP).

**Tabela 4.** Escala de valores de qualidade do marcador (QM)\*

<b>Escala</b>	<b>Qualidade do Marcador</b>
1,00	Boa qualidade; bandas ou picos únicos e nítidos.
0,75	Banda fraca ou pico baixo.
0,50	Bandas duplas ( <i>Stutters</i> )
0,25	Difícil avaliação (esforço para visualização e análise subjetiva)

\* Varshney et al. (2007)

**Tabela 5.** Valores de capacidade de documentação (DC) e Reprodutibilidade (PR) de diferentes marcadores moleculares.

<b>Parâmetro</b>	<b>Marcadores Moleculares</b>			
	<b>SNP e DaRTs</b>	<b>SSR</b>	<b>AFLP e ISSR</b>	<b>RAPD</b>
DC	1,00	0,75	0,50	0,25
PR	1,00	1,00	0,50	0,25

\*Adaptado de Varshney et al. (2007)

Para organismos poliploides, a avaliação da diversidade genética é comprometida, sobretudo, pelos complexos padrões de herança que podem ser

encontrados e pela dificuldade ou impossibilidade de se determinar o número de cópias exatas de cada alelo (Serang et al., 2012). Na prática, algumas soluções foram propostas para minimizar este problema. Dentre elas, a mais utilizada para marcadores codominantes é a codificação dos dados com avaliação na forma de matriz binária. Os resultados dessa adaptação em estudos de diversidade genética ainda não é bem relatado. Neste trabalho, os resultados mostraram que as codificações devem ser evitadas, pois resultam em perda de informação e menor acurácia. Se extrapolarmos essa constatação para poliploides, concluímos que o problema persiste e que a codificação continua sendo um problema. O grande diferencial nestes dois casos é que a codificação em poliploides não se trata de uma escolha, mas sim de uma necessidade do pesquisador uma vez que existe a impossibilidade de se determinar o número de cópias exatas de cada alelo. Assim, faz-se necessário o desenvolvimento de metodologias robustas que resolvam este problema.

Neste sentido, algumas soluções foram propostas com destaque para o trabalho de Serang et al. (2012), que apresentaram um modelo bayesiano gráfico para genotipagem de SNP, no qual os genótipos podem ser inferidos em populações onde o nível de ploidia é desconhecido. Esses conceitos foram implementados no *software* SuperMASSA® e trata-se de uma excelente alternativa para estudos de poliploides com SNPs. Metodologias e softwares específicos para organismos poliploides também devem ser preferidos em análises genéticas. Para o caso de estudos de diversidade destacamos o FDASH (Obbard et al., 2006), TETRASAT (Markwith et al., 2006) e o ATETRA (K et al., 2010).

## **CONCLUSÃO**

(i) Em dados simulados, a transformação dos dados resultou em perda de informação para os dois cenários propostos. Assim, a codificação mostrou-se problemática em estudos com marcadores bialélicos e multialélicos, sugerindo que a transformação de dados deve ser preterida nos estudos de diversidade genética

(ii) Nos estudos com dados reais utilizando SSR, a perda de informação ao transformar os dados resulta em menor eficiência dos programas de melhoramento genético de plantas, sobretudo, naqueles que buscam a identificação de grupos heteróticos e avaliação de recursos genéticos armazenados em bancos de germoplasmas. Para os

estudos com *C. canephora*, os dados codificados não foram eficientes na discriminação de todos os subgrupos por nenhuma das abordagens consideradas (agrupamento por índices de dissimilaridade e análise bayesiana).

(iii) Para os estudos de diversidade genética utilizando diferentes classes de marcadores moleculares, ao realizar a junção dos dados, é importante que as características intrínsecas de cada ensaio sejam consideradas de modo que a informatividade, cobertura e qualidade dos marcadores sejam ponderadas. Assim, o índice de ponderação, proposto neste estudo, é uma metodologia que considera todos esses fatores, tornando-a uma ferramenta importante nos estudos genéticos.

## REFERÊNCIAS

- Belaj, A., Satovic, Z., Cipriani, G., Baldoni, L., Testolin, R., Rallo, L., and Trujillo, I. (2003). Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. *Theoretical and Applied Genetics* **107**, 736-744.
- Bonin, A., Ehrlich, D., and Manel, S. (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Mol Ecol* **16**, 3737-3758.
- Bruvo, R., Michiels, N. K., D'Souza, T. G., and Schulenburg, H. (2004). A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol Ecol* **13**, 2101-2106.
- Combes, M. C., Andrzejewski, S., Anthony, F., Bertrand, B., Rovelli, P., Graziosi, G., and Lashermes, P. (2000). Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Mol Ecol* **9**, 1178-80.
- Coulibaly, I., Revol, B., Noirot, M., Poncet, V., Lorieux, M., Carasco-Lacombe, C., Minier, J., Dufour, M., and Hamon, P. (2003). AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. heterocalyx* x *C. canephora*) x *C. canephora*]. *Theor Appl Genet* **107**, 1148-55.
- Cruz, C. D. (2006). "Programa Genes," Universidade Federal de Viçosa, Viçosa, MG.
- Cruz, C. D., Medeiros, F. F., and Pessoni, L. A. (2011). "Biometria aplicada ao estudo de diversidade genética," Viçosa, MG.
- De Silva, H. N., Hall, A. J., Rikkerink, E., McNeilage, M. A., and Fraser, L. G. (2005). Estimation of allele frequencies in polyploids under certain patterns of inheritance. *Heredity (Edinb)* **95**, 327-334.
- Dice, L. R. (1945). Measures of the Amount of Ecologic Association between Species. *Ecology* **26**, 297-302.
- Diniz, E. C., Sakiyama, N. S., Lashermes, P., Caixeta, E. T., Oliveira, A. C. B., Zambolim, E. M., Loureiro, M. E., Pereira, A. A., and Zambolim, L. (2005). Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. *Crop Breeding and Applied Biotechnology* **5**, 387-393.

- Earl, D., and vonHoldt, B. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611-20.
- Ewens, W. J., and Spielman, R. S. (1995). The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* **57**, 455-64.
- Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**, 574-578.
- Ferrão, L. F. V., Caixeta, E. T., Souza, F. d., Zambolim, E. M., Cruz, C. D., Zambolim, L., and Sakiyama, N. S. (2013). Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. *Plant Systematics and Evolution* **299**, 225-238.
- Ferrão, R. G., Ferrão, M. G. A., Fonseca, A. F. A., and Pacova, B. E. V. (2007). Melhoramento genético de *Coffea canephora*. In "Café Conilon" (R. G. Ferrão, A. F. A. Fonseca, S. M. Bragança, M. G. A. Ferrão and L. H. De Muner, eds.), Vol. 1, pp. 123-173. Incaper, Vitória, Brazil.
- Gallego, F. J., Perez, M. A., Nunez, Y., and Hidalgo, P. (2005). Comparison of RAPDs, AFLPs and SSR markers for the genetic analysis of yeast strains of *Saccharomyces cerevisiae*. *Food Microbiol* **22**, 561-568.
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Leger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., and Petit, R. J. (2011). Current trends in microsatellite genotyping. *Mol Ecol Resour* **11**, 591-611.
- Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise Des Sciences Naturelles* **44**, 223-270.
- Jaccoud, D., Peng, K., Feinstein, D., and Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* **29**, E25.
- K, V. P., A, V. G., and Triest, L. (2010). atetra, a new software program to analyse tetraploid microsatellite data: comparison with tetra and tetrasat. *Mol Ecol Resour* **10**, 331-4.
- Karp, A., Kresovich, S., Bhat, K., Ayad, W., and Hodgkin, T. (1997). "Molecular tools in plant genetic resources conservation: a guide to the technologie." International Plant Genetic Resources Institute, Rome, Italy.
- Kosman, E., and Leonard, K. J. (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Mol Ecol* **14**, 415-24.
- Lamia, K., Hedia, B., Jean-Marc, A., and Neila, T. F. (2010). Comparative analysis of genetic diversity in Tunisian apricot germplasm using AFLP and SSR markers. *Scientia Horticulturae* **127**, 54-63.
- Laurentin, H. (2009). Data analysis for molecular characterization of plant genetic resources. *Genetic Resources and Crop Evolution* **56**, 277-292.
- Leroy, T., Marraccini, P., Dufour, M., Montagnon, C., Lashermes, P., Sabau, X., Ferreira, L. P., Jourdan, I., Pot, D., Andrade, A. C., Glaszmann, J. C., Vieira, L. G. E., and Piffanelli, P. (2005). Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theoretical and Applied Genetics* **111**, 1032-1041.

- Markwith, S. H., Stewart, D. J., and Dyer, J. L. (2006). TETRASAT: a program for the population analysis of allotetraploid microsatellite data. *Molecular Ecology Notes* **6**, 586-589.
- Missio, R. F., Caixeta, E. T., Zambolim, E. M., Zambolim, E. M., Cruz, C. D., and Sakiyama, N. S. (2010). Polymorphic information content of SSR markers for *Coffea* spp. *Crop Breeding and Applied Biotechnology* **10**, 89-94.
- Mohammadi, S. A., and Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants - Salient statistical tools and considerations. *Crop Science* **43**, 1235-1248.
- Moncada, P., and McCouch, S. (2004). Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome* **47**, 501-9.
- Nei, M. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proc Natl Acad Sci U S A* **70**, 3321-3323.
- Obbard, D. J., Harris, S. A., and Pannell, J. R. (2006). Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. *Heredity (Edinb)* **97**, 296-303.
- Poncet, V., Hamon, P., Minier, J., Carasco, C., Hamon, S., and Noirot, M. (2004). SSR cross-amplification and variation within coffee trees (*Coffea* spp.). *Genome* **47**, 1071-81.
- Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., and Rafalski, A. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2**, 225-238.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-959.
- Rambaut, A. (2006). Tree Figure Drawing Tool Version 1.3.1. Institute of Evolutionary Biology, University of Edinburgh, UK.
- Ramos, H. C. C., Pereira, M. G., Goncalves, L. S. A., do Amaral, A. T., and Scapim, C. A. (2011). Comparison of multiallelic distances for the quantification of genetic diversity in the papaya. *Acta Scientiarum-Agronomy* **33**, 59-66.
- Russell, J. R., Fuller, J. D., Macaulay, M., Hatz, B. G., Jahoor, A., Powell, W., and Waugh, R. (1997). Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theoretical and Applied Genetics* **95**, 714-722.
- Serang, O., Mollinari, M., and Garcia, A. A. (2012). Efficient exact maximum a posteriori computation for bayesian SNP genotyping in polyploids. *PLoS One* **7**, e30906.
- Smouse, P. E., and Peakall, R. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity (Edinb)* **82 ( Pt 5)**, 561-73.
- Sokal, R. R., and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* **38**, 1409-1438.
- Souframanien, J., and Gopalakrishna, T. (2004). A comparative analysis of genetic diversity in blackgram genotypes using RAPD and ISSR markers. *Theor Appl Genet* **109**, 1687-93.
- Souza, F. F. d. (2011). Estudos sobre a diversidade, estrutura populacional, desequilíbrio de ligação e mapeamento associativo em *Coffea canephora* Pierre ex Froehner, Universidade Federal de Viçosa, Viçosa, MG.
- Varshney, R. K., Chabane, K., Hendre, P. S., Aggarwal, R. K., and Graner, A. (2007). Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. *Plant Science* **173**, 638-649.

- Wright, S. (1965). The Interpretation of Population-Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution* **19**, 395-420.
- Wright, S. (1978). "Evolution and the genetics of populations," Univ. Chicago Press, Chicago, USA.

## ANEXO 1

**Anexo 1.** Coeficiente de ancestralidade (Q) dos oitenta e sete acessos de *Coffea canephora* avaliados para as metodologias de codificação *Cod*, *Bin* e *Dom*, considerando valor de  $k=4$  no software STRUCTURE. A classificação em subgrupos (SG) foi feita considerando os maiores valores de Q (em cinza no gráfico) e a análise dos dendrogramas da Figura 2 (apresentada no capítulo). Genótipos do grupo varietal Conilon foram classificados em seus respectivos subgrupos, de acordo com a localidade (RO.1, RO.2 e ES), os Robustas foram denominados de ROB e os híbridos naturais HIB.

Metodologias de codificações																	
<i>Cod</i>						<i>Bin</i>						<i>Dom</i>					
ID*	I	II	III	IV	SG	ID	I	II	III	IV	SG	ID	I	II	III	IV	SG
ES03	0.003	0.137	0.067	0.793	ES	ES03	0.002	0.01	0.057	0.93	RO.1	ES03	0.005	0.141	0.567	0.287	Conilon
ES110A	0.001	0.224	0.297	0.478	ES	ES110A	0.001	0.003	0.419	0.576	RO.1	ES110A	0.005	0.386	0.475	0.134	Conilon
ES028	0.001	0.829	0.03	0.14	ES	ES028	0.013	0.062	0.1	0.825	RO.1	ES028	0.013	0.056	0.097	0.834	Conilon
ES16	0.005	0.897	0.013	0.085	ES	ES16	0.024	0.206	0.008	0.762	RO.1	ES16	0.025	0.019	0.118	0.838	Conilon
ES45	0.002	0.18	0.006	0.813	RO.1	ES45	0.002	0.003	0.005	0.991	RO.1	ES45	0.012	0.053	0.602	0.333	Conilon
K661	0.004	0.974	0.01	0.012	ES	K661	0.026	0.381	0.018	0.575	RO.1	K661	0.035	0.025	0.064	0.876	Conilon
ESV1	0.001	0.247	0.057	0.695	ES	ESV1	0.001	0.004	0.038	0.957	RO.1	ESV1	0.003	0.068	0.259	0.67	Conilon
ESV2	0.003	0.752	0.051	0.194	ES	ESV2	0.011	0.134	0.042	0.813	RO.1	ESV2	0.021	0.061	0.457	0.461	Conilon
ESV4	0.003	0.28	0.035	0.681	ES	ESV4	0.009	0.134	0.033	0.824	RO.1	ESV4	0.031	0.104	0.413	0.452	Conilon
R1675	0.986	0.002	0.009	0.003	ROB	R1675	0.526	0.466	0.006	0.002	ROB	R1675	0.935	0.02	0.015	0.03	ROB
R2259	0.994	0.002	0.003	0.001	ROB	R2259	0.985	0.012	0.002	0.001	ROB	R2259	0.986	0.004	0.004	0.006	ROB
R22571	0.622	0.177	0.007	0.194	ROB	R22571	0.093	0.863	0.005	0.039	ROB	R22571	0.454	0.031	0.091	0.424	ROB
R22572	0.993	0.001	0.005	0.001	ROB	R22572	0.983	0.006	0.008	0.003	ROB	R22572	0.952	0.026	0.009	0.012	ROB
R6401	0.95	0.012	0.01	0.028	ROB	R6401	0.876	0.109	0.006	0.008	ROB	R6401	0.94	0.017	0.025	0.018	ROB
R6402	0.996	0.002	0.001	0.001	ROB	R6402	0.97	0.025	0.003	0.002	ROB	R6402	0.991	0.003	0.003	0.003	ROB
R22581	0.997	0.001	0.001	0.001	ROB	R22581	0.992	0.005	0.001	0.001	ROB	R22581	0.989	0.004	0.003	0.004	ROB
R22582	0.996	0.001	0.001	0.001	ROB	R22582	0.988	0.009	0.002	0.002	ROB	R22582	0.988	0.003	0.004	0.005	ROB
R22583	0.997	0.001	0.001	0.001	ROB	R22583	0.996	0.002	0.001	0.001	ROB	R22583	0.991	0.003	0.003	0.003	ROB
RO006	0.001	0.672	0.279	0.047	RO.2	RO006	0.021	0.145	0.063	0.771	RO.1	RO006	0.008	0.126	0.243	0.623	Conilon
RO047	0.002	0.392	0.044	0.561	RO.1	RO047	0.002	0.027	0.02	0.951	RO.1	RO047	0.006	0.396	0.233	0.364	Conilon
RO199	0.639	0.012	0.181	0.168	HIB	RO199	0.08	0.842	0.044	0.034	HIB	RO199	0.049	0.022	0.077	0.852	Conilon
RO052	0.001	0.008	0.006	0.984	RO.1	RO052	0.001	0.002	0.005	0.992	RO.1	RO052	0.004	0.132	0.768	0.096	Conilon
RO151	0.001	0.008	0.006	0.984	RO.1	RO151	0.001	0.002	0.007	0.99	RO.1	RO151	0.007	0.059	0.82	0.114	Conilon
RO172	0.001	0.004	0.005	0.991	RO.1	RO172	0.001	0.001	0.011	0.987	RO.1	RO172	0.004	0.102	0.822	0.072	Conilon
RO003	0.005	0.005	0.011	0.98	RO.1	RO003	0.005	0.005	0.018	0.972	RO.1	RO003	0.014	0.193	0.645	0.148	Conilon
RO058	0.004	0.021	0.134	0.841	RO.1	RO058	0.003	0.003	0.064	0.931	RO.1	RO058	0.003	0.138	0.822	0.038	Conilon

<b>RO059</b>	0.001	0.009	0.017	0.973	RO.1	<b>RO059</b>	0.001	0.002	0.031	0.966	RO.1	<b>RO059</b>	0.005	0.481	0.44	0.075	Conilon
<b>RO088</b>	0.001	0.003	0.008	0.988	RO.1	<b>RO088</b>	0.001	0.001	0.009	0.989	RO.1	<b>RO088</b>	0.004	0.098	0.85	0.048	Conilon
<b>RO096</b>	0.001	0.013	0.058	0.928	RO.1	<b>RO096</b>	0.001	0.002	0.1	0.897	RO.1	<b>RO096</b>	0.005	0.327	0.569	0.099	Conilon
<b>RO183</b>	0.001	0.006	0.005	0.989	RO.1	<b>RO183</b>	0.001	0.002	0.005	0.992	RO.1	<b>RO183</b>	0.007	0.086	0.765	0.142	Conilon
<b>RO015</b>	0.001	0.005	0.042	0.952	RO.1	<b>RO015</b>	0.001	0.001	0.086	0.912	RO.1	<b>RO015</b>	0.006	0.309	0.475	0.21	Conilon
<b>RO140</b>	0.001	0.006	0.005	0.987	RO.1	<b>RO140</b>	0.002	0.002	0.008	0.988	RO.1	<b>RO140</b>	0.006	0.179	0.583	0.233	Conilon
<b>RO001</b>	0.002	0.011	0.12	0.867	RO.1	<b>RO001</b>	0.001	0.002	0.11	0.886	RO.1	<b>RO001</b>	0.005	0.333	0.607	0.055	Conilon
<b>RO016</b>	0.006	0.016	0.048	0.931	RO.1	<b>RO016</b>	0.002	0.008	0.036	0.954	RO.1	<b>RO016</b>	0.003	0.129	0.844	0.024	Conilon
<b>RO044</b>	0.001	0.009	0.521	0.469	RO.1	<b>RO044</b>	0.001	0.002	0.648	0.349	RO.1	<b>RO044</b>	0.002	0.243	0.735	0.02	Conilon
<b>RO101</b>	0.002	0.004	0.15	0.844	RO.1	<b>RO101</b>	0.001	0.003	0.323	0.674	RO.1	<b>RO101</b>	0.004	0.064	0.868	0.064	Conilon
<b>RO119</b>	0.001	0.003	0.004	0.992	RO.1	<b>RO119</b>	0.001	0.001	0.006	0.992	RO.1	<b>RO119</b>	0.003	0.053	0.91	0.034	Conilon
<b>RO155</b>	0.002	0.035	0.016	0.947	RO.1	<b>RO155</b>	0.004	0.013	0.049	0.933	RO.1	<b>RO155</b>	0.03	0.112	0.744	0.114	Conilon
<b>RO004</b>	0.002	0.129	0.621	0.248	RO.1	<b>RO004</b>	0.005	0.022	0.582	0.392	RO.1	<b>RO004</b>	0.011	0.067	0.325	0.597	Conilon
<b>RO042</b>	0.001	0.004	0.136	0.859	RO.1	<b>RO042</b>	0.001	0.002	0.19	0.807	RO.1	<b>RO042</b>	0.005	0.425	0.514	0.056	Conilon
<b>RO098</b>	0.001	0.006	0.006	0.987	RO.1	<b>RO098</b>	0.001	0.002	0.015	0.982	RO.1	<b>RO098</b>	0.003	0.195	0.688	0.115	Conilon
<b>RO160</b>	0.001	0.002	0.006	0.991	RO.1	<b>RO160</b>	0.001	0.002	0.016	0.981	RO.1	<b>RO160</b>	0.003	0.097	0.885	0.015	Conilon
<b>RO184</b>	0.001	0.005	0.005	0.988	RO.1	<b>RO184</b>	0.001	0.001	0.005	0.993	RO.1	<b>RO184</b>	0.003	0.089	0.867	0.04	Conilon
<b>RO018</b>	0.001	0.003	0.139	0.856	RO.1	<b>RO018</b>	0.004	0.006	0.212	0.779	RO.1	<b>RO018</b>	0.008	0.379	0.049	0.563	Conilon
<b>RO045</b>	0.001	0.003	0.015	0.981	RO.1	<b>RO045</b>	0.001	0.001	0.029	0.969	RO.1	<b>RO045</b>	0.003	0.08	0.89	0.027	Conilon
<b>RO146</b>	0.001	0.002	0.007	0.99	RO.1	<b>RO146</b>	0.001	0.001	0.017	0.98	RO.1	<b>RO146</b>	0.003	0.074	0.906	0.016	Conilon
<b>RO194</b>	0.53	0.007	0.048	0.415	HIB	<b>RO194</b>	0.015	0.929	0.019	0.037	HIB	<b>RO194</b>	0.131	0.046	0.11	0.714	Conilon
<b>RO017</b>	0.001	0.01	0.848	0.141	RO.2	<b>RO017</b>	0.001	0.001	0.897	0.101	RO.2	<b>RO017</b>	0.004	0.834	0.109	0.053	Conilon
<b>RO043</b>	0.001	0.003	0.983	0.013	RO.2	<b>RO043</b>	0.001	0.001	0.979	0.019	RO.2	<b>RO043</b>	0.003	0.316	0.658	0.023	Conilon
<b>RO120</b>	0.001	0.002	0.986	0.011	RO.2	<b>RO120</b>	0.001	0.002	0.989	0.009	RO.2	<b>RO120</b>	0.003	0.824	0.138	0.035	Conilon
<b>RO010</b>	0.001	0.003	0.989	0.006	RO.2	<b>RO010</b>	0.001	0.002	0.98	0.017	RO.2	<b>RO010</b>	0.003	0.29	0.576	0.131	Conilon
<b>RO064</b>	0.005	0.015	0.973	0.006	RO.2	<b>RO064</b>	0.014	0.018	0.959	0.009	RO.2	<b>RO064</b>	0.011	0.697	0.211	0.081	Conilon
<b>RO086</b>	0.004	0.004	0.988	0.005	RO.2	<b>RO086</b>	0.003	0.004	0.985	0.008	RO.2	<b>RO086</b>	0.004	0.781	0.032	0.183	Conilon
<b>RO103</b>	0.001	0.007	0.963	0.029	RO.2	<b>RO103</b>	0.001	0.002	0.956	0.041	RO.2	<b>RO103</b>	0.005	0.465	0.408	0.121	Conilon
<b>RO203</b>	0.208	0.025	0.762	0.004	RO.2	<b>RO203</b>	0.014	0.304	0.667	0.015	RO.2	<b>RO203</b>	0.006	0.321	0.058	0.614	Conilon
<b>RO022</b>	0.002	0.005	0.989	0.004	RO.2	<b>RO022</b>	0.002	0.008	0.985	0.005	RO.2	<b>RO022</b>	0.019	0.427	0.037	0.516	Conilon
<b>RO024</b>	0.002	0.014	0.974	0.01	RO.2	<b>RO024</b>	0.002	0.01	0.941	0.047	RO.2	<b>RO024</b>	0.008	0.423	0.044	0.526	Conilon
<b>RO127</b>	0.002	0.005	0.986	0.006	RO.2	<b>RO127</b>	0.002	0.003	0.991	0.005	RO.2	<b>RO127</b>	0.003	0.461	0.513	0.023	Conilon
<b>RO147</b>	0.002	0.094	0.803	0.101	RO.2	<b>RO147</b>	0.001	0.003	0.914	0.082	RO.2	<b>RO147</b>	0.008	0.464	0.289	0.239	Conilon
<b>RO164</b>	0.006	0.002	0.969	0.023	RO.2	<b>RO164</b>	0.004	0.007	0.96	0.029	RO.2	<b>RO164</b>	0.003	0.311	0.656	0.03	Conilon
<b>RO189</b>	0.003	0.076	0.837	0.084	RO.2	<b>RO189</b>	0.002	0.003	0.918	0.076	RO.2	<b>RO189</b>	0.009	0.59	0.292	0.109	Conilon
<b>RO190</b>	0.934	0.015	0.014	0.037	<b>ROB</b>	<b>RO190</b>	0.463	0.526	0.006	0.005	<b>ROB</b>	<b>RO190</b>	0.885	0.031	0.033	0.051	<b>ROB</b>

<b>RO036</b>	0.001	0.004	0.984	0.011	RO.2	<b>RO036</b>	0.001	0.001	0.985	0.012	RO.2	<b>RO036</b>	0.005	0.779	0.157	0.06	Conilon
<b>RO089</b>	0.001	0.004	0.971	0.023	RO.2	<b>RO089</b>	0.001	0.002	0.957	0.04	RO.2	<b>RO089</b>	0.003	0.52	0.405	0.072	Conilon
<b>RO045B</b>	0.001	0.002	0.987	0.011	RO.2	<b>RO045B</b>	0.001	0.001	0.983	0.015	RO.2	<b>RO045B</b>	0.002	0.708	0.253	0.036	Conilon
<b>RO077</b>	0.002	0.008	0.984	0.006	RO.2	<b>RO077</b>	0.002	0.003	0.991	0.004	RO.2	<b>RO077</b>	0.005	0.804	0.124	0.067	Conilon
<b>RO138</b>	0.002	0.004	0.965	0.028	RO.2	<b>RO138</b>	0.003	0.005	0.964	0.028	RO.2	<b>RO138</b>	0.01	0.538	0.375	0.077	Conilon
<b>RO142</b>	0.002	0.008	0.938	0.052	RO.2	<b>RO142</b>	0.001	0.003	0.948	0.048	RO.2	<b>RO142</b>	0.002	0.421	0.56	0.016	Conilon
<b>RO196</b>	0.002	0.008	0.984	0.007	RO.2	<b>RO196</b>	0.002	0.004	0.982	0.012	RO.2	<b>RO196</b>	0.01	0.42	0.421	0.15	Conilon
<b>RO193</b>	0.64	0.039	0.297	0.023	HIB	<b>RO193</b>	0.017	0.953	0.022	0.008	HIB	<b>RO193</b>	0.19	0.095	0.145	0.57	Conilon
<b>RO049</b>	0.003	0.011	0.892	0.095	RO.2	<b>RO049</b>	0.001	0.003	0.938	0.058	RO.2	<b>RO049</b>	0.004	0.686	0.264	0.047	Conilon
<b>RO030</b>	0.002	0.004	0.982	0.012	RO.2	<b>RO030</b>	0.002	0.002	0.986	0.01	RO.2	<b>RO030</b>	0.007	0.583	0.386	0.023	Conilon
<b>RO032</b>	0.002	0.002	0.988	0.008	RO.2	<b>RO032</b>	0.001	0.002	0.992	0.005	RO.2	<b>RO032</b>	0.012	0.887	0.045	0.056	Conilon
<b>RO076</b>	0.01	0.043	0.903	0.045	RO.2	<b>RO076</b>	0.004	0.017	0.905	0.074	RO.2	<b>RO076</b>	0.003	0.665	0.297	0.035	Conilon
<b>RO161</b>	0.001	0.002	0.98	0.016	RO.2	<b>RO161</b>	0.001	0.001	0.981	0.017	RO.2	<b>RO161</b>	0.002	0.484	0.499	0.015	Conilon
<b>RO035</b>	0.001	0.005	0.989	0.005	RO.2	<b>RO035</b>	0.001	0.002	0.992	0.005	RO.2	<b>RO035</b>	0.007	0.865	0.061	0.067	Conilon
<b>RO038</b>	0.003	0.004	0.986	0.007	RO.2	<b>RO038</b>	0.001	0.003	0.978	0.017	RO.2	<b>RO038</b>	0.004	0.71	0.161	0.124	Conilon
<b>RO073</b>	0.001	0.004	0.962	0.033	RO.2	<b>RO073</b>	0.002	0.003	0.936	0.059	RO.2	<b>RO073</b>	0.013	0.699	0.141	0.147	Conilon
<b>RO139</b>	0.001	0.002	0.991	0.006	RO.2	<b>RO139</b>	0.001	0.002	0.987	0.01	RO.2	<b>RO139</b>	0.005	0.904	0.025	0.065	Conilon
<b>RO141</b>	0.001	0.002	0.981	0.016	RO.2	<b>RO141</b>	0.001	0.002	0.975	0.023	RO.2	<b>RO141</b>	0.004	0.543	0.412	0.041	Conilon
<b>RO171</b>	0.002	0.006	0.988	0.005	RO.2	<b>RO171</b>	0.002	0.005	0.988	0.005	RO.2	<b>RO171</b>	0.008	0.308	0.025	0.66	Conilon
<b>RO197</b>	0.002	0.039	0.943	0.015	RO.2	<b>RO197</b>	0.003	0.007	0.918	0.073	RO.2	<b>RO197</b>	0.006	0.231	0.076	0.687	Conilon
<b>RO025</b>	0.001	0.003	0.954	0.042	RO.2	<b>RO025</b>	0.001	0.002	0.979	0.018	RO.2	<b>RO025</b>	0.004	0.726	0.03	0.24	Conilon
<b>RO026</b>	0.003	0.01	0.938	0.05	RO.2	<b>RO026</b>	0.001	0.004	0.946	0.049	RO.2	<b>RO026</b>	0.005	0.658	0.253	0.085	Conilon
<b>RO072</b>	0.004	0.019	0.964	0.013	RO.2	<b>RO072</b>	0.003	0.007	0.982	0.009	RO.2	<b>RO072</b>	0.002	0.869	0.055	0.073	Conilon
<b>RO075</b>	0.005	0.003	0.987	0.005	RO.2	<b>RO075</b>	0.004	0.036	0.956	0.004	RO.2	<b>RO075</b>	0.027	0.423	0.039	0.512	Conilon
<b>RO115</b>	0.004	0.02	0.966	0.011	RO.2	<b>RO115</b>	0.005	0.006	0.974	0.016	RO.2	<b>RO115</b>	0.01	0.614	0.037	0.339	Conilon

\*Adaptado de Souza (2011) e Ferrão et al. (2013).

## 5. CAPÍTULO 2: ANÁLISES GENÔMICAS E COMPONENTES DE RESISTÊNCIA DE *Coffea canephora* à *Hemileia vastatrix*.

### RESUMO

Pertencente ao gênero *Coffea*, a espécie *Coffea canephora*, é a segunda mais cultivada no mundo e destaca-se pelo papel relevante no desenvolvimento social e econômico nos países produtores. Atualmente, o Brasil é o segundo maior produtor de grãos da espécie, sendo o Estado do Espírito Santo o maior produtor nacional. Dentre os cafeeiros cultivados no Estado, a grande maioria, são variedades clonais que dentre as vantagens, agregam elevada produtividade, estabilidade e, na maioria das vezes, resistência vertical à ferrugem (*Hemileia vastatrix*), principal doença foliar do cafeeiro. Na concepção patogênica, esse tipo de resistência em variedades clonais é considerado um risco, pois é uma característica determinada por poucos genes o que aumenta as chances da resistência ser suplantada. Assim, uma alternativa para aumentar a durabilidade da resistência seria combinar genes de efeito maior (resistência vertical) e genes de efeito menor (resistência horizontal). Neste contexto, o uso de técnicas moleculares merece destaque, pois possuem potencial para acelerar o desenvolvimento de variedades que agreguem características agrônômicas de interesse. Com isso, o presente estudo objetivou identificar fatores genéticos associados à resistência horizontal para *H. vastatrix* em híbridos de *C. canephora*, utilizando análises genômicas. Para isso, a fenotipagem foi realizada considerando dados de campo, coletados em dois ambientes pelo programa de melhoramento do Incaper (Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural); e inoculações controladas realizadas no Laboratório de Fitopatologia da Universidade Federal de Viçosa (UFV) para com a raça II e XXXIII de *H. vastatrix*. Esses dados foram associados às informações de genotipagem utilizando a metodologia de mapeamento de QTLs. A análise dos resultados fenotípicos mostrou a existência de expressão diferencial dos níveis de doença para as duas localidades em avaliação, o que resultou na seleção de genótipos resistentes, potencialmente úteis na composição de novas variedades ou formação de populações base para ciclos de seleção recorrente. Os dados genotípicos foram utilizados na construção do mapa de ligação genético e, posteriormente, na análise QTL. Regiões genômicas associadas à resistência de dados de campo e de laboratório foram identificados, abrindo novas perspectivas para os programas de seleção assistida

(SAM) e, futuramente, na clonagem posicional. Este é o primeiro relato de identificação de QTLs associados a resistência à *H. vastatrix* em *C. canephora*.

**Palavras chaves:** Análise de QTLs; café robusta; marcadores moleculares; SSR.

## INTRODUÇÃO

Com origem na África, o cafeeiro pertence à família Rubiaceae, que contém aproximadamente 500 gêneros e mais de seis mil espécies. Dentre todos estes gêneros, particular atenção tem sido dada ao gênero *Coffea*, ao qual se inclui duas espécies de grande importância econômica: *Coffea arabica* L. e *Coffea canephora* Pierre ex Froehner. Praticamente todo o café produzido e consumido no mundo é originado de variedades dessas duas espécies, que são conhecidas popularmente como café arábica e robusta, respectivamente. O café arábica, é a espécie mais cultivada e comercializada no mundo, é alotetraplóide ( $2n=4x=44$ ), autógama, adaptada aos locais de altitudes elevadas e clima ameno. Os cafeeiros da espécie *C. canephora* são diplóides ( $2n=2x=22$ ), com autoincompatibilidade genética do tipo gametofítica, reproduzem-se por fecundação cruzada e são adaptados a regiões de altitudes mais baixas e temperaturas mais elevadas (Berthaud and Charrier, 1988; Lashermes et al., 1997)

Tradicionalmente, a produção mundial de café concentrava-se apenas na espécie *C. arabica*, devido à qualidade dos grãos que agregavam aroma e sabor adocicado. No entanto, com a ascensão do café solúvel na década de 50, a espécie *C. canephora* passou a ser explorada comercialmente, sobretudo, por apresentar menor acidez e maior quantidade de sólidos solúveis, o que dava corpo aos *blends* (misturas de grãos das duas espécies) e maior eficiência nas indústrias de torrado e moído (Ferrão et al., 2007).

O resultado da comercialização e industrialização de *C. canephora* foi o progressivo aumento na produção mundial, liderado principalmente por países como Vietnã, Brasil e Indonésia. Atualmente, a espécie representa cerca de 36% de todo o café produzido e comercializado no mundo e possui papel de destaque no desenvolvimento econômico e social dos países produtores. Além da importância econômica já citada, a espécie *C. canephora* caracteriza-se por ser uma planta rústica, de alta produtividade, maior teor de cafeína e menos susceptíveis a fatores bióticos e abióticos. Neste sentido, a espécie tem sido alvo de estudos desde o século XIX, especialmente, por apresentarem maiores níveis de resistência à ferrugem, causada pelo

fungo *Hemileia vastatrix*. Na cafeicultura do arábica, de maior susceptibilidade ao fungo, estima-se que a doença seja responsável pela redução de 35 -50% da produção de grãos anuais, gerando perda na ordem de bilhões de dólares (Rozo et al., 2012; Zambolim et al., 2005) .

Em café, a resistência à ferrugem segue o conceito gene-a-gene proposto por Flor, que é caracterizado pela interação entre genes avirulentos (*Avr*) do patógeno e genes de resistência (*R*) do hospedeiro (Eskes, 1983). Estudos específicos em cafeeiro determinaram que a resistência as raças fisiológicas de *H. vastatrix* são condicionados por nove genes dominantes ( $S_H1$  a  $S_H9$ ), simples ou associados, que foram identificadas nas espécies *C. arabica*, *C. canephora* e *C. liberica* (Bettencourt et al., 1992; Bettencourt and Rodrigues, 1988; Prakash et al., 2005; Wagner and Bettencourt, 1965). São relatados cafeeiros totalmente resistentes como também susceptíveis (resistência vertical), além de níveis intermediários de doença (resistência horizontal), o que caracteriza a existência das duas categorias de resistência, segundo classificação de Van der Plank (1975) (Cadena, 2005; Capucho, 2012). Em programas de melhoramento genético, a resistência vertical tem sido preferencialmente focada, sobretudo, pela aceitação imediata dos cafeicultores.

No Brasil, o Estado do Espírito Santo é o maior produtor nacional de grãos de *C. canephora* e a grande maioria das lavouras plantadas nas últimas décadas são provenientes de variedades clonais (Ferrão et al., 2007). Em termos genéticos a utilização da propagação vegetativa nos programas de melhoramento significa o aproveitamento integral da variância genética, o que possibilita a fixação de caracteres de interesse; já em termos práticos, essa opção, representa economia de tempo e de áreas experimentais, o que é menos custoso. Além disso, o desenvolvimento de variedades policlonais é de grande interesse também dos produtores, visto que o agrupamento de genótipos com diferentes épocas de maturação permite o escalonamento da colheita (Ferrão, 2004; Ferrão et al., 2007). No entanto, na concepção de fitopatologia, o uso de materiais clonais com resistência vertical é um risco, pois se trata de uma característica determinada por poucos genes, o que aumenta as chances da resistência ser suplantada (Samils et al., 2011). Exemplos de variedades que tiveram sua resistência suplantada já foram descritas no gênero *Coffea* (Várzea et al., 2002). Assim, uma alternativa para aumentar a durabilidade da resistência seria combinar genes de efeito maior (resistência vertical) e genes de efeito menor (resistência horizontal)

(Cadena, 2005; Jorge et al., 2005). Apesar de ser uma alternativa interessante, a introgressão de vários genes em uma única variedade (piramidação de genes) não é um processo fácil, especialmente, para o cafeeiro que é uma planta perene e de período juvenil longo.

Neste contexto, os avanços das técnicas moleculares merecem destaque, pois possuem grande potencial para acelerar o desenvolvimento de variedades resistentes. Para isso, o mapeamento de QTLs (*Quantitative trait loci*) é uma ferramenta eficiente, pois além de auxiliar nos estudos da evolução, organização e estrutura genética, permite compreender como o caráter está atuando através das estimativas de posição, efeito e interações entre as diferentes regiões genômicas (Mackay et al., 2009). Atualmente, poucos trabalhos deste tipo encontram-se relatados na literatura para *C. canephora*, sendo que, nenhum deles é focado para a identificação de fatores genéticos associados à ferrugem. Dentre os genes/QTLs identificados na espécie, destacam-se aqueles para os locos S da incompatibilidade (Lashermes et al., 1996); características morfológicas (Amidou et al., 2007; Souza, 2011); capacidade de embriogênese somática (Priyono et al., 2010); e, mais recentemente, qualidade de bebida (Leroy et al., 2011).

Para o caso específico de resistência a doença o mapeamento de QTLs é uma metodologia particularmente adequada, sobretudo, para estudos de resistência quantitativa na qual se assume a existência de vários genes governando uma característica (Samils et al., 2011). Assim, a compreensão de como atua o caráter de interesse (fenótipo) e sua associação com marcadores moleculares (genótipo) é o primeiro passo para o estabelecimento de programas de seleção assistida por marcadores (MAS), que promete acelerar o desenvolvimento de variedades e piramidar genes de resistência para a ferrugem. Nesse sentido, este trabalho tem como objetivo (i) auxiliar no entendimento dos componentes de resistência à *H. vastatrix*; (ii) fornecer subsídios para identificação e seleção de genótipos resistentes, considerando diferentes métodos de avaliação fenotípica; (iii) construir um mapa de ligação genético para posicionamento de marcadores moleculares; e (iv) analisar o número de genes e/ou QTLs associados a resistência de *C. canephora* à *H. vastatrix*, considerando dados fenotípicos de campo e de laboratório.

## MATERIAIS E MÉTODOS

### *Material Vegetal*

Para os estudos de resistência à ferrugem, uma família de 126 irmãos completos (*full-sib*), derivada do cruzamento entre os clones 02 e 83 de *C. canephora*, foi utilizada. Essa população foi desenvolvida e mantida pelo Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper), no Estado do Espírito Santo, Brasil. Dentre os mais de 1.000 materiais genéticos mantidos na Instituição os clones 02 e 83 foram selecionados, pois apresentam elevada produtividade, estabilidade de produção, bom porte e, especialmente, contraste quanto a resistência a ferrugem. Os experimentos de campo foram conduzidos pelo Incaper e instalados nas Fazendas Experimentais de Marilândia (FEM) e de Sooretama (FES), Espírito Santo, Brasil; que possuem, respectivamente, 74 e 50 das progênies.

Os locais dos experimentos apresentam as seguintes características agroclimáticas: Sooretama (FES) situa-se na latitude de 15°47' sul e longitude de 43°18' oeste e altitude de 40 m; o solo é classificado como Latossolo Vermelho-Amarelo distrófico arenoso (80% de areia) e de baixa fertilidade; a precipitação pluviométrica média anual é de 1.200 mm e mal distribuída; a temperatura média anual é de 24 °C; a umidade relativa média do ar é de 80%; e a topografia plana com predominância de ventos sul. Marilândia (FEM) situa-se na latitude de 19°24' sul e longitude de 40°31' oeste, a uma altitude de 70 m; o solo é classificado como cristalino e de baixa fertilidade; a precipitação pluviométrica anual é de 1.100 mm; a temperatura média anual é de 24 °C; a umidade relativa média do ar é de 74%; e a topografia é ondulada acidentada. Em ambos os locais, os meses de janeiro, novembro e dezembro são úmidos, enquanto março, abril e outubro são parcialmente úmidos. Os meses mais secos são os de maio, junho, julho, agosto e setembro.

### *Coleta dos dados fenotípicos*

**Dados de campo.** A ferrugem foi avaliada visualmente considerando uma escala de severidade da doença, descrita por Eskes e Toma-Braghini (1981), com modificações. Nesta metodologia, a doença foi avaliada na forma de notas, que variaram de 1 a 9 de acordo com os sintomas observados na região foliar (Tabela 1). Assim, os genótipos com maiores notas são susceptíveis, enquanto que os resistentes receberam notas menores. As avaliações de campo foram realizadas nos anos de 2008, 2009 e 2010 nas

localidades de FEM e FES e a classificação em susceptível e resistente foi realizada considerando a maior nota dos genótipos nesses anos de avaliação. Os dados de campo foram cedidos pelo programa de melhoramento de café do Incaper (Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural).

**Tabela 2.** Escala de severidade (notas) de doença para dados de campo\*.

Notas	Descrição	Notas	Descrição
1	Nenhum sinal visível da infecção.	6	Semelhante a 5, mas com produção de esporos e esporulação de cerca de 75% das lesões.
2	Pequenos pontos cloróticos, frequentemente associados com pequenas tumefações e sem produção de uredosporos.	7	Semelhante a 6, mas com esporulação de praticamente todas as lesões e com abundante produção de esporos em muitas delas.
3	Pequenos e médios pontos cloróticos, frequentemente associados com tumefações e sem produção de uredosporos.	8	Mistura de lesões com graus variáveis de esporulação ou um nível médio de esporulação em todas as lesões, às vezes associadas com poucas tumefações.
4	Mistura de pontos cloróticos de vários tamanhos, com alguma formação de uredosporos nas maiores lesões e esporulação de no máximo 25% das lesões.	9	Lesões com abundante esporulação, sem clorose nas bordas.
5	Semelhante a nota 4, porém com mais esporos produzidos por lesão e esporulação de cerca de 50% das lesões.		

\* Eskes e Toma-Braghini (1981), com modificações.

**Avaliação dos componentes de resistência.** As avaliações das inoculações em laboratório foram realizadas com o objetivo de se ter maior acurácia nos dados fenotípicos, além de controle do patógeno. Para isso, avaliaram-se os componentes de resistência dos genitores e das 126 progênies de irmãos completos para as raças II e XXXIII. A raça II foi escolhida, pois é a mais amplamente distribuída no Brasil (Cabral et al., 2009); enquanto que a raça XXXIII foi isolada mais recentemente e pode representar um risco para áreas cultivadas (Capucho et al., 2012a). A avaliação da severidade da doença foi realizada conforme metodologia de discos de folha de Eskes (1982) descrita por Neto (2011) e Capucho (2012). O *software* QUANT (Vale et al., 2003) foi utilizado para estimar a área esporulada e os seguintes componentes de resistência foram avaliados: (i) Período de Incubação (PI): período (dias) entre a inoculação até a observação de 50% da incidência final de discos com sintomas; (ii)

Período Latente (PL): período (dias) entre a inoculação até a observação de 50% da incidência final de discos com lesões esporulantes; **(iii)** Porcentagem de Discos com Sintomas (DS): porcentagem final de discos com sintomas visíveis da doença; **(iv)** Porcentagem de Discos com Esporulação (DE): porcentagem final de discos contendo lesões com esporulação; **(v)** Porcentagem de Área Esporulada (AE): porcentagem final da área foliar com lesões esporulantes; **(vi)** Produção de Uredósporos por Área (EAT): quantidade final de uredosporos produzidos por área inoculada (uredosporos/cm<sup>2</sup>). Os dados de laboratório foram oriundos de um trabalho em parceria entre o Incaper e a Universidade Federal de Viçosa (UFV) e cedidos pelo Departamento de Fitopatologia da Universidade.

#### *Análises Moleculares*

**Extração de DNA e genotipagem utilizando SSRs.** Folhas jovens e completamente desenvolvidas de cada progênie de irmãos completos foram coletadas nos campos experimentais do Incaper, congeladas a -80 °C, liofilizadas, trituradas e conservadas a -20 °C no laboratório de Biotecnologia do Cafeeiro (BioCafé/Bioagro) na UFV. O DNA genômico foi extraído conforme protocolo descrito por Diniz et al. (2005). Para a genotipagem, foi realizado uma triagem de marcadores SSRs polimórficos. Para isso, SSRs genômicos descritos na literatura (Baruah et al., 2003; Bhat et al., 2005; Coulibaly et al., 2003; Leroy et al., 2005; Moncada and McCouch, 2004; Poncet et al., 2004), SSRs desenvolvidos para *C. arabica* (Missio et al., 2009) e EST-SSRs minerados do Projeto Brasileiro do Genoma Café (Capítulo 3) foram testados quanto ao polimorfismo, padrão de segregação e qualidade dos produtos de amplificação. A amplificação dos microssatélites foi realizada conforme relatado por Missio et al. (2010).

#### *Análises estatísticas*

**Dados Fenotípicos.** A correlação entre os componentes de resistência mensurados em laboratório foram calculados considerando o coeficiente de correlação de Pearson e a significância foi testada pelo teste *t*, considerando 5% e 1% de probabilidade. A amplitude dos dados e observação da existência de segregação transgressiva foi realizada considerando os valores codificados de componentes de resistência. A análise

multivariada de componentes principais (PCA) teve como objetivo resumir as informações dos seis componentes de resistência avaliados, de modo a facilitar a seleção de genótipos resistentes. Todos esses procedimentos foram realizados pelo *software* GENES (Cruz, 2006).

**Dados Moleculares.** Após a leitura dos locos polimórficos, foi feita a verificação do padrão de segregação de cada loco, de forma a observar a ocorrência de distorção. Para isso, os locos foram submetidos ao teste de aderência qui-quadrado ( $X^2$ ) e visando manter um nível de significância  $\alpha$  para o experimento, utilizou-se a proteção de Bonferroni. Para construção do mapa de ligação os marcadores moleculares foram integrados em grupos de ligação por intermédio do programa computacional GQMOL (<http://www.ufv.br/dbg/gqmol/gqmol.htm>). Este programa permite análise conjunta de marcadores com diferentes padrões de segregação (1:1; 1:2:1 e 1:1:1:1), para construção do mapa. A classificação desses padrões de segregação foi feito de acordo com Wu et al. (2002). Os grupos de ligação foram formados utilizando LOD 2 e máximo de 40% de recombinação. A representação gráfica do mapa genético foi feita utilizando o *software* MapChart (Voorrips, 2002).

**Identificação de QTLs.** O mapeamento dos QTLs foi realizado considerando os dados fenotípicos de campo e os componentes de resistência PL e EAT, devido à importância biológica. Para os dados de campo a marcação fenotípica utilizada nos estudos de QTLs foi à nota máxima recebida nos três anos de avaliação. *A priori*, a totalidade das informações foi considerada e os dados de campo das localidades de FEM e FES foram avaliados conjuntamente, de modo a investigar a existência de QTLs comuns aos dois ambientes. Posteriormente, os estudos foram realizados individualmente em cada ambiente, de modo a elucidar a existência de interação entre genótipo x ambiente (GxA). Para os dados de laboratório, as análises de QTLs foram realizadas considerando os componentes de resistência PL e EAT para a raça II e XXXIII de *H. vastatrix*. Para isso, foram utilizadas 94 progênies de irmãos completos e a metodologia de regressão por intervalo simples de Fulker and Cardon (1994). Derivado do método de regressão por IBD (*identical by descent*) de Haseman and Elston (1972), a regressão de Fulker e Cardon permite a análise investigativa de QTLs entre pares de marcadores de forma similar aos métodos de mapeamento por intervalo, o que permite o

isolamento da variância do QTL e da taxa de recombinação. Proposto inicialmente para estudos em humanos é uma metodologia específica para estudos em populações não experimentais e pouco numerosas. Os parâmetros da regressão, herdabilidade ( $h^2$ ) e proporção de IBD foram calculados segundo descrito por Rosado et al. (2010). A detecção de QTL foi realizada considerando os valores de teste de razão de verossilhança superiores a 13, o que corresponde, aproximadamente, a  $LOD > 3$ . As análises foram feitas utilizando o *software* GQMOL (<http://www.ufv.br/dbg/gqmol/gqmol.htm>).

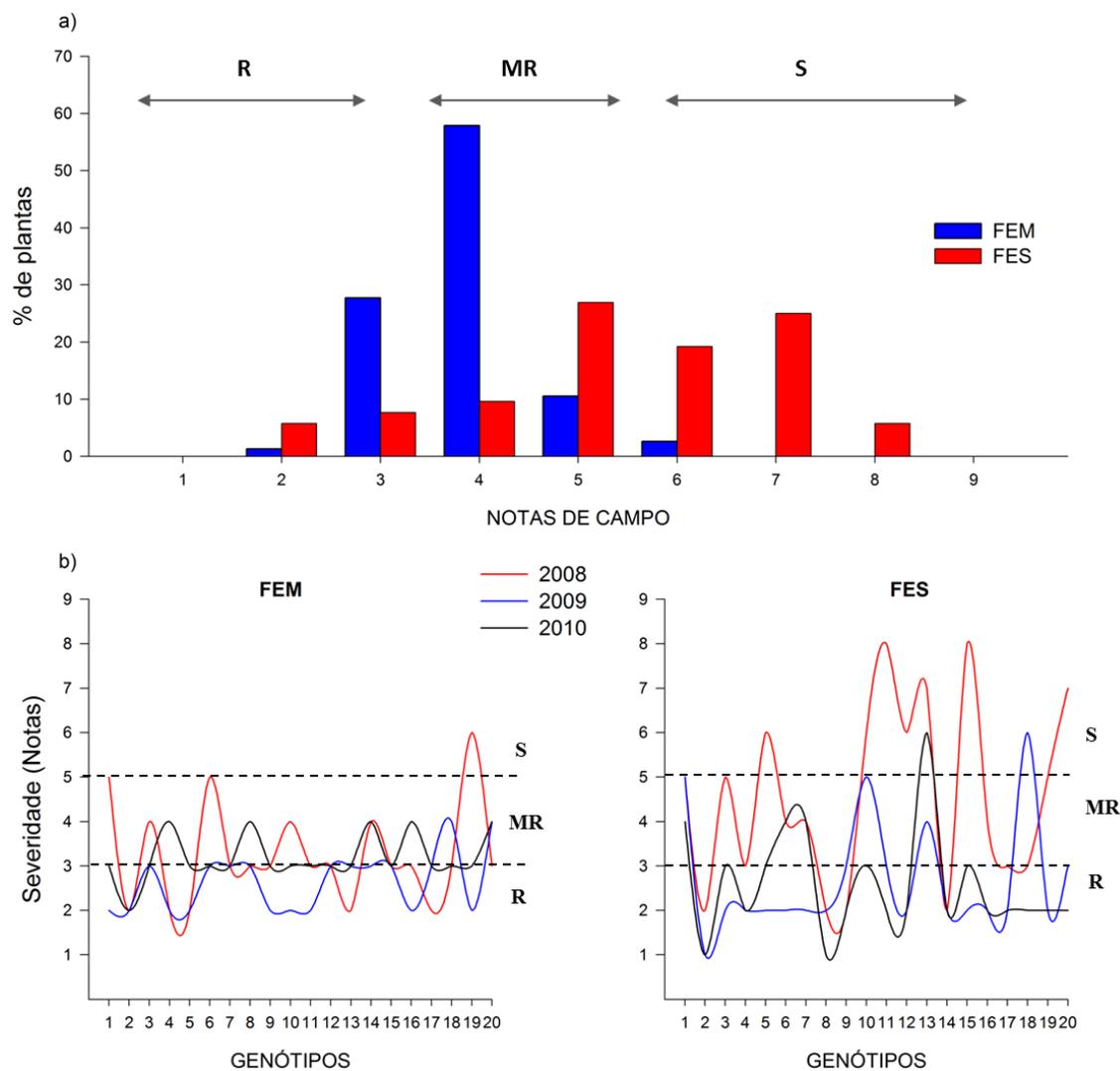
## RESULTADOS

### *Dinâmica da doença para os dados de campo*

A distribuição das frequências e a variação das notas (severidade) nos anos de avaliação em FEM e FES estão representadas na Figura 1. A classificação dos genótipos em classes fenotípicas (Figura 1a) foi feita considerando a nota máxima de doença obtida nos três anos de avaliação. Assim, genótipos com nota 1, 2 e 3 foram considerados resistentes (R), os genótipos com nota 4 e 5 foram classificados como moderadamente resistentes (MR) e, aqueles com nota 6 ou superior foram considerados susceptíveis (S). A localidade de FES apresentou maior porcentagem de genótipos susceptíveis do que em FEM. A amplitude das notas de campo nos diferentes anos de avaliação de ferrugem está representada na Figura 1b. Considerando uma amostragem de 20 genótipos, observa-se que, em FES, houve elevada variação de notas nos três anos de avaliação, enquanto que em FEM houve uma maior homogeneidade.

O resumo das avaliações de campo é apresentado na Tabela 2. Na localidade FES a incidência de doença foi significativamente maior do que em FEM (teste  $t$  à 5% de probabilidade). Para melhor entendimento da agressividade da ferrugem nas duas localidades, as avaliações de severidade (notas) foram associadas a variáveis influenciadoras do patossistema *H. vastatrix*-cafeeiro. Para isso são apresentados os dados climáticos e pluviométricos na Figura 2. Nas duas localidades observa-se semelhança nos valores de temperaturas máximas e mínimas nos três anos de avaliação (2008, 2009 e 2010), porém em FES foi observado menor precipitação pluviométrica (Tabela 2 e Figura 2). Quanto aos dados de produtividade, outra variável influenciadora do patossistema, observou-se que a produção em FES é, em média, maior do que em FEM. O ano de 2008 em FES foi o de maior produtividade e, conseqüentemente, o de

maior incidência de doença em campo (nota média de 5,15). Por fim, outra variável a ser considerada e que influência no progresso da doença é a umidade de cada localidade. Em FES o molhamento foliar, causado principalmente pelo orvalho nas primeiras horas do dia, é maior do que na localidade de FEM.

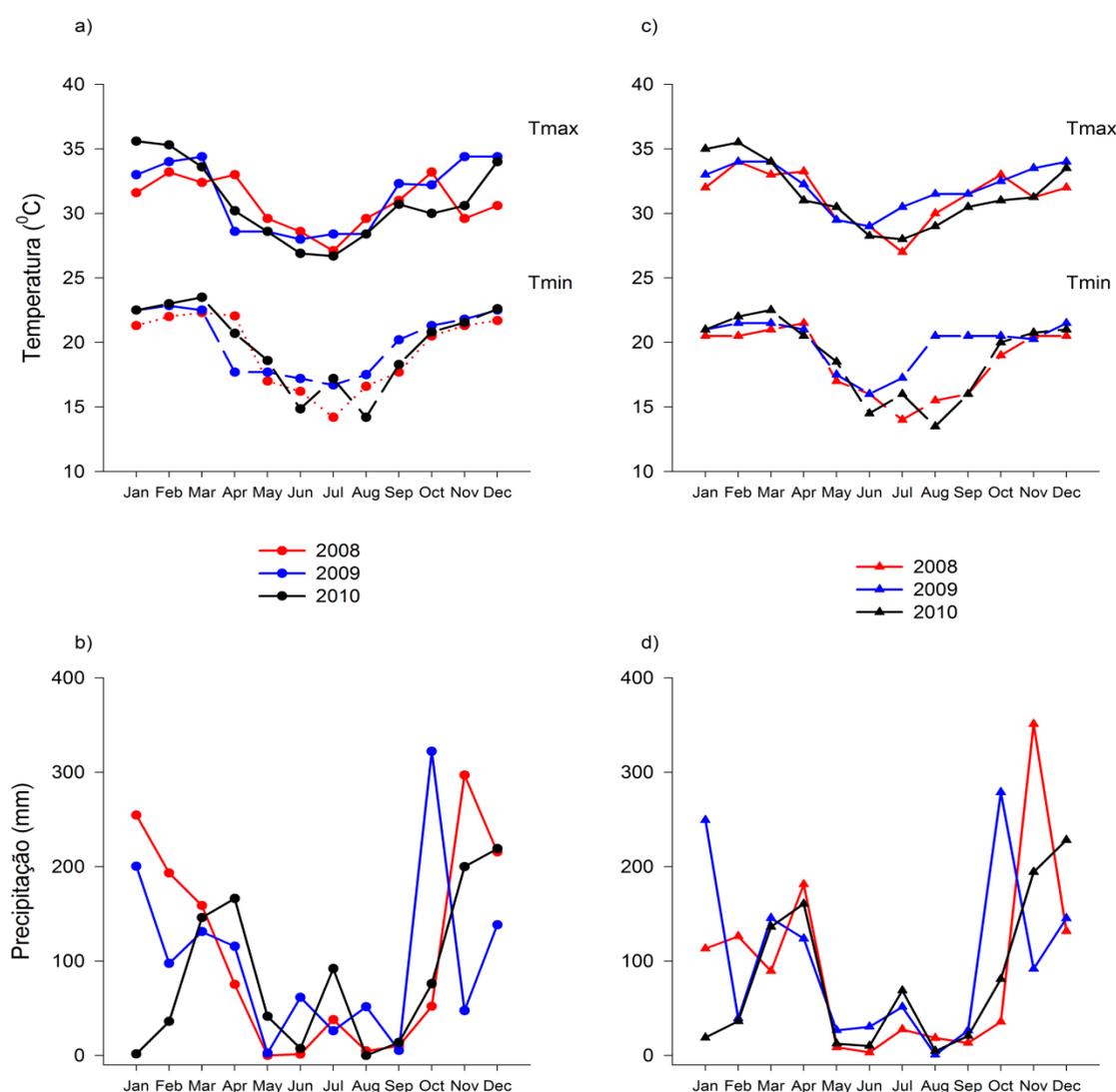


**Figura 1.** (a) Frequência das notas de campo para as avaliações de ferrugem realizadas na FEM (vermelho) e FES (azul) nas colheitas de 2008, 2009 e 2010. Para os três anos de avaliação, genótipos que apresentaram nota máxima entre 1 e 3 foram considerados resistentes (R), genótipos com nota entre 4 e 5 foram considerados moderadamente resistentes (MR) e aqueles com notas superiores a 6 foram considerados susceptíveis; (b) Variação das notas de campo nos anos de 2008 (vermelho), 2009 (azul) e 2010 (preto) considerando amostragem aleatória de 20 genótipos de FEM e FES.

**Tabela 2.** Análise descritiva dos dados de campos avaliados em três colheitas nas Fazendas Experimentais de Marilândia (FEM) e Sooretama (FES).

Variáveis	Parâmetro	FEM				FES			
		2008	2009	2010	Média	2008	2009	2010	Média
Severidade (notas) de Ferrugem	X	3,26	2,78	3,41	3,15	5,16	3,44	2,84	3,81
	Max	6	4	5	-	8	7	6	-
	Min	2	2	2	-	2	1	1	-
Clima (°C)	X <sub>MAX</sub>	30,79	31,4	30,9	31,03	31,3	32,4	31,5	31,7
	X <sub>MED</sub>	23,56	24,2	24,3	24,02	23,6	24,2	23,9	23,1
	X <sub>MIN</sub>	19,40	20	19,8	19,73	18,5	20,5	18,9	19,30
Chuva (mm)	X	108,39	100	83,3	97,23	91,7	71,6	81	81,43

**Legenda:** X = média aritmética; Max = valores máximos ; Min = valores mínimos.



**Figura 2.** Dados climáticos de FEM (a) e de FES (c) considerando as temperaturas máximas (Tmax) e temperaturas mínimas (Tmin) nos anos de 2008 (vermelho), 2009 (azul) e 2010 (preto). Dados pluviométricos de FEM (b) e FES (d) nos anos de 2008 (vermelho), 2009 (azul) e preto (2010).

### Avaliação dos componentes de resistência

As avaliações das inoculações em laboratório foram realizadas e os componentes de resistência para a raça II e XXXIII foram mensurados. O resumo dessas análises está apresentado na Tabela 3 e não foi observada diferença significativa da severidade entre os componentes de resistência avaliados para a raça II e XXXIII, pelo teste *t* a 5% de probabilidade.

A relação entre os componentes de resistência foi mensurada pela análise de correlação de Pearson (Tabela 4). Todas as correlações avaliadas foram significativas e os componentes AE e EAT foram as que apresentaram os maiores valores para a raça II e XXXIII (0,871 e 0,887, respectivamente). Valores negativos foram observados entre os componentes PI e PL em relação aos demais.

**Tabela 3.** Análise descritiva dos componentes de resistências avaliados para as raças II e XXXIII de *H. vastatrix*.

COMPONENTES DE RESISTÊNCIA							
Raça	Parâmetros	PI	PL	DS	DE	AE	EAT
RAÇA II	Média	19,76	28,88	91,02	52,39	2,25	746,69
	Máximo	28,00	36,00	100,00	100,00	22,22	5945,00
	Mínimo	15,00	21,00	2,10	0,00	0,00	0,00
	Clone 02	18,00	26,00	100,00	80,30	4,15	2204,00
	Clone 83	22,00	28,00	95,70	24,50	0,75	198,00
RAÇA XXXIII	Média	20,26	29,17	91,33	55,02	2,56	548,30
	Máximo	32,00	36,00	100,00	100,00	24,07	5184,00
	Mínimo	14,00	21,00	2,10	0,00	0,00	0,00
	Clone 02	18,00	24,00	100,00	55,30	4,53	961,00
	Clone 83	21,00	26,00	95,80	47,10	2,84	706,00

**Legenda:** PI = período de incubação; PL = período latente; DS = porcentagem de discos com sintomas; DE = porcentagem de discos com esporulação; AE = porcentagem de área esporulada; EAT = produção de uredóspero por área.

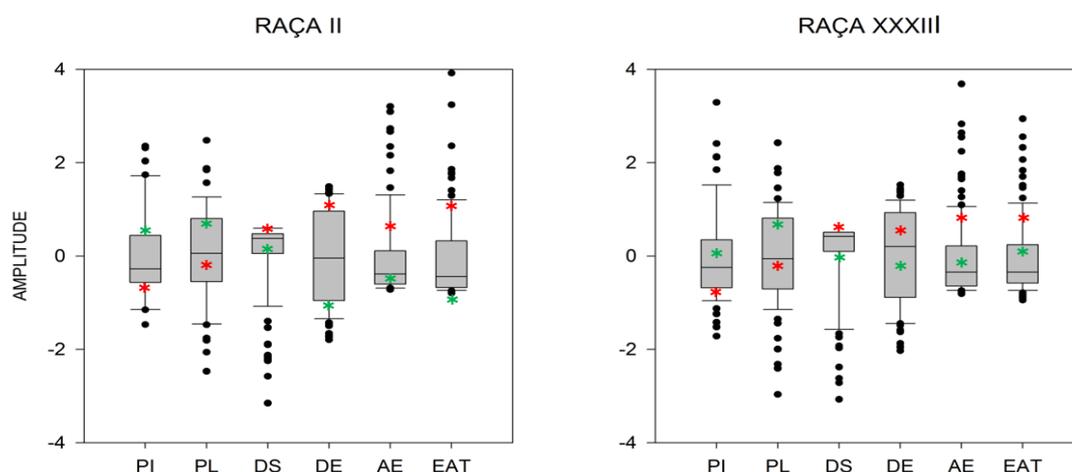
**Tabela 4.** Estimativa das correlações de Pearson entre os componentes de resistência e os dados de campo avaliados para as raças II (abaixo da diagonal) e XXXIII (acima da diagonal) de *H. vastatrix* na Fazenda Experimental de Marilândia (FEM).

Componentes	PI	PL	DS	DE	AE	EAT
PI	-	0,650**	-0,617**	-0,687**	-0,506**	-0,454**
PL	0,615**	-	-0,516**	-0,710**	-0,662**	-0,591**
DS	-0,535**	-0,553**	-	0,553**	0,246**	0,215*
DE	-0,669**	-0,751**	0,536**	-	0,595**	0,541**
AE	-0,452**	-0,639**	0,257**	0,651**	-	0,887**
EAT	-0,458**	-0,635**	0,274**	0,658**	0,871**	-

**Legenda:** \*significativo pelo teste *t* (P=0.05); \*\* significativo pelo teste *t* (P=0.01). PI = período de incubação; PL = período latente; DS = porcentagem de discos com sintomas; DE = porcentagem de discos com esporulação; AE = porcentagem de área esporulada; EAT = produção de uredóspero por área.

### Amplitude fenotípica e identificação de genótipos resistentes

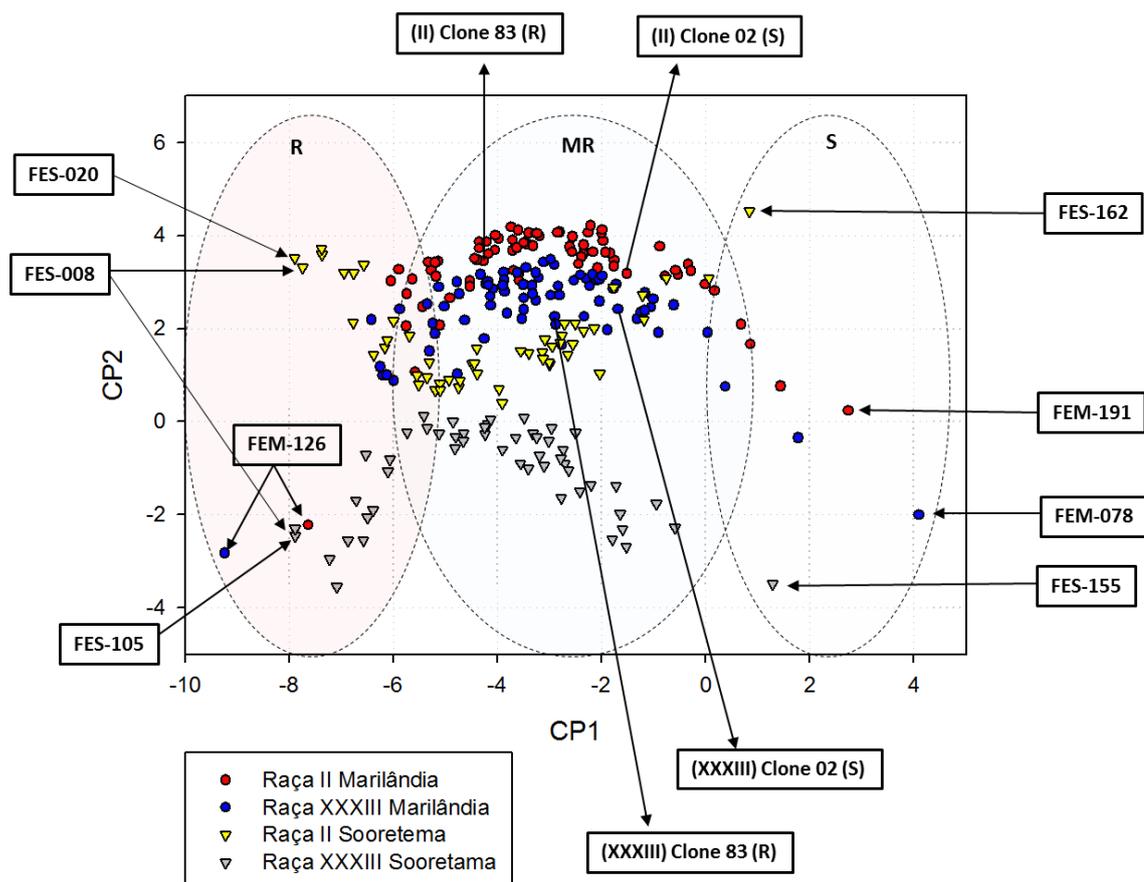
A amplitude fenotípica dos componentes de resistência para as raças II e XXXIII foram mensuradas considerando todos os 126 genótipos e estão representados na Figura 3. Nos resultados foi possível observar a existência de segregação transgressiva, que se caracteriza pela manifestação fenotípica nas progênes com maior ou menor intensidade do que nos genitores. Esse fenômeno pode ser observado naqueles indivíduos com valores fora dos limites de variação estabelecidos pelos genitores da população (Clone 02 e 83). O clone 83 mostrou-se resistente à *H. vastatrix*, devido aos maiores valores de PI e PL, e menores valores de DS, DE, AE e EAT; o clone 02 apresentou comportamento contrário e, por isso, foi considerado o genitor susceptível.



**Figura 3.** Amplitude fenotípica dos componentes de resistência para a raça II e XXXIII, considerando os valores codificados de PI (período de incubação), PL (período latente), DS (porcentagem de discos com sintomas), DE (porcentagem de discos com esporulação), AE (porcentagem de área esporulada) e EAT (produção de uredósporo por área). A posição dos genitores da população, Clone 02 e Clone 83, foram marcadas em vermelho e verde, respectivamente, permitindo a observação do fenômeno de segregação transgressiva. O clone 02 é o genitor susceptível e o 83 é o resistente.

A identificação de progênes resistentes para os dois ambientes foi feita utilizando as avaliações de laboratório e a análise multivariada de PCA. Os resultados destacaram os genótipos FEM-126 e FES-008 das localidades de Marilândia e Sooretama, respectivamente. Esses cafeeiros apresentaram bons níveis de resistência nas avaliações de campo e os melhores resultados nas avaliações de laboratório com as duas raças de *H. vastatrix*, o que torna esses genótipos promissores para o programa de

melhoramento genético da espécie. Por outro lado, para a raça II os genótipos FEM-191 e FES-162 e para a raça XXXIII os genótipos FEM-078 e FES-155 foram susceptíveis a doença e devem ser preterido nos processos de seleção. O resultado dos 126 genótipos avaliados para as duas raças está apresentado na Figura 4, no qual foram plotadas os dois primeiros componentes principais que explicaram mais de 80% da variação fenotípica do conjunto de dados.



**Figura 4.** Análise de componentes principais (PCA) dos 126 genótipos de Marilândia (FEM) e Sooretama (FES) avaliados para os componentes de resistência associados a raça II e XXXIII de *H. vastatrix*. Os componentes principais I e II explicaram mais de 80% da variação fenotípica observada. Grupos de genótipos resistentes (R), moderadamente resistente (MR) e susceptíveis (S) foram determinados e os genótipos mais contrastantes foram indicados. A posição do genitor resistente (Clone 83) e susceptível (Clone 02) para as raças II e XXXIII foram indicados.

A análise de PCA foi utilizada para identificação dos dez genótipos que apresentaram as maiores e menores taxas de severidade da doença, considerando as avaliações para as raças II e XXXIII (Tabela 5). Na localidade de Marilândia (FEM)

50% dos genótipos selecionados foram considerados resistentes para as duas raças avaliadas (FEM-126, FEM-019, FEM-023, FEM-127 e FEM-073); enquanto para Sooretama (FES) esse valor foi de 90% (FES-020, FES-008, FES-019, FES-011, FES-014, FES-017, FES-012, FES-013 e FES-105). O coeficiente de coincidência entre os dados de campo e os de laboratório para FEM e FES foi de 20 e 42,5%, respectivamente.

Ao considerar as avaliações de campo e de laboratório simultaneamente, em Marilândia, os genótipos FEM-126, FEM-019, FEM-023, FEM-127 e FEM-073 se destacaram pelos elevados níveis de resistência; enquanto em Sooretama, resultados semelhantes foram observados para os genótipos FES-020, FES-008, FES-019, FES-011 e FES-017.

**Tabela 5.** Lista dos dez genótipos que apresentaram as maiores e menores taxas de incidência de doença, considerando os 76 genótipos de FEM e os 50 de FES avaliados nas análises de PCA. Os resultados da nota de campo também foram incluídos.

<b>I. MARILÂNDIA (FEM)</b>						
<b>Ranking</b>	<b>ID. Campo</b>	<b>Raça II</b>	<b>Av. Campo</b>	<b>Id. Campo</b>	<b>Raça XXXIII</b>	<b>Av.Campo</b>
1	FEM-126	R	R	FEM-126	R	R
2	FEM-013	R	MR	FEM-073	R	MR
3	FEM-019	R	MR	FEM-136	R	R
4	FEM-133	R	R	FEM-023	R	MR
5	FEM-017	R	MR	FEM-144	R	MR
6	FEM-137	R	R	FEM-127	R	MR
7	FEM-023	R	MR	FEM-019	R	MR
8	FEM-127	R	MR	FEM-080	R	R
9	FEM-007	R	MR	FEM-138	R	R
10	FEM-073	R	MR	FEM-130	R	MR
67	FEM-200	S	MR	FEM-131	S	MR
68	FEM-140	S	MR	FEM-006	S	MR
69	FEM-188	S	MR	FEM-015	S	R
70	FEM-197	S	MR	FEM-200	S	MR
71	FEM-139	S	R	FEM-193	S	MR
72	FEM-006	S	MR	FEM-201	S	MR
73	FEM-083	S	S	FEM-081	S	R
74	FEM-067	S	MR	FEM-188	S	MR
75	FEM-081	S	R	FEM-067	S	MR
76	FEM-191	S	MR	FEM-078	S	MR
<b>II. SOORETAMA (FES)</b>						
<b>Ranking</b>	<b>ID. Campo</b>	<b>Raça II</b>	<b>Av. Campo</b>	<b>Id. Campo</b>	<b>Raça XXXIII</b>	<b>Av.Campo</b>
1	FES-020	R	R	FES-105	R	S
2	FES-008	R	MR	FES-008	R	MR
3	FES-019	R	MR	FES-013	R	S
4	FES-011	R	R	FES-012	R	S
5	FES-014	R	S	FES-014	R	S

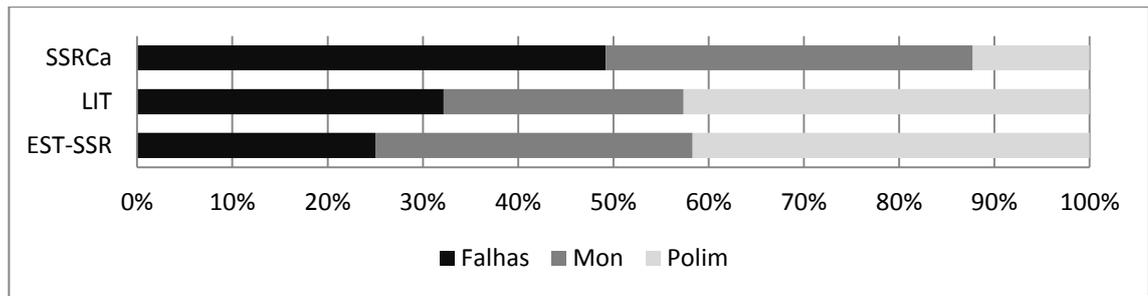
6	FES-017	R	R	FES-019	R	MR
7	FES-012	R	S	FES-011	R	S
8	FES-013	R	S	FES-002	R	R
9	FES-105	R	S	FES-017	R	R
10	FES-063	R	S	FES-020	R	R
43	FES-061	S	MR	FES-165	S	S
44	FES-101	S	S	FES-158	S	R
45	FES-070	S	S	FES-061	S	MR
46	FES-110	S	MR	FES-162	S	S
47	FES-112	S	MR	FES-051	S	S
48	FES-106	S	MR	FES-006	S	MR
49	FES-156	S	MR	FES-068	S	S
50	FES-170	S	S	FES-156	S	MR
51	FES-155	S	S	FES-117	S	S
52	FES-162	S	S	FES-155	S	S

**Legenda:** ID. Campo = identificação de campo das progênies da localidade de Marilândia (FEM) e Sooretama (FES); Av. campo = nota máxima obtida pela progênie nos três anos de avaliação (2008, 2009 e 2010); R = resistente; MR = moderadamente resistente; S = susceptível. Os genótipos marcados são aqueles que apresentaram correspondência nas avaliações com as raças II e XXXIII de *H. vastatrix*.

#### *Mapa de ligação genético e análise de QTLs*

Para a construção do mapa de ligação genético, um total de 258 SSR foram testados nos genitores da população de mapeamento, de modo a selecionar os marcadores polimórficos. Deste total, 115 foram SSRs genômicos (LIT), 96 foram EST-SSRs e 47 foram desenvolvidos para *C. arabica* (SSRCa).

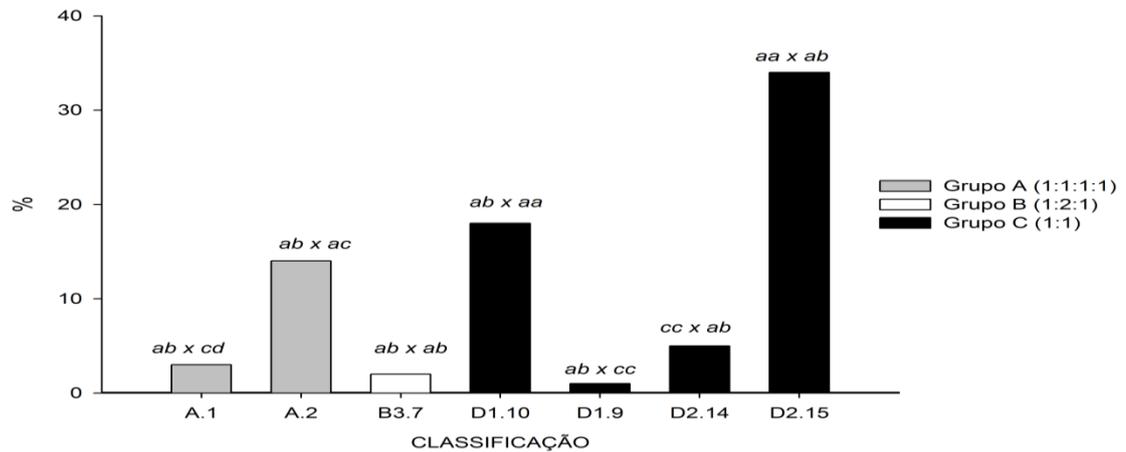
O conjunto de marcadores que apresentou maior porcentagem de polimorfismo foram os LIT (42,60%), seguido dos EST-SSRs (41,66%) e dos SSRCa (12,18%). No total foram selecionados, 95 potenciais marcadores para estudos de mapeamento genético. Marcadores que apresentaram falhas de amplificação, locos duplicados ou outros tipos de artefatos foram classificados como “falhas”. Os SSRCa foram os que apresentaram maior porcentagem de falhas (48,90%), seguido dos LIT e os EST-SSRs (32,17% e 25%, respectivamente). O resumo da triagem de marcadores polimórficos esta representado na figura 5.



**Figura 5.** Triagem de marcadores SSR para estudos de mapeamento genético, considerando *primers* desenvolvidos de biblioteca genômica de *Coffea arabica* (SSRCa), obtidos da literatura (LIT) e minerados do Projeto Brasileiro do Genoma Café (EST-SSRs).

Para o mapeamento genético, 76 SSRs polimórficos foram utilizados na genotipagem das progênies (Anexo 1). A classificação do tipo de polimorfismo acessado por cada marcador foi feita de acordo com Wu et al. (2002), que para as análises em progênies de irmãos completos descreveu os seguintes grupos de configurações genéticas. Grupo A e B são formados pelos marcadores heterozigotos em ambos os genitores e que segregam nas progênies nas proporções de 1:1:1:1 e 1:2:1, respectivamente; grupo C são os marcadores que segregam nas progênies com proporção de 3:1; e o grupo D apresenta a configuração de cruzamentos teste, segregando nas progênies na proporção de 1:1. Os SSR utilizados neste estudo foram predominantemente pertencentes ao grupo D (75,5%), seguido do grupo A (22%) e do Grupo B (2,5%). Marcadores do grupo C não foram observados (Figura 6).

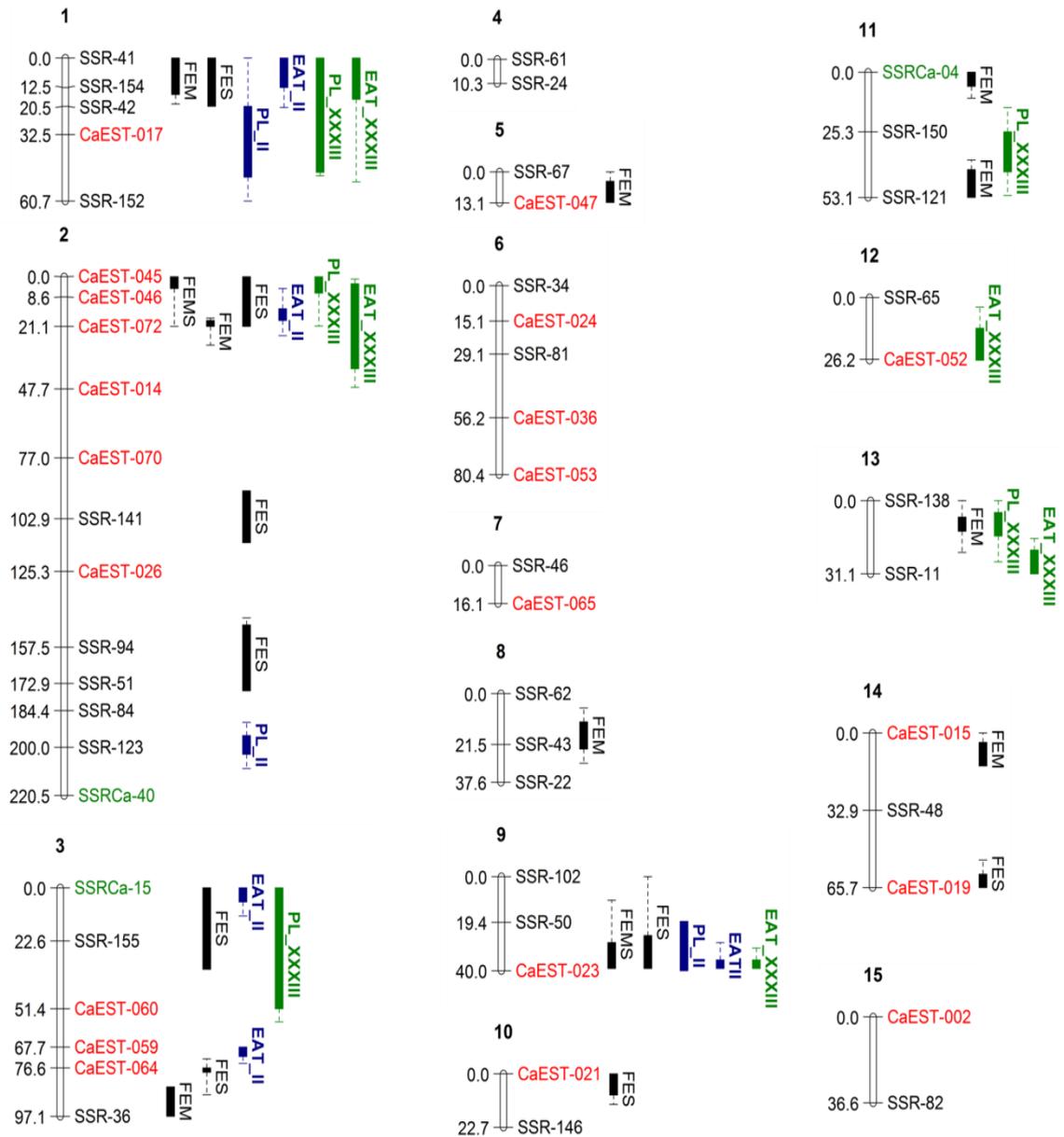
Do total de 76 marcadores utilizados, seis apresentaram distorção de segregação após a correção de Bonferroni (SSR-11, CaEST-004, CaEST-015, CaEST-023, CaEST-044, CaEST-077). O mapa genético foi construído utilizando a totalidade de marcas e 22 SSRs foram declarados não ligados. O restante foi posicionado em 15 grupos de ligação independentes, que variaram em tamanho e número de marcadores ligados. O comprimento total do genoma analisado foi de 818,50 cM. O menor grupo de ligação foi o 4, com 10,3 cM e dois marcadores ligados; enquanto que o maior foi o grupo de ligação 2, com 226,7 cM e doze marcadores ligados.



**Figura 6.** Configuração genética dos 76 marcadores SSR selecionados para o estudo de mapeamento genético em *C. canephora*. Segundo Wu et al. 2002, as letras simbolizam cada um dos alelos do marcador.

O mapa genético foi utilizado na análise de QTLs pela metodologia de regressão por IBD em intervalos simples, proposto por Fulker e Cardon (1994). No total foram mapeados 38 QTLs, considerando os dados fenotípicos de campo das localidades de FEM e FES; e os dados de laboratório avaliados para a raça II e XXXIII de *H. vastatrix*, considerando os componentes de resistência PL e EAT. O mapa genético e a posição dos QTLs estão representados na Figura 7.

Para os dados de campo foram mapeado um total de 20 QTLs distribuídos em dez grupos de ligação diferentes. Deste total, dois QTLs foram identificados ao considerar os dois ambientes simultaneamente. As evidências de interação GxA motivaram as análises considerando os ambientes em separados; e foram identificados nove QTLs para cada uma das localidades. Em FEM os QTLs foram mapeados em oito grupos de ligação diferentes e o QTL de maior significância foi encontrado no grupo de ligação 3 a 0,205 cM do marcador CaEST-064. Em FES os QTLs foram identificados em seis grupos de ligação diferentes com destaque para o marcador SSR-50 localizado no grupo de ligação 9 a 0,15 cM do QTL de maior significância. A posição e os efeitos dos QTLs mapeados para os dados de campo encontram-se detalhados na Tabela 5.I; e QTLs exclusivos para cada localidade foram identificados.



**Figura 7.** Mapa genético de *C. canephora* constituído de 54 marcadores alocados em 15 grupos de ligação independentes. Os CaEST (vermelho) correspondem aos marcadores minerados do Projeto Brasileiro do Genoma Café, os SSR (preto) são SSRs genômicos relatados na literatura e os SSRCa são marcadores desenvolvidos para a espécie *C. arabica*. Os QTLs para os dados de campo (barras pretas), componentes de variância associados a raça II (barras azuis) e raça XXXIII (barras verdes) encontram-se representados. As siglas FEMS referem-se aos QTLs identificados quando considerado os dados de campo das duas localidades simultaneamente; FEM aos QTLs localizados apenas para os dados de campo de Marilandia; FES aos QTLs localizados apenas para os dados de Sooretama; PL e EAT são os QTLs referentes ao período latente e quantidade de uredósporo, respectivamente.

**Tabela 6. I.** Identificação de locos controladores de características quantitativas (QTLs) considerando os dados de notas de campo, avaliados nos dois ambientes simultaneamente (FEMS); e separadamente nas localidades de Marilândia (FEM) e Sooretama (FES); **II.** Identificação de QTLs considerando os componentes de resistência PL e EAT avaliados para a raça II e XXXIII.

I. DADOS DE CAMPO									
CARAC	GL	Marcador	Posição (cM)	$\beta_0$	$\beta_1$	QMDesv	$-\ln(1-R^2)$	pF	$h^2$ ajust
FEMS	2	CaEST-045	0	3,0828	1,5485	32,8494	19,636	43,7368	-16,4567
FEMS	9	SSR-50	0,14	3,3985	1,719	34,7157	15,3	13,3913	-18,2687
FEM	1	SSR-41	0,08	0,7052	0,4596	1,9905	21,386	16,568	-24,5372
FEM	2	CaEST-046	0,125	1,1646	-0,3918	2,0708	20,207	15,6525	20,9209
FEM	3	CaEST-064	0,205	1,4423	-0,8704	2,1151	93,123	63,8249	46,4708
FEM	5	SSR-67	0,131	1,199	-0,4151	2,1499	21,876	16,9486	22,1617
FEM	8	SSR-62	0,2	1,0712	-0,2633	1,9845	16,499	14,4427	14,0555
FEM	11	SSRCa-04	0	1,1588	-0,3593	2,1075	16,701	11,8632	19,1855
FEM	11	SSR-150	0,278	1,0653	-0,36	1,8992	18,553	14,369	19,2201
FEM	13	SSR-138	0,11	1,1267	-0,349	2,0183	13	10,3791	18,6333
FEM	14	CaEST-015	0,03	0,878	-0,242	0,9629	13,056	9,6832	12,9195
FES	1	SSR-41	0,01	6,2926	-1,7981	48,9696	16,886	9,034	16,3549
FES	2	CaEST-045	0	3,1346	4,2949	48,0668	100,591	36,7359	-39,064
FES	2	SSR-141	0	7,1251	-3,7576	44,9472	83,182	9,7408	34,1769
FES	2	SSR-94	0,05	7,3659	-3,2973	52,4267	44,211	23,7276	29,9903
FES	3	SSR-155	0	4,0851	2,9929	48,9899	50,187	6,5131	-27,222
FES	3	CaEST-064	0	5,8118	-1,0975	47,281	13,526	4,8907	9,9819
FES	9	SSR-50	0,15	1,9573	6,9815	55,4671	151,126	19,8412	-63,5002
FES	10	CaEST-021	0	6,4164	-1,3151	54,2977	18,224	6,5931	11,9614
FES	14	SSR-48	0,3	6,2076	-1,7053	47,2468	13,461	6,8382	15,5101
II. COMPONENTES DE RESISTÊNCIA									
CARAC	GL	Marcador	Posição (cM)	$\beta_0$	$\beta_1$	QMDesv	$-\ln(1-R^2)$	pF	$h^2$ ajust
PL (II)	1	SSR-41	0,05	31,53	-15,9489	905,8125	57,698	101,8109	33,0758
PL (II)	2	SSR-84	0,14	30,0829	-9,6163	1071,616	20,845	34,605	19,9429
PL (II)	9	SSR-50	0,11	19,3827	16,9294	1265,833	37,806	25,7475	-35,1094
EAT (II)	1	SSR-41	0,05	5434700	-4709997	4,94 10 <sup>13</sup>	91,895	162,7889	82,2323
EAT (II)	2	CaEST-046	0,08	3907968	-1849863	5,14 10 <sup>13</sup>	13,714	22,0982	32,2969
EAT (II)	3	SSRCa-15	0	5242558	-2141348	7,27 10 <sup>13</sup>	17,23	12,7849	37,386
EAT (II)	3	CaEST-059	0	3941188	-1492995	5,39 10 <sup>13</sup>	22,201	32,7367	26,0664
EAT (II)	9	SSR-50	0,196	1983493	1606638	4,63 10 <sup>13</sup>	14,62	9,9308	-28,0505
PL(XXXIII)	1	SSR-42	0,04	27,3153	-13,0778	801,4043	52,96	104,1502	31,2854
PL (XXXIII)	2	CaEST-45	0,03	25,84	-7,54	866,40	14,89	24,00	18,03
PL(XXXIII)	3	SSR-155	0,12	28,9276	-16,72	666,7026	48,669	46,2202	39,9984
PL(XXXIII)	11	SSR-150	0,09	27,6347	-10,6983	903,1157	17,839	27,9241	25,5931
PL(XXXIII)	13	SSR-138	0,11	24,122	-6,9212	813,0509	14,037	26,0447	16,5574
EAT (XXXIII)	1	SSR-41	0,04	2599237	-1831895	2,13 10 <sup>13</sup>	33,36	58,7017	57,6801
EAT (XXXIII)	2	CaEST-046	0,07	2966779	-2425854	2,25 10 <sup>13</sup>	52,578	85,0981	76,3818
EAT (XXXIII)	9	SSR-50	0,196	1423250	1544096	3,36 10 <sup>13</sup>	18,632	12,6616	-48,6183
EAT (XXXIII)	12	SSR-65	0,18	788230,1	745708,9	9,83 10 <sup>13</sup>	14,267	20,3816	-23,4798
EAT (XXXIII)	13	SSR-138	0,31	1123172	1003423	2,11 10 <sup>13</sup>	25,462	47,3035	-31,5944

**Legenda:** CARAC = característica; GL = grupo de ligação; Posição = posição do marcador em cM;  $\beta_0$  e  $\beta_1$  = parâmetros da regressão; QMDesv = quadrado médio do desvio da regressão;  $-\ln(1-R^2)$  = teste de razão de verossimilhança;  $h^2$  = herdabilidade ajustada; MS = dados de campos da FEM e FES analisados conjuntamente; FEM = dados fenotípicos da Fazenda Experimental de Marilândia; FES = dados fenotípicos da Fazenda Experimental de Sooretama; PL II e XXXIII = período latente avaliados para a raça II e XXXIII de *H. vastatrix*; EAT II e XXXIII = produção de uredóporo por área para a raça II e XXXIII.

Ao considerar as avaliações de laboratório e os locos envolvidos com os componentes de resistência para a raça II, foram detectados oito QTLs, sendo três para PL e cinco para EAT. Para o componente PL, o QTL de maior significância foi

mapeado no grupo de ligação 9 a uma distância de 0,11 cM do marcador SSR-150; para o componente EAT, o QTL de maior significância foi mapeado no grupo de ligação 1 à 0,05 cM do marcador SSR-41. Para a raça II ressalta-se ainda que QTLs para ambos os componentes foram co-localizados nos grupos de ligação 1 e 9 (Tabela 6-II).

Regiões genômicas associadas aos componentes de resistência para a raça XXXIII também foram localizados e totalizaram dez, sendo cinco para PL e cinco para EAT. Ao considerar o componente PL, o QTL de maior significância foi identificado no grupo de ligação 3 a 0,13 cM do marcador SSR-155; e para o componente EAT, o QTL mais significativo foi mapeado no grupo de ligação 2 a 0,07 cM do marcador CaEST-046. No grupo de ligação 1 foram co-localizados QTLs para ambos os componentes (Tabela 6-II).

A existência de QTLs exclusivos para as raças II e XXXIII foi observada. No grupo de ligação 1, marcadores significativamente ligados a todas as características em avaliação foram mapeados. Assim, neste grupo de ligação, o SSR-41 constitui-se em um marcador promissor para os programas de seleção assistida por marcadores (MAS).

## **DISCUSSÃO**

### *Dados fenotípicos: dinâmica da doença e componentes de resistência*

Nos programas de melhoramento de plantas, a seleção de plantas resistentes passa pela necessidade de quantificação da doença, de forma precisa e operacionalmente simples, uma vez que número de ensaios conduzidos é grande. O parâmetro severidade é o mais apropriado para medir doenças foliares, já que a porcentagem da área de tecido coberto por sintomas é um bom indicativo da quantidade de doença. Com isso, a utilização de escala de notas é uma alternativa comumente utilizada nos programas de melhoramento genético, visto que permite avaliações rápidas e relativamente fáceis da doença observada em experimentos de campo.

As avaliações de campo utilizando escalas de notas (severidade) evidenciaram a existência de resistência nos genótipos avaliados, o que resultou em elevada amplitude fenotípica com a observação de plantas resistentes, moderadamente resistentes e susceptíveis à *H. vastatrix*. A expressão diferencial de níveis de doença também foi observada nos dois ambientes avaliados. Para melhor compreensão dessa diferença, os dados fenotípicos foram associados à variáveis explicativas, que sabidamente influenciam o patossistema cafeeiro-ferrugem. Dentre essas, a maior produtividade e a

elevada umidade em FES foram considerados, possivelmente, os fatores causadores dos maiores níveis de doença observada na localidade. A elevada umidade dessa região faz com que o molhamento foliar seja maior, sobretudo, nas horas iniciais do dia. No campo isso pode ocorrer em um dia nublado seguido de uma noite com temperaturas amenas, o que resulta na presença de água livre na superfície das folhas do cafeeiro, devido a formação de orvalho. A elevada umidade e, conseqüentemente, o longo período de molhamento foliar são consideradas condições ideais para germinação dos uredósporos e estabelecimento da doença (Capucho et al., 2012b).

A associação dos dados climáticos e de doença é de fundamental importância para o entendimento do progresso da doença, pois irão determinar o sucesso da interação planta-patógeno e a agressividade em cada localidade (Rozo et al., 2012). A maior incidência de doença em FES sugeriu a existência de interação genótipo x ambiente (GXA) e forneceu indícios importantes de que os programas de melhoramento genético para resistência a doença devem considerar cada localidade separadamente, principalmente nas definições de estratégias de melhoramento e manejo da doença. Todavia, ao considerar o estudo de um patossistema é importante que além do hospedeiro e do ambiente favorável, seja considerado também o patógeno. Dessa forma, avaliações em laboratório foram realizadas no sentido de especificar a resistência dos genótipos às raças fisiológicas II, que apresenta o gene de virulência v5; e raça XXXIII, que apresenta os genes de virulência v5, v7 ou v5, v7 e v9, que é considerada uma raça com genes de virulência complexos (Zambolim et al., 2005).

Nos estudos de laboratório seis componentes de resistência foram utilizados para mensuração da resistência horizontal observada nos genótipos. A utilização dessa escala de severidade fornece ao pesquisador resultados mais acurados, sobretudo, pelas condições controladas que são impostas nos experimentos. Assim, ao considerar as avaliações para as duas raças não se observou diferença significativa de agressividade da doença, ou seja, os valores de PI, PL, DS, DE, AE e EAT foram estatisticamente iguais nas avaliações com ambas às raças. Ao considerar a correlação dos componentes de resistência nas avaliações de severidade, foi observado para a raça II e XXXIII significância entre todos os componentes, sendo o AE e EAT os que apresentaram os maiores valores. A existência de valores negativos para PI e PL é explicada pela associação inversa com os outros componentes, em outras palavras, maiores valores de PI e PL significa maior tempo para o aparecimento de sintomas e, conseqüentemente,

maiores níveis de resistência; por outro lado, valores elevados dos outros componentes de resistência indicam uma maior incidência de doença, o que neste caso significa maior susceptibilidade do genótipo.

Além da similaridade dos valores para as avaliações de severidade e correlação, para as raças II e XXXIII foi possível observar também ampla variação fenotípica e o fenômeno de segregação transgressiva, que se caracteriza pela existência de progênies que ultrapassaram o limite fenotípico estabelecido pelos genitores da população (Clone 02 e 83). A existência desse fenômeno evidencia o potencial dessa população para seleção de genótipos resistentes nos programas de melhoramento genético de cafeeiro. Neste sentido, para cada localidade, dez genótipos promissores foram identificados, sendo que para FEM cinco deles foram resistentes para as avaliações de campo e de laboratório, considerando as duas raças fisiológicas (FEM-126, FEM-019, FEM-023, FEM-127 e FEM-073). Para FES, cinco genótipos, que também apresentaram resistência para as avaliações de campo e de laboratório foram identificados (FES-020, FES-008, FES-019, FES-011 e FES-017). Assim, se os genótipos aqui mencionados agregarem outras características agrônômicas importantes, tais como produtividade, padrão de maturação, porte e qualidade de bebida; são materiais promissores para compor novas variedades de *C. canephora*, ou então, serem selecionados para inclusão em ciclos de seleção recorrente.

Na localidade de FES, foi observado também um maior valor de porcentagem de coincidência ao comparar os dados de campo e os de laboratório. A explicação para isso esta associada, possivelmente, as condições climática da localidade, que é mais propício ao estabelecimento da doença o que o aproxima das condições ideais que são impostas nos experimentos realizados em laboratório.

#### *Mapas de ligação genética em C. canephora.*

Ao se observar as propriedades biológicas dos materiais em estudo, para os trabalhos de mapeamento genéticos em plantas, dois tipos básicos de populações são utilizadas: as oriundas de cruzamentos controlados (endogâmicas) e as populações naturais (exôgâmicas). O primeiro tipo engloba as populações F<sub>2</sub>, retrocruzamentos, duplo-haplóides e linhagens endogâmicas recombinantes (RILs); enquanto que o segundo inclui, sobretudo, as famílias de irmãos completos e meios irmãos. A escolha de uma população em detrimento da outra deve ser feita levando em consideração o

número de indivíduos amostrados, as características de interesse dos genitores e, sobretudo, o sistema reprodutivo da espécie em estudo (Liu, 1998).

Grande parte dos mapas genéticos já desenvolvidos são originados de populações de cruzamentos controlados, derivados de  $F_1$ , onde os genitores são linhas puras. Para este caso, as análises de ligação são relativamente simples, pois o desequilíbrio e as fases de ligação são conhecidas (Lin et al., 2003). No entanto, em culturas como maracujá, eucalipto, cana de açúcar, citros, espécie florestais e, sobretudo, em *C. canephora*; não é possível a obtenção de linhas puras devido a diferentes fatores biológicos. Para o caso da espécie *C. canephora* essa característica é causada pela autoincompatibilidade gametofítica, que consiste na impossibilidade das flores de uma planta ser fecundadas pelo seu próprio pólen (Berthaud and Charrier, 1988). Nesse tipo de situação os mapas genéticos são construídos a partir da genotipagem de progênies originadas de cruzamento biparental entre indivíduos não homocigóticos (*outbred cross*), gerando  $F_1$  segregantes formada por indivíduos que são irmãos completos (*full sib*) ou meios irmãos (*half sib*).

Para esse tipo de população o estabelecimento de mapas genéticos apresenta algumas particularidades que não são observadas no mapeamento de populações controladas. A primeira delas é o número de alelos do marcador e os diferentes padrões de segregação. Para a população de mapeamento desse estudo, as progênies de irmãos completos foram avaliados e classificados segundo Wu et al. (2002). Os marcadores do grupo D, que segregam como um retrocruzamento (1:1) foram os mais abundantes (73,33%), seguido dos marcadores da classe A (22%), que segregam na proporção de 1:1:1:1; e da classe B (3,63%), que segregam como uma  $F_2$  na proporção de 1:2:1. A possibilidade de determinação da origem de cada um dos alelos dos parentais faz com que alguns cruzamentos sejam mais informativos do que outros. A segunda particularidade das populações exogâmicas, é o desconhecimento *a priori* da fase de ligação dos genitores da população, o que faz necessário o desenvolvimento de algoritmos para caracterização da fase mais provável. Esse conhecimento é requerido para a detecção de eventos de recombinação e define a configuração dos alelos de um par de locos heterocigotos em cromossomos homólogos de um parental (Maliepaard et al., 1997; Wu et al., 2002).

Uma alternativa para as análises neste tipo de populações foi proposta por Grattapaglia e Sederoff (1994) e foi denominada de duplo “*pseudo-testcross*” ou

“*pseudocruzamento teste*”. Este método propõe a utilização apenas daqueles marcadores com segregação 1:1 (grupo D) nas análises de ligação genética. Para isso, são construídos dois mapas de ligação, um para cada genitor, em função de qual deles possui o alelo heterozigoto para o loco marcador. A configuração dos cruzamentos não precisa ser definida *a priori* como em cruzamentos testes clássicos, mas sim inferida *a posteriori* após a análise de segregação dos marcadores nas progênies. Esta é uma abordagem muito utilizada na prática (Bundock et al., 2008; Fabbrini et al., 2012; Rae et al., 2008; Rohde et al., 2011), pois permite o pronto uso das metodologias e *softwares* desenvolvidos para populações experimentais. A facilidade de execução, fez com que a metodologia fosse preferida em estudos de mapeamento com *C. canephora* (Lashermes et al., 1996; Leroy et al., 2011; Priyono et al., 2010). Todavia, a desvantagem do “*pseudo-testcross*” está no fato de não se utilizar marcadores com outros tipos de segregação e não conseguir integrar em um único mapa os marcadores que segregam para cada genitor (Garcia et al., 2006; Gazaffi, 2009).

Neste contexto, a possibilidade de construção de um único mapa genético, integrando todos os marcadores moleculares é vantajoso, principalmente, por possibilitar o aumento da saturação do mapa de ligação e estender a caracterização da variação polimórfica em todo o genoma, o que pode facilitar a localização de QTLs (Garcia et al., 2006; Maliepaard et al., 1997). A percepção dessas vantagens fez com que o uso integrado de todos os marcadores ganhasse destaque em estudos genéticos (Garcia et al., 2006; Hanley et al., 2011; Moraes, 2005; Pereira, 2010; Rosado et al., 2010; Tomaz, 2008), e por isso foi utilizado no presente trabalho para a construção do mapa genético. Assim, 54 marcadores codominantes foram alocados em 15 grupos de ligação independentes, quatro a mais do que o número de cromossomos da espécie. Dentre os SSR utilizados os EST-SSRs foram os mais polimórficos e que apresentaram menor porcentagem de falhas. Esse é o primeiro relatado da utilização dessa abordagem de mapeamento em estudos na espécie *C. canephora*.

#### *Análises de QTLs para resistência à H. vastatrix*

Assim como para as análises de ligação, o mapeamento de QTLs em famílias de irmãos completos apresenta algumas particularidades que não são observadas nos estudos de populações controladas. Dentre essas os diferentes padrões de segregação, o número incerto de alelos segregando e o desconhecimento da fase de ligação; são

considerados os principais complicadores genético-estatísticos das análises (Gazaffi, 2009; Lin et al., 2003). Neste contexto, os métodos de mapeamento baseados em IBD são ferramentas poderosas e partem do pressuposto de que indivíduos parecidos fenotipicamente compartilham um maior número de alelos do QTL do que indivíduos fenotipicamente discordantes (Almasy and Blangero, 2009; Haseman and Elston, 1972). Com isso, modelos de regressões podem ser estabelecidas, na qual a diferença fenotípica entre pares de irmãos é regredida em função no número de alelos IBD compartilhadas.

Neste estudo, por meio da regressão por intervalor simples proposta por Fulker e Cardon (1994) foi identificado um total de 38 regiões genômicas potencialmente envolvidas com a resistência do cafeeiro à ferrugem. Com base nos dados de campo, avaliados sem distinção de localidade, foram mapeados QTLs nos grupos de ligação 2 e 9. O CaEST-045 no grupo de ligação 2 foi o de maior significância e trata-se de um marcador importante para estudos simultâneos nas localidades de FEM e FES. A evidencia de interação GxA, observada na diferença significativa de níveis de doença, motivou as análises de QTL considerando as duas localidade separadamente.

Para FEM, foram mapeados nove QTLs, nos grupos de ligação 1, 2, 3, 5, 8, 11, 13 e 14. Dentre esses, o marcador CaEST-064, no grupo de ligação 3, foi posicionado a 0,205 cM do QTL de maior significância. Já para as análises dos dados de FES, foram mapeados um total de nove QTLs, distribuído nos grupos de ligação 1, 2, 3, 9, 10 e 14. O marcador SSR-50, no grupo de ligação 9, foi mapeado a uma distancia de 0,15 cM do QTL mais significativo. A identificação de regiões genômicas exclusivas para as avaliações em FEM e FES evidencia a existência de expressão gênica diferencial de acordo com a localidade de avaliação. A explicação para isso está associada a duas hipóteses, não necessariamente mutuamente exclusiva. A primeira refere-se à possibilidade de existência de patógenos distintos nas duas localidades. Dessa forma, a incidência de diferentes raças de ferrugem, com os respectivos genes de virulência, irão acionar distintas respostas de defesa do cafeeiro, e diferentes QTLs serão identificados e posicionados no mapa genético. A segunda hipótese, já evidenciada neste estudo, é a influência de fatores ambientais no sucesso da interação planta-patógeno. Segundo Roza et al. (2012), em estudos com cafeeiros na Colômbia, a ocorrência de epidemias de ferrugem está mais relacionada às condições ambientais favoráveis do que a existência de novas raças mais agressivas.

Ao considerar as avaliações de laboratório para a raça II e XXXIII, QTLs para os componentes de resistência PL e EAT foram mapeados nos grupos de ligação 1, 2, 3, 9, 11, 12 e 13. Esses resultados demonstraram a existência de diferentes QTLs para as raças e os componentes de resistência avaliados, embora algumas sobreposições tenham sido observadas. Na concepção dos programas de melhoramento, que objetivam a SAM, essas sobreposições são alvos de interesse, pois permitem a identificação de locos de efeitos significantes para diferentes fatores de resistência. No grupo de ligação 1, onde está alocado o marcador SSR-41 foram co-localizados QTLs que desempenham funções importantes para os níveis de resistência, mensurados em campo, e resistência para ambos as raças consideradas nas avaliações de laboratório. Resultados semelhantes foram relatados em outros estudos (Hanley et al., 2011; Jorge et al., 2005; Samils et al., 2011), no qual QTLs para diferentes escalas de severidades foram co-localizados na mesma região genômica. Segundo Hanley et al. (2011), este é um exemplo claro da aplicabilidade do SAM em programas de melhoramento, pois permite a seleção de locos de resistência mesmo em situações onde o fenótipo não é observado, devido a ausência de pressão da doença no campo.

Com isso, essa região genômica representa um alvo de interesse para programas de melhoramento genético de plantas, que objetivam a identificação de regiões associadas à resistência a ferrugem. Assim, acredita-se que a melhor caracterização dessas regiões possa auxiliar no entendimento da interação planta-patógeno e, conseqüentemente, no desenvolvimento de variedades de *C. canephora* com resistência durável. Para a espécie *C. arabica*, espécie do gênero *Coffea* no qual os estudos com ferrugem estão mais avançados, marcadores ligados aos genes de resistência  $S_{H3}$  já foram localizados e caracterizados, tornando-se ferramentas importantes para os programas de melhoramento da espécie (Prakash et al., 2004). Em *C. canephora*, este é o primeiro relato de marcadores de DNA ligados a genes que conferem resistência à ferrugem. A identificação desses marcadores é considerada um ponto de partida importante para estudos dentro da espécie que objetivam a piramidação gênica, seleção assistida por marcadores (SAM) e clonagem de QTLs. Todavia, para que esses resultados possam ser utilizados eficientemente em programas de melhoramento genético é importante que seja realizado o mapeamento fino da região, aumentando a densidade de marcadores e o número de progênies genotipados. Os resultados encontrados além de serem aplicados diretamente na seleção de genitores e progênies

podem ser extrapolados para estudos comparativos dentro do gênero *Coffea*, fornecendo informações importantes se o QTL encontrado é aplicável em outros *backgrounds* genéticos.

## CONCLUSÃO

Os resultados discutidos neste estudo são de grande valia nos programas de melhoramento genético para resistência durável de *H. vastatrix* em *C. canephora*. Dentre as implicações desse trabalho na cafeicultura, destacamos: (i) a observação de níveis de doença diferenciada para cada localidade, o que sugere manejo e estratégias de melhoramento diferenciadas para FEM e FES; (ii) seleção de genótipos resistentes para as avaliações de campo e as avaliações de laboratório para as raças II e XXXIII, que podem compor novas variedades ou constituírem populações base de seleção recorrente; (iii) disponibilização de mapa genético para espécie utilizando marcadores SSRs com diferentes padrões de segregação; (iv) primeiro estudo focado na detecção de regiões genômicas associadas a resistência à ferrugem em *C. canephora*; e (v) identificação de marcadores candidatos para utilização em programas de seleção assistida (SAM) e clonagem posicional.

## REFERENCIAS

- Almasy, L., and Blangero, J. (2009). Human QTL linkage mapping. *Genetica* **136**, 333-340.
- Amidou, N., Michel, N., Serge, H., and Valerie, P. (2007). Genetic basis of species differentiation between *Coffea liberica* Hiern and *C. canephora* Pierre: Analysis of an interspecific cross. *Genetic Resources and Crop Evolution* **54**, 1011-1021.
- Baruah, A., Naik, V., Hendre, P. S., Rajkumar, R., Rajendrakumar, P., and Aggarwal, R. K. (2003). Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. *Molecular Ecology Notes* **3**, 647-650.
- Berthaud, J., and Charrier, A. (1988). Genetics resources of *Coffea*. In "Coffee" (R. J. Clarke and R. Macrae, eds.), Vol. 4, pp. 1-42. Elsevier Applied Science, London,.
- Bettencourt, A. J., Lopes, J., and Palma, S. (1992). Factores genéticos que condicionam a resistência às raças de *Hemileia vastatrix* Berk. et Br. dos clones-tipo dos grupos 1, 2 e 3 de derivados de Híbrido de Timor. *Broteria Genética* **XIII**, 185-194.
- Bettencourt, A. J., and Rodrigues, J. C. J. (1988). "Principles and practice of coffee breeding for resistance to rust and other disease," *Agronomy*, Elsevier Applied Science, London.
- Bhat, P. R., Krishnakumar, V., Hendre, P. S., Rajendrakumar, P., Varshney, R. K., and Aggarwal, R. K. (2005). Identification and characterization of expressed

- sequence tags-derived simple sequence repeats markers from robusta coffee variety 'CxR' (an interspecific hybrid of *Coffea canephora* × *Coffea congensis*). *Molecular Ecology Notes* **5**, 80-83.
- Bundock, P. C., Potts, B. M., and Vaillancourt, R. E. (2008). Detection and stability of quantitative trait loci (QTL) in *Eucalyptus globulus*. *Tree Genetics & Genomes* **4**, 85-95.
- Cabral, P. G. C., Zambolim, E. M., Zambolim, L., Lelis, T. P., Capucho, A. S., and Caixeta, E. T. (2009). Identification of a new race of *Hemileia vastatrix* in Brazil. *Australasian Plant Disease Notes* **4**, 129-130.
- Cadena, G. C. (2005). Searching for expression of durable resistance to the coffee leaf rust disease in conilon (*Coffea canephora*) germoplasm in the late 70's. In "Durable Resistance to Coffee Leaf Rust" (L. Zambolim, E. Maciel-Zambolim and V. Varzea, eds.), Vol. 1. Universidade Federal de Viçosa, Viçosa, MG.
- Capucho, A. S. (2012). Epidemiologia e resistência do cafeeiro à ferrugem, Universidade Federal de Viçosa, Viçosa, MG.
- Capucho, A. S., Zambolim, E. M., Freitas, R. L., Haddad, F., Caixeta, E. T., and Zambolim, L. (2012a). Identification of race XXXIII of *Hemileia vastatrix* on *Coffea arabica* Catimor derivatives in Brazil. *Australasian Plant Dis. Notes*, 189-191.
- Capucho, A. S., Zambolim, L., Cabral, P. G. C., Maciel-Zambolim, E., and Caixeta, E. T. (2012b). Climate favourability to leaf rust in Conilon coffee. *Australasian Plant Pathol.* **Published online: 20 December 2012.**
- Caranta, C., Lefebvre, V., and Palloix, A. (1997). Polygenic resistance of pepper to potyviruses consists of a combination of isolate-specific and broad-spectrum quantitative trait loci. *Molecular Plant-Microbe Interactions* **10**, 872-878.
- Chantret, N., Mingeot, D., Sourdille, P., Bernard, M., Jacquemin, J. M., and Doussinault, G. (2001). A major QTL for powdery mildew resistance is stable over time and at two development stages in winter wheat. *Theoretical and Applied Genetics* **103**, 962-971.
- Coulibaly, I., Revol, B., Noirot, M., Poncet, V., Lorieux, M., Carasco-Lacombe, C., Minier, J., Dufour, M., and Hamon, P. (2003). AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. heterocalyx* × *C. canephora*) × *C. canephora*]. *Theor Appl Genet* **107**, 1148-55.
- Cruz, C. D. (2006). "Programa Genes," Universidade Federal de Viçosa, Viçosa, MG.
- Cruz, C. D. (2008). GQMOL: Programa para análise de genética quantitativa molecular ([www.ufv.br/dbggqmol/gqmol.htm](http://www.ufv.br/dbggqmol/gqmol.htm)). Universidade Federal de Viçosa, Viçosa, MG.
- Diniz, E. C., Sakiyama, N. S., Lashermes, P., Caixeta, E. T., Oliveira, A. C. B., Zambolim, E. M., Loureiro, M. E., Pereira, A. A., and Zambolim, L. (2005). Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. *Crop Breeding and Applied Biotechnology* **5**, 387-393.
- Eskes, A. B. (1982). The use of leaf disk inoculations in assessing resistance to coffee leaf rust (*Hemileia vastatrix*). *Netherlands Journal of Plant Pathology*, 127-141.
- Eskes, A. B. (1983). Incomplete resistance to coffee leaf rust., Agricultural University of Wageningen, Landbouwhogeschool.
- Eskes, A. B., and Toma-Braghini, M. (1981). Assessment methods for resistance to coffee leaf rust (*Hemileia vastatrix* Berk. et. Br.). *Plant Protection* **29**, 56-66.
- Fabbrini, F., Gaudet, M., Bastien, C., Zaina, G., Harfouche, A., Beritognolo, I., Marron, N., Morgante, M., Scarascia-Mugnozza, G., and Sabatti, M. (2012). Phenotypic

- plasticity, QTL mapping and genomic characterization of bud set in black poplar. *BMC Plant Biol* **12**, 47.
- Ferrão, R. G. (2004). Biometria aplicada ao melhoramento genético de café Conilon, Universidade Federal de Viçosa, Viçosa, MG.
- Ferrão, R. G., Ferrão, M. G. A., Fonseca, A. F. A., and Pacova, B. E. V. (2007). Melhoramento genético de *Coffea canephora*. In "Café Conilon" (R. G. Ferrão, A. F. A. Fonseca, S. M. Bragança, M. G. A. Ferrão and L. H. De Muner, eds.), Vol. 1, pp. 123-173. Incaper, Vitória, Brazil.
- Fulker, D. W., and Cardon, L. R. (1994). A sib-pair approach to interval mapping of quantitative trait loci. *Am J Hum Genet* **54**, 1092-103.
- Garcia, A. A. F., Kido, E. A., Meza, A. N., Souza, H. M. B., Pinto, L. R., Pastina, M. M., Leite, C. S., da Silva, J. A. G., Ulian, E. C., Figueira, A., and Souza, A. P. (2006). Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theoretical and Applied Genetics* **112**, 298-314.
- Gazaffi, R. (2009). Desenvolvimento de modelo genético-estatístico para mapeamento de QTLs em progênies de irmãos completos, com aplicação em cana-de-açúcar, ESALQ, Piracicaba, SP.
- Grattapaglia, D., and Sederoff, R. (1994). Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* **137**, 1121-37.
- Hanley, S. J., Pei, M. H., Powers, S. J., Ruiz, C., Mallott, M. D., Barker, J. H. A., and Karp, A. (2011). Genetic mapping of rust resistance loci in biomass willow. *Tree Genetics & Genomes* **7**, 597-608.
- Haseman, J. K., and Elston, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* **2**, 3-19.
- Jorge, V., Dowkiw, A., Faivre-Rampant, P., and Bastien, C. (2005). Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytologist* **167**, 113-127.
- Lashermes, P., Combes, M. C., Trouslot, P., and Charrier, A. (1997). Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical and Applied Genetics* **94**, 947-955.
- Lashermes, P., Couturon, E., Moreau, N., Paillard, M., and Louarn, J. (1996). Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. *Theoretical and Applied Genetics* **93**, 458-462.
- Leroy, T., De Bellis, F., Legnate, H., Kananura, E., Gonzales, G., Pereira, L. F., Andrade, A. C., Charmetant, P., Montagnon, C., Cubry, P., Marraccini, P., Pot, D., and de Kochko, A. (2011). Improving the quality of African robustas: QTLs for yield- and quality-related traits in *Coffea canephora*. *Tree Genetics & Genomes* **7**, 781-798.
- Leroy, T., Marraccini, P., Dufour, M., Montagnon, C., Lashermes, P., Sabau, X., Ferreira, L. P., Jourdan, I., Pot, D., Andrade, A. C., Glaszmann, J. C., Vieira, L. G. E., and Piffanelli, P. (2005). Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theoretical and Applied Genetics* **111**, 1032-1041.

- Lin, M., Lou, X. Y., Chang, M., and Wu, R. (2003). A general statistical framework for mapping quantitative trait loci in nonmodel systems: issue for characterizing linkage phases. *Genetics* **165**, 901-13.
- Liu, B. H. (1998). "Statistical genomics: linkage, mapping, and QTL analysis," CRC Press, Boca Raton, Florida, USA.
- Mackay, T. F., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**, 565-77.
- Maliepaard, C., Jansen, J., and Van Ooijen, J. W. (1997). Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genet Res* **70**, 237-250.
- Missio, R. F., Caixeta, E. T., Zambolim, E. M., Zambolim, L., Cruz, C. D., and Sakiyama, N. S. (2010). Polymorphic information content of SSR markers for *Coffea* spp. *Crop Breeding and Applied Biotechnology* **10**, 89-94.
- Missio, R. F., Caixeta, E. T., Zambolim, E. M., Zambolim, L., and Sakiyama, N. S. (2009). Development and validation of SSR markers for *Coffea arabica* L. *Crop Breeding and Applied Biotechnology* **9**, 361-371.
- Moncada, P., and McCouch, S. (2004). Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome* **47**, 501-9.
- Moraes, M. C. (2005). Mapas de ligação e mapeamento de QTL ("Quantitative Trait loci") em maraujá-amarelo (*Passiflora edulis* Sims f. *flavicarpa* Deg.), ESALQ/USP, Piracicaba, SP.
- Neto, P. N. S. (2011). Resistência de híbridos de *Coffea canephora* ex Froehner às raças II e XXXIII de *Hemileia vastatrix* Berk. et Br., Universidade Federal de Viçosa, Viçosa, MG.
- Pereira, G. S. (2010). Desenvolvimento de marcadores SSR, M-AFLP e SNP visando a integração de mapas genético-moleculares de *Passiflora alata* Curtis, ESALQ/USP, Piracicaba, SP.
- Poncet, V., Hamon, P., Minier, J., Carasco, C., Hamon, S., and Noirot, M. (2004). SSR cross-amplification and variation within coffee trees (*Coffea* spp.). *Genome* **47**, 1071-81.
- Prakash, N. S., Ganesh, D., and Bhat, S. S. (2005). Population dynamics of coffee leaf rust (*Hemileia vastatrix*) and recent advances in India. In "Durable resistance to coffee leaf rust" (L. Zambolim, E. M. Zambolim and V. M. P. Várzea, eds.), pp. 411-442. Universidade Federal de Viçosa, Viçosa, Brazil.
- Prakash, N. S., Marques, D. V., Várzea, V. M. P., Silva, M. C., Combes, M. C., and Lashermes, P. (2004). Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *C. arabica* L. *Theoretical and Applied Genetics* **109**, 1311-1317.
- Priyono, Florin, B., Rigoreau, M., Ducos, J. P., Sumirat, U., Mawardi, S., Lambot, C., Broun, P., Petiard, V., Wahyudi, T., and Crouzillat, D. (2010). Somatic embryogenesis and vegetative cutting capacity are under distinct genetic control in *Coffea canephora* Pierre. *Plant Cell Rep* **29**, 343-57.
- Rae, A. M., Pinel, M. P. C., Bastien, C., Sabatti, M., Street, N. R., Tucker, J., Dixon, C., Marron, N., Dillen, S. Y., and Taylor, G. (2008). QTL for yield in bioenergy *Populus*: identifying GxE interactions from growth at three contrasting sites. *Tree Genetics & Genomes* **4**, 97-112.
- Rohde, A., Storme, V., Jorge, V., Gaudet, M., Vitacolonna, N., Fabbrini, F., Ruttink, T., Zaina, G., Marron, N., Dillen, S., Steenackers, M., Sabatti, M., Morgante, M.,

- Boerjan, W., and Bastien, C. (2011). Bud set in poplar--genetic dissection of a complex trait in natural and hybrid populations. *New Phytol* **189**, 106-21.
- Rosado, T. B., Tomaz, R. S., Junior, M. F. R., Guimarães, L. M. d. S., Araújo, E. F., Alfenas, A. C., and Cruz, C. D. (2010). Detection of QTL associated with rust resistance using IBD-based methodologies in exogamic *Eucalyptus* spp. populations. *Crop Breed. Appl. Biotechnol.* **10**, 321-328.
- Rozo, Y., Escobar, C., Gaitan, A., and Cristancho, M. (2012). Aggressiveness and Genetic Diversity of *Hemileia vastatrix* During an Epidemic in Colombia. *Journal of Phytopathology* **160**, 732-740.
- Samils, B., Ronnberg-Wastljung, A. C., and Stenlid, J. (2011). QTL mapping of resistance to leaf rust in *Salix*. *Tree Genetics & Genomes* **7**, 1219-1235.
- Souza, F. F. d. (2011). Estudos sobre a diversidade, estrutura populacional, desequilíbrio de ligação e mapeamento associativo em *Coffea canephora* Pierre ex Froehner, Universidade Federal de Viçosa, Viçosa, MG.
- Tomaz, R. S. (2008). Mapeamento de loco de resistência à *Puccinia psidii* em *Eucalyptus* sp. por meio de marcadores moleculares microssatélites, UFV, Viçosa, MG.
- Vale, F. X. R., Fernandes, E. I., and Liberato, J. R. (2003). QUANT: A software plant disease severity assessment. In "Proceedings of the 8th International Congress of Plant Pathology" (R. Close, M. Braithwaite and I. Havery, eds.), Vol. 105, New Zealand.
- Van der Plank, J. E. (1975). "Principles of plant infection," Academic Press, Ney York.
- Várzea, V. M. P., Rodrigues, C. J., Silva, M. C. M. L., Gouveia, M. M., Marques, D. V., guerra-guimarães, I., and Ribeiro, A. (2002). Resistência do cafeeiro a *Hemileia vastatrix*. In "O estado da arte de tecnologias na produção de café" (L. Zambolim, ed.), Vol. 1. Universidade Federal de Viçosa, Viçosa, Brasil.
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* **93**, 77-8.
- Wagner, M., and Bettencourt, A. J. (1965). Inheritance of reaction to *Hemileia vastatrix* Berk & Br. in *Coffea arabica* L. *Coffee Rusts Research Center*.
- Wu, R., Ma, C. X., Painter, I., and Zeng, Z. B. (2002). Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theor Popul Biol* **61**, 349-63.
- Zambolim, L., Zambolim, E.M., Vale, F. X. R., Pereira, A. A., Sakiyama, N. S., and Caixeta, E. T. (2005). Physiological races of *Hemileia vastatrix* Berk. et Br. in Brazil: Physiological variability, current situation and future prospects. In "Durable Resistance to Coffee Leaf Rust" (L. Zambolim, E. M. Maciel-Zambolim and V. M. P. Várzea, eds.), Vol. 1. Universidade Federal de Viçosa, Viçosa, Brasil.

## ANEXO 1

**Anexo 1.** Marcadores SSRs utilizados na genotipagem de progênies de *Coffea canephora*.

PRIMERS MICROSSATÉLITES UTILIZADOS NO ESTUDO DE MAPEAMENTO					
PRIMERS EST-SSRs DESENVOLVIDOS NO BIOCAFÉ/BIOAGRO, UNIVERSIDADE FEDERAL DE VIÇOSA					
	Forward primer (5' - 3')	Reverse primers (5'-3')	Tm (°C)	Tamanho (pb)	Referência
CCG-03	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	
AACC-15	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	
CaEST-002	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-004	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-005	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-014	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-015	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-017	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-018	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-019	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-021	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-022	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-023	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-024	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-026	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-036	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-044	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-045	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-046	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-047	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-048	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-052	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-053	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-054	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-055	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-058	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-059	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-060	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-064	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-065	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-070	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3
CaEST-072	XXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXX	XXX	XXXXXX	Capítulo 3

CaEST-077	XXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXXXXXX	xxx	xxxxxxx	Capítulo 3
CaEST-086	XXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXXXXXX	xxx	xxxxxxx	Capítulo 3
CaEST-089	XXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXXXXXXXX	xxx	xxxxxxx	Capítulo 3

**PRIMERS SSR DA LITERATURA**

	<b>Forward primer (5' – 3')</b>	<b>Reverse primers (5'-3')</b>	<b>Tm (°C)</b>	<b>Tamanho (pb)</b>	<b>Referência</b>
SSR-005	AAGGGGAGTGGATAAGAAGG	GGCTGGATTTGTGCTTTAAG	60-55	128	Rovelli et al (2000)
SSR-011	TGCTTAGGCACTTGATATAGGA	CACGTGCAAGTCACATACTTTA	60-55	150	Rovelli et al (2000)
SSR-022	AACTCTCCATTCCC GCATT C	CTGGGTTTTCTGTGTTCTCG	60-55	89–135	Combes et al (2000)
SSR-024	TGATGGACAGGAGTTGATGG	TGCCAATCTACCTACCCCTT	60-55	100–132	Combes et al (2000)
SSR-026	CATCCGTCATAATCCAGCGTC	AGGCCAGGAAGCATGAAAGG	60–55	?	Coulibaly et al (2003)
SSR-034	GGAGACGCAGGTGGTAGAAG	TCGAGAAGTCTTGGGGTGTT	58	294	Poncet et al (2004)
SSR-036	ATGCCAAGTCGGAAGAAAGAA	CACTAGCATAATTGCGTGGA	55–50	157	Poncet et al (2004)
SSR-041	CCAACCTTCGATGGTTGTCTG	TGATAGGAGGCACTTGACACA	60–55	121	Coulibaly et al (2003)
SSR-042	TTGCTTGCTTGTCTGTTAT	TGACACGAGGTTAGAAATGA	60–55	126	Coulibaly et al (2003)
SSR-043	TTTTCTGGGTTTTCTGTGTTCTC	TAACTCTCCATTCCC GCATT C	60–55	134	Poncet et al (2004)
SSR-046	AATGAAGAAGAGGGTGGTG	CGAGGGTATTGTTTTCCAG	60–55	228	Poncet et al (2004)
SSR-048	AGCAAGTGGAGCAGAAGAAG	CGGTGAATAAGTCGCAGTC	60–55	144	Coulibaly et al (2003)
SSR-050	AGGTGCCTATCACCGTCCTC	AATCCGATGCCAAAAGTACG	60–55	169	Poncet et al (2004)
SSR-051	GGTACATGCTCCTTCCAAGA	TCCTCTGTATCTCTTTACCTTACC	55–50	148	Coulibaly et al (2003)
SSR-061	GCACACATGAAAAGATGCTTG	TGATGGACAGGAGTTGATGG	60–55	194	Coulibaly et al (2003)
SSR-062	AAAGACAAGCACTGGCACAT	CATTGATGCAAAACAAATCCA	60–55	110	Poncet et al (2004)
SSR-065	CTCACACACACTGTCACG	CGAATGAGGCTCCATCAC	60–55	108	Poncet et al (2004)
SSR-067	CGTCTCGTTTTACGCTCTCT	GATCTGCATGTA CTGGTGCTTC	60–55	237	Poncet et al (2004)
SSR-081	AGTAATGAACCTGCCGCCTCTTT	TTGTCACTTCTGTGTTTTCCATCC	55	230	Moncada e McCouch (2004)
SSR-082	CAAAATGAAGGAGAAAAGTGGACA	TGGCTTCATCTCAACCTTCCCTC	55	220	Moncada e McCouch (2004)
SSR-084	AAGTAGATTGGTGAAGGGAAGC	TCCTTCATTTCTCCTCTTG GTT	55	130	Moncada e McCouch (2004)
SSR-087	ATTCGACGACTCCAAAGCATA	CCTTGCTGGCCCTTCCCTT	55	229	Moncada e McCouch (2004)
SSR-094	TTTCTCGTCTTCCATTCCAGT	TGTACCACGTCTATCACAATG	50	219	Moncada e McCouch (2004)
SSR-102	CTCAGAGCTGCGGTGGTGTC A	CCGGACGATCTTTCTTTCTTCA	55	282	Moncada e McCouch (2004)
SSR-121	CGACACTTTCTTTGGCACTC	AGACACCCACCCATCCAC	55–50	177	Poncet et al (2004)
SSR-123	GATAGGAACCCAACCACGAA	GAGGGAGTCAAGTTTTCAAGGA	60–55	165	Poncet et al (2004)
SSR-138	GCACTTCCCTCTCAACCAAC	ACTAGGACAGCAAATAGCATAACC	60–55	213	Poncet et al (2004)
SSR-141	ACACATACATACCCCAAATGC	TTTCAGCAATACCTTGTCTATCAA	60–55	154	Poncet et al (2004)
SSR-146	GGTCGCTCACTCATATCTTCCAG	GCCTGGAAGCAAACGTCTCA	57	137–151	Bhat et al (2005)
SSR-148	CAGGCACAGAAGGAATGAAGAGC	TGGTGGTATGGA AACAGGAAGG	57	129–133	Bhat et al (2005)
SSR-149	GGCCGCTCGTATCTGCTATGG	GGTAACCAGAGCCCCGTGA ACT	57	193–199	Bhat et al (2005)
SSR-150	TCCATAACACAGAAATTTCCAGAGAGA	ACCGTAACCTCCGTCTTCGA ACTG	57	142–145	Bhat et al (2005)
SSR-152	CGTGCTACTCCCTCTCCTCTCACT	GAGGCTGAGGCTTGAAGGAAATTAAT	57	114–120	Bhat et al (2005)
SSR-154	TGAGGGCAAAGGAGTAAGAAAAG	TCAAACCTTCAACAATCAAATACCC	55	306	Moncada e McCouch (2004)
SSR-155	ATTCGACGACTCCAAAGCATA	CCTTGCTGGCCCTTCCCTT	55	229	Moncada e McCouch (2004)

**PRIMERS SSR DESENVOLVIDOS PARA *Coffea arabica***

	<b>Forward primer (5' – 3')</b>	<b>Reverse primers (5'-3')</b>	<b>T<sub>m</sub> (°C)</b>	<b>Tamanho (pb)</b>	<b>Referência</b>
SSRCa-04	CCATGAGCACTTGTCCATAAA	ATCAAAGAACAAACCCGACA	58	287	Missio et al. 2009
SSRCa-11	ATCCAACCAACCATTGAAAC	CATCCACTTTTTCCACCTTC	57	347	Missio et al. 2009
SSRCa-15	TCGCAATAACCAATCACAAG	AGCTATTGACCCCACTGAAA	57	273	Missio et al. 2009
SSRCa-19	GGGTTAGATAGAGCAAGAATGA	CTGTGAAGGTGTGGAGTTTT	55	329	Missio et al. 2009
SSRCa-40	AGGGATGTAGAACCAGCAAA	CCAATAGCTACAACAAAGG	56	283	Missio et al. 2009
SSRCa-46	ATGAAGAGGGGTCCATCA	CATAGACTTTTCTTGCCTCCT	57	254	Missio et al. 2009

## **6. CAPÍTULO 3: DESENVOLVIMENTO E VALIDAÇÃO DE EST-SSR PARA *Coffea arabica*: TRANSFERABILIDADE E APLICAÇÃO EM ESTUDOS DE CARACTERIZAÇÃO MOLECULAR E MAPEAMENTO GENÉTICO.**

### **RESUMO**

Marcadores microssatélites (SSRs) têm ampla utilidade em estudos genéticos de plantas, especialmente, pela elevada taxa de polimorfismo, natureza codominante, multialelismo e reprodutibilidade. Os EST-SSR são marcadores derivados de sequências expressas do genoma e representam genes transcritos que são altamente conservados nas espécies correlacionadas. Apesar da importância, um número reduzido de marcadores EST-SSR são listados na literatura para a espécie *Coffea arabica*. Dessa forma, este estudo teve como finalidade a mineração, desenvolvimento e validação de marcadores EST-SSR a partir do Projeto Brasileiro do Genoma Café. Para isso, foram investigadas 130.792 sequências ESTs (*Expressed Sequence Tags*), sendo identificadas 37.826 contendo microssatélites, que após análise de agrupamento e de redundância resultaram em 24.031. Foram minerados todos os SSR perfeitos com di-, tri- e tetranucleotídeos de tamanho mínimo de 12pb. Os tetranucleotídeos foram os mais frequentes. A partir das sequências mineradas e após o teste de estabilidade e amplificação, foram validados neste estudo 101 novos marcadores EST-SSR para *C. arabica*. Visando avaliar a eficiência desses marcadores, os EST-SSR foram utilizados em estudos de diversidade, caracterização molecular e mapeamento genético. A taxa de transferabilidade em espécies correlacionadas foi mensurada e o valor de 88% mostrou que esses marcadores são úteis em estudos genéticos dentro do gênero *Coffea*. O resultado do nível de polimorfismo e do estudo de diversidade foi coerente com a história evolutiva das espécies, confirmando a grande similaridade genética de *C. arabica* e dos Híbridos de Timor. Por outro lado, na espécie *C. canephora* os EST-SSR foram muito polimórficos e capazes de discriminar grupos varietais em acessos de Bancos de Germoplasma. Nos estudos de mapeamento genético, foi possível mapear 17 e 23 EST-SSR nos grupos de ligação de *C. arabica* e *C. canephora*, respectivamente, demonstrando que esses marcadores podem ser úteis em estudos de mapeamento de QTL e seleção assistida. Os resultados obtidos demonstraram que os novos marcadores EST-SSR constituem uma ferramenta molecular valiosa com potencial uso em estudo

genético, seleção de genitores em programas de melhoramento e no manejo de recursos genéticos em bancos de germoplasma, bem como estudos evolutivos e de filogenia molecular dentro do gênero *Coffea*.

**Palavras chaves:** *Coffea sp*; café; microssatélites; Projeto Brasileiro do Genoma Café; marcadores moleculares.

## INTRODUÇÃO

O cafeeiro é membro da família *Rubiaceae* e pertencente ao gênero *Coffea* que compreende centenas de espécies tropicais. Dentro dessas, duas merecem destaque por serem comercialmente cultivadas: *Coffea arabica* ( $2n=4x=44$ ) plantadas em regiões altas e de clima ameno; e *Coffea canephora* ( $2n=2x=22$ ) de baixas altitudes e de clima quente. A bebida produzida por essas duas espécies, normalmente, são servidas na forma de *blends* e são consumidas regularmente por 40% da população mundial. Esse fato faz do café uma das bebidas não alcoólicas mais populares do mundo, ocupando posição estratégica no desenvolvimento social e econômico dos países produtores.

Esforços globais têm sido realizados para melhorar a qualidade do café. Neste sentido, os programas de melhoramento genético ganham destaque, sobretudo, pela melhoria dos aspectos agrônômicos envolvidos na produção e comercialização de grãos. Apesar do sucesso notável, os avanços têm se mostrado demasiadamente lentos devido a diferentes fatores, tais como: composição genética e fisiológica (baixa diversidade genética em *C. arabica* e auto-incompatibilidade em *C.canephora*), longos ciclos de geração, dependência de grandes áreas para experimentação e escassez de técnicas fáceis e acessíveis de ferramentas genéticas para avaliação e seleção (Hendre et al., 2008). Este cenário propicia a busca de novas ferramentas que possam direcionar os programas de melhoramento, visando alcançar uma maior eficiência na transferência de fatores genéticos. Neste contexto, o uso de marcadores moleculares torna-se imperativo, pois permite acessar informações em nível de DNA com confiabilidade e clareza. Dentre os diferentes tipos de marcadores, os microssatélites ou SSR (*simple sequence repetitives*) destacam-se por agregarem características importantes, tais como: elevado polimorfismo, natureza codominante, multialelismo, loco especificidade, reprodutibilidade e facilidade de automatização. Essas características fazem deste

marcador o preferido para estudos de mapeamento genético, seleção assistida (MAS), diversidade genética, análise de QTL e manutenção de germoplasma. Apesar dessas vantagens, o número de marcadores SSRs para o gênero *Coffea* é limitado, especialmente, pelas dificuldades de desenvolvimento de *primers* específicos.

A identificação de locos microssatélites, normalmente, ocorre por duas abordagens: via bibliotecas genômicas de DNA ou por sequenciamento de cDNA. O uso de bibliotecas genômicas de DNA, apesar de relativamente simples apresenta baixa eficiência, custo elevado e é laborioso; o que limita o uso na maioria das espécies vegetais (Guichoux et al., 2011). Por outro lado, nas últimas décadas, o avanço nas técnicas de sequenciamento permitiu que milhares de sequências expressas de DNA (ESTs) fossem sequenciadas e depositadas em bancos públicos e privados de DNA. Sequências ESTs são fragmentos sequenciados de DNA complementares às sequências de RNA mensageiro e representa parte de uma região transcrita do genoma em certas condições fisiológicas. Esse panorama abriu a oportunidade para o desenvolvimento de marcadores SSRs com maior facilidade e menores custos. Assim, EST-SSR passaram a ser desenvolvidos com a vantagem de estarem relacionados com porções funcionais do genoma e apresentarem altos níveis de transferabilidade entre espécies relacionadas (Guichoux et al., 2011; Kalia et al., 2010).

Para o gênero *Coffea* está disponível um banco de dados com mais de 200 mil ESTs proveniente do Projeto Brasileiro do Genoma Café (Vieira et al., 2006). Nesse Projeto foram sequenciados milhares de clones escolhidos aleatoriamente em bibliotecas de cDNA das espécies *C. arabica*, *C. canephora* e *C. racemosa* representando estádios específicos do desenvolvimento de células e de tecidos do cafeeiro. As sequências disponibilizadas no Projeto Brasileiro do Genoma Café constitui-se em um valioso banco de dados para desenvolvimento de marcadores EST-SSR. Dessa forma, o presente trabalho, objetivou a mineração de microssatélites e a caracterização de novos marcadores EST-SSR para *C.arabica*. A aplicação desses marcadores na caracterização molecular e no mapeamento genético, além da taxa de transferabilidade dentro do gênero *Coffea*, também foi mensurada.

## MATERIAIS E MÉTODOS

### *Material Vegetal*

Para estudos de caracterização molecular, diversidade genética e transferabilidade dos EST-SSR foram utilizados 24 genótipos (Tabela 1-I), os quais incluem as espécies *C. arabica*, *C. canephora*, *C. eugenioides*, *C. racemosa*, *C. congensis*; e híbridos interespecíficos (Híbridos de Timor e um triplóide natural). Todos os genótipos pertencem ao Banco de Germoplasma da Universidade Federal de Viçosa (UFV), Brasil; e são de interesse do programa de melhoramento. Os genótipos da espécie *C. arabica*, apresentam origens diferentes e por isso foram codificados da seguinte maneira: CAR são os acessos que pertencem à coleção de germoplasma; e os CAR-V são variedades comerciais resultante de hibridizações nos programas de melhoramento.

A avaliação dos EST-SSR em estudos específicos com *C. canephora* foi realizada utilizando 40 acessos do Banco de Germoplasma da Embrapa Rondônia, Brasil (Tabela 1-II). A origem e classificação dos acessos em Conilon e Robusta foi relatado por Ferrão et al. (2013).

Os estudos de mapeamento genético foram realizados utilizando 247 progênies F<sub>2</sub> de *C. arabica* (Pestana, 2010) e 96 progênies de irmãos completos de *C. canephora* (Tabela 1-III). Essas duas populações são mantidas, respectivamente, pela Universidade Federal de Viçosa e pelo Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper), Brasil.

**Tabela 1.** Genótipos utilizados na validação dos EST-SSR.

I Caracterização Molecular, Diversidade Genética e Transferabilidade		
Genótipos	Espécie	Código
UFV 2144 (Catuaí Vermelho IAC 44)	<i>Coffea arabica</i>	CAR.1
Típica UFV 2945	<i>Coffea arabica</i>	CAR.2
Bourbon UFV 2952	<i>Coffea arabica</i>	CAR.3
Bourbon Amarelo UFV 535-1	<i>Coffea arabica</i>	CAR.4
Arábica UFV 10832	<i>Coffea arábica</i>	CAR.5
Bourbon Amarelo UFV 10745	<i>Coffea arábica</i>	CAR.6
T 3751 (Robusta)	<i>Coffea canephora</i>	CCA.1
T 3580 (Robusta)	<i>Coffea canephora</i>	CCA.2
Conillon UFV 513 (Conillon)	<i>Coffea canephora</i>	CCA.3
Guarini UFV 514 (Robusta)	<i>Coffea canephora</i>	CCA.4
Apoatã IAC 2258 (Robusta)	<i>Coffea canephora</i>	CCA.5
Híbrido de Timor CIFC 832/2	<i>C. arabica</i> x <i>C. canephora</i>	HT.1
Híbrido de Timor CIFC 4106	<i>C. arabica</i> x <i>C. canephora</i>	HT.2
Híbrido de Timor CIFC 1343/269	<i>C. arabica</i> x <i>C. canephora</i>	HT.3

UFV 557-2	Triplóide ( <i>C. arabica</i> x <i>C. racemosa</i> )	TRIP
UFV Racemosa	<i>Coffea racemosa</i>	CRAC
UFV Eugeniodes	<i>Coffea eugeniodes</i>	CEUG
UFV Congensis	<i>Coffea congensis</i>	CCON
Catiguá MG2*	<i>C. arabica</i> x HT	CAR.V1
IAPAR 59*	<i>C. arabica</i> x HT	CAR.V2
Oeiras MG6851*	<i>C. arabica</i> x HT	CAR.V3
Sacramento MG1*	<i>C. arabica</i> x HT	CAR.V4
Catucaí Amarelo 2SL*	<i>C. arabica</i> x Icatu vermelho	CAR.V5
Obatã Amarelo IAC 4932*	<i>C. arabica</i> x HT	CAR.V6

### II Estudo de Diversidade em *Coffea canephora*

Genótipo	Grupo	Código	Genótipo	Grupo	Código
Conilon Incaper 03	Conilon	ES03	Cpafro 138	Conilon	RO138
Conilon Incaper 16	Conilon	ES028	Cpafro 139	Conilon	RO139
Conilon Incaper 28	Conilon	ES110A	Cpafro 141	Conilon	RO141
Conilon Incaper 45	Conilon	ES16	Cpafro 142	Conilon	RO142
Conilon Incaper 110	Conilon	ES45	Cpafro 147	Conilon	RO147
Conilon Incaper V.1	Conilon	ESV1	Cpafro 161	Conilon	RO161
Conilon Incaper V.2	Conilon	ESV2	Cpafro 164	Conilon	RO164
Conilon Incaper V.3	Conilon	ESV4	Cpafro 167	Conilon	RO167
Cpafro 017	Conilon	RO017	Cpafro 171	Conilon	RO171
Cpafro 025	Conilon	RO025	Cpafro 196	Conilon	RO196
Cpafro 030	Conilon	RO030	Cpafro 197	Conilon	RO197
Cpafro 035	Conilon	RO035	Cpafro 203	Conilon	RO203
Cpafro 036	Conilon	RO036	Robusta IAC 1675	Robusta	R1675
Cpafro 038	Conilon	RO038	Robusta IAC 2257.2	Robusta	R22572
Cpafro 072	Conilon	RO072	Robusta IAC 2258.1	Robusta	R22581
Cpafro 073	Conilon	RO073	Robusta IAC 2258.2	Robusta	R22582
Cpafro 075	Conilon	RO075	Robusta IAC 2258.3	Robusta	R22583
Cpafro 096	Conilon	RO096	Robusta IAC 2259	Robusta	R2259
Cpafro 101	Conilon	RO101	Robusta IAC 640.1	Robusta	R6401
Cpafro 125	Conilon	RO125	Robusta IAC 640.2	Robusta	R6402

### III Parental e População de Mapeamento Genético

Espécie	Parental	Numero de Progênes
<i>Coffea arabica</i>	Híbrido de Timor (UFV443-03) x Catuaí Amarelo (UFV 2148-57)	247 progênes F <sub>2</sub>
<i>Coffea canephora</i>	Conilon02 x Conilon83	96 progênes de irmãos completos

\* Variedades comerciais.

### Mineração e caracterização de EST-SSR do Projeto Brasileiro do Genoma Café

As sequências EST utilizadas no trabalho foram geradas pelo Projeto Brasileiro do Genoma Café, como descrito por Vieira et al. (2006). Neste projeto, um total de 214.964 clones das espécies *C. arabica*, *C. canephora* e *C. racemosa* foram sequenciados e após processo de triagem dos dados, foram disponibilizados para cada espécie, 130.790, 12.381 e 10.566 sequências, respectivamente. No presente estudo,

apenas as EST de *C. arabica* foram investigadas. Para identificação dos EST-SSR foram consideradas todas as combinações de repetições com dois (DNRs), três (TNRs) e quatro (TtNRs) nucleotídeos, totalizando 46 classes diferentes (4 DNRs, 10 TNRs e 32 TtNRs). As minerações foram realizadas considerando SSRs com tamanhos mínimos que variavam de 12pb a 20pb. Em todas as minerações foram considerados apenas SSRs perfeitos.

A mineração dos dados foi realizada na plataforma de bioinformática do Projeto Brasileiro do Genoma Café (<http://lge.ibi.unicamp.br/cafe>). A identificação e localização dos microssatélites foram realizadas pelo *software* Gramene (<http://gramene.org/db/searches/ssrtool>). Os *softwares* Primer3 (Rozen and Skaletsky, 2000) e PrimerSelect (Plasterer, 1997) foram utilizados, respectivamente, para o desenho de *primers* e avaliação da estabilidade.

#### *Extração do DNA gênomico e amplificação de PCR.*

Para as análises moleculares, folhas jovens e completamente desenvolvidas de cada cafeeiro foram coletadas, congeladas a -80 °C, liofilizadas, trituradas e conservadas a -20 °C no Laboratório de Biotecnologia do Cafeeiro (BioCafé/Bioagro), na Universidade Federal de Viçosa (UFV). O DNA genômico foi extraído conforme protocolo descrito por Diniz et al. (2005). A amplificação dos microssatélites foi realizada conforme relatado por Missio et al. (2010).

#### *Validação dos marcadores: análises de polimorfismo e transferabilidade dos EST-SSR*

O teste de *primers* e a análise de polimorfismos dos EST-SSR foram realizados utilizando doze genótipos da espécie *C. arabica*, cinco genótipos de *C. canephora* e três Híbridos de Timor (Tabela 1- I). O *software* TETRASAT (Markwith and Scanlon, 2007), específico para tetraploides, foi utilizado nas análises em *C. arabica* e Híbrido de Timor. Para a espécie *C. canephora*, diploide, foi utilizado o *software* PowerMarker v3.25 (Liu and Muse, 2005). Os seguintes parâmetros foram calculados:

- 1) Número de alelos acessado por cada marcador (Na);
- 2) Heterozigidade: dada pela fórmula  $He = 1 - \sum pi^2$ , onde  $pi$  é a frequência alélica para  $i$ ésimo alelo.
- 3) Índice de Diversidade de Shannon-Weiner ( $H'$ ): calculada pela fórmula  $H' = - \sum pi^2 \log_2 pi$ , onde  $pi$  é a frequência alélica para  $i$ ésimo alelo.

4) Conteúdo Médio de Informação Polimórfica (PIC): calculado de acordo com Botstein et al. (1980), pela fórmula  $PIC = 1 - [\sum_{i=1}^a pi^2 - \sum_{i,j=1}^a \sum_{(i \neq j)} pi^2 pj^2]$ , onde  $pi$  é a frequência alélica para  $i$ -ésimo alelo.

5) Poder Discriminante (DP): proposto por Tessier et al. (1999), calculado para o  $k$ -ésimo marcador de acordo com a fórmula  $DP_K = 1 - \sum_{j=1}^i p_j \frac{N p_j - 1}{N - 1}$ , onde  $N$  é o número de indivíduos e  $p_j$  é a frequência do  $j$ -ésimo padrão.

O PIC foi utilizado para mensurar o conteúdo informacional de cada loco marcador. Assim, aqueles EST-SSR que apresentaram valores nulos de PIC foram considerados monomórficos. O DP foi o parâmetro utilizado para diferenciação de genótipos. Assim, marcadores com valores nulos de DP apresentaram um único padrão de banda e não possibilitaram a diferenciação dos genótipos amostrados.

A taxa de transferabilidade dos EST-SSR ( $T_{MAR}$ ) é uma medida das espécies e foi calculada considerando a proporção de *primers* e o sucesso de amplificação em cada espécie, de acordo com a fórmula:  $T_{MAR} = SA_{esp} / NM_{esp}$ , onde  $SA_{esp}$  é o sucesso de

amplificação para cada espécie e  $NM_{esp}$  é o número total de *primers* utilizado nas análises de cada espécie. A taxa de conservação ( $C_{MAR}$ ) é uma medida dos marcadores e foi calculada considerando a proporção de espécies e o sucesso de amplificação de cada marcador, de acordo com a fórmula:  $C_{MAR} = SA_{marc} / NE_{marc}$ , onde  $SA_{marc}$  é o sucesso

de amplificação para cada marcador e  $NE_{marc}$  é o número total de espécies avaliadas para cada marcador. Para essas análises, foram considerados todos os genótipos da Tabela1-I e uma amostragem aleatória de 74 EST-SSR. Os marcadores que apresentaram locos duplicados foram considerados como ampliações independentes.

#### *Caracterização molecular de acessos de cafeeiro*

Para as análises de caracterização molecular e diversidade genética foram realizadas análises de agrupamentos baseados em coeficientes de similaridade genética e abordagem bayesiana. Em ambas as análises optaram-se pela avaliação binária dos dados de genotipagem, devido aos diferentes níveis de ploidia das espécies estudadas (Tabela1-I). O coeficiente de similaridade genética foi calculado utilizando o Índice de

Similaridade de Jaccard (1908) no módulo SIMQUAL do *software* NTSYS-pc Version 2.10 (Rohlf, 2008). Um dendrograma *Neighbor-Joining* (NJ) foi construído baseado na matriz de distância genética de Jaccard utilizando o módulo NJoin do mesmo programa. As análises de agrupamento bayesiana foram realizadas pelo *software* STRUCTURE (Pritchard et al., 2000), no qual foi considerado valores de K variando de 1 a 10 com modelo de mistura genética e 15 repetições. Cada corrida foi implementada com um período de 50.000 *burn-in* seguido por 100.000 MCMC. O verdadeiro número de grupos genéticos foi estimado pelo cálculo de  $\Delta K$ , como descrito por Evanno et al. (2005), utilizando o *software* STRUCTURE HARVESTER (Earl and vonHoldt, 2012).

A eficácia dos EST-SSR em estudos específicos com *C. canephora* foi avaliada considerando 40 acessos (Tabela1-II) e 10 EST-SSR tomados aleatoriamente (CaEST-013, CaEST-017, CaEST-025, CaEST-053, CaEST-058, CaEST-059, CaEST-060, CaEST-063, CaEST-068 e CaEST-072) . Os estudos de caracterização molecular e diversidade genética dentro da espécie foram realizados conforme descrito por Ferrão et al. (2013).

#### *Estudos de mapeamento genético*

O posicionamento dos EST-SSR em mapas de ligação foi realizado nas espécies *C. arabica* e *C. canephora* (Tabela1-III). O mapa de ligação de *C. arabica* foi desenvolvido por Pestana (2010). Para a espécie *C. canephora*, o mapa de ligação foi construído considerando LOD 2 e frequência de recombinação de 40%. O cálculo das distâncias genéticas (cM) foi realizado no *software* R utilizando o pacote *Onemap* (Margarido et al., 2007). Para o ordenamento dos marcadores dentro dos grupos de ligação foi utilizando o algoritmo *Unidirectional Growth* (Tan and Fu, 2006). A representação gráfica do mapa de ligação foi feita no *software* MapChart (Voorrips, 2002).

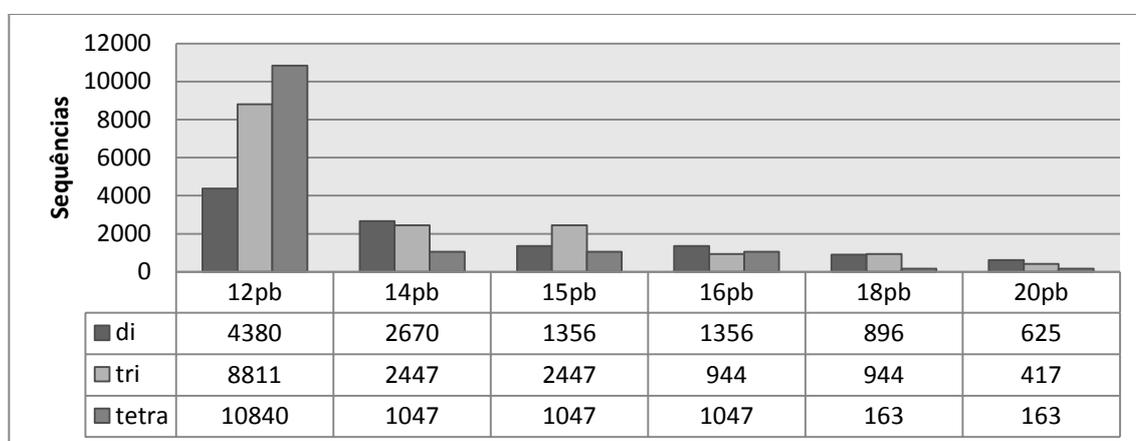
## **RESULTADOS**

### *Mineração e caracterização de EST-SSR do Projeto Brasileiro do Genoma Café*

Foram analisadas 130.792 EST de *C. arabica*, totalizando 59,6mega pares de bases, com uma média de 456pb por EST (Vieira et al., 2006). A maioria (58,9%) das ESTs avaliadas apresentaram comprimentos superiores a 500pb. O número de unidades repetitivas dos EST-SSR usado para a mineração variou de 12 a 20 pb (Figura 1). Todas

as sequências mineradas foram armazenadas em banco de dados. Para o desenvolvimento dos marcadores foram utilizadas as sequências cujo critério de mineração foi repetição com número mínimo de 12pb.

O resultado da mineração dos dados do Projeto Brasileiro do Genoma Café, utilizando o critério de repetição mínima de 12 pb, permitiu a identificação de 37.826 EST contendo microssatélites. Após análise de agrupamento e de redundância, o número de sequências obtidas foi reduzida para 24.031 (Figura 1). Dentre essas, 4380 (18,23%) continham DNRs, 8811 (36,67%) continham TNRs e 10840 (45,11%) TtNTRs. As classes mais abundantes de DNRs, TNRs e TtNTRs encontradas foram, respectivamente, (AG)n; (AGG)n; e (AGGG)n (Anexo 1).



**Figura 1.** Avaliação de diferentes critérios de mineração dos dados do Projeto Brasileiro do Genoma Café. Número mínimo de SSRs variando de 12 a 20pb.

#### *Validação dos marcadores: análises de polimorfismo e transferabilidade dos EST-SSR*

No total, 146 pares de *primers* específicos para *C. arabica* foram desenhados, sendo que desses, 101 (36,6% DNRs; 37,6% TNRs; 25,8% TtNTRs) apresentaram amplificação com bandas nítidas em cafeeiros da espécie *C. arabica*. Os 101 marcadores selecionados foram validados e os níveis de polimorfismo foram mensurados nas espécies cultivadas *C. arabica* e *C. canephora*; e no híbrido interespecífico Híbrido de Timor (Tabela 2). A definição de marcadores monomórficos foi realizada considerando valores nulos de  $H_e$ ,  $H'$  e PIC. O poder dos EST-SSR em discriminar genótipos foi mensurado de acordo com o parâmetro DP. A frequência de marcadores que não funcionaram (NW) e de locos duplicados (LD) também foram mensurados.

Os genótipos de *C. arabica* e Híbridos de Timor apresentaram o mesmo número de locos duplicados (4) e 31 e 25 locos monomórficos, respectivamente. Todos os marcadores funcionaram para os híbridos interespecíficos e apresentaram bons padrões de amplificação, contendo bandas nítidas. A espécie *C. canephora* apresentou o menor número de locos monomórficos (16), contudo apresentou maior frequência de locos duplicados e marcadores que não amplificaram (15 e 9, respectivamente), o que era esperado pelo fato dos EST-SSR terem sido desenvolvidos para a espécie *C. arabica*.

**Tabela 2.** Diversidade alélica dos novos EST-SSR avaliados em *C. arabica*, Híbridos de Timor e *C. canephora*.

Primers	<i>C. arabica</i> (N=12)					Híbrido de Timor (N=3)					<i>C. canephora</i> (N=5)							
	Na	He	H'	PIC	DP	Na	He	H'	PIC	DP	Na	He	H'	PIC	DP			
CaEST-001 <sup>B</sup>	4	0,75	2,00	0,70	0,17	4	0,74	1,96	0,69	0,67	4	0,64	1,33	0,58	0,90			
CaEST-002 <sup>AB</sup>	4	0,71	1,88	0,68	0,68	4	0,70	1,85	0,67	0,67	loco duplicado							
CaEST-003	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	loco duplicado							
CaEST-004 <sup>A</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	3	0,62	1,33	0,55	0,90			
CaEST-005	4	0,75	2,00	0,70	0,00	4	0,75	2	0,70	0,00	3	0,54	0,95	0,47	0,70			
CaEST-006 <sup>AB</sup>	4	0,68	1,70	0,62	0,17	6	0,82	2,54	0,79	0,67	4	0,58	1,33	0,54	0,90			
CaEST-007	3	0,66	1,57	0,59	0,00	3	0,66	1,57	0,59	0,00	2	0,50	0,00	0,38	0,00			
CaEST-008	4	0,71	1,88	0,68	0,68	5	0,70	1,98	0,70	1	4	0,58	1,33	0,54	0,90			
CaEST-009	4	0,68	1,74	0,61	0,17	4	0,70	1,85	0,67	0,67	2	0,32	0,50	0,27	0,40			
CaEST-010	6	0,78	2,30	0,73	0,32	4	0,75	2	0,70	0,00	3	0,46	1,05	0,41	0,80			
CaEST-011	1	<b>monomórfico</b>					3	0,49	1,21	0,45	0,67	5	0,68	1,33	0,64	0,90		
CaEST-012	3	0,55	1,28	0,43	0,17	2	0,49	0,98	0,38	0,00	2	0,42	0,95	0,33	0,60			
CaEST-013	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	3	0,34	0,95	0,31	0,60			
CaEST-014 <sup>A</sup>	2	0,49	0,98	0,38	0,00	4	0,71	1,90	0,69	0,67	3	0,61	0,64	0,54	0,90			
CaEST-015 <sup>A</sup>	1	<b>monomórfico</b>					1	<b>monomórfico</b>					3	0,61	1,10	0,54	1	
CaEST-016	1	<b>monomórfico</b>					1	<b>monomórfico</b>					3	0,54	1,33	0,47	0,90	
CaEST-017 <sup>A</sup>	3	0,53	1,18	0,43	0,17	2	0,49	0,98	0,38	0,00	4	0,66	1,33	0,61	0,90			
CaEST-018 <sup>A</sup>	loco duplicado					loco duplicado					loco duplicado							
CaEST-019 <sup>A</sup>	2	0,49	0,98	0,38	0,00	3	0,65	1,56	0,59	0,00	4	0,70	1,33	0,65	0,90			
CaEST-020	1	<b>monomórfico</b>					1	<b>monomórfico</b>					1	<b>monomórfico</b>				
CaEST-021 <sup>A</sup>	2	0,49	0,98	0,38	0,00	3	0,63	1,50	0,58	0,67	4	0,74	1,33	0,69	0,90			
CaEST-022 <sup>AB</sup>	1	<b>monomórfico</b>					2	0,27	0,63	0,16	0,67	5	0,74	1,33	0,70	0,90		
CaEST-023 <sup>A</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	4	0,66	1,05	0,61	0,80			
CaEST-024 <sup>AB</sup>	2	0,49	0,98	0,38	0,00	2	0,43	0,89	0,36	0,67	3	0,46	1,05	0,41	0,80			
CaEST-025	1	<b>monomórfico</b>					1	<b>monomórfico</b>					3	0,64	1,05	0,56	0,80	
CaEST-026 <sup>A</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	2	0,32	0,50	0,27	0,40			
CaEST-027	1	<b>monomórfico</b>					1	<b>monomórfico</b>					1	<b>monomórfico</b>				
CaEST-028	3	0,53	1,18	0,43	0,17	3	0,58	1,40	0,54	0,67	loco duplicado							
CaEST-029	3	0,61	1,45	0,59	0,41	3	0,63	1,50	0,58	0,67	4	0,74	1,33	0,69	0,90			
CaEST-030	3	0,64	1,53	0,59	0,17	3	0,66	1,57	0,59	0,00	loco duplicado							
CaEST-031 <sup>B</sup>	3	0,40	1,05	0,36	0,41	3	0,27	0,63	0,24	0,67	2	0,32	0,50	0,27	0,40			
CaEST-032 <sup>A</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	4	0,58	0,95	0,54	0,70			
CaEST-033	loco duplicado					loco duplicado					loco duplicado							
CaEST-034 <sup>B</sup>	3	0,55	1,28	0,46	0,17	3	0,63	1,51	0,55	1	3	0,59	1,04	0,51	0,90			
CaEST-035	1	<b>monomórfico</b>					1	<b>monomórfico</b>					5	0,80	1,61	0,77	1	
CaEST-036 <sup>A</sup>	3	0,55	1,28	0,46	0,49	2	0,63	1,51	0,55	0,67	2	0,42	0,67	0,33	0,60			
CaEST-037	1	<b>monomórfico</b>					2	0,63	1,51	0,55	0,67	loco duplicado						
CaEST-038	1	<b>monomórfico</b>					1	<b>monomórfico</b>					não funciona					
CaEST-039	1	<b>monomórfico</b>					1	<b>monomórfico</b>					1	<b>monomórfico</b>				
CaEST-040 <sup>B</sup>	4	0,60	1,49	0,52	0,56	3	0,63	1,50	0,55	0,67	4	0,70	1,33	0,65	0,90			
CaEST-041	2	0,48	0,96	0,36	0,17	2	0,49	0,98	0,38	0,00	1	<b>monomórfico</b>						
CaEST-042	1	<b>monomórfico</b>					1	<b>monomórfico</b>					1	<b>monomórfico</b>				
CaEST-043	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	1	<b>monomórfico</b>						
CaEST-044 <sup>A</sup>	3	0,66	1,57	0,59	0,00	3	0,66	1,57	0,59	0,00	loco duplicado							
CaEST-045 <sup>AB</sup>	3	0,54	1,21	0,43	0,17	3	0,57	1,35	0,48	0,67	3	0,54	1,33	0,47	0,90			
CaEST-046 <sup>A</sup>	3	0,66	1,57	0,59	0,00	3	0,66	1,57	0,59	0,00	2	0,32	0,50	0,27	0,40			

CaEST-047 <sup>A</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	4	0,66	1,33	0,61	0,90
CaEST-048 <sup>AB</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	1	<b>monomórfico</b>			
CaEST-049	1	<b>monomórfico</b>				1	<b>monomórfico</b>				<b>não funciona</b>				
CaEST-050	loco duplicado					loco duplicado					loco duplicado				
CaEST-051 <sup>A</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	2	0,32	0,67	0,27	0,60
CaEST-052 <sup>A</sup>	1	<b>monomórfico</b>				5	0,53	1,58	0,50	0,67	loco duplicado				
CaEST-053 <sup>AB</sup>	3	0,36	0,88	0,31	0,44	3	0,65	1,54	0,58	0,67	5	0,78	1,33	0,74	0,90
CaEST-054 <sup>A</sup>	2	0,49	0,98	0,38	0,00	3	0,63	1,50	0,55	0,67	2	0,32	0,67	0,27	0,60
CaEST-055 <sup>A</sup>	1	<b>monomórfico</b>				1	<b>monomórfico</b>				<b>monomórfico</b>				
CaEST-056	2	0,44	0,89	0,37	0,32	1	<b>monomórfico</b>				2	0,22	0,56	0,19	0,70
CaEST-057	1	<b>monomórfico</b>				1	<b>monomórfico</b>				loco duplicado				
CaEST-058 <sup>AB</sup>	3	0,56	1,30	0,46	0,32	2	0,49	0,98	0,38	0,00	6	0,80	1,61	0,77	1
CaEST-059 <sup>A</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	6	0,76	1,33	0,73	1
CaEST-060 <sup>A</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	4	0,66	0,50	0,60	0,40
CaEST-061	3	0,51	1,13	0,40	0,17	4	0,71	1,90	0,66	0,67	4	0,58	1,05	0,54	0,70
CaEST-062 <sup>B</sup>	3	0,56	1,31	0,48	0,46	3	0,63	1,50	0,55	0,67	2	0,18	0,50	0,16	0,40
CaEST-063	loco duplicado					loco duplicado					loco duplicado				
CaEST-064 <sup>A</sup>	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	4	0,72	1,61	0,67	1
CaEST-065 <sup>A</sup>	1	<b>monomórfico</b>				4	0,42	1,21	0,39	0,67	4	0,70	1,33	0,65	0,90
CaEST-066	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	3	0,34	0,95	0,31	0,70
CaEST-067	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	3	0,58	0,95	0,49	0,70
CaEST-068	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	4	0,72	1,33	0,67	0,90
CaEST-069	1	<b>monomórfico</b>				1	<b>monomórfico</b>				loco duplicado				
CaEST-070 <sup>A</sup>	3	0,52	1,13	0,43	0,17	3	0,63	1,50	0,55	0,67	3	0,46	0,95	0,41	0,70
CaEST-071 <sup>B</sup>	1	<b>monomórfico</b>				2	0,49	0,98	0,38	0,00	2	0,48	1,05	0,36	0,80
CaEST-072 <sup>4B</sup>	2	0,49	0,98	0,38	0,00	3	0,63	1,50	0,55	0,67	1	<b>monomórfico</b>			
CaEST-073	3	0,53	1,18	0,40	0,17	3	0,57	1,35	0,50	0,67	4	0,48	0,95	0,45	0,70
CaEST-074	2	0,49	0,98	0,38	0,00	2	0,35	0,77	0,30	1	<b>não funciona</b>				
CaEST-075	1	<b>monomórfico</b>				1	<b>monomórfico</b>				1	<b>Monomórfico</b>			
CaEST-076	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	<b>não funciona</b>				
CaEST-077 <sup>A</sup>	1	<b>monomórfico</b>				1	<b>monomórfico</b>				<b>não funciona</b>				
CaEST-078	1	<b>monomórfico</b>				2	0,49	0,98	0,38	0,00	1	<b>monomórfico</b>			
CaEST-079	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	4	0,64	0,95	0,58	0,70
CaEST-080	3	0,53	1,18	0,40	0,17	3	0,58	1,40	0,50	0,67	3	0,58	0,50	0,49	0,40
CaEST-081 <sup>A</sup>	1	<b>monomórfico</b>				1	<b>monomórfico</b>				1	<b>monomórfico</b>			
CaEST-082	1	<b>monomórfico</b>				1	<b>monomórfico</b>				loco duplicado				
CaEST-083	3	0,66	1,57	0,59	0,00	3	0,66	1,57	0,59	0,00	loco duplicado				
CaEST-084	1	<b>monomórfico</b>				1	<b>monomórfico</b>				1	<b>monomórfico</b>			
CaEST-085	1	<b>monomórfico</b>				1	<b>monomórfico</b>				2	0,32	0,67	0,27	0,60
CaEST-086 <sup>A</sup>	1	<b>monomórfico</b>				1	<b>monomórfico</b>				<b>não funciona</b>				
CaEST-087	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	1	<b>monomórfico</b>			
CaEST-088	1	<b>monomórfico</b>				1	<b>monomórfico</b>				1	<b>monomórfico</b>			
CaEST-089 <sup>AB</sup>	3	0,55	1,28	0,46	0,17	3	0,57	1,35	0,50	0,67	2	0,22	0,56	0,19	0,70
CaEST-090	2	0,49	0,98	0,38	0,00	3	0,57	1,35	0,50	0,67	2	0,42	0,95	0,33	0,60
CaEST-091	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	loco duplicado				
CaEST-092	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	2	0,32	0,67	0,27	0,60
CaEST-093	2	0,49	0,98	0,38	0,00	3	0,58	1,40	0,50	0,67	3	0,62	0,67	0,55	0,60
CaEST-094	1	<b>monomórfico</b>				1	<b>monomórfico</b>				1	<b>monomórfico</b>			
CaEST-095	3	0,52	1,13	0,43	0,17	3	0,63	1,50	0,55	0,67	4	0,75	1,04	0,70	0,90
CaEST-096	2	0,49	0,98	0,38	0,00	3	0,57	1,35	0,50	0,67	2	0,50	0,00	0,38	0,00
CaEST-097	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	<b>não funciona</b>				
CaEST-098	4	0,75	2,00	0,70	0,00	4	0,75	2	0,70	0,00	3	0,62	0,67	0,55	0,60
CaEST-099	1	<b>monomórfico</b>				1	<b>monomórfico</b>				<b>não funciona</b>				
CaEST-100	1	<b>monomórfico</b>				1	<b>monomórfico</b>				1	<b>monomórfico</b>			
CaEST-101	2	0,49	0,98	0,38	0,00	2	0,49	0,98	0,38	0,00	2	0,18	0,50	0,16	0,40
<b>Média</b>	<b>2,11</b>	<b>0,54</b>	<b>1,20</b>	<b>0,45</b>	<b>0,12</b>	<b>2,30</b>	<b>0,56</b>	<b>1,30</b>	<b>0,48</b>	<b>0,33</b>	<b>2,78</b>	<b>0,54</b>	<b>0,97</b>	<b>0,48</b>	<b>0,72</b>
<b>Desv.Padrão</b>	<b>1,02</b>	<b>0,09</b>	<b>0,34</b>	<b>0,11</b>	<b>0,18</b>	<b>1,08</b>	<b>0,11</b>	<b>0,39</b>	<b>0,13</b>	<b>0,36</b>	<b>1,32</b>	<b>0,17</b>	<b>0,38</b>	<b>0,17</b>	<b>0,23</b>
<b>Min</b>	<b>1,0</b>	<b>0,36</b>	<b>0,88</b>	<b>0,31</b>	<b>0</b>	<b>1,0</b>	<b>0,27</b>	<b>0,63</b>	<b>0,16</b>	<b>0</b>	<b>1,0</b>	<b>0,18</b>	<b>0</b>	<b>0,16</b>	<b>0</b>
<b>Max</b>	<b>6,0</b>	<b>0,78</b>	<b>2,30</b>	<b>0,73</b>	<b>0,68</b>	<b>6,0</b>	<b>0,82</b>	<b>2,54</b>	<b>0,79</b>	<b>1,0</b>	<b>6,0</b>	<b>0,8</b>	<b>1,61</b>	<b>0,77</b>	<b>1,0</b>

**Legenda:** N= número de genótipos avaliados; Na = número de alelos acessado por cada marcador; He = heteroziguidade; H<sup>7</sup> =índice de diversidade de Shannon-Weider; PIC = conteúdo médio de informação polimórfica; DP = poder discriminante; Média= média aritmética; Desv.Padrão = desvio padrão; Min = valores mínimos observados; Max = valores máximos observados; <sup>A</sup> EST-SSR utilizado nos estudos de mapeamento da espécie *C. arabica*; <sup>B</sup> EST-SSR utilizados no estudo de mapeamento da espécie *C. canephora*.

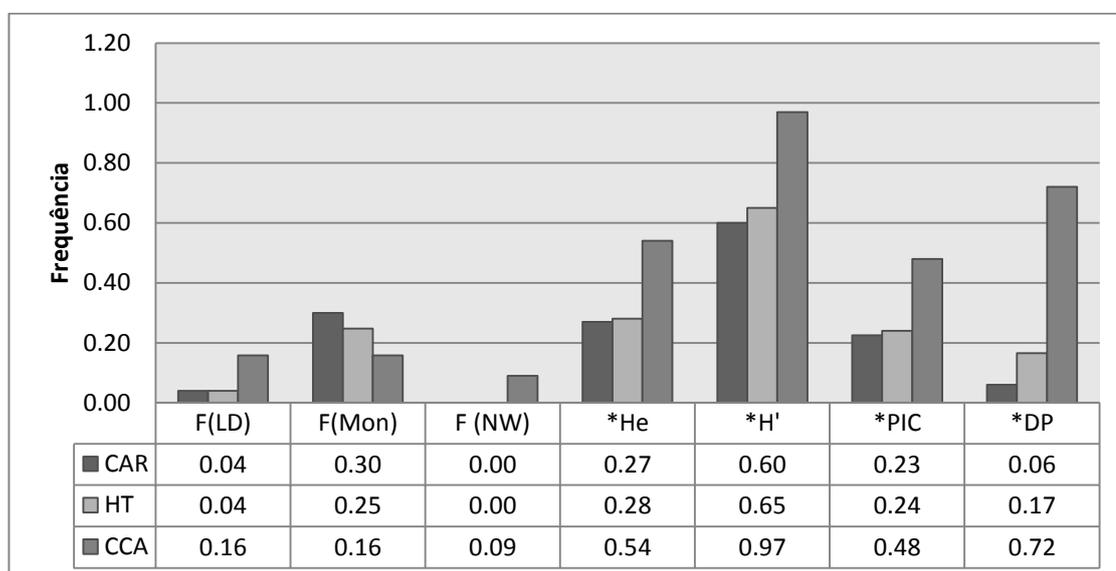
A espécie *C. arabica* apresentou valores de Na que variaram de 1 a 6, com valor médio de 2,11 alelos por loco. Valores médios de He, PIC, H' e DP foram, respectivamente, de 0,54; 0,45; 1,20 e 0,12. Dentre os EST-SSR analisados, o CaEST-010 destacou-se pelos maiores valores de Na (6), He (0,78) e PIC (0,73). Já os marcadores CaEST-002, CaEST-011 e CaEST-040 foram os que apresentaram os maiores valores de DP (0,68; 0,68 e 0,56; respectivamente). As análises do outro grupo de poliploides, os Híbridos de Timor, apresentaram valores médios de Na, He e PIC de 2,30; 0,56 e 0,48, respectivamente. Os valores médio de H' e DP (1,30 e 0,48, respectivamente) foram superiores, com destaque para o CaEST-008, CaEST-034 e o CaEST-074, que apresentaram valores máximos de DP.

Para a espécie *C. canephora* os valores de Na variaram de 1 a 6, com valor médio de 2,78 alelos por loco. Os valores médios de He, H', PIC e DP foram de 0,54; 0,97; 0,48 e 0,72; respectivamente. Dentre os marcadores, destacaram-se o CaEST-015, CaEST-035, CaEST-058, CaEST-059 e o CaEST-064 pelos valores elevados de PIC e os valores máximos de DP.

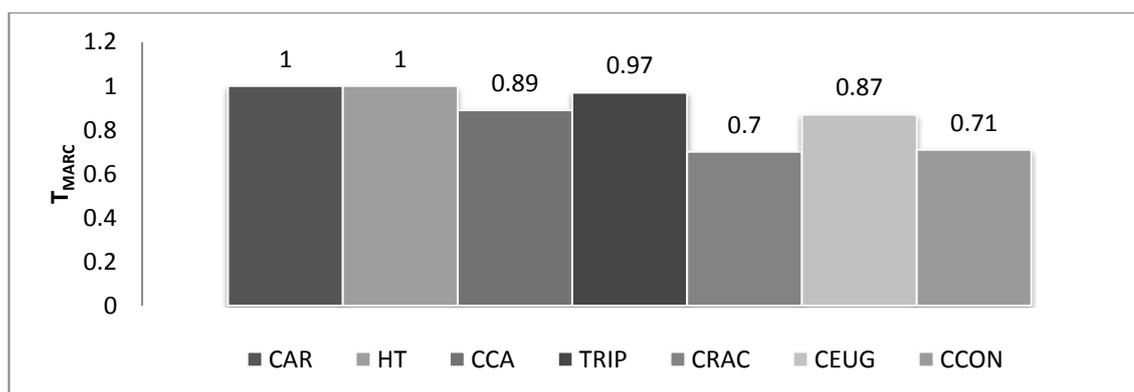
O resumo da análise comparativa dos níveis de polimorfismo dos marcadores EST-SSR para os três grupos de acessos de cafeeiro está apresentado na Figura 2. Os parâmetros He, H', PIC e DP sofrem influência da ploidia, uma vez que, o número e a frequência dos alelos são levados em consideração durante o cálculo desses parâmetros. Para critério de comparação, na Figura 3, são representados os valores médios de \*He, \*H', \*PIC e \*DP corrigidos para a ploidia de cada organismo. Assim, é possível observar que *C. canephora* é a espécie que apresentou os melhores níveis de polimorfismo, seguido dos Híbridos de Timor e de *C. arabica*. Por outro lado, a espécie *C. canephora* foi a que apresentou as maiores frequências de locos duplicados (LD) e de marcadores que não funcionaram (NW).

A taxa de transferabilidade ( $T_{MAR}$ ) dos novos marcadores EST-SSR foi testada em outras quatro espécies do gênero *Coffea* (*C. canephora*, *C. eugeniodeis*, *C. congensis* e *C. racemosa*) e em dois híbridos naturais interespecíficos (Triploide e o Híbrido de Timor) (Anexo 2). Os resultados estão resumidos na Figura 3 e foi observado um valor médio de  $T_{MAR}$  de 88%. Destaque é dado para os híbridos naturais HT e TRIP que apresentaram elevados valores de  $T_{MAR}$  (1,0 e 0,97, respectivamente). As espécies CRAC e CCON apresentaram os menores valores (0,70 e 0,71, respectivamente), enquanto que CCA (0,89) e CEUG (0,87) apresentaram valores

intermediários. Os resultados obtidos indicaram também que a taxa de conservação dos marcadores testados ( $C_{MAR}$ ) é elevado entre as espécies, sendo que 52,7% dos EST-SSR foram conservados. O menor valor de  $C_{MAR}$  foi observado para o marcador CaEST-015 (42,85%) indicando especificidade deste marcador (Anexo 2).



**Figura 2.** Análise comparativa dos níveis de polimorfismo considerando a frequência de locos duplicados (LD), locos monomórficos (Mon) e de marcadores que não funcionaram (NW); e os valores médios corrigidos para ploidia da Heterozigidade (\*He), Índice de Diversidade de Shannon-Weiner (\*H'), Conteúdo Médio de Informação Polimórfica (\*PIC) e Poder Discriminante (\*DP) de *C. arabica* (CAR), Híbrido Timor (HT) e *C. canephora* (CCA).



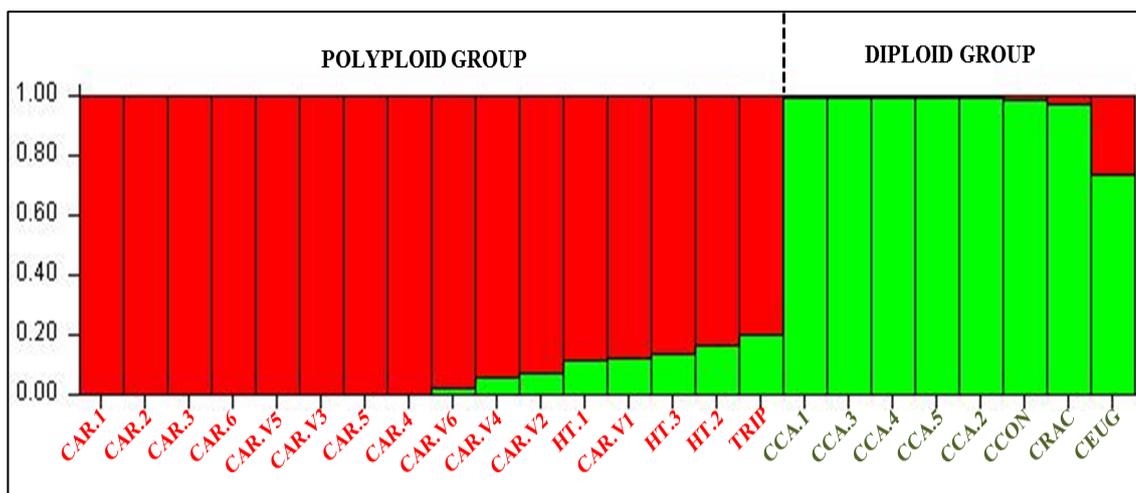
**Figura 3.** Transferabilidade ( $T_{mark}$ ) dos EST-SSR nos genótipos do gênero *C. arabica* (CAR), Híbrido de Timor (HT), *C. canephora* (CCA), Triploide natural (TRIP), *C. racemosa* (CRAC), *C. eugeniodeis* (CEUG) e *C. congensis* (CCON).

### *Caracterização molecular de acessos de cafeeiro*

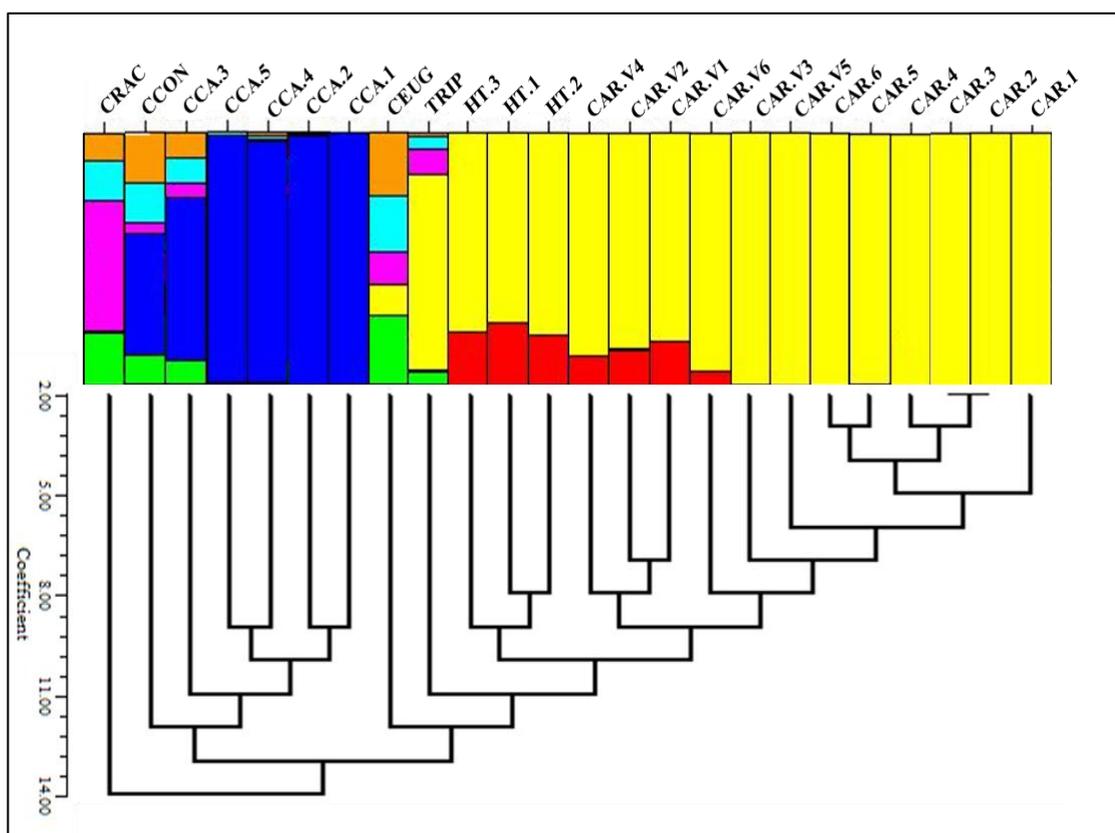
A caracterização molecular dos EST-SSR foi feita utilizando 25 genótipos do gênero *Coffea*, pertencentes a cinco espécies diferentes além de dois híbridos naturais interespecíficos. Para isso, foi utilizado a abordagem bayesiana implementada no software STRUCTURE e segundo metodologia proposta por Evanno et al. (2005), o maior valor de  $\Delta K$  (216,40) foi obtido para  $K=2$ , indicando a estruturação em duas populações (Figura 4). Agrupando com base na similaridade dos genótipos, obteve-se um primeiro grupo formado pelos genótipos poliploides (CAR, HT e TRIP) e um segundo grupo representado apenas por espécies diploides (CCA, CCON, CEUG e CRAC). Dentro do grupo dos poliploides, com exceção do genótipo CAR.V5 e CAR.V3, foi possível observar maior semelhança das variedades comerciais com os HT. Por outro lado, no grupo diploide a estruturação permitiu a distinção entre as espécies CCA, CCON, CRAC e CCON.

Um estudo mais detalhado da diversidade genética foi realizado utilizando a análise integrada da abordagem bayesiana ( $K=7$ ) e das distâncias genéticas apresentadas, respectivamente, na forma de gráfico de barras e dendrograma (Figura 5). Os resultados dessas duas abordagens foram compatíveis e mostrou a eficiência dos EST-SSR no estudo de diversidade dentro do gênero *Coffea*. As espécies poliploides foram novamente discriminadas, contudo, nessa abordagem integrada, foi possível observar clara distinção entre os HT, TRIP e CAR. Dentre os CAR foi possível distinguir as variedades comerciais (CAR-V) e dos acessos do germoplasma (CAR). Todos os cafeeiros foram discriminados, com exceção do CAR.2 e CAR.3. Considerando as espécies diploides houve clara distinção entre as espécies CCA, CEUG e CCON. Destaque é dado para o grupo dos CCA, no qual o genótipo CCA.3 apresentou comportamento diferencial dos demais. Este resultado motivou análises específicas dentro da espécie *C. canephora*, de modo a avaliar a eficiência dos EST-SSR em classificar acessos de Conilon e Robusta dentro de Bancos de Germoplasma.

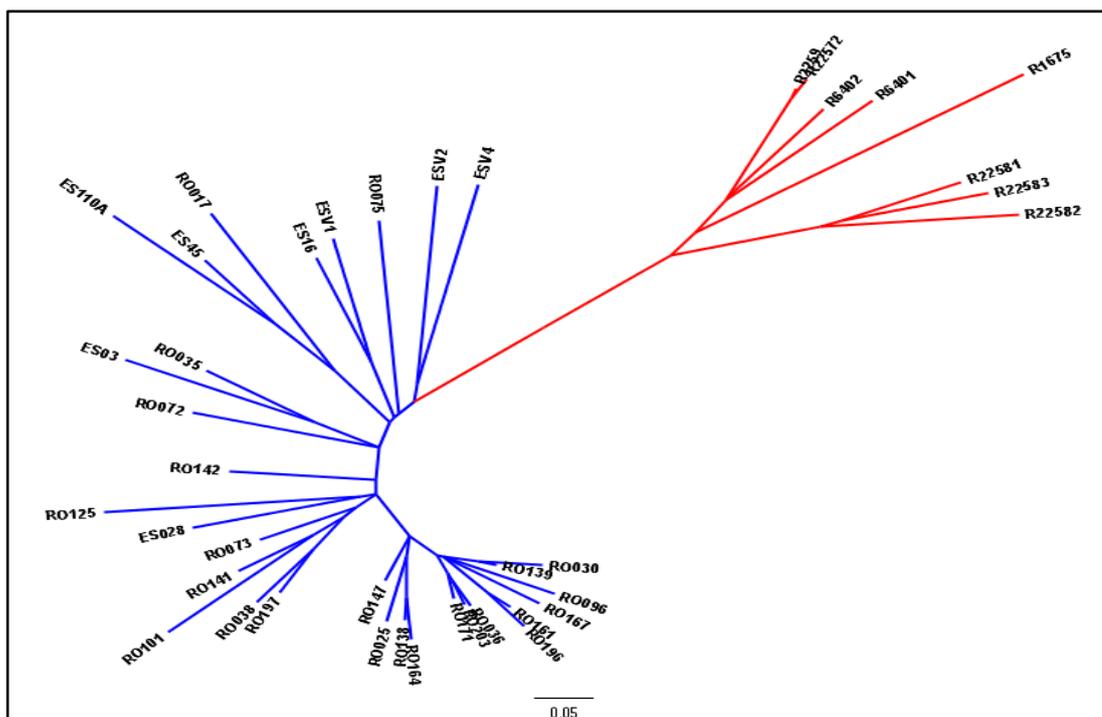
Os EST-SSR foram analisados em 20 acessos de *C. canephora* e possibilitou a separação completa dos grupos varietais Robusta e Conilon (Figura 6). Os resultados foram condizentes com aqueles apresentados por Ferrão et al. (2013), que utilizaram SSR genômico e os marcadores RAPD e AFLP para o estudo de relações genéticas na espécie.



**Figura 4.** Gráfico de barras do *software* STRUCTURE utilizado para estudo de diversidade no gênero *Coffea*. Os 24 genótipos foram divididos em dois grupos (K=2) de acordo com o nível de ploidia e foram agrupados de acordo com o valor de Q (coeficiente de ancestralidade).



**Figura 5.** Análise de diversidade genética de vinte e quatro genótipos do gênero *Coffea*, utilizando o gráfico de barras do *software* STRUCTURE (K=7) e o NJ dendrograma baseado no Índice de Similaridade de Jaccard (1908).

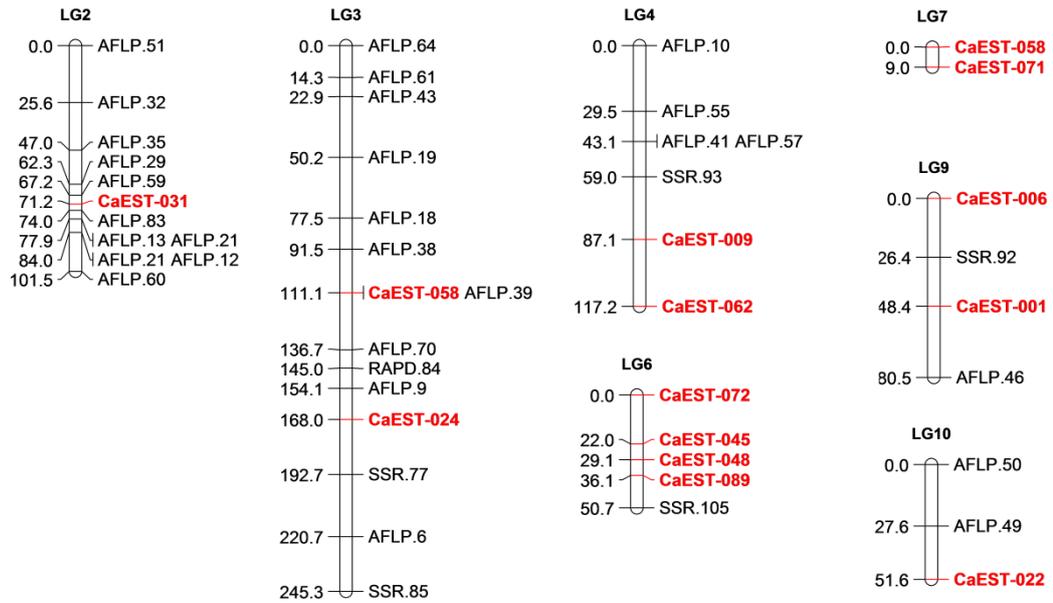


**Figura 6.** Análise de diversidade genética e taxonomia em *C. canephora*. O clado em vermelho representa os acessos do grupo varietal Robusta e o clado azul representa os acessos do grupo varietal Conilon.

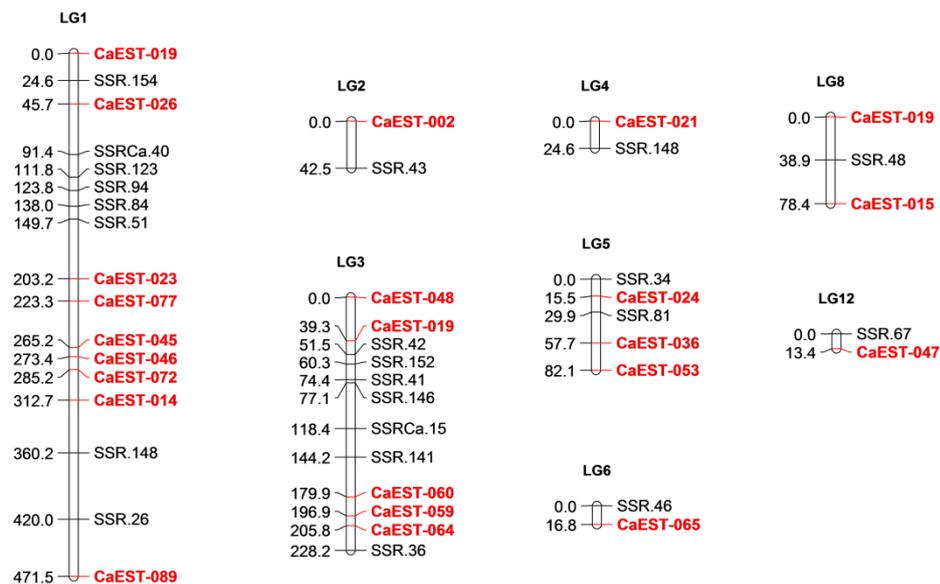
#### *Estudos de mapeamento genético*

Os novos EST-SSR foram testados em estudos de mapeamento genético nas espécies *C. arabica* e *C. canephora*. Em *C. arabica*, 17 EST-SSR (16,83%) apresentaram polimorfismos nos genitores e foram utilizados na genotipagem das progênes F<sub>2</sub>. Deste total, 14 foram alocados em sete grupos de ligação independentes (Figura 7). O grupo de ligação 6 (GL6) foi o que apresentou o maior número de EST-SSR ligados, enquanto que os grupos de ligação 1, 5, 8 e 11 não apresentaram nenhum marcador alocado (Pestana, 2010).

Na espécie *C. canephora*, 36 EST-SSR (35,64%) apresentaram polimorfismo nos genitores e foram utilizados no estudo de mapeamento genético (Tabela 2). Deste total, 23 (56%) foram posicionados em oito grupos de ligação independentes (Figura 8), enquanto que os outros foram declarados não ligados (Tabela 2). O grupo de ligação 1 (GL1) foi o que apresentou o maior número de EST-SSR (9 EST-SSR). O número de EST-SSR polimórficos em *C. canephora* foi cerca de 2 vezes maior do que em *C. arabica*.



**Figura 7.** Posição em cM dos 14 EST-SSR (em vermelho) mapeados em 246 progênies  $F_2$  de *C. arabica*. Os grupos de ligação (GL) 1, 5, 8 e 11 não possuem nenhum EST-SSR ligado e por isso não foram representados. Marcadores SSR genômicos, AFLP e RAPDs (em preto) foram mantidos no mapa de ligação e são relatados por Pestana (2010).



**Figura 8.** Posição em cM dos 23 EST-SSR (em vermelho) mapeados em 96 progênies de irmãos completos de *C. canephora*. Os grupos de ligação (GL) 7, 9, 10 e 11 não apresentaram nenhum EST-SSR ligado e por isso não foram representados.

## DISCUSSÃO

### *Mineração e caracterização de EST-SSR do Projeto Brasileiro do Genoma Café*

A realização do Projeto Brasileiro do Genoma Café disponibilizou para a comunidade científica um banco de sequências EST da espécie *C. arabica*, no qual 24.031 sequências contendo microssatélite foram identificadas. Desse total, 4380 (18,23%) são DNRs, 8811(36,67%) TNRs e 10840 (45,11%) TtNTRs. Dentre as classes de repetições mais frequentes destacaram-se os (AGGG)<sub>N</sub> e os (AGG)<sub>N</sub>. Na literatura é comumente reportada a predominância de *motifs* TNRs em espécies como o citros, arroz, cevado, trigo, cana de açúcar, uva (Huang et al., 2011; Kantety et al., 2002; La Rota et al., 2005; Luro et al., 2008; Oliveira et al., 2009) e, em especial, café (Poncet et al., 2006). Contudo, resultados diferentes foram observados neste estudo, uma vez que os TtNTRs (45,11%) constituíram a classe de *motifs* mais abundante. Apesar de menos comum, alguns estudos relatam situação semelhante (Rungis et al., 2004; Wang et al., 2011; Zhang et al., 2012), indicando que os *motifs* mais frequentes dependem da espécie, do número de sequências avaliadas e do critério de mineração adotado em cada trabalho.

A análise criteriosa das sequências ESTs resultou no desenvolvimento de 101 novos EST-SSR para *C. arabica*. Esse número torna-se expressivo na medida em que se compara com o número de SSRs para café que são descritos na literatura. Atualmente, são listados para a espécie *C. canephora* e *C. arabica* cerca de 400 SSRgenômicos (Baruah et al., 2003; Combes et al., 2000; Cristancho and Gaitán, 2008; Hendre et al., 2008; Leroy et al., 2005; Missio et al., 2009; Moncada and McCouch, 2004; Poncet et al., 2007; Poncet et al., 2004), e 68 EST-SSR (Aggarwal et al., 2007; Bhat et al., 2005; Geromel et al., 2006; Poncet et al., 2006). Esses números reforçam a importância do desenvolvimento e validação de novos marcadores SSR devido sua importância nos estudos genéticos da espécie.

### *Validação dos marcadores: análises de polimorfismo e transferabilidade dos EST-SSR*

Foram desenvolvidos 146 EST-SSR, sendo que desses, 101 foram validados como eficientes marcadores microssatélites para *C. arabica*. Essa taxa de sucesso de 69,17% é semelhante àquelas encontradas para outras espécies, tais como, seringueira (Mantello et al., 2012); cana de açúcar (Oliveira et al., 2009); cevada (Castillo et al.,

2008) e trigo (Nicot et al., 2004). Uma possível explicação para o insucesso de alguns marcadores são os sítios de *splicing* que podem ser encontrados dentro das sequências ESTs. Esses sítios contém *introns* o que, provavelmente, dificulta o anelamento dos *primers* e comprometem a amplificação das regiões microssatélites (Oliveira et al., 2009).

As avaliações de polimorfismos dos novos EST-SSR foram realizadas nas duas espécies cultivadas do gênero *Coffea* (*C. arabica* e *C. canephora*) e no Híbrido de Timor; um híbrido natural interespecífico utilizado como genitor em programas de melhoramento de arábica, sobretudo, por ser resistente à ferrugem do cafeeiro (*Hemileia vastatrix*). A análise comparativa de polimorfismos revelou moderada diversidade genética/alélica nos Híbridos de Timor e em *C. arabica*, enquanto que a espécie *C. canephora* apresentou índices elevados. Resultados semelhantes foram descritos por Lashermes et al. (1999), Poncet et al. (2006) e Hendre et al. (2008); e são coerentes, na medida em que se considera a forma de reprodução, a evolução do genoma e o processo de domesticação dos materiais estudados.

Endêmica de terras de altitude do sudoeste da Etiópia, mas também com registros de ocorrência no Sudão e Quênia, o início do cultivo de *C. arabica* deu-se no Iêmen há cerca de quinhentos anos. No início do século 18, progênies de uma única planta foram levadas da Indonésia para a Europa e depois para o continente Americano, constituindo-se na base genética das principais variedades plantadas tanto no Brasil, como em outros países (Anthony et al., 2002; Orozco-Castillo et al., 1994). O resultado dessa domesticação restrita somada ao modo de reprodução da espécie (autógama) é uma base genética estreita, que refletiu diretamente no baixo nível de diversidade observado. Contudo, se por um lado o processo de domesticação em *C. arabica* foi restrito, o mesmo não é válido para a espécie *C. canephora*. Endêmica das regiões ocidental, centro-tropical e subtropical do continente Africano, registros históricos mostram que o cultivo da espécie se iniciou no Congo, em 1870, estendendo-se para região central da África e, posteriormente, por intermédio de intercâmbios, para Indonésia, Java e continente Americano. O resultado disso, somado a polinização cruzada (alogamia), é a formação de populações com alta variabilidade genética e elevados níveis de heterozigose (Ferrão et al., 2007; Lashermes et al., 2000). Essas características de variabilidade genética descritas anteriormente foram facilmente

visualizadas nos valores superiores de He, PIC e DP dos marcadores EST-SSR analisados em *C. canephora* quando comparado com os analisados em *C. arabica*.

O Híbrido de Timor é um híbrido interespecífico resultante do cruzamento natural entre *C. arabica* x *C. canephora*, que foi encontrado originalmente em uma plantação de café arábica na Ilha de Timor, Ásia (Berthaud and Charrier, 1988; Bettencourt, 1973). Admite-se que suas populações tenham sido originadas a partir de uma única planta, provavelmente, estabelecida no início do século 19 (Bettencourt, 1973). No trabalho de caracterização usando EST-SSR, apesar de ser um híbrido interespecífico, foi observada grande similaridade genética dos Híbridos de Timor com *C. arabica*. Essa semelhança também foi relatada por Setotaw (2009), que confirmou a similaridade entre esses dois grupos pela maior porcentagem de alelos compartilhados (cerca de 2/3 do seu genoma). A explicação para isso, segundo o autor, está na origem do Híbrido de Timor, pois se assume que nesses genótipos houve seleção direcional e retrocruzamentos para *C. arabica*. Esses eventos naturais fizeram com que a introgressão do genoma de *C. canephora* nos Híbridos de Timor fossem reduzidas, tornando-o assim mais similar à espécie *C. arabica*.

As análises de transferabilidade ( $T_{\text{MARK}}$ ) mostraram que marcadores EST-SSR apresentam potencial para uso em outras espécies do gênero *Coffea*. No geral, os EST-SSR apresentam uma maior taxa de transferência entre espécies relacionadas do que os tradicionais SSRgenômicos (Ellis and Burke, 2007). A explicação para isso é que os *primers* ESTs flanqueiam regiões codificadoras do genoma que são mais susceptíveis a serem conservadas entre as espécies. Essa característica já foi demonstrada em diferentes estudos com espécies cultivadas, incluindo uva (Decroocq et al., 2003), cevada (Castillo et al., 2008), trigo (Gupta et al., 2003) e cana de açúcar (Cordeiro et al., 2001). Em café, valores de  $T_{\text{MARK}}$  superiores a 85%, como o obtido no presente estudo, foram já relatados (Aggarwal et al., 2007; Bhat et al., 2005). No presente estudo, observou-se que muitos marcadores foram monomórficos em *C. arabica*, no entanto, apresentaram altas taxas de polimorfismos em espécies relacionadas. Resultados semelhantes foram apresentados por Hendre et al. (2008), porém utilizando marcadores SSRgenômicos. Em relação à taxa de conservação dos marcadores ( $C_{\text{MARK}}$ ), o menor valor foi observado para o CaEST-015 o que indicou especificidade deste marcador para estudos com *C. arabica*.

### *Caracterização molecular de acessos de cafeeiro*

A eficácia dos novos EST-SSR em estudos de diversidade e taxonomia foi mensurada utilizando duas abordagens distintas. A primeira delas, implementada no *software* STRUCTURE, foi à bayesiana e destaca-se pela ampla utilidade em estudos de estrutura de população. Utilizando esta metodologia, os genótipos foram classificados em duas subpopulações ( $k=2$ ), segundo o valor de  $\Delta K$  definido por Evanno et al. (2005), indicando que o maior nível de hierarquia está relacionado com a ploidia dos materiais avaliados. Assim, os tetraploides (HT e CAR) e o triploide (TRIP) foram alocados em um grupo; enquanto os diploides (CCAN, CCON, CRAC e CEUG) constituíram outro grupo distinto. Resultados semelhantes foram encontrados nos estudos de López-Gartner et al. (2009), no qual os CAR e HT apresentaram agrupamento diferenciado das espécies diploides. Posteriormente, os genótipos foram avaliados considerando a análise integrada de similaridade genética e da abordagem bayesiana. Para isso, foi considerada a formação de sete subpopulações ( $k=7$ ); sendo quatro destas correspondentes às espécies diploides (CCAN, CCON, CRAC e CEUG) e três formados por poliploides (CAR, HT e TRIP).

O grupo dos poliploides é um grupo homogêneo formado por materiais que apresentam elevada similaridade genética. Com relação à espécie *C. arabica*, já foi citado que o gargalo genético sofrido durante o processo de domesticação resultou em populações de base genética estreita e com baixos níveis de polimorfismos. Essa característica foi observada, primeiramente, nos resultados de nível de polimorfismo e foi confirmada pela elevada similaridade genética dos genótipos nos estudos de diversidade. Para as análises de caracterização molecular, os novos EST-SSR foram eficientes em discriminar os genótipos CAR.V, que são variedades comerciais provenientes de hibridizações em programas de melhoramento e que contêm, portanto, uma pequena porção de genoma de outras espécies. Para o caso dos acessos CAR, que são genótipos que contêm apenas genoma de *C. arabica*, apesar da base genética estreita, os EST-SSR também permitiram a diferenciação genotípica.

Para os Híbridos de Timor, foi comprovada a similaridade com a espécie *C. arabica*. Essa semelhança é de grande importância nas hibridações em programas de melhoramento, uma vez que permite o desenvolvimento de variedades que agreguem a qualidade de bebida do café arábica com a resistência a doenças dos Híbridos de Timor, que apresenta genes de resistência originados da ancestralidade com a espécie *C.*

*canephora* (Lashermes et al., 2000; Setotaw et al., 2010). Neste trabalho, cinco das seis variedades comerciais (CAR.V) utilizadas são resultantes da hibridação entre *C. arabica* e Híbridos de Timor (Tabela 1-I). Isso explica a grande similaridade observada dos genótipos CAR.V com os HT (Figura 5 e 6). A única exceção foi observado para o genótipo CAR.V5, que é justamente a única variedade que não apresentou introgressões com HT em sua genealogia. A utilização da abordagem de distância genéticas, representou melhor essa tendência do que o gráfico de barras da abordagem bayesiana, confirmando assim a importância do emprego de diferentes metodologias estatísticas para estudo da diversidade. Com relação ao outro híbrido natural estudado (TRIP), foi possível observar uma maior similaridade com os CAR e CRAC, o que demonstrou que o TRIP foi resultado do cruzamento natural entre uma espécie tetraploide (*C. arabica*) e uma espécie diploide (*C. racemosa*). Dessa forma, os resultados apresentados confirmam a eficiência dos novos EST-SSR em estudos de caracterização molecular dentro de genótipos poliploides do gênero *Coffea*.

Considerando o grupo das espécies diploides, o estudo de diversidade genética foi eficiente e todos os genótipos foram discriminados. Segundo Lashermes et al. (1997) e Maurin et al. (2007), as espécies do gênero *Coffea* são divididas em grandes grupos de acordo com a correspondência geográfica. Assim, já era esperado que as espécies de mesmo centro de origem, como CCAN e CCON, apresentassem maior homologia; enquanto que as espécies de centro de origem diferentes, como CRAC e CEUG, compartilhassem menor similaridade genética. Esse mesmo resultado foi encontrado na análise dos marcadores EST-SSR deste estudo.

Para a espécie *C. canephora* foi possível observar o comportamento diferencial do genótipo CCAN.3 em relação aos demais. Estudos de diversidade dentro da espécie identificaram a existência de subgrupos de acordo com o isolamento geográfico e o curso evolutivo sofrido nos últimos séculos (Maurin et al., 2007). O genótipo CCAN. 3, dentro da espécie *C. canephora*, pertence ao subgrupo SG1 que reúne os genótipos popularmente chamados de Conilon. Os Conilons apresentam como características adaptativas folhas e frutos menores, menor porte e vigor, tolerância à seca e susceptibilidade à ferrugem. Por outro lado, os demais CCAN pertencem aos subgrupos SG2, B e C e são popularmente conhecidos como Robustas. Dentre as características do Café Robusta, destaca-se o porte elevado das plantas, folhas e frutos maiores, maior

vigor, resistência a ferrugem e sensibilidade a seca (Cubry et al., 2008; Ferrão et al., 2007; Musoli et al., 2009).

A possibilidade dos EST-SSR discriminarem subgrupos em *C. canephora* motivou análises mais detalhadas dentro da espécie, pois apesar das características morfológicas que separaram os subgrupos, a classificação não é uma tarefa fácil, visto que, as populações são formadas por plantas com elevada heterozigose e grande amplitude fenotípica. Para isso, os EST-SSR foram testados em avaliações de acessos do Banco de Germoplasma da Embrapa. Os resultados obtidos demonstraram que os marcadores puderam separar com clareza os acessos pertencentes ao grupo varietal Conilon e Robusta. Isso reforça a importância dos marcadores EST-SSR, especialmente, para seleção de genitores em programas de melhoramento e no manejo de recursos genéticos em bancos de germoplasmas.

#### *Estudos de mapeamento genético*

O desenvolvimento de mapas genéticos é considerado uma das aplicações de maior impacto dos marcadores moleculares na análise genética de espécies. No contexto do melhoramento de plantas, os estudos de mapeamento são importantes, sobretudo, para localização e quantificação de regiões genômicas que controlam caracteres de interesse. Para que as informações moleculares possam ser utilizadas em programas de seleção assistida (MAS), o primeiro passo é a existência de mapas genéticos robustos com alta densidade de marcadores. Neste trabalho, confirmamos a utilidade dos novos EST-SSR em estudos de mapeamento genético com *C. arabica* e *C. canephora*. O mapa genético para *C. arabica* foi descrito por Pestana (2010) e possui 94 marcadores alocados em 11 grupos de ligação. Deste total, 14 são EST-SSR e estão posicionados em sete grupos de ligação. O pequeno número de EST-SSR observados no mapa se deve ao baixo nível de polimorfismo dos genitores da população de mapeamento. Como já discutido anteriormente, esse fato é resultado da base genética estreita decorrente do processo de domesticação da espécie. Uma alternativa para aumentar a saturação do mapa foi a utilização de outras classes de marcadores, como o AFLP e o RAPD. Para a espécie *C. canephora*, o número de EST-SSR polimórficos nos genitores foi superior, e por esse motivo observa-se um maior número de marcadores alocados no mapa de ligação da espécie. Até o presente momento, o mapa genético da espécie possui 54 marcadores distribuídos em 15 grupos de ligação. Ambos os trabalhos estão em

desenvolvimento e o objetivo é localizar genes e/ou QTLs associados a caracteres de importância agrônômica para utilização em programas de MAS e clonagem posicional.

## CONCLUSÃO

O presente estudo disponibiliza para comunidade científica um conjunto de 101 novos EST-SSR para a espécie *C. arabica*, minerados a partir do Projeto Brasileiro do Genoma Café. A caracterização e validação desses marcadores demonstraram a sua utilidade em estudos de caracterização molecular, diversidade genética, classificação taxonômica e mapeamento genético. A taxa de transferabilidade foi mensurada e mostrou que esses estudos podem ser ampliados para outras espécies dentro do gênero *Coffea*, incluindo estudos evolutivos e de filogenia molecular. Em programas de melhoramento, os EST-SSR surgem como uma ferramenta importante na seleção de genitores, manejo de recursos genéticos, seleção assistida e elucidação de caracteres complexos (QTLs). Dessa forma, este estudo espera aumentar o número de ferramentas moleculares disponíveis para o gênero *Coffea*, de forma a enriquecer pesquisas envolvendo a cultura do café. Por fim, ressalta-se que um banco de dados com 24 mil sequências contendo microssatélites está disponível para o desenvolvimento e validação de novos marcadores EST-SSR.

## REFERÊNCIAS

- Aggarwal, R. K., Hendre, P. S., Varshney, R. K., Bhat, P. R., Krishnakumar, V., and Singh, L. (2007). Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor Appl Genet* **114**, 359-72.
- Anthony, F., Combes, C., Astorga, C., Bertrand, B., Graziosi, G., and Lashermes, P. (2002). The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theor Appl Genet* **104**, 894-900.
- Baruah, A., Naik, V., Hendre, P. S., Rajkumar, R., Rajendrakumar, P., and Aggarwal, R. K. (2003). Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. *Molecular Ecology Notes* **3**, 647-650.
- Berthaud, J., and Charrier, A. (1988). Genetics resources of *Coffea*. In "Coffee" (R. J. Clarke and R. Macrae, eds.), Vol. 4, pp. 1–42. Elsevier Applied Science, London,.
- Bettencourt, A. (1973). "Cosiderações gerais sobre o "Híbrido de Timor", Instituto Agronomico de Campinas, Campinas, Brazil.
- Bhat, P. R., Krishnakumar, V., Hendre, P. S., Rajendrakumar, P., Varshney, R. K., and Aggarwal, R. K. (2005). Identification and characterization of expressed sequence tags-derived simple sequence repeats markers from robusta coffee

- variety 'CxR' (an interspecific hybrid of *Coffea canephora* × *Coffea congensis*). *Molecular Ecology Notes* **5**, 80-83.
- Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* **32**, 314-31.
- Castillo, A., Budak, H., Varshney, R. K., Dorado, G., Graner, A., and Hernandez, P. (2008). Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense*. *BMC Plant Biol* **8**, 97.
- Combes, M. C., Andrzejewski, S., Anthony, F., Bertrand, B., Rovelli, P., Graziosi, G., and Lashermes, P. (2000). Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Mol Ecol* **9**, 1178-80.
- Cordeiro, G. M., Casu, R., McIntyre, C. L., Manners, J. M., and Henry, R. J. (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci* **160**, 1115-1123.
- Cristancho, M., and Gaitán, A. L. (2008). Isolation, characterization and amplification of simple sequence repeat loci in coffee. *Crop Breeding and Applied Biotechnology* **8**, 321-329.
- Cubry, P., Musoli, P., Legnate, H., Pot, D., de Bellis, F., Poncet, V., Anthony, F., Dufour, M., and Leroy, T. (2008). Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. *Genome* **51**, 50-63.
- Decroocq, V., Fave, M. G., Hagen, L., Bordenave, L., and Decroocq, S. (2003). Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theor Appl Genet* **106**, 912-22.
- Diniz, E. C., Sakiyama, N. S., Lashermes, P., Caixeta, E. T., Oliveira, A. C. B., Zambolim, E. M., Loureiro, M. E., Pereira, A. A., and Zambolim, L. (2005). Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. *Crop Breeding and Applied Biotechnology* **5**, 387-393.
- Earl, D., and vonHoldt, B. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.
- Ellis, J. R., and Burke, J. M. (2007). EST-SSR as a resource for population genetic analyses. *Heredity (Edinb)* **99**, 125-32.
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611-20.
- Ferrão, L. F. V., Caixeta, E. T., Souza, F. d., Zambolim, E. M., Cruz, C. D., Zambolim, L., and Sakiyama, N. S. (2013). Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. *Plant Systematics and Evolution* **299**, 225-238.
- Ferrão, R. G., Ferrão, M. G. A., Fonseca, A. F. A., and Pacova, B. E. V. (2007). Melhoramento genético de *Coffea canephora*. In "Café Conilon" (R. G. Ferrão, A. F. A. Fonseca, S. M. Bragança, M. G. A. Ferrão and L. H. De Muner, eds.), Vol. 1, pp. 123-173. Incaper, Vitória, Brazil.
- Geromel, C., Ferreira, L. P., Guerreiro, S. M., Cavalari, A. A., Pot, D., Pereira, L. F., Leroy, T., Vieira, L. G., Mazzafera, P., and Marraccini, P. (2006). Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. *J Exp Bot* **57**, 3243-58.

- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Leger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., and Petit, R. J. (2011). Current trends in microsatellite genotyping. *Mol Ecol Resour* **11**, 591-611.
- Gupta, P. K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., and Balyan, H. S. (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics* **270**, 315-323.
- Hendre, P. S., Phanindranath, R., Annapura, V., Lalremruata, A., and Aggarwal, R. K. (2008). 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 Biol* **8**, 51.
- Huang, H., Lu, J., Ren, Z., Hunter, W., Dowd, S. E., and Dang, P. (2011). Mining and validating grape (*Vitis* L.) ESTs to develop EST-SSR markers for genotyping and mapping. *Mol Breed* **28**, 241-254.
- Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudoise Des Sciences Naturelles* **44**, 223-270.
- Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., and Dhawan, A. K. (2010). Microsatellite markers: an overview of the recent progress in plants. *Euphytica* **177**, 309-334.
- Kantety, R. V., La Rota, M., Matthews, D. E., and Sorrells, M. E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* **48**, 501-10.
- La Rota, M., Kantety, R. V., Yu, J. K., and Sorrells, M. E. (2005). Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* **6**, 23.
- Lashermes, P., Andrzejewski, S., Bertrand, B., Combes, M. C., Dussert, S., Graziosi, G., Trouslot, P., and Anthony, F. (2000). Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). *Theoretical and Applied Genetics* **100**, 139-146.
- Lashermes, P., Combes, M. C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F., and Charrier, A. (1999). Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol Gen Genet* **261**, 259-66.
- Lashermes, P., Combes, M. C., Trouslot, P., and Charrier, A. (1997). Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical and Applied Genetics* **94**, 947-955.
- Leroy, T., Marraccini, P., Dufour, M., Montagnon, C., Lashermes, P., Sabau, X., Ferreira, L. P., Jourdan, I., Pot, D., Andrade, A. C., Glaszmann, J. C., Vieira, L. G. E., and Piffanelli, P. (2005). Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theoretical and Applied Genetics* **111**, 1032-1041.
- Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128-9.
- López-Gartner, G., Cortina, H., McCouch, S. R., and Moncada, M. D. P. (2009). Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. *Tree Genetics & Genomes* **5**, 435-446.
- Luro, F. L., Costantino, G., Terol, J., Argout, X., Allario, T., Wincker, P., Talon, M., Ollitrault, P., and Morillon, R. (2008). Transferability of the EST-SSR developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other *Citrus* species and their effectiveness for genetic mapping. *BMC Genomics* **9**, 287.
- Mantello, C. C., Suzuki, F. I., Souza, L. M., Goncalves, P. S., and Souza, A. P. (2012). Microsatellite marker development for the rubber tree (*Hevea brasiliensis*):

- characterization and cross-amplification in wild *Hevea* species. *BMC Res Notes* **5**, 329.
- Margarido, G. R., Souza, A. P., and Garcia, A. A. (2007). OneMap: software for genetic mapping in outcrossing species. *Hereditas* **144**, 78-9.
- Markwith, S. H., and Scanlon, M. J. (2007). Multiscale analysis of *Hymenocallis coronaria* (Amaryllidaceae) genetic diversity, genetic structure, and gene movement under the influence of unidirectional stream flow. *Am J Bot* **94**, 151-60.
- Maurin, O., Davis, A. P., Chester, M., Mvungi, E. F., Jaufeerally-Fakim, Y., and Fay, M. F. (2007). Towards a Phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences. *Ann Bot* **100**, 1565-83.
- Missio, R. F., Caixeta, E. T., Zambolim, E. M., Zambolim, L., Cruz, C. D., and Sakiyama, N. S. (2010). Polymorphic information content of SSR markers for *Coffea* spp. *Crop Breeding and Applied Biotechnology* **10**, 89-94.
- Missio, R. F., Caixeta, E. T., Zambolim, E. M., Zambolim, L., and Sakiyama, N. S. (2009). Development and validation of SSR markers for *Coffea arabica* L. *Crop Breeding and Applied Biotechnology* **9**, 361-371.
- Moncada, P., and McCouch, S. (2004). Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome* **47**, 501-9.
- Musoli, P., Cubry, P., Aluka, P., Billot, C., Dufour, M., De Bellis, F., Pot, D., Bieysse, D., Charrier, A., and Leroy, T. (2009). Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda. *Genome* **52**, 634-46.
- Nicot, N., Chiquet, V., Gandon, B., Amilhat, L., Legeai, F., Leroy, P., Bernard, M., and Sourdille, P. (2004). Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). *Theor Appl Genet* **109**, 800-5.
- Oliveira, K. M., Pinto, L. R., Marconi, T. G., Mollinari, M., Ulian, E. C., Chabregas, S. M., Falco, M. C., Burnquist, W., Garcia, A. A., and Souza, A. P. (2009). Characterization of new polymorphic functional markers for sugarcane. *Genome* **52**, 191-209.
- Orozco-Castillo, C., Chalmers, K. J., Waugh, R., and Powell, W. (1994). Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theoretical and Applied Genetics* **87**, 934-940.
- Pestana, K. N. (2010). Caracterização fenotípica e molecular da resistência do cafeeiro Hibrido de Timor a *Hemileia vastatrix*, Universidade Federal de Viçosa, Viçosa, MG, Brazil.
- Plasterer, T. N. (1997). PRIMERSELECT. Primer and probe design. *Methods Mol Biol* **70**, 291-302.
- Poncet, V., Dufour, M., Hamon, P., Hamon, S., de Kochko, A., and Leroy, T. (2007). Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. *Genome* **50**, 1156-61.
- Poncet, V., Hamon, P., Minier, J., Carasco, C., Hamon, S., and Noirot, M. (2004). SSR cross-amplification and variation within coffee trees (*Coffea* spp.). *Genome* **47**, 1071-81.
- Poncet, V., Rondeau, M., Tranchant, C., Cayrel, A., Hamon, S., de Kochko, A., and Hamon, P. (2006). SSR mining in coffee tree EST databases: potential use of EST-SSR as markers for the *Coffea* genus. *Mol Genet Genomics* **276**, 436-49.

- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945–959.
- Rohlf, F. J. (2008). NTSYSpc: Numerical Taxonomy System. In "Exeter Publishing" (v. 2.20, ed.). Setauket, NY.
- Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-86.
- Rungis, D., Bérubé, Y., Zhang, J., Ralph, S., Ritland, C., Ellis, B., Douglas, C., Bohlmann, J., and Ritland, K. (2004). Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theoretical and Applied Genetics* **109**, 1283-1294.
- Setotaw, T. A. (2009). Genetic diversity and genome introgression in coffee, Universidade Federal de Viçosa, Viçosa, MG.
- Setotaw, T. A., Caixeta, E. T., Pena, G. F., Zambolim, E. M., Pereira, A. A., and Sakiyama, N. S. (2010). Breeding potential and genetic diversity of “Híbrido do Timor” coffee evaluated by molecular markers. *Crop Breeding and Applied Biotechnology* **10**, 298-304.
- Tan, Y. D., and Fu, Y. X. (2006). A novel method for estimating linkage maps. *Genetics* **173**, 2383-90.
- Tessier, C., David, J., This, P., Boursiquot, J. M., and Charrier, A. (1999). Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theoretical and Applied Genetics* **98**, 171-177.
- Vieira, L. G. E., Andrade, A. C., Colombo, C. A., Moraes, A. H. d. A., A., d. O. C., Labate, C. A., Marino, C. L., Vitorello, C. B. M., Monte, D. C., Giglioti, E., Kimura, E. T., Romano, E., Kuramae, E. E., Lemos, E. G. M., Almeida, E. R. P., Jorge, E. C., Albuquerque, E. V., da Silva, F. R., Vinecky, F., Sawazaki, H. E., Dorry, H. F. A., Carrer, H., Abreu, I. N., Batista, J. A. N., Teixeira, J. B., Kitajima, J. P., Xavier, K. G., de Lima, L. M., de Camargo, L. E. A., Pereira, L. F., Coutinho, L. L., Lemos, M. V. F., Romano, M. R., Machado, M. A., Costa, M. M. C., de Sá, M. F. G., Goldman, M. H. S., Ferro, M. I. T., Tinoco, M. L. P., Oliveira, M. C., Filho, M. A. G., Shimizu, M. M., Maluf, M. P., Eira, M. T. S., de Oliveira, O. L. B. C., Harakava, R., Balbao, S. F., Tsai, S. M., Formighieri, S. M. Z., Carazzolle, M. F., and Pereira, G. A. G. (2006). Brazilian coffee genome project: an EST-based genomic resource. *Braz. J. Plant Physiol.*, **18**, 95-108.
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* **93**, 77-8.
- Wang, Z., Li, J., Luo, Z., Huang, L., Chen, X., Fang, B., Li, Y., Chen, J., and Zhang, X. (2011). Characterization and development of EST-derived SSR markers in cultivated sweetpotato (*Ipomoea batatas*). *BMC Plant Biol* **11**, 139.
- Zhang, H., Wei, L., Miao, H., Zhang, T., and Wang, C. (2012). Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics* **13**, 316.

## ANEXO 1

**Anexo1.** Número de EST-SSRS minerados do Projeto Brasileiro Genoma Café de acordo com a classe e/ou categoria de motifs repetidos.

Repeat motif				Number of repeat							Total		
Dinucleotide				Classes	3	4	5	6	7	8	9	≥10	
AC/TG		CA/GT	(AC)n	-	-	-	405	144	56	39	67	711	
AG/TC		GA/CT	(AG)n	-	-	-	711	953	295	179	362	2500	
AT		TA	(AT)n	-	-	-	332	144	79	46	193	794	
CG		GC	(CG)n	-	-	-	262	73	30	7	3	375	
Total				-	-	-	1710	1314	460	271	625	4380	
Trinucleotide				Classes	3	4	5	6	≥7	Total			
AAC/TTG	ACA/TGT	CAA/GTT	(AAC)n	-	304	53	20	10	387				
AAG/TTC	AGA/TCT	GAA/CTT	(AAG)n	-	1341	309	132	109	1891				
AAT/TTA	ATA/TAT	TAA/ATT	(AAT)n	-	423	68	25	35	551				
ACC/TGG	CAC/GTG	CCA/GGT	(ACC)n	-	581	157	52	37	827				
ACG/TGC	GAC/CTG	CGA/GCT	(ACG)n	-	298	60	31	25	414				
ACT/TGA	CTA/GAT	TAC/ATG	(ACT)n	-	392	90	31	19	532				
AGC/TCG	CAG/GTC	GCA/CGT	(AGC)n	-	279	112	62	37	490				
AGG/TCC	GAG/CTC	GGA/CCT	(AGG)n	-	1995	503	102	113	2713				
AGT/TCA	GTA/CAT	TAG/ATC	(AGT)n	-	277	66	39	25	407				
CCG/GGC	CGC/GCG	GCC/CGG	(CCG)n	-	474	85	33	7	599				
Total				-	6364	1503	527	417	8811				
Tetranucleotide				Classes	3	4	≥5	Total					
AAAC/TTTG	AACA/TTGT	ACAA/TGTT	CAAA/GTTT	(AAAC)n	438	33	6	477					
AAAG/TTTC	AAGA/TTCA	AGAA/TCTT	GAAA/CTTT	(AAAG)n	1659	150	37	1846					
AAAT/TTTA	AATA/TTAT	ATAA/TATT	TAAA/ATTT	(AAAT)n	614	58	15	687					
AACC/TTGG	ACCA/TGGT	CCAA/GGTT	CAAC/GTTG	(AACC)n	160	9	1	170					
AACG/TTGC	AGCA/TCGT	CGAA/GCTT	GAAC/CTTG	(AACG)n	78	1	0	79					
AACT/TTGC	ACTA/TGAT	CTAA/GATT	TAAC/ATTG	(AACT)n	83	5	4	92					

AAGC/TTCG	AGCA/TCGT	GCAA/CGTT	CAAG/GTTC	(AAGC)n	395	4	0	399
AAGG/TTCC	AGGA/TCCT	GGAA/CCTT	GAAG/CTTC	(AAGG)n	845	49	15	909
AAGT/TTCA	AGTA/TCAT	GTAA/CATT	TAAG/ATTC	(AAGT)n	57	7	0	64
AATC/TTCG	ATCA/TAGT	TCAA/AGTT	CAAT/GTTA	(AATC)n	62	6	0	68
AATG/TTAC	ATGA/TACT	TGAA/ACTT	GAAT/CTTA	(AATG)n	65	6	1	72
AATT/TTAA	ATTA/TAAT			(AATT)n	98	3	2	103
ACAG/TGTC	CAGA/GTCT	AGAC/TCTG	GACA/CTGT	(ACAG)n	72	2	3	77
ACAT/TGTA	CATA/GTAT	ATAC/TATG	TACA/ATGT	(ACAT)n	151	12	5	168
ACCC/TGGG	CACC/GTGG	CCAC/GGTG	CCCA/GGGT	(ACCC)n	483	31	1	515
ACCG/TGGC	CCGA/GGCT	CGAC/GCTG	GACC/CTGG	(ACCG)n	33	3	0	36
ACCT/TGGA	CCTA/GGAT	CTAC/GATG	TACC/ATGG	(ACCT)n	78	9	0	87
ACGC/TGCG	CGCA/GCGT	GCAC/CGTG	CACG/GTGC	(ACGC)n	85	3	0	88
ACGG/TGCC	CGGA/GCCT	GGAC/CCTG	GACG/CTGC	(ACGG)n	33	3	2	38
ACGT/TGCA	CGAT/GCAT	GTAC/CATG	TACG/ATGC	(ACGT)n	68	4	2	74
ACTC/TGAG	CTCA/GAGT	TCAC/AGTG	CACT/GTGA	(ACTC)n	117	8	3	128
ACTG/TGAC	CTGA/GACT			(ACTG)n	20	3	0	23
AGAT/TCTA	GATA/CTAT	ATAG/TATC	TAGA/ATCT	(AGAT)n	118	3	2	123
AGCC/TCGG	GCCA/CGGT	CCAG/GGTC	CAGC/GTCG	(AGCC)n	30	5	0	35
AGCG/TCGC	GCGA/CGCT	CGAG/GCTC	GAGC/CTCG	(AGCG)n	81	3	2	86
AGCT/TCGA	GCTA/CGAT	CTAG/GATC	TAGC/ATCG	(AGCT)n	37	7	0	44
AGGC/TCCG	GGCA/CCGT	GCAG/CGTC	CAGG/GTCC	(AGGC)n	29	6	1	36
AGGG/TCCC	GGGA/CCCT	GGAG/CCTC	GAGG/CTCC	(AGGG)n	3145	405	53	3603
AGGT/TCCA	GGTA/CCAT	GTCG/CATC	TACC/ATCC	(AGGT)n	64	7	3	74
AGTC/TCAG	GTCA/CAGT			(AGTC)n	13	0	3	16
CCCG/GGGC	CCGC/GGCG	CGCC/GCGG	GCCC/CGGG	(CCCG)n	489	36	2	527
CCGG/GGCC	CGGC/GCCG			(CCGG)n	93	3	0	96
Total					9793	884	163	10840

## ANEXO 2

**Anexo 2.** Taxa de transferibilidade ( $T_{MAR}$ ) e de conservação de 74 EST-SSRs, avaliados nas espécies *Coffea arabica* (CAR), *Coffea canephora* (CCA), *Coffea racemosa* (CRAC), *Coffea eugenioides* (CEUG), *Coffea congensis* (CCON); e nos híbridos interespecíficos Híbrido de Timor (HT) e no Triplóide natural (TRIP).

Primer	CAR	HT	CCA	TRIP	CRAC	CEUG	CCON	$C_{MAR}$
CaEST-001	1	1	1	1	1	1	1	1.00
CaEST-003	1	1	1	1	1	1	1	1.00
CaEST-004	1	1	1	1	1	1	1	1.00
CaEST-005	1	1	1	1	1	1	1	1.00
CaEST-008	1	1	1	1	1	1	1	1.00
CaEST-009	1	1	1	1	1	1	-	0.86
CaEST-011	1	1	1	1	-	-	-	0.57
CaEST-015	1	1	1	-	-	-	-	0.43
CaEST-016	1	1	1	1	-	-	1	0.71
CaEST-017	1	1	1	1	-	1	1	0.86
CaEST-019	1	1	1	1	1	1	1	1.00
CaEST-020	1	1	1	1	1	1	1	1.00
CaEST-021	1	1	1	1	-	1	1	0.86
CaEST-022	1	1	1	1	1	-	1	0.86
CaEST-023	1	1	1	1	1	1	-	0.86
CaEST-024	1	1	1	1	1	1	1	1.00
CaEST-025	1	1	1	1	-	1	1	0.86
CaEST-026	1	1	1	1	-	1	1	0.86
CaEST-027	1	1	1	1	1	1	-	0.86
CaEST-028	1	1	1	1	-	1	1	0.86
CaEST-029	1	1	1	1	1	1	1	1.00
CaEST-030	1	1	1	1	1	1	1	1.00
CaEST-031	1	1	1	1	-	1	-	0.71
CaEST-033	1	1	1	1	1	1	1	1.00
CaEST-034	1	1	1	1	-	1	1	0.86
CaEST-035	1	1	1	1	-	-	-	0.57
CaEST-036	1	1	1	1	-	1	-	0.71
CaEST-037	1	1	1	1	1	1	1	1.00
CaEST-038	1	1	-	1	-	1	-	0.57
CaEST-039	1	1	1	1	1	1	1	1.00
CaEST-040	1	1	1	-	-	1	1	0.71
CaEST-041	1	1	1	1	1	1	1	1.00
CaEST-042	1	1	1	1	-	1	1	0.86
CaEST-043	1	1	1	1	1	1	1	1.00
CaEST-044	1	1	1	1	1	1	1	1.00
CaEST-045	1	1	1	1	1	1	1	1.00
CaEST-046	1	1	1	1	1	1	1	1.00
CaEST-047	1	1	1	1	1	1	1	1.00
CaEST-048	1	1	1	1	1	1	1	1.00
CaEST-049	1	1	-	1	-	-	1	0.57
CaEST-050	1	1	1	1	1	1	1	1.00
CaEST-055	1	1	1	1	1	1	-	0.86
CaEST-057	1	1	1	1	1	1	-	0.86
CaEST-058	1	1	1	1	1	1	1	1.00
CaEST-062	1	1	1	1	1	1	1	1.00
CaEST-063	1	1	1	1	1	1	1	1.00
CaEST-064	1	1	1	1	1	1	-	0.86
CaEST-074	1	1	-	1	-	1	-	0.57

CaEST-076	1	1	-	1	1	1	-	0.71
CaEST-077	1	1	-	1	1	1	-	0.71
CaEST-078	1	1	1	1	1	1	1	1.00
CaEST-079	1	1	1	1	1	1	1	1.00
CaEST-080	1	1	1	1	1	1	1	1.00
CaEST-081	1	1	1	1	1	1	1	1.00
CaEST-082	1	1	1	1	1	1	1	1.00
CaEST-083	1	1	1	1	1	1	1	1.00
CaEST-084	1	1	1	1	1	1	1	1.00
CaEST-085	1	1	1	1	1	1	1	1.00
CaEST-086	1	1	-	1	-	1	-	0.57
CaEST-087	1	1	1	1	1	1	1	1.00
CaEST-088	1	1	1	1	1	1	-	0.86
CaEST-089	1	1	1	1	-	-	-	0.57
CaEST-090	1	1	1	1	1	-	1	0.86
CaEST-091	1	1	1	1	1	1	1	1.00
CaEST-092	1	1	1	1	1	1	1	1.00
CaEST-093	1	1	1	1	1	1	1	1.00
CaEST-094	1	1	1	1	1	1	1	1.00
CaEST-095	1	1	1	1	-	-	1	0.71
CaEST-096	1	1	1	1	1	1	-	0.86
CaEST-097	1	1	-	1	-	1	-	0.57
CaEST-098	1	1	1	1	1	1	1	1.00
CaEST-099	1	1	-	1	1	1	-	0.71
CaEST-100	1	1	1	1	1	1	1	1.00
CaEST-101	1	1	1	1	-	1	1	0.86
<b>T<sub>MAR</sub></b>	<b>1.00</b>	<b>1.00</b>	<b>0.89</b>	<b>0.97</b>	<b>0.70</b>	<b>0.88</b>	<b>0.72</b>	<b>0.88</b>

## 7. CONCLUSÕES GERAIS

Nos estudos de diversidade, a codificação de dados moleculares representou um problema nas análises genéticas para dados simulados e reais. Nos dados simulados, a perda de informações foi observado nas avaliações realizadas para marcadores bialélicos e multialélicos. Nos estudos com dados reais em acessos de *C. canephora*, a codificação dos dados de SSR sugerem menor eficiência dos programas de melhoramento genético de plantas, sobretudo, naqueles que buscam a identificação de grupos heteróticos e avaliação de recursos genéticos armazenados em bancos de germoplasmas. Para aqueles estudos que utilizam mais de um marcador molecular e objetivam a junção dos dados nos estudos de diversidade, foi proposto um índice de ponderação que leva em consideração as características intrínsecas de cada ensaio sejam consideradas de modo que a informatividade, cobertura e qualidade dos marcadores sejam ponderadas no cálculo da distância genética.

Nas análises de QTLs para resistência à *H. vastatrix*, o estudo apresentou resultados de grande valia para programas de melhoramento genético, principalmente, naqueles que objetivam a identificação de genótipos resistentes para as raças fisiológicas II e XXXIII do patógeno; e identificação de regiões genômicas associadas a resistência à doença, que podem ser utilizados, futuramente, em programas de seleção assistida por marcadores e clonagem posicional. Além disso foi apresentado o primeiro mapa de ligação da espécie utilizando a abordagem de marcadores com diferentes padrões de segregação, o que é uma ferramenta importante em estudos comparativos e de evolução dentro do gênero *Coffea*.

Por fim, foram disponibilizados um conjunto de 101 novos EST-SSR para a espécie *C. arabica*, minerados a partir do Projeto Brasileiro do Genoma Café. A caracterização e validação desses marcadores demonstraram a sua utilidade em estudos de caracterização molecular, diversidade genética, classificação taxonômica e mapeamento genético. A taxa de transferabilidade foi mensurada e mostrou que esses estudos podem ser ampliados para outras espécies dentro do gênero *Coffea*, incluindo estudos evolutivos e de filogenia molecular. Os marcadores apresentaram grande aplicabilidade nos estudos genéticos dentro da espécie *C. canephora*.