

ROBSON FERNANDO MISSIO

MARCADORES MICROSSATÉLITES PARA *Coffea arabica* L.

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

VIÇOSA
MINAS GERAIS – BRASIL
2008

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

T

M678m
2008 Missio, Robson Fernando, 1979-
Marcadores microssatélites para *Coffea arabica* L. /
Robson Fernando Missio. – Viçosa, MG, 2008.
x, 73f.: il. ; 29cm.

Orientador: Ney Sussumu Sakiyama.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Café - Melhoramento genético. 2. Genética molecular.
3. Marcadores genéticos. 4. *Coffea arabica*. I. Universidade
Federal de Viçosa. II. Título.

CDD 22.ed. 633.73

ROBSON FERNANDO MISSIO

MARCADORES MICROSSATÉLITES PARA *Coffea arabica* L.

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

APROVADA: 21 de julho de 2008

Prof. Cosme Damião Cruz

Prof. Messias Gonzaga Pereira

Pesq. Eveline Teixeira Caixeta
(Co-Orientadora)

Pesq. Eunize Maciel Zambolim
(Co-Orientadora)

Prof. Ney Sussumu Sakiyama
(Orientador)

Aos meus pais, Sergito e Nelita, pelo apoio em minha vida pessoal e acadêmica, meus eternos educadores...

Aos meus irmãos Graciela e Leonardo...

MINHA HOMENAGEM

À minha esposa Vivian Carré Missio, pelo companheirismo, apoio e compreensão, com amor...

OFEREÇO E DEDICO

AGRADECIMENTOS

A Deus, pela vida, saúde e força concebida para superar as dificuldades enfrentadas durante este caminho.

À Universidade Federal de Viçosa e ao Programa de Pós-graduação em Genética e Melhoramento pela oportunidade de realização deste curso.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudo e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro para o desenvolvimento deste projeto (Processo Nº: CAG-675/03).

À minha esposa Doutora Vivian Carré Missio, pelo companheirismo, apoio e compreensão fundamentais para meu crescimento pessoal e acadêmico.

Ao Professor Ney Sussumu Sakiyama, pela amizade, orientação, apoio, incentivo, e, sobretudo, pela confiança depositada para o desenvolvimento deste trabalho.

Ao Laboratório de Biotecnologia do Cafeeiro, em especial à Doutora Eunize Maciel Zambolim e Professor Laércio Zambolim pela confiança, amizade e pela disponibilização das condições adequadas de trabalho, fundamental para a realização desta tese.

À Doutora Eveline Teixeira Caixeta, pesquisadora da Embrapa-Café pelo aconselhamento, amizade, auxílio, incentivo e sugestões cruciais durante a realização desta tese.

Ao Professor Luiz Antônio dos Santos Dias, responsável pela minha vinda à Viçosa, pela amizade, apoio, co-orientação, e sugestões.

Aos Professores Everaldo Gonçalves de Barros e Vagner Tebaldi de Queiroz, pelo auxílio no desenvolvimento de protocolo, disponibilização de materiais e sugestões para a realização deste trabalho.

Ao Professor Cosme Damião Cruz pelo auxílio e colaboração fundamental para o desenvolvimento deste trabalho.

A todos os amigos do Laboratório Biocafé que compartilharam idéias, trocaram experiências e também contribuíram para o desenvolvimento deste trabalho.

Aos amigos e colegas do Programa de Pós-Graduação em Genética e Melhoramento.

Aos meus pais Sergito e Nelita R. Missio, e meus irmãos Graciela e Leonardo Missio, pelo amor, apoio, incentivo e preocupações inerentes a minha formação pessoal e acadêmica.

BIOGRAFIA

Robson Fernando Missio, Filho de Sergito Missio e Nelita Ritterbucsh Missio, nasceu em Caiçara estado do Rio Grande do Sul, no dia 11 de dezembro de 1979.

Cursou o primeiro grau na Escola Municipal Sagrado Coração de Jesus, Linha Lazzari, no município de Caiçara, RS. Em dezembro de 1996 formou-se Técnico em Agropecuária, pelo Colégio Agrícola do Oeste do Paraná, em Palotina, Paraná.

Em Junho de 2002, concluiu o curso de Agronomia na Universidade Estadual do Oeste do Paraná (Unioeste), em Marechal Cândido Rondon, Paraná.

Em Julho de 2004, concluiu o curso de Mestrado em Sistemas de Produção pela Universidade Estadual Paulista “Julio de Mesquita Filho”, Ilha Solteira, São Paulo. Em agosto de 2004 iniciou o curso de Doutorado em Genética e Melhoramento pela Universidade Federal de Viçosa.

SUMÁRIO

RESUMO	vii
ABSTRACT	ix
INTRODUÇÃO GERAL.....	1
OBJETIVO GERAL.....	4
REFERÊNCIAS BIBLIOGRÁFICAS	5

CAPÍTULO I

DESENVOLVIMENTO E VALIDAÇÃO DE MARCADORES MICROSSATÉLITES PARA *Coffea arabica* L

Resumo	9
Introdução.....	10
Material e Métodos	11
Material genético e extração de DNA.....	11
Construção das bibliotecas genômicas enriquecidas	11
Digestão do DNA genômico.....	11
Remoção das extremidades coesivas, desfosforilação e ligação dos fragmentos de DNA aos adaptadores	11
Reação de <i>nick-ligation</i>	12
Enriquecimento dos fragmentos de DNA genômicos	12
Clonagem, transformação, seleção e sequenciamento dos clones positivos	13
Análise das seqüências, desenho e nomenclatura dos <i>primers</i>	13
Validação dos <i>primers</i> SSR.....	14
Resultados	15
Análise das seqüências dos clones positivos	15
Frequência dos SSR	15
Desenho e avaliação dos <i>primers</i> SSR.....	16
Discussão	16
Análise das seqüências e frequência dos SSR.....	16
Validação dos marcadores SSR	18
Referências	19

CAPÍTULO II

NOVOS MARCADORES MICROSSATÉLITES PARA *Coffea arabica*

Resumo	34
Marcadores SSR para <i>Coffea arabica</i>	35

Referências	39
--------------------------	-----------

CAPÍTULO III

COMPARAÇÃO DA EFICIÊNCIA DE MARCADORES EST-SSR E SSR GENÔMICO PARA ESTUDO DA DIVERSIDADE GENÉTICA EM *Coffea*

Resumo	45
Introdução	46
Material e Métodos	47
Material genético e extração de DNA	47
Marcadores microssatélites	47
Análise estatística.....	48
Resultados	48
Número de alelos	48
Nível de polimorfismo e diversidade genética entre e dentro de populações de <i>Coffea</i>	49
Análise de agrupamento e distância genética	50
Número de marcas e análise discriminante	51
Discussão	52
Referências	57

RESUMO

MISSIO, Robson Fernando, D.Sc., Universidade Federal de Viçosa, julho de 2008.

Marcadores microssatélites para *Coffea arabica* L. Orientador: Ney Sussumu Sakiyama. Co-Orientadores: Eveline Teixeira Caixeta, Eunize Maciel Zambolim e Luiz Antônio dos Santos Dias.

Os marcadores microssatélites (SSR) são desejáveis para estudos genéticos e seleção assistida por marcadores. A limitada disponibilidade destes marcadores para *Coffea arabica*, uma das principais culturas em importância sócio-econômica do mundo, justifica novos esforços para o desenvolvimento de marcadores SSR. Portanto, o objetivo deste trabalho foi desenvolver, caracterizar e avaliar a utilização de marcadores SSR para estudos da diversidade genética em *Coffea*. Duas bibliotecas genômicas enriquecidas para as sequências repetidas (GT)₁₅ e (AGG)₁₀ foram construídas, utilizando-se o DNA genômico do genótipo Bourbon Amarelo UFV 570. Um total de 756 clones positivos, com sequências de boa qualidade, foi utilizado para a localização de 287 SSR e o desenho de 96 pares de *primers* SSR. Aproximadamente 15% dos clones foram redundantes. As repetições de dinucleotídeos foram mais frequentes que os tri- e tetranucleotídeos, sendo as classes AG (32,8%), AC (12,9%), AAG (9,8%) e AGG (5,6%) as mais abundantes na porção analisada do genoma de *C. arabica*. Em média, foi encontrado um SSR a cada 1,45kb. Do total de 96 *primers* SSR desenhados e sintetizados, 90 (94%) foram validados como marcadores SSR úteis para estudos genéticos em *C. arabica*, sendo 21 (23%) polimórficos entre os cafeeiros Híbrido de Timor UFV 445-46 e Catuaí UFV 2143-235. Trinta e três *primers* SSR foram analisados em 24 acessos de *Coffea* importantes para o programa de melhoramento da UFV/EPAMIG. O número médio de alelos por *primer* foi de 5,1 e os valores médios de PIC foram de 0,56 para *C. canephora* e de 0,22 para *C. arabica*, variando de 0,08 a 0,79 entre todos os acessos avaliados. Neste trabalho, também, foi comparada a eficiência de 18 marcadores SSR genômicos e 17 EST-SSR para o estudo da diversidade genética em *Coffea*. O número médio de alelos/*primer* foi de 5,1 e 5,3 para os marcadores EST-SSR e SSR genômico, respectivamente. Os marcadores SSR genômicos apresentaram maior número de alelos exclusivos e foram mais polimórficos dentro e entre os grupos genéticos, quando comparado com os EST-SSR. Os SSR genômicos apresentaram também maiores coeficientes de dissimilaridade genética e maior número de alelos discriminantes e foram eficientes em distinguir todos os acessos estudados. Os resultados demonstraram o potencial dos marcadores SSR genômicos na detecção do nível de polimorfismo, dos alelos exclusivos e da dissimilaridade genética,

tornando-os úteis para estudos de diversidade genética em *Coffea*, principalmente em *C. arabica* que apresenta pequena variabilidade genética. Este trabalho aumentou o número de marcadores SSR disponíveis para estudos genéticos em *C. arabica* e espécies relacionadas do gênero *Coffea* importantes para o melhoramento genético.

ABSTRACT

MISSIO, Robson Fernando, D.Sc., Universidade Federal de Viçosa, July of 2008.

Microsatellites markers for *Coffea arabica* L. Adviser: Ney Sussumu Sakiyama. Co-Advisers: Eveline Teixeira Caixeta, Eunize Maciel Zambolim and Luiz Antônio dos Santos Dias.

The molecular markers microsatellites (SSR) are desirable for genetic study and in molecular marker assisted selection for genetic improvement. The limited availability of this marker for *Coffea arabica*, which is one of the principal crops in socio economic activity in the world, justified new efforts for the development of SSR markers. Therefore, the objective of this work was to develop, characterize and evaluate the utilization of the SSR markers, for the study of genetic diversity in *Coffea*. Two enriched genomic libraries for the repetitive sequences (GT)₁₅ and (AGG)₁₀, using genomic DNA from Bourbon Amarelo (UFV 570) genotype, were constructed. A total of 756 positive clones with a sequence of good quality were utilized for the localization of 287 SSR and to design 96 SSR primer pairs. Approximately 15% of those clones were redundant. The repeated sequence of dinucleotides were more frequent than trinucleotides and tetranucleotides, being as classes AG (32.8%), AC (12.9%), AAG (9.8%) and AGG (5.6%) were more abundant in the portion of the *C. arabica* genome analyzed. In average one SSR was found in each 1.45kb. From a total of 96 SSR primers designed and synthesized, 90 (94%) were tested as SSR marker useful for genetic study in *C. arabica*, being 21 (23%) polymorphic among progenitors of Híbrido de Timor UFV 445-46 and Catuaí UFV 2143-235. Thirty three SSR primers were tested in 24 accessions of *Coffea* which are important for the breeding program of UFV/EPAMIG. The average number of alleles per primer was 5.1 and the average value of PIC were 0.56 for *C. canephora* and 0.22 for *C. arabica*, within range of 0.08 up to 0.79 among all the accessions evaluated. In this work the efficiency of 18 genomic-SSR markers and 17 EST-SSR markers for genetic diversity study in *Coffea*, were also compared. The number of alleles/primer was 5.1 and 5.3 for EST-SSR markers and genomic-SSR markers, respectively. Those markers of genomic-SSR produced a higher number of exclusive alleles and were more polymorphic within and among the genetic groups, when compared with EST-SSR markers. The genomic-SSR markers also showed higher genetic dissimilarity and higher number of discriminate alleles, and were efficient to distinguish all the accessions studied. The result obtained demonstrated the potencial of genomic-SSR markers in detecting the level of polymorphism, the exclusive alleles, and the genetic dissimilarity, that make these markers usefull for genetic

diversity study in *Coffea*, mainly in *C. arabica* which present low genetic variability. This work increased the number of SSR markers available for the genetic study in *C. arabica* and related species of the genera *Coffea* which are important for the genetic improvement.

INTRODUÇÃO GERAL

O Brasil é um dos principais países produtores e exportadores de café do mundo. A produção brasileira representa aproximadamente 29% da produção mundial de café. As exportações no período de maio de 2007 a abril de 2008 ultrapassaram os 28 milhões de sacas, segundo a Organização Internacional do Café (International Coffee Organization; <http://www.ico.org/>), colocando o Brasil como o principal país fornecedor de café mundial.

O gênero *Coffea* L. (família Rubiácea, subfamília Cinchonoidae, tribo Cofeeae) é constituído de 103 espécies conhecidas e distribuídas em regiões tropicais e subtropicais (Davis et al., 2006). As duas espécies de maior importância comercial e econômica são *Coffea arabica* L. e *Coffea canephora* Pierre ex Froehmer. A espécie *C. arabica* é a única tetraplóide ($2n=4x=44$) do gênero, enquanto as outras espécies são diplóides ($2n=2x=22$).

Estima-se que o tamanho do genoma de *C. arabica* e *C. canephora* seja de aproximadamente $2,56 \times 10^9$ pb e $1,38 \times 10^9$ pb, respectivamente (Clarindo, 2008). Estudos citogenéticos recentes revelaram que *C. arabica* é um verdadeiro alotetraplóide (Clarindo, 2008), e não alotetraplóide segmental como suposto anteriormente.

Os programas de melhoramento genético do cafeeiro no Brasil, têm-se voltado principalmente para obtenção de cultivares com elevado potencial produtivo, elevada qualidade de bebida, boa arquitetura da planta e maior resistência a estresses bióticos e abióticos. Por se tratar de uma cultura perene e de período juvenil longo, o melhoramento genético do cafeeiro é demorado, sendo estrategicamente interessante a implementação de técnicas, como as dos marcadores moleculares, para facilitar e acelerar a seleção e avaliação de genótipos superiores.

Os marcadores moleculares têm sido utilizados para auxiliar os programas de melhoramento genético de várias culturas. No melhoramento do cafeeiro, os marcadores moleculares são de grande utilidade para construções de mapas genéticos (Pailard et al., 1996; Ky et al., 2000; Lashermes et al., 2001; Pearl et al., 2004; Teixeira-Cabral et al., 2004; Oliveira et al., 2007; Hendre et al., 2008), orientações de cruzamentos interespecíficos (Coulibaly et al., 2003), determinação da porcentagem de recuperação dos genes no genitor recorrente (Oliveira et al., 2003), caracterização molecular dos genótipos (Etienne et al., 2002; Nunes et al., 2003) e identificação de marcas ligadas a genes de resistência a pragas e doenças (Moreno et al., 2000; Diniz et al., 2005; Mahé et al., 2008). Os marcadores moleculares mais utilizados nos programas de melhoramento genético do cafeeiro, principalmente no Brasil, são RAPD (Random Amplified Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism) e microssatélites.

Microsatélites ou SSR (*Simple Sequence Repeats*) correspondem a seqüências de DNA em que um único par ou um pequeno número de pares de bases (1-6) de DNA são repetidas em *tandem* (Litt e Luty 1989). A principal vantagem dos SSR como marcadores moleculares é o elevado nível de polimorfismo (Morgante e Olivieri 1993). Além disso, eles são marcadores multialélicos, podem ser amplificados por PCR, possuem alta reprodutibilidade e são abundantemente distribuídos por todo o genoma de organismos eucariotos e procariotos (Gur-Arie et al., 2000). Considerando suas diversas vantagens e ampla aplicabilidade em estudos genéticos e em programas de melhoramento, os SSR tornaram-se um dos principais marcadores moleculares utilizados em plantas e animais (Gupta e Varshney 2000).

A maior desvantagem desta classe de marcadores está no custo do desenvolvimento dos *primers* (Zane et al., 2002; Squirrell et al., 2003). O desenvolvimento de marcadores SSR envolve o conhecimento das seqüências únicas adjacentes à região contendo as repetições em tandem, para que seja possível o desenho dos *primers*. Diversas metodologias e estratégias podem ser utilizadas para a obtenção de *primers* SSR, incluindo biblioteca genômica parcial, biblioteca genômica enriquecida, bibliotecas de cDNA e BAC (*Bacterial Artificial Chromosome*), bem como seqüências de DNA depositadas em bancos de dados (Zane et al., 2002).

Atualmente a pesquisa por SSR em bancos de DNA públicos (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>) ou privados, tem sido utilizada como alternativa para diminuir custos e tempo no desenvolvimento dos *primers* SSR para diversas espécies de plantas (Gadaleta et al., 2007). As seqüências de DNA disponibilizadas em bancos públicos são fontes potencialmente atrativas para a identificação de novos marcadores SSR (Caixeta et al., 2006). Entretanto, a maior limitação desta metodologia consiste na pequena disponibilidade de seqüências depositadas nos bancos de dados, ficando limitada às espécies mais estudadas e de grande importância econômica e comercial.

As ESTs (*Expressed Sequence Tags*) são fragmentos seqüenciados do DNA complementar ao RNA mensageiro e representam parte de uma região transcrita do genoma em determinadas condições fisiológicas. Portanto, as ESTs correspondem às seqüências relativamente conservadas do genoma de um organismo. Devido à natureza conservada das seqüências codantes, é esperado um menor polimorfismo dos marcadores SSR obtidos das ESTs quando comparados a SSR de biblioteca genômica (Varshney et al., 2005). Uma análise comparativa entre os SSR genômicos (oriundos de bibliotecas genômicas enriquecidas) e EST-SSR (oriundos de seqüências do genoma funcional – ESTs) revelou vantagens para as duas classes de marcadores (Varshney et al., 2005). Entretanto, devido ao baixo nível de

polimorfismo apresentado para algumas espécies, os EST-SSRs não são tão eficientes quanto os genômicos para distinguir genótipos muito próximos geneticamente. As principais vantagens dos EST-SSR são a grande probabilidade de estarem associados a regiões funcionais do genoma e a maior possibilidade de utilização do mesmo conjunto de *primers* SSR em diferentes espécies do mesmo gênero (Poncet et al., 2006). Os marcadores EST-SSR tem sido utilizados para o estudo da diversidade genética em videira (Scott et al., 2000), cana-de-açúcar (Cordeiro et al., 2003) e café (Aggarwal et al., 2007; Poncet et al., 2006), na integração de mapa genético da soja (Song et al., 2004), no mapeamento genético em trigo (Gao et al., 2004), batata (Feingold et al., 2005) e algodão (Han et al., 2006) e na identificação de genes ligados a resistência do cafeeiro ao bicho mineiro (Pinto et al., 2007).

Atualmente, seqüências ESTs para o gênero *Coffea* têm sido geradas e podem ser utilizadas para o desenvolvimento de *primers* SSR. O Projeto Brasileiro do Genoma Café gerou mais de 214 mil seqüências ESTs (Vieira et al., 2006). Um outro banco que pode ser utilizado é o de seqüências de DNA da espécie *C. arabica* chamado “CoffeeDNA” (<http://www.coffeedna.net/index.php>), criado em 2004 e mantido pela Universidade de Trieste na Itália, que atualmente contém 13.686 seqüências ESTs. Entretanto, embora exista um significativo número de seqüências de ESTs disponíveis para o gênero *Coffea*, até a presente data apenas 58 *primers* EST-SSR foram desenvolvidos (Bhat et al., 2005; Poncet et al., 2006 e Aggarwal et al., 2007).

Devido os acessos e as variedades da espécie *C. arabica*, importantes para o melhoramento genético no Brasil, apresentarem baixa variabilidade genética (Silvestrini et al., 2007), é importante a utilização de marcadores SSR, principalmente os derivados de seqüências genômicas. Estes marcadores, em geral, apresentam maior nível de polimorfismo, tendo, portanto, maior potencial para acessar a variação genética existente. Entretanto, apenas 186 *primers* SSR genômicos estão disponíveis para o gênero *Coffea* (Combes et al., 2000; Rovelli et al., 2000; Baruah et al., 2003; Moncada e McCouch 2004; Leroy et al., 2005; Poncet et al., 2007; Hendre et al., 2008). Diante do exposto, novos esforços para o desenvolvimento de marcadores SSR genômicos são importantes para aumentar a disponibilidade desta classe de marcadores para estudos genéticos no gênero *Coffea*.

OBJETIVO GERAL

Desenvolver, caracterizar e avaliar a utilização de marcadores SSR genômicos em *Coffea*.

Objetivos específicos

- Desenvolver marcadores SSR, a partir de biblioteca genômica enriquecida;
- Caracterizar as diferentes classes de SSR encontradas no genoma de *C. arabica* analisado;
- Validar os marcadores SSR para estudos genéticos no gênero *Coffea*;
- Comparar a eficiência dos marcadores SSR genômicos e EST-SSR para o estudo da diversidade genética em acessos de cafeeiros importantes para o programa brasileiro de melhoramento.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. **Theoretical and Applied Genetic**, 114: 359-372.
- Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK, Aggarwal RK (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.
- Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P, Aggarwal RK (2003) Isolation and characterization of nine microsatellite markers from *Coffea arabica* (L.) showing wide crossspecies amplifications. **Molecular Ecology Notes**, 3: 647-650.
- Caixeta ET, Oliveira ACB, Brito GG, Sakiyama NS (2006) **Tipos de marcadores moleculares**. In: Borém A, Caixeta ET. Marcadores moleculares. Viçosa, MG, 374p.
- Clarindo WR (2008) **Utilização de técnicas citogenéticas, de citometria de fluxo e de imagem para caracterização do genoma de *Coffea canephora* e *C. arabica***. Tese (Doutorado – Universidade Federal de Viçosa), 113p.
- Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G, Lashermes P (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. **Molecular Ecology**, 9: 1171-1193.
- Cordeiro GM, Pan YB, Henry RJ (2003) Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. **Plant Science**, 165: 181-189.
- Coulibaly I, Revol B, Noirot M, Poncet V, Lorieux M, Carasco-Lacombe C, Minier J, Dufour M, Hamon P (2003) AFLP and SSR polymorphism in a *Coffea* interspecific backcross progeny [(*C. heterocalyx* x *C. canephora*) x *C. canephora*]. **Theoretical and Applied Genetic**, 107: 1148-1155.
- Davis AP, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). **Botanical Journal of the Linnean Society**, 152: 465-512.
- Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET, Oliveira ACB (2005) Analysis of AFLP marker associated to the Mex-1 resistance locus in Icatu progenies. **Crop Breeding and Applied Biotechnology**, 5: 387-393.

- Etienne H, Anthony F, Dussert S (2002) Biotechnological applications for the improvement of coffee (*Coffea arabica* L.). **In Vitro Cellular Development Biological Plant**, 38: 129-138.
- Feingold S, Lloyd J, Norero N, Bonierbale M, Lorenzen J (2005) Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). **Theoretical and Applied Genetic**, 111: 456-466.
- Gadaleta A, Mangini G, Mulè G, Blanco A (2007) Characterization of dinucleotide and trinucleotide EST-derived microsatellites in the wheat genome. **Euphytica**, 153: 73-85.
- Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. **Theoretical and Applied Genetics**, 108: 1392-400.
- Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. **Euphytica**, 113: 163-185.
- Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM, Kashi Y. (2000) Simple Sequence Repeats in *Escherichia coli*: Abundance, Distribution, Composition, and Polymorphism. **Genome Research**, 10: 62-71.
- Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. **Theoretical and Applied Genetic**, 112: 430-439.
- Hendre PS, Phanindranath R, Annapurna V, Lalremruata A, Aggarwal RK (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 Biology**, 51: doi:10.1186/1471-2229-8-51.
- Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Charrier A, Hamon S, Noirot M (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). **Theoretical and Applied Genetic**, 101: 669-676.
- Lashermes P, Combes MC, Prakash NS, Trouslot P, Lorieux M, Charrier A (2001) Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. **Genome**, 44: 589-596.
- Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Ferreira LP, Jourdan I, Pot D, Andrade AC, Glaszmann JC, Vieira LGE, Piffanelli P (2005) Construction and characterization of a *Coffea canephora*BAC library to study the organization of sucrose biosynthesis genes. **Theoretical and Applied Genetic**, 111: 1032-1041.

- Litt M, Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American Journal of Human Genetics**, 44: 397-401.
- Mahé L, Combes MC, Várzea VMP, Guilhaumon C, 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.
- Moncada P, McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome**, 47, 501-509.
- Moreno G, Cortina H, Alvarado G, Gaitan A (2000) Utilización de los recursos genéticos de café en el programa de mejoramiento genético de *C. arabica*, en Colombia. In: **Atelier sur l'amélioration durable du caféier *Coffea arabica***, 2000, CATIE, Turrialba, Costa Rica, p.33-38.
- Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. **The Plant Journal**, 3: 175-182.
- Nunes LM, Crochemore ML, Milach SCK, Oliveira AS, Vieira LGE (2003) **Distinção entre cultivares de cafeeiro por AFLP: perspectivas para o registro e proteção**. In: III Simpósio de Pesquisa dos Cafés do Brasil – Workshop Internacional de Café & Saúde. Anais...Porto Seguro-BA. Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café, p.92.
- Oliveira ACB, Caixeta ET, Zambolim EM, Diniz LEC (2003) **Avanços tecnológicos em biologia molecular: marcadores de DNA no melhoramento do cafeeiro**. In: Zambolim L. Produção integrada de café. Viçosa, 710p.
- Oliveira CB, Sakiyama NS, Caixeta ET, Zambolim EM, Rufino RN, Zambolim L (2007) Mapa parcial de ligação gênica de *Coffea arabica* L. e recuperação do genitor recorrente em progênies de retrocruzamentos. **Crop Breeding and Applied Biotechnology**, 7: 196-203.
- Paillard M, Lashermes P, Petiard V (1996) Construction of a molecular linkage map in coffee. **Theoretical and Applied Genetic**, 93: 41-47.
- Pearl HM, Nagai C, Moore PH, Steiger DL, Osgood RV, Ming R (2004) Construction of a genetic map for arabica coffee. **Theoretical and Applied Genetic**, 108: 829-835.
- Pinto FO, Maluf MP, Guerreiro-Filho O (2007) Study of simple sequence repeat markers from coffee expressed sequences associated to leaf miner resistance. **Pesquisa Agropecuária Brasileira**, 42: 377-384.

- Poncet V, Dufour M, Hamon P, Hamon S, Kochko A, Leroy T (2007) Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. **Genome**, 50: 1156-1161.
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, De Kochko A, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. **Molecular and Genetics Genomics**, 276: 436-449.
- Rovelli P, Mettullo R, Anthony F (2000) **Microsatellites in *Coffea arabica* L.** In: Coffee Biotechnology and Quality (eds Sera T, Soccol CR, Pandey A, Roussos S), pp. 123-133. Kluwer Academic Publishers.
- Scott KD, Egger P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. **Theoretical and Applied Genetic**, 100: 723-726.
- Silvestrini M, Junqueira MG, Favarin AC, Guerreiro-Filho O, Maluf MP, Silvarolla MB, Colombo CA (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. **Genetics Resource and Crop Evolution**, 54: 1367-1379.
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. **Theoretical and Applied Genetic**, 109: 122-128.
- Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? **Molecular Ecology**, 12: 1339-1348.
- Teixeira-Cabral TA, Sakiyama NS, Zambolim L, Pereira AA, Barros EG, Silva DS (2004) Characterization of differential coffee tree hosts for *Hemileia vastatrix* Berk. et Br. with RAPD markers. **Crop Breeding and Applied Biotechnology**, 4: 68-73.
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. **TRENDS in Biotechnology**, 23: 48-55.
- Vieira LGE, Andrade AC, Colombo CA, et al. (2006) Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, 18: 95-108.
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. **Molecular Ecology**, 11: 1-16.

CAPÍTULO I

Desenvolvimento e validação de marcadores microssatélites para *Coffea arabica* L.

Resumo

Com o objetivo de desenvolver novos marcadores SSR para *Coffea arabica*, duas bibliotecas genômicas enriquecidas para as repetições SSR (GT)₁₅ e (AGG)₁₀ foram construídas, utilizando-se o DNA genômico do genótipo Bourbon Amarelo UFV 570. Os programas SSRIT e *Primer3* foram utilizados para localizar os SSRs e desenhar os *primers*, respectivamente. Um total de 835 clones positivos foi seqüenciado, dos quais 756 apresentaram seqüências de boa qualidade e foram utilizados para a localização de SSR e desenho de *primers*. Aproximadamente 15% dos clones foram redundantes. Um total de 287 SSR foi encontrado, sendo 170 (59,2%) repetições simples e 117 (40,8%) repetições compostas. As repetições de dinucleotídeos foram mais freqüentes que os tri- e tetranucleotídeos, sendo as classes AG (32,8%), AC (12,9%), AAG (9,8%) e AGG (5,6%) as mais abundantes na porção do genoma de *C. arabica* analisada. Em média foi encontrado um SSR a cada 1,45kb das seqüências analisadas. Um total de 96 *primers* SSR foram desenhados e sintetizados, dos quais 90 (94%) foram validados como marcadores SSR úteis para estudos genéticos em *C. arabica*, resultando numa elevada taxa de conversão *primer*/marcador. Dos 90 marcadores SSR desenvolvidos, 21 (23%) foram polimórficos entre o Híbrido de Timor UFV 445-46 e o Catuaí UFV 2143-235 e estão disponíveis para incorporação em mapa genético construído a partir de uma população F₂ proveniente do cruzamento desses progenitores. Este trabalho aumentou o número de marcadores SSR disponíveis para estudos genéticos em *C. arabica* e espécies relacionadas do gênero *Coffea* importantes para o melhoramento genético.

Palavras-chave: SSR genômico, biblioteca genômica enriquecida, café, marcador molecular.

Introdução

Microsatélites ou SSR (*Simple Sequence Repeats*) correspondem a seqüências de DNA em que um único par ou um pequeno número de pares de bases (1-6) são repetidas em *tandem* (Litt e Luty 1989). Os SSR estão presentes nas regiões codantes e não codantes do genoma de eucariotos e procariotos, sendo caracterizados pelo elevado nível de polimorfismo (Gur-Arie et al., 2000). Os SSR tornaram-se um dos principais marcadores moleculares para estudos genéticos, principalmente pelo elevado nível de polimorfismo, multialelismo e alta reprodutibilidade (Zane et al., 2002).

A maior desvantagem dos marcadores SSR é o elevado custo, tempo e trabalho necessários até o desenvolvimento dos *primers* (Zane et al., 2002; Squirrell et al., 2003). Diferentes estratégias para o desenvolvimento dos *primers* SSR estão disponíveis atualmente, dentre as quais se destaca a metodologia de biblioteca genômica enriquecida por meio de hibridização seletiva (Zane et al., 2002). As duas estratégias mais utilizadas para o desenvolvimento de *primers* SSR por hibridização seletiva são a utilização de sondas SSR ligadas à membranas de nylon (Armour et al., 1994; Edwards et al., 1996) e as marcadas com biotina, que são recuperadas por meio de esferas magnéticas cobertas com estreptavidina (Kijas et al., 1994; Kandpal et al., 1994; Hamilton et al., 1999). A metodologia de hibridização seletiva permite obter grande quantidade de fragmentos de DNA contendo regiões SSR. Nesta metodologia os fragmentos de DNA são hibridizados com sondas complementares, aumentando significativamente o número de clones contendo seqüências SSR para o desenho de *primers*. O desenvolvimento de *primers* SSR por meio de biblioteca genômica enriquecida tem sido amplamente utilizada para muitas espécies de plantas como eucalipto (Brondani et al., 1998), piqui (Collevatti et al., 1999), pimenta (Buso et al., 2000), cana-de-açúcar (Cordeiro et al., 2000), feijão (Queiroz, 2004; Pereira, 2005), arroz (Brondani et al., 2001), abacate (Ashworth et al., 2004), lichia (Viruel e Hormaza 2004), melão (Ritschel et al., 2004), lúpulo (Stajner et al., 2005), amora (Zhao et al., 2005) e trigo (Song et al., 2005).

Para o gênero *Coffea*, poucos *primers* SSR foram desenvolvidos e estão disponíveis para estudos genéticos (Tabela 1), quando comparado com culturas como milho, soja, arroz, trigo e cevada (<http://www.gramene.org/>). Em soja, Song et al. (2004) desenvolveram um mapa integrado contendo 1.015 marcadores SSR. Em milho, 2.095 marcadores SSR já foram mapeados (<http://www.maizegdb.org/ssr.php>). Em café, até a presente data, 251 marcadores SSR estão disponíveis (Combes et al., 2000; Rovelli et al., 2000; Baruah et al., 2003; Moncada e McCouch 2004; Bhat et al., 2005; Poncet et al., 2006; Poncet et al., 2007; Aggarwal et al., 2007; Tesfaye et al., 2007; Hendre et al., 2008) e poucos foram mapeados

(Tabela 1). Do total de *primers* SSR disponíveis, 153 (61%) são derivados da espécie *Coffea canephora* e 98 (39%) de *C. arabica*.

O objetivo deste trabalho foi desenvolver e validar novos marcadores SSR para café, a partir de biblioteca genômica enriquecida, para serem incorporados aos programas de pesquisa do café.

Material e Métodos

Material genético e extração de DNA

O genótipo de *C. arabica*, Bourbon Amarelo (BA), acesso UFV 570, foi utilizado para construção das bibliotecas genômica enriquecida para repetições SSR. O DNA genômico foi extraído de folhas jovens e completamente desenvolvidas seguindo procedimento descrito por Diniz et al. (2005). A qualidade do DNA foi analisada em gel de agarose e a sua quantificação em espectrofotômetro Smart Spec da BioRad.

Construção das bibliotecas genômicas enriquecidas

A construção das bibliotecas genômicas enriquecidas foi realizada seguindo o protocolo descrito por Hamilton et al. (1999) e Queiroz (2004). Resumidamente, as seguintes etapas foram seguidas:

Digestão do DNA genômico

O DNA genômico (50µg) do genótipo BA foi digerido utilizando 50U de cada uma das enzimas *EcoRI*, *NheI*, *HaeIII* e *RsaI* em um volume final de 200µL. Os fragmentos de tamanho entre 200-1000pb foram excisados do gel e purificados com o Kit Gel Extraction (QIAGEN, Valencia, CA, EUA), segundo recomendações do fabricante.

Remoção das extremidades coesivas, desfosforilação e ligação dos fragmentos de DNA aos adaptadores

A reação para remoção das extremidades coesivas foi realizada com 45µg (225ng/µL) de DNA digerido, 15U da enzima MBN (*Mung Bean Nuclease*), tampão MBN 1X e água deionizada para completar o volume de 240µL. A reação foi transferida para um microtubo de 1,5µL e incubada a 30°C por 40 minutos. Os fragmentos de DNA com extremidades abruptas foram purificados com o Kit PCR Purification (QIAGEN, Valencia, CA, EUA), segundo recomendações do fabricante e eluídos em 50 µL de tampão EB (Tris-HCl 10 mM, pH 8,5).

A desfosforilação foi realizada com 40µg (200ng/µL) dos fragmentos de DNA com extremidades abruptas, 40U da enzima CIP (*Calf Intestinal Alkaline Phosphatase*), tampão NEB 3 1X e água deionizada para completar 200µL. A reação foi incubada a 37°C por 4 horas. Os fragmentos de DNA desfosforilados foram purificados com o Kit PCR Purification e eluídos em 28µL de tampão EB, segundo recomendações do fabricante.

A ligação do adaptador aos fragmentos de DNA foi realizada em um volume de 50µL, contendo 3µg (60ng/µL) dos fragmentos desfosforilados, 2µM do adaptador SNX (Hamilton et al., 1999) fita dupla (SNX_F: 5'-CTAAGGCCTTGCTAGCAGAAGC-3' e SNX_R: 5'-GCTTCTGCTAGCAAGGCCTTAGAAAA-3'), tampão OPA 1X, rATP 1 mM, 100µg/mL de BSA, 10U de *XmnI* e 15U da T4 DNA ligase (Promega, Madison, WI, EUA). A reação de ligação foi transferida para um termociclador PTC 200 (MJ Research), programado para 19 ciclos a 16°C/30 minutos e 37°C/10 minutos, com uma etapa final a 65°C/20 minutos para inativar as enzimas *XmnI* e T4 DNA ligase. O conteúdo da reação permaneceu a 22°C/30 minutos e em seguida, foi armazenado a -20°C.

A eficiência da reação de ligação foi confirmada por PCR, utilizando-se o oligonucleotídeo SNX_F como *primer*. Os produtos da reação foram aplicados em gel de agarose (1,2%) a 90 volts por aproximadamente 1 hora e corado com brometo de etídeo.

Reação de nick-ligation

Esta reação é importante para corrigir o *nick* formado entre a extremidade 5' desfosforilada dos fragmentos e a extremidade 3'-OH da fita SNX_F do adaptador.

O conteúdo da reação de ligação foi dividido em alíquotas de 3,5µL. Para cada alíquota foram adicionados MgCl₂ 2mM, 25µg/mL de BSA, 160µM de cada um dos desoxinucleotídeos (dNTPs) e 1U da enzima *Taq* DNA polimerase e tampão 1X, (Promega, Madison, WI, EUA). As reações foram realizadas em termociclador, programado para 72°C/30 minutos. Posteriormente, o conteúdo de todas as reações foi reunido, purificado e eluído em 28µL de tampão EB do Kit PCR Purification, segundo recomendações do fabricante.

Enriquecimento dos fragmentos de DNA genômicos

O enriquecimento dos fragmentos contendo microssatélites foi realizado pela hibridização com as sondas de microssatélites biotinizadas (GT)₁₅ e (AGG)₁₀. Após a desnaturação do DNA a 95°C/15 minutos, os fragmentos foram hibridizados com 2 pmols de cada sonda em tampão SSC 3X e SDS 0,05%, a 65°C/14 horas. O processo de captura foi

realizado com partículas magnéticas (Dynabeads M-270) cobertas com estreptoavidina, que se liga por força iônica à biotina (Kijas et al., 1994). Após a hibridização, os fragmentos foram lavados duas vezes em baixa estringência (SSC 2X e SDS 0,1% por cinco minutos a temperatura ambiente) e duas vezes em alta estringência (SSC 1X e SDS 0,1% por 15 minutos a 65°C).

Após as lavagens, os fragmentos foram amplificados por PCR num volume de 50µL contendo 0,8µM do *primer* SNX_F, 2mM de MgSO₄, 160µM de cada dNTP, 2U da enzima *Taq* Platinum High Fidelity e tampão 1X (Invitrogen Brasil Ltda) e 15µL dos fragmentos de DNA capturados. A reação foi realizada em termociclador, programado para 96°C/5 minutos e 40 ciclos a 96°C/45 segundos, 62°C/60 segundos e 72°C/90 segundos, seguidos de uma extensão final a 72°C/7 minutos.

Clonagem, transformação, seleção e sequenciamento dos clones positivos

Os fragmentos enriquecidos e amplificados foram digeridos com 2U de *Nhe*I a 37°C por 14 horas, e ligados ao plasmídeo *pBluescript* SK+ (Stratagene) previamente digerido com *Xba*I (New England Biolabs). Células competentes de *Escherichia coli* DH5α foram transformadas com os plasmídeos recombinantes por choque térmico, de acordo com protocolo-padrão (Ausubel, 1998). As células transformadas foram plaqueadas em meio de cultura LB contendo ampicilina (100µg/mL), X-Gal (40µg/mL) e IPTG (10µg/mL) e incubadas a 37°C durante a noite. A seleção e o diagnóstico das colônias transformadas foram realizadas por PCR, utilizando os *primers* T3 e T7 (Invitrogen). Os produtos da amplificação foram separados por eletroforese em gel de agarose 1,2%. As colônias, contendo transformantes com inserções superiores a 400pb foram selecionadas e cultivadas em meio LB líquido contendo ampicilina (100µg/mL) por 16 horas à 37°C sob agitação constante (160rpm).

A extração do DNA plasmidial foi realizada utilizando o kit Wizard Plus SV Minipreps (Promega, Madison, WI, EUA) seguindo as recomendações do fabricante. O DNA dos clones positivos foi diluído e sequenciado.

Análise das seqüências, desenho e nomenclatura dos primers

A análise das seqüências dos fragmentos de DNA foi realizada com auxílio dos programas CodonCode Aligner 1.6.3 (CodonCode Corporation) e SSRIT (<http://www.gramene.org>). O programa CodonCode Aligner foi utilizado para discriminar as regiões de fragmentos do genoma do café arábica, eliminar as seqüências do plasmídeo e verificar a presença de seqüências redundantes. O SSRIT foi usado para localizar as

repetições de microssatélite nas seqüências. Os critérios utilizados para a definição de um microssatélite foram: A) (NN)₄ - 4 repetições de dinucleotídeo; B) (NNN)₃ - 3 repetições de trinucleotídeo. C) (NNNN)₃ - 3 repetições de tetranucleotídeo. D) (NNNNN)₃ - 3 repetições de pentanucleotídeo e E) (NNNNNN)₃ - 3 repetições de hexanucleotídeo. Para repetições imperfeitas, adotou-se a diferença máxima entre dois motivos como sendo de 10pb.

Primers flanqueando os locos microssatélites identificados foram desenhados utilizando o programa Primer3 (Rozen e Skaletsky 2000) seguindo os seguintes critérios: 1) tamanho dos *primers* de 18 a 24pb; 2) T_m de 55 a 60°C; 3) concentração de sal de 50mM; 4) produto de amplificação de 100 a 600pb; 5) porcentagem de GC de 40 a 60%. A qualidade e estabilidade dos *primers* foram confirmadas pelo programa Netprimer (<http://www.premierbiosoft.com/netprimer/netprimer.html>).

Como ainda não há na literatura uma nomenclatura padronizada para *primers* SSR para o cafeeiro, definimos para este trabalho os seguintes termos: (a) o termo SSR foi colocado no início, para diferenciar o tipo de marcador que está sendo utilizado; (b) as letras “Ca” foram empregadas para designar *primers* desenvolvidos para *C. arabica*; (c) um número seqüencial foi adicionado no final para diferenciar um par de *primer* de outro; (d) as letras F e R foram acrescentadas para indicar *primers forward* e *reverse*, respectivamente. Portanto, a representação SSRCa 001F refere-se ao *primer* microssatélite *forward*, número 001, desenvolvido para *C. arabica*.

Validação dos primers SSR

Os *primers* SSR desenvolvidos neste trabalho foram utilizados para amplificação e avaliação do polimorfismo entre os acessos Híbrido de Timor UFV 445-46 e Catuaí UFV 2143-235. O DNA de cada um dos acessos foi extraído de folhas jovens, seguindo o protocolo descrito por Diniz et al. (2005). A reação de PCR foi realizada em um volume total de 20µL contendo 50ng do DNA genômico, 0,6 unidades de *Taq* DNA polimerase e tampão 1x (Promega), 1mM de MgCl₂, 150µM de cada dNTP e 0,1µM de cada *primer*. A amplificação foi efetuada em termociclador PTC 200 (MJ Research), utilizando-se o procedimento *touchdown*-PCR, que consistiu de desnaturação inicial a 94°C/2 minutos, seguido de 13 ciclos a 94°C/30 segundos, 67°C a 55°C/30 segundos, reduzindo 1°C a cada ciclo, e 72°C/30 segundos. Os 13 ciclos foram seguidos de mais 30 ciclos a 94°C/30 segundos, 55°C/30 segundos e 72°C/30 segundos e extensão final a 72°C/8 minutos. O polimorfismo foi visualizado em gel de poliacrilamida desnaturante a 6% e corado com prata conforme protocolo descrito por Creste et al. (2001).

Resultados

Análise das seqüências dos clones positivos

Neste trabalho os fragmentos de DNA das bibliotecas enriquecidas para as repetições (GT)₁₅ e (AGG)₁₀, foram seqüenciados e analisados. Um total de 835 clones positivos foi seqüenciado, sendo que 64 apresentaram problemas no seqüenciamento e 15 clones tiveram seqüências inferiores a 100pb e foram descartados da avaliação (Tabela 2). Desse modo, 756 clones com tamanho suficiente (acima de 100pb) para o desenho de *primers* foram analisados, sendo 530 oriundos da biblioteca (GT)₁₅ e 226 da biblioteca (AGG)₁₀. Utilizando a plataforma SSRIT um total de 287 SSR foi encontrado (Tabela 2). Aproximadamente 15% das seqüências foram redundantes

Frequência dos SSR

O tamanho médio de cada clone seqüenciado foi de 500pb. Foram analisados cerca de 418Kb do genoma de *C. arabica*, nos quais se observaram, em média, um SSR a cada 1,45Kb. Mesmo sendo uma amostragem pequena em relação ao genoma completo do cafeeiro, e de terem sido analisadas bibliotecas genômicas enriquecidas, foram encontrados um dinucleotídeo a cada 2,9Kb, um trinucleotídeo a cada 4,4Kb e um tetranucleotídeo a cada 18,2Kb de seqüência analisada, indicando uma alta frequência de SSR no genoma de *C. arabica*.

Na Tabela 3 e Figura 1 estão apresentadas as classes e frequências de SSR encontradas nas seqüências e analisadas neste trabalho. Os SSR encontrados variaram quanto à classe e número de repetições (Tabela 3). Dos 287 SSR encontrados, 170 (59,2%) foram repetições simples e 117 (40,8%) foram repetições compostas. Aproximadamente 98,2% das repetições simples foram do tipo perfeita (SP), enquanto que para as repetições compostas, 44,4% foram do tipo imperfeita (CI).

Do total de SSR encontrados, 51% foram de dinucleotídeos e 33% de trinucleotídeos (Tabela 3). Mono, tetra, penta e hexanucleotídeos somaram menos de 17% dos SSR encontrados. Dentre os dinucleotídeos, a classe AG/CT foi a mais freqüente (64,4%), seguida por AC/GT (25,3%) e AT/AT (9,6%). A classe CG/CG foi a menos freqüente com 0,7% (Tabela 3 e Figura 1). Entre os trinucleotídeos, a classe AAG/CTT foi a mais freqüente (29,5%), seguida pelas classes AGG/CCT (16,8%), AAC/GTT (15,8%) e ACC/GGT (14,7%), enquanto que a classe AGC/GCT foi a menos freqüente com pouco mais de 1%.

Desenho e avaliação dos *primers* SSR

Dos 287 SSR encontrados, um total de 96 pares de *primers*, denominados SSRCa, foram desenhados e sintetizados (Tabela 4). Entre os 96 pares de *primers* desenhados, 51 (53%) foram repetições de dinucleotídeos e 25 (26%) de trinucleotídeos, sendo seis complementares a sonda GT/CA e 16 a sonda AGG/TCC (Tabela 4). O número máximo de repetições para dinucleotídeos foi de 38 para a classe CT/GA (SSRCa 081), enquanto que para repetições de trinucleotídeos foram encontrados menores números de repetições (Tabela 4).

Todos os *primers* SSR desenvolvidos neste trabalho foram utilizados para avaliar o nível de polimorfismo entre o Híbrido de Timor UFV 445-46 e o Catuaí UFV 2143-235. Dos 96 *primers* SSR desenvolvidos, 90 foram validados como marcadores úteis para estudos genéticos em *C. arabica* (Tabela 4). Destes, 21 (23,3%) apresentaram polimorfismo entre os acessos e 69 (76,7%) foram monomórficos (Tabela 5). O número de alelos, considerando os acessos Híbrido de Timor UFV 445-46 e Catuaí UFV 2143-235, variou de 1 a 4, com uma média de 1,86 alelos/*primer* (Tabela 4).

Discussão

Análise das seqüências e frequência dos SSR

Marcadores SSR tem sido desenvolvidos utilizando-se o procedimento de biblioteca genômica (parcial ou enriquecida) para diversas espécies de plantas (Tabela 5). A eficiência de cada metodologia para o desenvolvimento de marcadores SSR é indicada pela proporção de clones contendo SSR em relação ao número total de clones examinados (Zane et al., 2002). Neste trabalho, 756 clones com seqüências de boa qualidade foram analisadas e 287 SSR foram localizados, resultando numa eficiência de aproximadamente 38%. A eficiência média na detecção de SSR encontrada neste trabalho foi superior a encontrada em amora, 26% (Zhao et al., 2005), *Brassica*, 18,5% (Cui et al., 2008), piqui, 14,4% (Collevatti et al., 1999), *Corchorus capsularis*, 34,5% (Mir et al., 2008) e *Citrus*, 25% (Novelli et al., 2006).

A metodologia empregada para construção das bibliotecas genômica enriquecidas apresentou uma boa eficiência, demonstrado pela ausência de contaminação das seqüências com DNA de *E. coli* ou seqüências sem o inserto (Tabela 2). Aproximadamente 15% das seqüências foram redundantes, que pode ser o resultado da reação de PCR utilizada no processo de conversão dos fragmentos em fita dupla após o procedimento de captura, a qual aumenta milhões de vezes o número de um fragmento para ser clonado.

A frequência, distribuição e abundância dos SSRs são altamente variáveis entre diferentes organismos, principalmente devido aos diferentes critérios de buscas, origem das

seqüências e tamanho do genoma amostrado (Varshney et al., 2005). As maiores freqüências de SSR no genoma de *C. arabica* foram encontradas para repetições dinucleotídeos (51%) e trinucleotídeos (33%), sendo as classes AG/CT (32,8%), AC/GT (12,9%), AAG/CTT (9,8%) e AGG/CCT (5,6%) as mais freqüentes. Este resultado foi similar ao encontrado por Hendre et al. (2008), que utilizando uma biblioteca genômica parcial para a espécie *C. canephora*, encontraram as classes AG/CT (26,7%), AC/GT (12,9%) e AGC/GCT (7,8%) como as mais freqüentes no genoma desta espécie. Analisando ESTs para o gênero *Coffea*, Aggarwal et al. (2007) encontraram 46% de dinucleotídeos, onde a classe AG/CT foi a mais abundante (53%) e 26% de trinucleotídeos, sendo a classe AAG/CTT a mais freqüente (29%). Por outro lado, Poncet et al. (2006) analisando ESTs de *C. canephora*, encontraram 34% de repetições trinucleotídeos, 26% de dinucleotídeos e 22% de hexanucleotídeos, onde as classes AGG (23%) e GA (62,2%) foram as mais freqüentes dentre os tri- e dinucleotídeos, respectivamente.

As repetições de dinucleotídeos foram mais freqüentes em SSR derivados de seqüências genômicas em quinoa (Jarvis et al., 2008), amendoim (Cuc et al., 2008), melão (Ritschel et al., 2004), *citrus* (Novelli et al., 2006), feijão (Benchimol et al., 2007) e cana-de-açúcar (Cordeiro et al., 2000). Entretanto, as repetições de trinucleotídeos foram as mais freqüentes em SSR derivados de ESTs em *C. canephora* (Poncet et al., 2006), soja e arroz (Cardle et al., 2000; Gao et al., 2003), milho, tomate e algodão (Cardle et al., 2000). A maior freqüência de SSR trinucleotídeos em ESTs foi atribuída a menor taxa de mutações em regiões codantes do genoma (Metzgar et al., 2000). As repetições de trinucleotídeos correspondem a um pequeno aminoácido hidrofílico e são mais toleradas em regiões codantes, enquanto que a elevada pressão de seleção provavelmente elimina a expansão desta classe nas regiões não codantes (Katti et al., 2001). Os resultados obtidos neste trabalho mostram a elevada freqüência de di- e trinucleotídeos no genoma da espécie *C. arabica*, em conformidade com o obtido com outras espécies de plantas (Tabela 5), em que as classes AG/CT, AC/GT e AAG/CTT são predominantemente as mais freqüentes tanto em seqüências genômicas quanto em seqüências expressas (ESTs) do genoma.

Neste trabalho, foram analisados cerca de 418Kb do genoma de *C. arabica*, nos quais foram encontrados, em média, um SSR a cada 1,45Kb. Mesmo sendo uma amostragem pequena em relação ao genoma completo do cafeeiro, e ter sido analisado seqüências de biblioteca genômica enriquecida, foram encontrados um dinucleotídeo a cada 2,9Kb, um trinucleotídeo a cada 4,4Kb e um tetranucleotídeo a cada 18,2Kb, indicando uma alta freqüência de SSR no genoma de *C. arabica*. Em ESTs de *C. canephora*, Poncet et al. (2006) encontraram 1 SSR a cada 7,73Kb. Para seqüências ESTs do gênero *Coffea*, Aggarwal et al.

(2007) encontraram 1SSR/2,16Kb do genoma analisado. Em *Eucalyptus*, Ceresini et al. (2005) encontraram 1SSR/2,7Kb de seqüências ESTs analisadas. Maiores freqüências foram obtidas por Morgante et al. (2002), com 1SSR/2,1Kb, 1SSR/1,1Kb e 1SSR/1,3Kb em seqüências ESTs de arroz, milho e trigo, respectivamente. No geral, os resultados obtidos neste trabalho indicam uma maior freqüência de SSR em seqüências de DNA genômico de *C. arabica* do que os obtidos para as espécies descritas acima.

Validação dos marcadores SSR

Para o gênero *Coffea*, poucos *primers* SSR foram desenvolvidos (Combes et al., 2000; Rovelli et al., 2000; Baruah et al., 2003; Moncada & McCouch 2004; Bhat et al., 2005; Poncet et al., 2006; Aggarwal et al., 2007, Poncet et al., 2007, Hendre et al., 2008), quando comparado com culturas como milho, arroz, trigo e cevada (<http://www.gramene.org/>). Neste trabalho, 90 novos marcadores SSR foram validados para *C. arabica*, e podem ser utilizados para estudos genéticos. A proporção de *primers* desenhados que produzem produtos de amplificação, nítidos e bem definidos, é utilizada para avaliar o sucesso na taxa de conversão de *primers* SSR desenhados em marcadores funcionais (Hendre et al., 2008). Dos 96 *primers* SSR desenhados, 90 produziram produtos de amplificação nos acessos Híbrido de Timor UFV 445-46 e Catuaí UFV 2143-235 e foram validados em *C. arabica*, resultando numa taxa de conversão de *primer*/marcador de 94%. Valores similares foram obtidos para *Corchorus capsularis* (91%; Mir et al., 2008), *Lolium multiflorum* (90,4%; Hirata et al., 2006), *Cucurbitáceas* (94%; Gong et al., 2008), *Humulus lupulus* (92,2%; Stajner et al., 2005) e *Avena sativa* (95,5%; Li et al., 2000). Para muitas espécies de plantas como amora, abacate, cana-de-açúcar, arroz, amendoim, *Citrus* e *C. canephora* (Tabela 5), a eficiência de conversão *primers*/marcador foi menor que a encontrada neste trabalho. Uma das menores taxas de conversão *primer*/marcador foi encontrada para *Pinus* (4,1%; Hicks et al., 1998). Estudos realizados por Garner (2002) indicam que a porcentagem de *primers* SSR que não apresentam amplificação é alta e positivamente correlacionada com o tamanho do genoma da espécie estudada. Entretanto, apesar de *C. arabica* possuir um elevado tamanho do genoma, estimado em $2,56 \times 10^9$ pb (Clarindo, 2008), uma alta eficiência na conversão *primer*/marcador foi obtida (94%) neste trabalho, quando comparada ao estudo de Hendre et al. (2008) em *C. canephora* (75,8%) que apresenta um menor tamanho estimado do genoma ($1,38 \times 10^9$ pb; Clarindo, 2008).

Uma das maiores aplicações dos marcadores moleculares é a utilização para construção de mapas genéticos e identificação de marcas ligadas a genes que controlam características de interesse agrônômico, promovendo possibilidades de utilização na seleção assistida por marcadores. Mapas genéticos para o gênero *Coffea* tem sido desenvolvidos

(Pailard et al., 1996; Ky et al., 2000; Lashermes et al., 2001; Pearl et al., 2004; Teixeira-Cabral et al., 2004; Oliveira et al., 2007), entretanto, apenas 37 marcadores SSR foram mapeados para *C. canephora* (Hendre et al., 2008) e nenhum marcador foi mapeado para a espécie *C. arabica*. Portanto, esforços no sentido de ampliar a disponibilidade de marcadores SSR para o gênero *Coffea* são de suma importância para o melhoramento genético. Neste trabalho, 23% dos *primers* SSR que apresentaram produto de amplificação (Tabela 4), foram polimórficos entre os progenitores Híbrido de Timor UFV 445-46 e Catuaí UFV 2143-235. O nível de polimorfismo encontrado com os marcadores SSR desenvolvidos neste trabalho foi superior ao encontrado por Lashermes et al. (2001), Teixeira-Cabral et al. (2004) e Oliveira et al. (2007) utilizando marcadores RAPD. Em um trabalho desenvolvido por Hendre et al. (2008), 20,5% dos SSR desenvolvidos para *C. canephora* foram polimórficos entre os pais e foram mapeados.

Neste trabalho, 90 novos marcadores SSR úteis para estudos genéticos em *C. arabica* foram desenvolvidos por meio de biblioteca genômica enriquecida. Esta metodologia, embora de elevado custo, tempo e dificuldades experimentais, apresentou uma significativa taxa de sucesso no desenvolvimento de marcadores SSR para *C. arabica*. A análise dos clones positivos permitiu observar a relativa abundância e distribuição de diferentes classes de SSR comparada a outras espécies do gênero *Coffea* e também às demais espécies de plantas. Este trabalho aumentou em 36% o número de marcadores SSR atualmente disponíveis para estudos genéticos em *C. arabica* e espécies relacionadas do gênero *Coffea*.

Referências

- Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. **Theoretical and Applied Genetic**, 114: 359-372.
- Armour JA, Neumann R, Gobert S, Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridization selection. **Human Molecular Genetics**, 3: 599-565.
- Ashworth VETM, Kobayashi MC, De La Cruz M, Clegg MT (2004) Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. **Scientia Horticulturae**, 101: 255-267.
- Ausubel F (1998) **Current Protocols in Molecular Biology**. New York: John Wiley & Sons, Tenton Data Systems.

- Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P, Aggarwal RK (2003) Isolation and characterization of nine microsatellite markers from *Coffea arabica* (L.) showing wide crossspecies amplifications. **Molecular Ecology Notes**, 3: 647-650.
- Benchimol LL, Campos T, Carbonell SAM, Colombo CA, Chioratto AF, Formighieri EF, Gouvêa LRL, Souza AP (2007) Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. **Genetics Resources and Crop Evolution**, 54: 1747-1762.
- Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK, Aggarwal RK (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.
- Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical and Applied Genetic**, 97: 816-827.
- Brondani C, Brondani RPV, Rangel PH, Ferreira ME (2001) Development and mapping of *Oryza glumaepatula*-derived microsatellite markers in the interspecific cross *Oryza glumaepatula* x *O. sativa*. **Hereditas**, 134: 59-71.
- Bryan GJ, Collins AJ, Stephenson P, Orry A, Smith JB, Gale MD (1997) Isolation and characterisation of microsatellites from hexaploid bread wheat. **Theoretical and Applied Genetic**, 94: 557-563.
- Buso GSC, Brondani RV, Amaral ZPS, Reis AMM, Ferreira ME (2000) Desenvolvimento de primers SSR para análise genética de pimentas e pimentões (*Capsicum* spp.) utilizando biblioteca genômica enriquecida. **Boletim Técnico da EMBRAPA/EPAGRI**, 15: 01-27.
- Butcher PA, Decroocq S, Gray Y, Moran GF (2000) Development, inheritance and cross-species amplification of microsatellite markers from *Acacia mangium*. **Theoretical and Applied Genetic**, 101: 1282-1290.
- Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. **Genetics**, 156: 847-854.
- Ceresini PC, Silva CLSP, Missio RF, Souza EC, Fischer CN, Guillherme IR, Gregorio I, Silva EHT, Cicarelli RMB, Silva MTA, Garcia JF, Avelar GA, Porto Neto LR, Marçon AR, Bacci Junior M, Marini DC (2005) Satelyptus: Analysis and database of microsatellites from ESTs of *Eucalyptus*. **Genetics and Molecular Biology**, 28: 589-600.

- Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa*L.). **Theoretical and Applied Genetic**, 95: 553-567.
- Clarindo WR (2008) **Utilização de técnicas citogenéticas, de citometria de fluxo e de imagem para caracterização do genoma de *Coffea canephora* e *C. arabica***. Tese (Doutorado – Universidade Federal de Viçosa), 113p.
- Collevatti RG, Brondani RV, Grattapaglia D (1999) Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. **Heredity**, 83: 748-756.
- Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G, Lashermes P (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. **Molecular Ecology**, 9: 1171-1193.
- Cordeiro GM, Taylor GO, Henry RJ (2000) Characterization of microsatellite markers from sugarcane (*Saccharum* sp), a highly polyploid species. **Plant Science**, 155: 161-168.
- Creste S, Tulmann-Neto A, Figueira A (2001) Detection of simple sequence repeat polymorphisms in denaturing polyacrilamide sequencing gels by silver staining. **Plant Molecular Biology Report**, 4: 299-306.
- Cuc LM, Mace ES, Crouch JH, Quang VD, Long TD, Varshney RK (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). **BMC Plant Biology**, 8: 55 doi:10.1186/1471-2229-8-55.
- Cui X, Dong Y, Hou X, Cheng Y, Zhang J, Jin M (2008) Development and characterization of microsatellite markers in *Brassica rapa* ssp. *chinensis* and transferability among related species. **Agricultural Sciences in China**, 7: 19-31.
- Davis AP, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). **Botanical Journal of the Linnean Society**, 152: 465-512.
- Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET, Oliveira ACB (2005) Analysis of AFLP marker associated to the *Mex-1* resistance locus in Icatu progenies. **Crop Breeding and Applied Biotechnology**, 5: 387-393.
- Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite Libraries Enriched for Several Microsatellite Sequences in Plants. **BioTechniques**, 20: 758-760.
- Gao LF, Tang J, Li H, Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. **Molecular Breeding**, 12: 245-261.

- Garner TW (2002) Genome size and microsatellites: the effect of nuclear size on amplification potential. **Genome**, 45: 212-215.
- Gong L, Stift G, Kofler R, Pachner M, Lelley T (2008) Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L. **Theoretical and Applied Genetic**, doi: 10.1007/s00122-008-0750-2.
- Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM, Kashi Y. (2000) Simple Sequence Repeats in *Escherichia coli*: Abundance, Distribution, Composition, and Polymorphism. **Genome Research**, 10: 62-71.
- Hamilton MB, Pincus EL, Di Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. **Biotechniques**, 27: 500-507.
- Hendre PS, Phanindranath R, Annapurna V, Lalremruata A, Aggarwal RK (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 Biology**, doi: 10.1186/1471-2229-8-51.
- Hicks M, Adams D, O'Keefe S, Macdonald E, Hodgetts R (1998) The development of RAPD and microsatellite markers in lodgepole pine (*Pinus contorta* var. *latifolia*). **Genome**, 41: 797-805.
- Hirata M, Cai H, Inoue M, Yuyama N, Miura Y, Komatsu T, Takamizo T, Fujimori M (2006) Development of simple sequence repeat (SSR) markers and construction of an SSR-based linkage map in Italian ryegrass (*Lolium multiflorum* Lam.). **Theoretical and Applied Genetic**, 113: 270-279.
- Hopkins MS, Casa AM, Wang T, Mitchell SE, Dean RE, Kochert GD, Kresovich S (1999) Discovery and Characterization of Polymorphic Simple Sequence Repeats (SSRs) in Peanut. **Crop Science**, 39: 1243-1247.
- Jarvis DE, Kopp OR, Jellen EN, Mallory MA, Pattee J, Bonifacio A, Coleman CE, Stevens MR, Fairbanks DJ, Maughan PJ (2008) Simple sequence repeat marker development and genetic mapping in quinoa (*Chenopodium quinoa* Willd.). **Journal of Genetics**, 87: 39-51.
- Kandpal RP, Kandpal G, Weissman SM (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. **Proceedings of the National Academy of Sciences of the USA**, 91: 88-92.
- Katti MV, Rangeekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. **Molecular Biology and Evolution**, 18: 1161-1167.

- Kijas JHM, Fowler JCS, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. **BioTechniques**, 16: 657-662.
- Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Charrier A, Hamon S, Noirot M (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). **Theoretical and Applied Genetic**, 101: 669-676.
- L'taief B, Horres R, Jungmann R, Molina C, SiW B, Lachaâl M, Winter P, Kahl G (2008) Locus-speciWc microsatellite markers in common bean (*Phaseolus vulgaris* L.): isolation and characterization. **Euphytica**, doi: 10.1007/s10681-007-9577-1.
- Lashermes P, Combes MC, Prakash NS, Trouslot P, Lorieux M, Charrier A (2001) Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. **Genome**, 44:589-596.
- Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Ferreira LP, Jourdan I, Pot D, Andrade AC, Glaszmann JC, Vieira LGE, Piffanelli P (2005) Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. **Theoretical and Applied Genetic**, 111: 1032-1041.
- Li CD, Rosnagel BG, Scoles GJ (2000) The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat cultivars. **Theoretical and Applied Genetic**, 101: 1259-1268.
- Litt M, Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, 44: 397-401.
- Lowe AJ, Moule C, Trick M, Edwards KJ (2004) Efficient large-scale development of microsatellites for marker and mapping applications in *Brassica* crop species. **Theoretical and Applied Genetic**, 108: 1103-1112.
- Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. **Genome Research**, 10: 72-80.
- Mir RR, Rustgi S, Sharma S, Singh R, Goyal A, Kumar J, Gaur A, Tyagi AK, Khan H, Sinhá MK, Balyan HS, Gupta PK (2008) A preliminary genetic analysis of fibre traits and the use of new genomic SSRs for genetic diversity in jute. **Euphytica**, doi: 10.1007/s10681-007-9597-x.
- Moncada P, McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome**, 47, 501-509.
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics**, 30: 194-200.

- Novelli VM, Cristofani M, Souza AA, Machado MA (2006) Development and characterization of polymorphic microsatellite markers for the sweet orange (*Citrus sinensis* L. Osbeck). **Genetics and Molecular Biology**, 29: 90-96.
- Oliveira CB, Sakiyama NS, Caixeta ET, Zambolim EM, Rufino RN, Zambolim L (2007) Mapa parcial de ligação gênica de *Coffea arabica* L. e recuperação do genitor recorrente em progênies de retrocruzamentos. **Crop Breeding and Applied Biotechnology**, 7: 196-203.
- Paillard M, Lashermes P, Petiard V (1996) Construction of a molecular linkage map in coffee. **Theoretical and Applied Genetic**, 93: 41-47.
- Pearl HM, Nagai C, Moore PH, Steiger DL, Osgood RV, Ming R (2004) Construction of a genetic map for arabica coffee. **Theoretical and Applied Genetic**, 108: 829-835.
- Pereira MR (2005) **Desenvolvimento de marcadores microssatélites para o feijoeiro comum (*Phaseolus vulgaris* L.)**. Dissertação (Mestrado - Universidade Federal de Viçosa), Viçosa, 37p.
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, De Kochko A, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. **Molecular and Genetics Genomics**, 276: 436-449.
- Poncet V, Dufour M, Hamon P, Hamon S, Kochko A, Leroy T (2007) Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. **Genome**, 50: 1156-1161.
- Queiroz VT (2004) **Obtenção de primers microssatélites e desenvolvimento, validação e mapeamento de marcadores SCAR em feijoeiro-comum**. Tese (Doutorado – Universidade Federal de Viçosa), Viçosa, 155p.
- Ritschel PS, Lins TCL, Tristan RL, Buso GSC, Buso JA, Ferreira ME (2004) Development of microsatellite markers from an enriched genomic library for genetic analysis of melon (*Cucumis melo* L.). **BMC Plant Biology**, 4: 1-14.
- Rovelli P, Mettullo R, Anthony F (2000) **Microsatellites in *Coffea arabica* L.** In: *Coffee Biotechnology and Quality* (eds Sera T, Soccol CR, Pandey A, Roussos S), pp. 123-133. Kluwer Academic Publishers.
- Rozen S, Skaletsky HJ (2000) **Primer 3 on the WWW for general users and for biologist programmers**. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386.
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. **Theoretical and Applied Genetic**, 109: 122-128.

- Song QJ, Shi JR, Singh S, Fickus EW, Costa JM, Lewis J, Gill BS, Ward R, Cregan PB (2005) Development and mapping of microsatellite (SSR) markers in wheat. **Theoretical and Applied Genetic**, 110:550-560.
- Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? **Molecular Ecology**, 12: 1339-1348.
- Stajner N, Jakse J, Kozjak P, Javornik B (2005) The isolation and characterisation of microsatellites in hop (*Humulus lupulus* L.). **Plant Science**, 168: 213-221.
- Teixeira-Cabral TA, Sakiyama NS, Zambolim L, Pereira AA, Schuster I (2004) Single-locus inheritance and partial linkage map of *Coffea arabica* L. **Crop Breeding and Applied Biotechnology**, 4: 416-421.
- Tesfaye K, Borsch T, Govers K, Bekele E (2007) Characterization of *Coffea* chloroplast microsatellites and evidence for the recent divergence of *C. arabica* and *C. eugenioides* chloroplast genomes. **Genome**, 50: 1112-1129.
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. **TRENDS in Biotechnology**, 23: 48-55.
- Viruel MA, Hormaza JI (2004) Development, characterization and variability analysis of microsatellites in lychee (*Litchi chinensis* Sonn., Sapindaceae). **Theoretical and Applied Genetic**, 108: 896-902.
- Watcharawongpaiboon N, Chunwongse J (2008) Development and characterization of microsatellite markers from an enriched genomic library of cucumber (*Cucumis sativus*). **Plant Breeding**, 127: 74-81.
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. **Molecular Ecology**, 11: 1-16.
- Zhao W, Miao X, Jia S, Pany, Huang Y (2005) Isolation and characterization of microsatellite loci from the mulberry, *Morus* L. **Plant Science**, 168: 519-525.

Tabela 1. Marcadores SSR existentes para o gênero *Coffea*.

Referência	Espécie	Nº primers	Origem	Método	Polimorfismo (%)	Ho	Nº SSR mapeados*
Combes et al. (2000)	Ca	11	SSRg	BGE	45,4	0,47	18
Rovelli et al. (2000)	Ca	13	SSRg	BGE	60,0	-	-
Baruah et al. (2003)	Ca	9	SSRg	BG	-	0,37	-
Moncada e McCouch (2004)	Ca	34	SSRg	BG	19,0-70,0	0,31	-
Bhat et al. (2005)	Cc	9	EST-SSR	EST	-	0,38	-
Leroy et al. (2005)	Cc	35	SSRg	BAC	71,0	-	10
Poncet et al. (2006)	Cc	25	EST-SSR	EST	75,0-86,0	0,59	-
Poncet et al. (2007)	Cc	40	SSRg	BGE	100,0	0,51	-
Aggarwal et al. (2007)	Cc, Ca	24	EST-SSR	EST	75,0	0,51	-
Tesfaye et al. (2007)	Ca	7	SSRcp	-	-	-	-
Hendre et al. (2008)	Cc	44	SSRg	BG	90,2	0,52	9
Missio (2008)**	Ca	90	SSRg	BGE	14,7-89,2	0,43	-
Total		341					37

Ca: *Coffea arabica*; Cc: *Coffea canephora*; SSRg: microssatélite genômico; EST-SSR: microssatélite de EST (*Expressed Sequence Tags*); SSRcp: microssatélites de cloroplasto; BGE: Biblioteca genômica enriquecida; BG: Biblioteca genômica; BAC: Cromossomo artificial de levedura; Ho: Heterozigose observada; (-) não avaliado; *mapeados em *C. canephora*; **dados obtidos do presente trabalho.

Tabela 2. Análise das seqüências.

Resultados dos clones seqüenciados	Número	%
Total de clones seqüenciados	835	100
Problemas no seqüenciamento	64	7,6
Clones sem inserto	0	-
Contaminação com DNA de <i>Escherichia Coli</i>	0	-
Seqüências com tamanho insuficiente para desenho de <i>primers</i>	15	1,8
Total de seqüências disponíveis para desenho de <i>primers</i> SSR	756	90,5
Seqüências da biblioteca (GT) ₁₅	530	70,1
Seqüências da biblioteca (AGG) ₁₀	226	29,9
Seqüências redundantes	113	14,9
Seqüências com SSR	218	28,8
Seqüências com mais de um SSR	69	9,1
Total de SSR encontrados	287	38,0
Seqüências enriquecidas com (GT) ₁₅	37	7,0
Seqüências enriquecidas com (AGG) ₁₀	16	7,1
<i>Primers</i> desenhados	96	12,7

Tabela 3. Frequência e distribuição das classes de SSR nas seqüências utilizadas para o desenvolvimento de *primers* SSR em *Coffea arabica*.

Classes	Número de repetições									Total	% ¹	% ²
	3	4	5	6	7	8	9	10	>10			
Mononucleotídeo										3	1,1	
(A/T)	-	-	-	-	-	-	-	1	2	3	100	1,0
Dinucleotídeo										146	50,9	
(AC/GT)	3	9	2	3	-	3	4	-	13	37	25,3	12,9
(AG/CT)	1	25	13	21	5	2	-	2	25	94	64,4	32,8
(AT/AT)	-	5	8	1	-	-	-	-	-	14	9,6	4,9
(CG/CG)	-	1	-	-	-	-	-	-	-	1	0,7	0,4
Trinucleotídeo										95	33,1	
(AAC/GTT)	11	2	-	-	2	-	-	-	-	15	15,8	5,2
(AAG/CTT)	20	7	1	-	-	-	-	-	-	28	29,5	9,8
(AAT/ATT)	6	3	-	1	-	-	-	-	-	10	10,5	3,5
(ACC/GGT)	7	5	-	-	1	-	1	-	-	14	14,7	4,9
(ACG/CGT)	1	-	-	-	1	-	-	-	-	2	2,1	0,7
(ACT/AGT)	3	-	-	-	-	-	-	-	-	3	3,2	1,0
(AGC/GCT)	1	-	-	-	-	-	-	-	-	1	1,1	0,4
(AGG/CCT)	11	2	1	1	1	-	-	-	-	16	16,8	5,6
(AGT/ACT)	4	-	-	-	-	-	-	-	-	4	4,2	1,4
(CCG/CGG)	1	-	-	1	-	-	-	-	-	2	2,1	0,7
Tetranucleotídeo										23	8,0	
(AAAC/GTTT)	2	-	-	-	-	-	-	-	-	2	8,7	0,7
(AAAG/CTTT)	9	-	1	-	-	-	-	-	-	10	43,5	3,5
(AAAT/ATTT)	4	-	-	-	-	-	-	-	-	4	17,4	1,4
(AACC/GGTT)	1	-	-	-	-	-	-	-	-	1	4,4	0,4
(AACT/AGTT)	1	-	-	-	-	-	-	-	-	1	4,4	0,4
(AAGG/CCTT)	2	-	-	-	-	-	-	-	-	2	8,7	0,7
(AAGT/ACTT)	1	-	-	-	-	-	-	-	-	1	4,4	0,4
(AGGG/CCCT)	-	1	-	-	-	-	-	-	-	1	4,4	0,4
(AGGT/ACCT)	1	-	-	-	-	-	-	-	-	1	4,4	0,4
Pentanucleotídeo	9	1	-	-	-	-	-	-	-	10	3,5	3,5
Hexanucleotídeo	10	-	-	-	-	-	-	-	-	10	3,5	3,5
Total	109	61	26	28	10	5	5	3	40	287	100	

¹Porcentagem de cada repetição dentro da respectiva classe em relação ao total de SSR; ²Porcentagem de cada repetição em relação ao total de SSR.

Tabela 4. Descrição dos 96 pares de *primers* SSR para *Coffea arabica*.

<i>Primer</i>	Repetição	<i>Forward primer (5'→3')</i>	<i>Reverse primer (5'→3')</i>	Tm °C	Tam. frag. esperado	Nº de alelos
SSRCa 001	(CCCTTT) ₃(TC) ₃ C(CTT) ₃	CCCACTACTCCATTCCATTC	AGCAGATTCCATCCTTATCCT	57	173	2
SSRCa 002*	(TTCC) ₃(GT) ₁₇	CTGTCCCACCAACCAAAA	CTTCAACCCCCAACACAC	57	258	2
SSRCa 003	(GT) ₁₂	ATGATTTCGTAGGTGGAGTGG	CTAAGCCGCAAATGACAGA	57	196	1
SSRCa 004	(CT) ₈ CG(CT) ₄	CCATGAGCACTTGTCCATAAA	ATCAAAGAACAAACCCGACA	58	287	1
SSRCa 005	(CT) ₅	TGTCACCTCCTTGTGGATT	GCTTGATTGAGATGATTTGC	55	201	2
SSRCa 006	(CT) ₆	CTTGCTCAGTGAACCATCC	TGCCTCTTATGCCACTACTAAA	56	209	4
SSRCa 007	(GGA) ₃ N ₃ (AT) ₂	GTTCTTTCATTCCAGGTAAAGC	TAGAAGGAATCGGTGGAGAA	57	178	2
SSRCa 008	(AG) ₆	TTACCCACTTTTTCCACCTC	TTTGGCTTCAATCTTGCTC	56	373	2
SSRCa 009	(TTTA) ₃	CAGTTTGGAAATGCTTGAGTG	CCGGAACTTAACCTTATTGG	56	352	2
SSRCa 010	(CT) ₆	GTTGATTGGTGGAGTGATTG	AAGCATCAAGTAAGGGAGGA	56	105	2
SSRCa 011	(CT) ₆	ATCCAACCAACCATTGAAAC	CATCCACTTTTTCCACCTTC	57	347	2
SSRCa 012*	(CT) ₄ N ₆ (CT) ₄(TTTTTC) ₃(AAT) ₄	TCTCCTCTATTCGCTGTTCTC	TCTGTGCTCGTTTTTTTTCAC	56	595	1
SSRCa 013	(AG) ₆	TCAAAAACAACCACACCATC	CCATTTCACTCAATCTTCCA	56	317	2
SSRCa 014	(TA) ₅	ATTCTCTTTCTCCCACACA	AGCGGAAAACATCCAAAAC	57	214	1
SSRCa 015	(AT) ₅	TCGCAATAACCAATCACAAG	AGCTATTGACCCCACTGAAA	57	273	1
SSRCa 016	(GAA) ₃(GGAAAG) ₃	AGCAGATTCCATCCTTATCCT	CCACTAATCCATTCCATTCC	56	172	3
SSRCa 017	(ATTTT) ₃	TATGATTGGTTGCTTGGATG	ATCCTACAAGGCGGTGTG	57	205	2
SSRCa 018*	(GT) ₁₈ (GA) ₁₀	GTCTCGTTTCACGCTCTCTC	ATTTTTGGCACGTTATGTTC	57	115	3
SSRCa 019	(GA) ₁₁	GGTTAGATAGAGCAAGAATGA	CTGTGAAGGTGTGGAGTTTT	55	329	2
SSRCa 020	(AGA)G(AGA) ₃ ...(TG) ₄(ATT) ₆	GGTAGGCGAAGGACAGATAA	TGGGGCAGAGTGAAGATAAG	57	264	2
SSRCa 021*	(GGA) ₃ N ₄ (AAG) ₂	GCTGAGAGTTTTGAGGGAAA	CCGACGTAGTTGATGATTGA	57	232	4
SSRCa 022	(GA) ₅ ...(AAT) ₃	GGGAGCCATTCTGTGGA	CCCCATCTGGAAACCAA	57	445	2
SSRCa 023	(AATG) ₃	GACCCTTGCCTTTTGTG	GCCATTCATCCATTTCATTC	56	259	2
SSRCa 024	(AG) ₃ (CT) ₃	CCACTTACCGCTCTACCACT	CTTGGCTTGTCTCAGTCCTT	57	299	2
SSRCa 025	(TAA) ₂ (TCT) ₃	CTGCAACTTGTGAAATGGAC	ATACGGAGGATGAAGAAGCA	56	176	1
SSRCa 026	(T) ₁₆ N ₁₂ (TC) ₇(CAC) ₄	GAATCTGGTGGGCTTTGA	AAGGAGAGGGGAAGAAAATG	57	289	2
SSRCa 027*	(AC) ₆	TGACCTCTCTTTTCATTTGG	CATCACTGCCTTTCTTTTTG	55	221	1
SSRCa 028	(AGG) ₃ ...(CT) ₆	GCTTGGTTGAGGTTGAAAAA	GCCGAAATACGAAAATGTGT	57	328	2
SSRCa 029*	(ACA) ₃ ...(AAC) ₃ ...(AAG) ₃	AATGCACGAGAACAAGATG	TAGCACCAAAATCAATCCAC	56	344	1
SSRCa 030*	(CCAT) ₃	GAGGAATCGAGAACCAGTGT	GTTTAGGGTTGCATTTTTCC	56	189	1
SSRCa 031	(AG) ₆	TCGGACAGATTAGGGGTTTC	TGGTGGAGTTTGTGTTGAAGAG	57	350	1
SSRCa 032	(GAA)G(GAA) ₃ (GCA) ₂	TCACACCATCCATACATTCC	ACATCCCACATTTTCAGCAC	56	328	1
SSRCa 033*	(AAT) ₃ ...(GA) ₄ N ₅ (CA) ₂ N ₃ (CG) ₃ N ₃ (GC) ₃	GTTTTTACGCGCACGATTA	TTCAAAGTCAACTCATTCTCC	57	179	2
SSRCa 034	(CT) ₃ N ₃ (CT) ₅ ...(TC) ₄ (AC) ₂	TGGACAAGAAATTGAAGTGG	GGGTTTAAATTATCGGGTGT	55	257	2
SSRCa 035	(TC) ₅ N ₃ (CT) ₃	GCTTAGTGGTTCCTTCTCCA	CAAGCCATTTCTTCTTCTC	56	192	1
SSRCa 036	(CA) ₈	ATGTTTCGTGAAACACACGTC	GGTTTGCCTTCATCTTTGTT	56	128	1
SSRCa 037	(CT) ₆	TTTTGGCTTCAATCTTGCTC	TTACCCACTTTTTCCACCTC	57	374	1
SSRCa 038*	(AAGA) ₃(A) ₁₈	CGCAGGAATCATCAAGAA	ATAAGGAAGCAGGCTAATGG	56	312	1
SSRCa 039	(AG) ₆	GAGTCAAAGCCCCTTATTACC	AGTTTGGTGGAGTTTGTGTTG	56	263	1

Continuação...

<i>Primer</i>	<i>Repetição</i>	<i>Forward primer (5'→3')</i>	<i>Reverse primer (5'→3')</i>	<i>Tm °C</i>	<i>Tam. frag. esperado</i>	<i>Nº de alelos</i>
SSRCa 040*	(GAG) ₃ A(AG) ₃	AGGGATGTAGAACCAGCAAA	CCAATAGCTCACAACAAAGG	56	283	2
SSRCa 041	(AC) ₄ N ₉ (TC) ₃	TCCCATGATTTCTCCACTTT	TTGAGCACTGGTATGGTTTG	57	195	1
SSRCa 042	(AG) ₆ ...(AGG) ₃ N ₄ (TTC) ₂	TTGTTACCTTTCCCACCT	AATCAGCAAAACCAACCATC	57	352	1
SSRCa 043*	(CAA) ₂ (TCT) ₂ ...(GAA) ₄	GCCAAAATCCTTGTCTTCAC	GTCTTCCTGTTTGTGGTTC	57	270	1
SSRCa 044	(CT) ₇	CCCAATCTCACAACTAACCA	CTTCATCACCTCAACCACAA	57	248	1
SSRCa 045*	(TTTAC) ₃ ...(AC) ₃ T(CA) ₃	GACTTGTTCATTCCCCTA	GCGCATGTGAAGAGAAAGT	56	303	2
SSRCa 046	(AAAT) ₃	ATGAAGAGGGGTTCCATCA	CATAGACTTTTCTTGCCTCCT	57	254	1
SSRCa 047	(AT) ₅ ...(AAAGA) ₃	TAGAGGGTCTTTTCGAGTTT	AAAACCTTTCCGTCCACTT	56	466	1
SSRCa 048	(AAAAT) ₃	TAGTCTACAAGGCGGTGTG	TATGATTGGTTGCTTGGATG	57	207	1
SSRCa 049	(CTT) ₄ ...(GT) ₆	TTGCATTCTACCCAACAAAG	CCCATCCACTTCAAATAACA	56	216	2
SSRCa 050	(CA) ₄ (GA) ₂ (CA) ₂ (GAGG) ₄ ...(GA) ₆	AGCAATACATGCAGAGACCA	AATGTCGTTCCAACCAGAAG	57	133	2
SSRCa 051	(ATC) ₃ (ACA) ₃ ...(CAC) ₃ (TGC) ₂	GAACAAGAACAGCAGACACAA	GAAAAGTTGGTGGGAAGAGA	56	383	2
SSRCa 052*	(TTG) ₇	GATGGAAACCCAGAAAGTTG	TAGAAGGGCTTTGACTGGAC	57	129	3
SSRCa 053	(ATA) ₂ (TCT) ₂ (CT) ₂ ...(AAGA) ₃ N ₃ (AAAG) ₂ (GGT) ₄	ACCACTTGACCACATTTTT	TTTTCTCCTTGATGCTCTC	56	259	1
SSRCa 054*	(AAAG) ₃	CCGAACCCAACTAACATCTC	GCAGGTCTTCCATTGTCTGT	57	354	2
SSRCa 055	(ATC) ₃ ...(AAGG) ₂ N ₉ (CT) ₄ N ₆ (CT) ₃	AAGGAAAACAACCCCAAGA	CGAGACAAGAGAGGGGAAA	57	294	4
SSRCa 056	(GGT) ₃ (TTGG) ₂ ...(GTTT) ₂ (GTT) ₂ (GAT) ₂ T(GAT)(GTAAAA) ₂ ...(CGGAG) ₃	CGTATTGATGGCTGATGGT	AGGTCTGGTCCCTTTCTTCT	56	412	3
SSRCa 057	(TTTT) ₂ N ₃ (TG) ₃ ...(TTG) ₃ N ₇ (TTG) ₂	GCGGGCTAGATGAAAACCTC	ATCTCACGCGACAGCAAC	57	169	2
SSRCa 058*	(CATC) ₂ (AT) ₃ ...(CA) ₅	ATCATTACCTTGCCCAAATC	ACCCTTGACTGCCATAAATC	56	364	1
SSRCa 059	(GA) ₃ (AAG) ₂ ...(TCT) ₃	AGTCTCATGCACGGTTTTG	ACGTTTCATGCTTGTGTTGAG	56	249	1
SSRCa 060	(CT) ₆	AGTTTGGTGGAGTTTGTGTTG	GAGTCAAAGCCCCTTATTACC	56	263	2
SSRCa 061	(CCAA) ₂ (CT) ₅	GCAGGTGCAAGTGATAAAAG	CGTCTTGTGATGTGTTAGGG	56	242	4
SSRCa 062	(CAA) ₂ G(AGAA) ₂ (AG) ₄ N ₈ (GA) ₄	AAGTTATTAGGGCAAGAGTGA	AAGCTCCAAGACCAAAGATG	57	275	2
SSRCa 063	(TG) ₃ A(GT) ₃ N(TG) ₄	CTCCGCTGATTTTGTCTTTT	ACCACTTTTCTCCTCTC	57	222	-
SSRCa 064	(TTCT) ₃	TGCAGTAAGTGAGACCAACC	TGGACTATCCCATACATAACCA	56	242	2
SSRCa 065	(AG) ₂ (AAG) ₃	ATCTAACAAAATCCCCGTCA	ATCGGTCGCCCTTCTAAT	57	142	4
SSRCa 066*	(GAA) ₃ (AG) ₃	GTGTGTCTTGAGGGCAGTTT	TCTTGATAGGTCTCAGCATC	57	205	1
SSRCa 067	(GAA) ₄	TCTCCTCCATGACCTAAAA	CGAACAAAGCTGAAGTGAAA	56	107	2
SSRCa 068	(AGG) ₇ ...(GAA) ₄	ATGTTGTTGGAGGCATTTTC	AGGAGCAGTTGTTGTTTTCC	57	236	1
SSRCa 069*	(AG) ₄ (AT) ₂ (GA) ₂	GATTGGGCATAAGTTTTCCA	TGAATCCTCCAAGAATAGCC	57	146	1
SSRCa 070	(AG) ₆	AAGCATCAAGTAAGGGAGGA	GATTGGTGGAGTGATTGGA	56	102	2
SSRCa 071	(AC) ₅	TTCCTCCTTCCTTTCTTCTTC	GGGAGTGTGTTGGTTCATTT	56	114	2
SSRCa 072	(AC) ₆ N ₂ (TC) ₅ N ₃ (TC) ₂ (CT) ₃	GCCACATTTGTGCGATTTT	GCACAACAACCATCCATCTT	58	221	-
SSRCa 073	(TC) ₂ (TTC) ₃ (CT) ₃	GCTGTGTGAGAAGCAAAGAA	CCAACAACCCTAAAGAAGC	56	298	1
SSRCa 074	(AAG) ₃ G(GA) ₃ ...(GGAAAG) ₃	CCACTACTCCATTCCATTCC	AGCAGATTCCATCCTTATCCT	56	172	2

Continuação...

<i>Primer</i>	Repetição	<i>Forward primer (5'→3')</i>	<i>Reverse primer (5'→3')</i>	Tm °C	Tam. frag. esperado	Nº de alelos
SSRCa 075	(AC) ₁₅	TTCCCATGTCAAGCAAATC	CATCGCTAGTGCAGTCAAAG	57	106	-
SSRCa 076	(TA) ₅	GTGTGTGCAAATGAATGAAG	AGGGAAATGAGCGAGTGT	55	297	2
SSRCa 077	(TCA) ₃ ...(GCA) ₃	TGTTCTTGGCATACTTCATC	GTTTCATGTGGGTATCTTTCCT	56	297	1
SSRCa 078	(TCC) ₅	AGCCTCCCTTAGTTTGTCTC	GGAAAGTCGTCAGATTGGTT	56	210	1
SSRCa 079	(CCCT) ₂ N ₅ (GAAAA) ₃	AAGTGGAGGAGTTTTGTGGA	CCAAGTGGATAGGTGTGAGAG	57	287	2
SSRCa 080*	(CA) ₉ N ₈ (CT) ₃₀	GTTCTTTCCGCCGTCAAT	GAGAAGAGAGAGGAAGGGAAA	57	250	3
SSRCa 081	(CT) ₃₈	ACCGTTGTTGGATATCTTTG	GGTTGAACCTAGACCTTATTT	53	229	1
SSRCa 082	(CT) ₁₇ CG(CT) ₆	GCTTGTTCATCGCTAAA	TTACACGTCAACCCACAAAC	56	178	2
SSRCa 083	(TC) ₃₂	TCCAACAACATTAAGCGTATTC	GACAAACCTGAGGGAAAAGA	56	223	1
SSRCa 084	(CCA) ₄ ...(CAC) ₇	ATCGGAAAGATGTCAACCAT	CAAATTGAAGCCAGTGGTG	57	157	3
SSRCa 085	(TC) ₂₄	ATGTGAAAATGGGAAGGATG	CACAGGAAAGTGACACGAAG	57	105	4
SSRCa 086	(AC) ₁₁	AGAGAGAAGCCATGATTTGA	TCAGTCCCAGAGAATAAGGA	54	105	-
SSRCa 087	(TC) ₂₂	TCACTCTCGCAGACACACTAC	GCAGAGATGATCACAAGTCC	56	143	4
SSRCa 088	(TTTTCT) ₃	TACCTCTCCTCCTCCTCCT	ATTTCTATGGACCGGCAAC	57	180	3
SSRCa 089	(TC) ₁₉	GAAATGGTGAACCTCTCTTGG	ATTTGCATGGCTTTGGTG	58	185	1
SSRCa 090	(GA) ₂₁	TGACTCGATTACATCCCTAATG	GTATTTTGGTTCCCATGTT	56	120	-
SSRCa 091*	(GT) ₈ (GA) ₁₀	CGTCTCGTATCACGCTCTC	TGTTCCCTCGTTCCTCTCTCT	56	110	4
SSRCa 092	(CCA) ₇ CT(TCCACC) ₅	ATAGCCTGAGCCGTAACCA	GGGTAATTATGACGAGGGACA	58	142	4
SSRCa 093	(CT) ₃₇	TTGCCTACAATACCTGTCTCC	CCCAATTCCTCTCCATTCT	56	196	-
SSRCa 094	(TC) ₄ (TTCT) ₃ ...(TTTCCT) ₃ (TTTC) ₅	GTGTCTAGGGAAGGGTAAG	GAGTGCTAGGAGAGGGAGAG	55	195	1
SSRCa 095*	(TG) ₁₁	GAGAGAGCCGAGTGAAGAGA	GAGAGAGAAGCCATGATTTGA	57	185	1
SSRCa 096*	(CT) ₁₈	GAAATGGTGAACCTCTCTTGG	ATTTGCATGGCTTTGGTG	57	183	1

(-) Não apresentou produto de amplificação nas condições estabelecidas; **Primers* polimórficos entre Híbrido de Timor (UFV 445-46) e Catuaí (UFV 2145-235).

Tabela 5. Marcadores SSR desenvolvidos para diversas espécies de plantas.

Espécie	Nº clones sequenciados	Sondas	Nº de SSR	Eficiência (%)*	Nº primers desenhados	Repetição mais frequente	% polimorfismo	% não amplificação	Primer/marcador (%)**	Referência
<i>Chenopodium quinoa</i>	1172	GA,CAA,AAT	436	37,2	402	GA (49%) e CAA (35,6%)	54	41,0	59,0	Jarvis et al. (2008)
<i>Coffea canephora</i>	446	CA,GA,GC,CAA,CAT,ACT,GATA,AGA,CATA,CTG,GAC,AGG,GGT,GCC,	92	20,6	58	AG (27%), AC (13%), AGC (8%), CCG (1%)	75,9	24,0	76,0	Hendre et al. (2008)
<i>Arachis hypogaea</i>	720	GT,GA	490	68,0	170	GT (37,6%), GA (25,9%)	44,2	38,8	61,2	Cuc et al. (2008)
<i>Arachis hypogaea</i>	120	GT,CT	66	55,0	26	-	19	-	-	Hopkins et al. (1999)
<i>Cucumis melo</i>	254	CT	170	66,9	86	GA (93%)	52,3	22,1	77,9	Watcharawongpaiboon e Chunwongse (2008)
<i>Cucumis melo</i>	700	AG	450	64,3	144	AG (41%)	97	-	-	Ritschel et al. (2004)
<i>Cucurbitáceas</i>	2400	GA,CA,CAT,CCA,CAA,GAA	105	44,1	532	GA,GAA	81	6,0	94,0	Gong et al. (2008)
<i>Citrus</i>	1115	AG,GT,TCA,AAC	279	25,0	171	AG (69%), GT (16%), TCA (8%), AAC (6%)	29	34,0	66,0	Novelli et al. (2006)
<i>Phaseolus vulgaris</i>	25	GA	25	100,0	18	-	-	11,0	89,0	L'taief et al. (2008)
<i>Phaseolus vulgaris</i>	1209	CT,GT	714	59,0	123	GA,GT	73,7	2,4	97,6	Benchimol et al. (2007)
<i>C. capsularis</i>	194	-	67	34,5	45	-	91	9,0	91,0	Mir et al. (2008)
<i>Lolium multiflorum</i>	1544	CA	1044	67,6	395	CA (91%)	72,8	9,6	90,4	Hirata et al. (2006)
<i>Caryocar brasiliensis</i>	195	AG	28	14,4	28	-	35,7	46,4	53,6	Collevatti et al. (1999)
<i>Litchi chinensis</i>	42	CT	30	71,4	26	CT	46,1	30,8	69,2	Viruel e Hormaza (2004)
<i>Brassica spp.</i>	1041	CA,CT,CATA,CAA,GATA	-	-	398	GA (44%), GT (23,7%), GGC (7,9%)	62,3	11,7	88,3	Lowé et al. (2004)
<i>Brassica rapa spp. chinensis</i>	1024	-	190	18,5	143	AG (54,7%), AC (31,6%)	8	-	-	Cui et al. (2008)
<i>Humulus lupulus</i>	157	GA,GT,ACA,AGA,CAG,ACTC	151	96,2	60	GA, GT	-	7,8	92,2	Stajner et al. (2005)
<i>Morus L.</i>	96	CA	25	26,0	20	CA	50	40,0	60,0	Zhao et al. (2005)
<i>Persea americana</i>	726	GA,ATG	353	48,6	208	AG (81%), AC (19%)	51	49,0	51,0	Ashworth et al. (2004)
<i>Saccharum officinarum</i>	798	CT,CA,ACT,AGA,CAA,CTA,CTT,CTG,GAC,AGC,CAT,ACA	457	57,3	124	TG (29,5%), GA (9,7%), TAC (5,5%)	91	19,0	81,0	Cordeiro et al. (2000)
<i>Acacia mangium</i>	223	GA,CA,CGC,AGC,CTC,CAC,AAC	166	74,4	81	AC	-	46,9	53,1	Butcher et al (2000)
<i>Avena spp</i>	100	AC,AG,AAG	78	78	61	-	62	4,5	95,5	Li et al. (2000)
<i>Oryza sativa</i>	139	GA	122	87,8	89	GA (88%)	94	27,0	73,0	Chen et al. (1997)
<i>Triticum aestivum</i>	277	CA,GA,CAA,GAA,ACG,CAG	222	80,1	153	CA, GA	32	-	-	Bryan et al. (1997)
Média				56,3				24,0	76,0	

*Eficiência = Porcentagem de SSR em relação ao número total de clones sequenciados; **Porcentagem de primers SSR que produzem produto de amplificação.

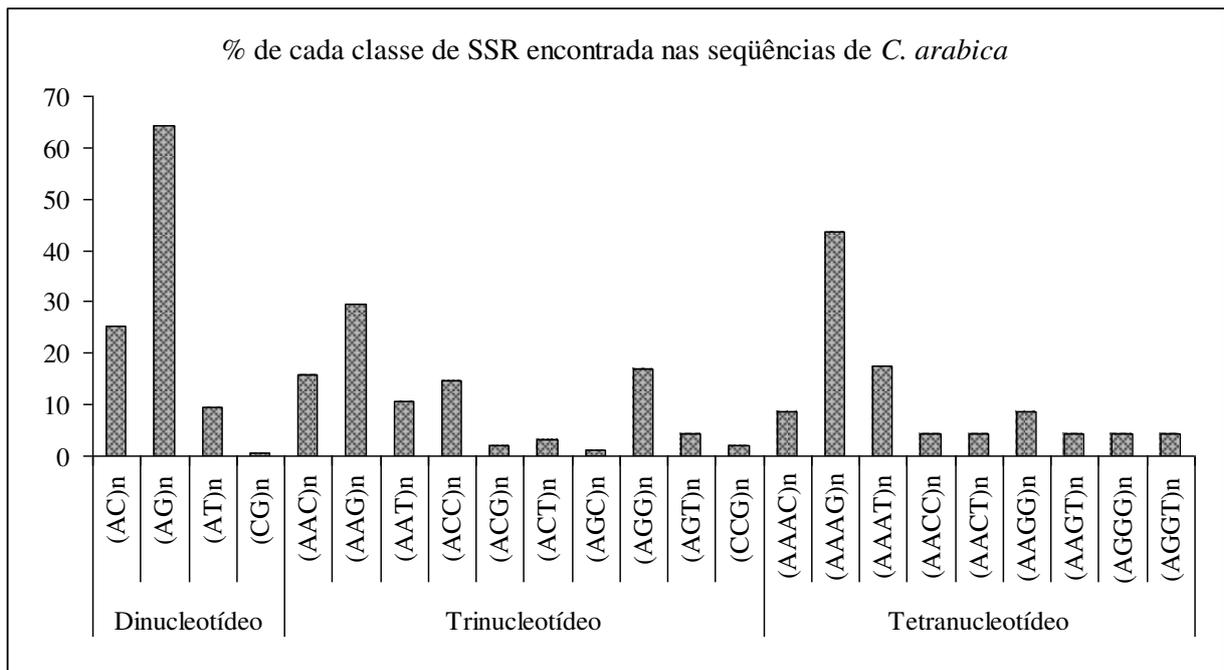


Figura 1. Frequência de cada repetição dentro da respectiva classe de SSR encontradas para *Coffea arabica*.

CAPÍTULO II

Novos marcadores microssatélites para *Coffea arabica*

Resumo

Trinta e três novos *primers* SSR para o gênero *Coffea* obtidos de biblioteca genômica enriquecida com repetições (GT)₁₅ e (AGG)₁₀ foram analisados em 24 acessos de cafeeiros. Vinte e dois *primers* apresentaram polimorfismo entre os acessos, com um número médio de alelos por *primer* de 5,1. Os valores de PIC variaram de 0,08 a 0,79. Os maiores valores médios de PIC foram encontrados para *C. canephora* (0,46) e os menores para *C. arabica* (0,22) e Triploides (0,22). Os marcadores SSRs polimórficos desenvolvidos neste trabalho possibilitarão avanços moleculares, especialmente nos estudos genéticos da espécie *C. arabica*.

Palavras-chave: SSR, café arábica, marcadores moleculares, PIC.

Marcadores SSR para *Coffea arabica*

O gênero *Coffea* é constituído de 103 espécies conhecidas (Davis *et al.*, 2006), sendo *Coffea arabica* L. a de maior importância comercial e a mais cultivada no mundo. Estudos anteriores, utilizando diferentes marcadores moleculares, mostraram baixos níveis de polimorfismo entre acessos de *C. arabica* (Baruah *et al.*, 2003; Poncet *et al.*, 2006; Aggarwal *et al.*, 2007; Hendre *et al.*, 2008), dificultando o avanço da pesquisa genética nesta espécie. Nestas situações, os marcadores microssatélites (SSR) são potencialmente úteis, especialmente por explorar regiões do genoma geralmente muito variáveis entre indivíduos ou populações da mesma espécie. Além disso, os SSRs possuem outras características importantes, como a alta reprodutibilidade, herança co-dominante e possibilidade de automação.

Entretanto, com todas as vantagens dos marcadores SSR para estudos genéticos, poucos *primers* SSR foram desenvolvidos para o café (Combes *et al.*, 2000; Rovelli *et al.*, 2000; Baruah *et al.*, 2003; Moncada & McCouch 2004; Bhat *et al.*, 2005; Poncet *et al.*, 2006; Aggarwal *et al.*, 2007), quando comparado com culturas como milho, arroz, trigo e cevada (<http://www.gramene.org/>). Neste trabalho foram desenvolvidos 33 novos marcadores microssatélites para *C. arabica*, a partir de bibliotecas genômicas enriquecidas, os quais podem ser utilizados para diferentes estudos genéticos do cafeeiro.

O DNA genômico do genótipo de *C. arabica* Bourbon Amarelo (BA), acesso UFV 570, foi extraído de folhas jovens e completamente desenvolvidas, de acordo com o procedimento descrito por Diniz *et al.* (2005). O DNA genômico (50µg) foi digerido em fragmentos de aproximadamente 200-1000bp, utilizando as enzimas de restrição *EcoRI*, *NheI*, *HaeIII* e *RsaI* (New England BioLabs). Os fragmentos foram convertidos em extremidades abruptas pelo tratamento com a enzima *Mung Bean Nuclease* (MBN), desfosforilados com *Calf intestinal phosphatase* (CIP) e ligados aos adaptadores SNX dupla fita (Hamilton *et al.*, 1999). O enriquecimento foi realizado pela hibridização do DNA com duas sondas SSR biotinizadas, (GT)₁₅ e (AGG)₁₀. Após as lavagens, os fragmentos foram amplificados por PCR utilizando o adaptador SNX_F como *primer*. Os fragmentos enriquecidos foram então digeridos com *NheI* e ligados ao plasmídeo *pBluescript SK+* (Stratagene), previamente digerido com *XbaI* (New England Biolabs). Células competentes de *Escherichia coli* DH5α foram transformadas com os plasmídeos recombinantes por meio de choque térmico. A seleção e o diagnóstico das colônias brancas foram realizados por PCR, utilizando os *primers* T3 e T7 (Invitrogen). Os produtos da amplificação foram separados por eletroforese em gel de agarose

1,2%. As colônias, contendo transformantes com inserções superiores a 400pb foram selecionadas e cultivadas em meio LB líquido contendo ampicilina (100µg/mL). Os clones positivos foram seqüenciados e a análise das seqüências dos fragmentos de DNA foi realizada com auxílio dos programas CodonCode Aligner 1.6.3 (CodonCode Corporation) e SSRIT (<http://www.gramene.org/db/searches/ssrtool>). Pares de *primers* flanqueando os 33 locos SSR identificados foram desenhados utilizando o programa Primer3 (Rozen and Skaletsky 2000) e sintetizados.

Os *primers* SSRs foram testados em reação de PCR em um volume de 20µL contendo 50ng de DNA, 0,6 U de *Taq* DNA polimerase e tampão 1 X (Promega, Madison, WI, EUA), 1mM de MgCl₂, 150µM de cada dNTP e 0,1µM de cada *primer*. A amplificação foi efetuada em termociclador PTC 200 (MJ Research) utilizando-se o procedimento *touchdown* PCR, que consistiu de desnaturação inicial a 94°C/2min, seguido de 13 ciclos de desnaturação a 94°C/30s, anelamento de 67°C a 55°C/30s, reduzindo 1°C a cada ciclo, e extensão a 72°C/30s. Os 13 ciclos foram seguidos de mais 30 ciclos a 94°C/30s, 55°C/30s e 72°C/30s e extensão final a 72°C/8min. O polimorfismo dos marcadores SSR foi verificado em gel de poliacrilamida desnaturante a 6% e corado com prata (Creste et al., 2001).

O potencial dos 33 pares de *primers* desenvolvidos, como marcadores moleculares, foi avaliado em 24 acessos do gênero *Coffea*, importantes para o programa de melhoramento genético brasileiro. Os acessos incluíram seis genótipos de *C. arabica*, cinco de *C. canephora*, três Híbridos de Timor (*C. arabica* x *C. canephora*), três Triplóides (*C. arabica* x *C. racemosa*), seis variedades comerciais de *C. arabica* resistentes à ferrugem e um acesso de *Coffea racemosa*. Foram analisados a amplificação e o PIC (*Polymorphism Information Content*) para cada marcador SSR. Para o cálculo do PIC foi utilizado o software CENES (Cruz, 2007) que possibilita considerar até quatro alelos para o mesmo indivíduo, como é o caso de alguns acessos alotetraplóides de *Coffea*.

Os 33 pares de *primers* desenvolvidos nesse trabalho foram denominados SSRCa e as seqüências correspondentes encontram-se na Tabela 1. Desses *primers*, 26 (79%) amplificaram bandas nítidas e bem definidas, sendo que 22 (67%) apresentaram polimorfismo entre os 24 acessos analisados e quatro foram monomórficos. Cinco *primers* não apresentaram nenhum produto de amplificação e dois apresentaram um padrão multi-bandas (Tabela 1). Os 22 *loci* polimórficos amplificaram um total de 112 alelos, com uma média de 5,1 alelos por *primer*.

Considerando os 24 acessos de *Coffea* e os *loci* polimórficos, os valores de PIC variaram de 0,08 a 0,79. Os maiores valores médios de PIC foram encontrados para *C. canephora* (0,46) e os menores para os acessos de *C. arabica* (0,22) e Triplóides (0,22). O baixo polimorfismo encontrado para *C. arabica* está em concordância com o trabalho de Baruah et al. (2003) e pode ser explicado pela natureza autógama e a base genética estreita desta espécie. Todos os *primers* SSR foram capazes de detectar diferenças genéticas entre as populações de *C. arabica* (6 acessos) e de *C. canephora* (5 acessos), assim como entre *C. canephora* e as demais populações (dados não mostrados). Isto é atribuído a alogamia da espécie *C. canephora*, que resulta em elevada variabilidade genética. Os *primers* SSRCa 018 e SSRCa 091 apresentaram os maiores valores de PIC, 0,87 e 0,82, respectivamente, para os genótipos de *C. canephora* (Tabela 1). Para *C. arabica*, os *primers* que apresentaram os maiores valores de PIC foram SSRCa 094 (0,59), SSRCa 018 (0,54) e SSRCa 087 (0,54) (Tabela 1).

Os *primers* SSRCa 003, 016, 019, 020, 023, 026, 062, 078, 079, 082, 083, 091, 094 e 095 apresentaram o máximo de dois alelos/indivíduo em todos os acessos. Os *primers* SSRCa 018, 052, 068, 080, 087, 088 e 092 apresentaram mais de dois alelos/indivíduo em alguns acessos triplóides e alotetraplóides (Figura 1). Resultados semelhantes foram encontrados por Baruah et al. (2003), onde dois marcadores SSR CM11 e CM48 também apresentaram mais do que dois alelos/indivíduo em *C. arabica*. Esse resultado pode ser explicado pela ploidia dos cafeeiros analisados neste trabalho, ou seja, diplóides (*C. canephora* e *C. racemosa*), triplóides e alotetraplóides (acessos e variedade de *C. arabica* e Híbridos de Timor). Nos acessos alotetraplóides do gênero *Coffea*, os marcadores SSRs podem apresentar segregação diplóide e tetraplóide. O marcador se comporta como tetraplóide (apresenta até quatro alelos) quando amplifica regiões homólogas nos dois genomas ancestrais da espécie, e diplóide quando amplifica apenas um dos genomas. Neste trabalho, nenhum *primer* apresentou número maior de alelos/indivíduo do que o esperado para as espécies tetraplóides (até quatro alelos), triplóides (até três alelos) e diplóides (até dois alelos), o que os tornam adequados para os estudos genéticos e o melhoramento.

Os *primers* SSRCa 003, 052, 068 e 080 apresentaram alelos nulos para alguns acessos de *C. arabica* e *C. canephora*. Um total de 38 alelos exclusivos foi encontrado para os acessos de *C. canephora*, um para *C. arabica*, três para Híbridos de Timor e Triplóides e nove para *C. racemosa*, demonstrando a importância desses marcadores para distinção entre as populações. Os *primers* SSRCa 018 e SSRCa 091 permitiram distinguir todos os acessos de

C. canephora (T 3751; T 3580; Guarini UFV 514 e Apoatã IAC 2258, do grupo Robusta; e UFV 513, do grupo Conillon), e podem ser úteis para *fingerprint* genético e certificação de variedades comerciais clonais derivadas de *C. canephora*.

Todos os *loci* polimórficos obtidos neste trabalho, embora desenvolvidos de *C. arabica*, amplificaram fragmentos nos acessos de *C. canephora*, Híbridos de Timor e Triplóides. Treze (59%) desses foram validados também para *C. racemosa* (Tabela 1). Esses resultados demonstram o potencial dos marcadores SSR aqui desenvolvidos para estudos genéticos em espécies relacionadas do gênero *Coffea*. A transferência de *primers* entre espécies de *Coffea* foi anteriormente observada para marcadores EST-SSRs (Bhat et al., 2005; Poncet et al., 2006 e Aggarwal et al., 2007).

Os resultados obtidos neste trabalho demonstram que os 22 *loci* SSRs polimórficos desenvolvidos possibilitarão avanços moleculares, especialmente, nos estudos genéticos da espécie *C. arabica*. Além disso, este trabalho contribuiu para aumentar o número de marcadores SSR para estudos genéticos em *C. arabica* e espécies relacionadas.

Referências

- Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theoretical and Applied Genetic*, 114, 359-372
- Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P, Aggarwal RK (2003) Isolation and characterization of nine microsatellite markers from *Coffea arabica* (L.) showing wide crossspecies amplifications. *Molecular Ecology Notes*, 3, 647-650.
- Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK, Aggarwal RK (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.
- Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G, Lashermes P (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Molecular Ecology*, 9, 1171-1193.
- Creste S, Tulmann-Neto A, Figueira A (2001) Detection of simple sequence repeat polymorphisms in denaturing polyacrilamide sequencing gels by silver staining. *Plant Molecular Biology Report*, 4, 299-306.
- Cruz CD (2007) **Genes versão 2007.0.0**: Programa GENES versão Windows. Editora UFV, Viçosa, MG, (Programa Genes - Aplicativo computacional em genética e estatística, www.ufv.br/dbg/genes/genes.htm).
- Davis AP, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society*, 152, 465-512.
- Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET, Oliveira ACB (2005) Analysis of AFLP marker associated to the Mex-1 resistance locus in Icatu progenies. *Crop Breeding and Applied Biotechnology*, 5, 387-393.
- Hamilton MB, Pincus EL, Di Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques*, 27, 500-507.
- Hendre PS, Phanindranath R, Annapurna V, Lalremruata A, Aggarwal RK (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 Biology* 8:51 (doi: 10.1186/1471-2229-8-51).

- Moncada P, McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome*, 47, 501-509.
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, De Kochko A, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Molecular and Genetics Genomics*, 276, 436-449.
- Rovelli P, Mettullio R, Anthony F (2000) Microsatellites in *Coffea arabica* L. In: *Coffee Biotechnology and Quality* (eds Sera T, Soccol CR, Pandey A, Roussos S), pp. 123-133. Kluwer Academic Publishers.
- Rozen S, Skaletsky HJ (2000) Primer 3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (CBP&D) pelo apoio financeiro. Os autores agradecem ao Dr. Vagner Tebaldi de Queiroz e ao Professor Everaldo Gonçalves de Barros pelo auxílio no estabelecimento do protocolo.

Legenda de Figuras:

Figura 1: Variação alélica apresentada pelos *primers* SSRCa 068 (A), SSRCa 092 (B) e SSRCa 018 (C). M: Marcador de peso molecular; C.a: *Coffea arabica*; C.c: *Coffea canephora*; HT: Híbrido de Timor (*C. arabica* x *C. canephora*); T: Triplóides (*C. arabica* x *Coffea racemosa*); C.r: *Coffea racemosa*; V: Variedades comerciais resistentes à ferrugem.

Tabela 1. Caracterização de novos *loci* SSR para *Coffea arabica*.

Nº de acesso GenBank	Primer	Repetições	Sequência do primer (5' → 3')	Tam. frag. (pb)	Número de alelos/PIC						Número total de alelos/PIC (n=24)
					C.c (n=5)	C.a (n=6)	HT (n=3)	T (n=3)	C.r ¹	V (n=6)	
SSRCa 003		(GT) ₁₂	F: ATGATTCGTAGGTGGAGTGG R: CTAAGCCGCAAATGACAGA	196	2/0.31	1/0.00	1/0.00	1/0.00	2	1/0.00	4/0.21
SSRCa 010*		(CT) ₆	F: GTTGATTGGTGGAGTGATTG R: AAGCATCAAGTAAGGGAGGA	105	2/0.00	2/0.00	2/0.00	2/0.00	2	2/0.00	2/0.00
SSRCa 016		(GAA) ₃ / (GGAAAG) ₃	F: AGCAGATTCCATCCTTATCCT R: CCACTAATCCATTCCATTCC	172	1/0.00	2/0.38	2/0.38	2/0.38	2	2/0.38	3/0.54
SSRCa 017*		(ATTTT) ₃	F: TATGATTGGTTGCTTGGATG R: ATCCTACAAGGCGGTGTG	205	2/0.00	2/0.00	2/0.00	2/0.00	2	2/0.00	2/0.00
SSRCa 018		(GT) ₁₈ (GA) ₁₀	F: GTCTCGTTTCACGCTCTCTC R: ATTTTGGCACGGTATGTTT	115	9/0.87	3/0.54	2/0.38	3/0.54	2	3/0.54	13/0.74
SSRCa 019		(GA) ₁₁	F: GGGTTAGATAGAGCAAGAATGA R: CTGTGAAGGTGTGGAGTTTT	329	4/0.67	1/0.00	2/0.24	2/0.7	1	1/0.00	4/0.44
SSRCa 020		(AGA) ₃ (AGA) ₃ / (TG) ₄ (ATT) ₆	F: GGTAGGCGAAGGACAGATAA R: TGGGGCAGAGTGAAGATAAG	264	1/0.00	1/0.00	1/0.00	1/0.00	1	1/0.00	3/0.08
SSRCa 023		(AATG) ₃	F: GACCCTTGCTTTTGTG R: GCCATTCATCCATTCATTC	259	2/0.38	2/0.36	2/0.38	2/0.38	-	2/0.38	2/0.37
SSRCa 026		(T) ₁₆ N ₁₂ (TC) ₇ / (CAC) ₄	F: GAATCTGGTGGGCTTTGA R: AAGGAGAGGGGAAGAAAATG	289	4/0.60	2/0.38	2/0.38	2/0.38	-	2/0.38	5/0.60
SSRCa 036*		(CA) ₈	F: ATGTTTCGTGAAACACACGTC R: GGTTTGCCTTCATCTTTGTT	128	2/0.00	2/0.00	2/0.00	2/0.00	2	2/0.00	2/0.00
SSRCa 052		(TTG) ₇	F: GATGGAAACCCAGAAAGTTG R: TAGAAGGGCTTTGACTGGAC	129	3/0.34	2/0.38	3/0.15	2/0.38	1	3/0.36	4/0.46
SSRCa 062		(CAA) ₂ G(AGAA) ₂ (AG) ₄ N ₈ (GA) ₄	F: AAGTTATTAGGGCAAGAGTGGA R: AAGCTCCAAGACCAAAGATG	275	2/0.38	2/0.38	2/0.38	2/0.38	1	2/0.38	3/0.37
SSRCa 063		(TG) ₃ A(GT) ₃ N(TG) ₄	F: CTCCTGATTTTGTCTTTT R: ACCACTTTTTCCCTCCCTCTC	222	-	-	-	-	-	-	-
SSRCa 064*		(TTCT) ₃	F: TGCAGTAAGTGAGACCAACC R: TGGACTATCCATACATAACCA	242	2/0.00	2/0.00	2/0.00	2/0.00	2	2/0.00	2/0.00
SSRCa 068		(AGG) ₇ / (GAA) ₄	F: ATGTTGTTGGAGGCATTTTC R: AGGAGCAGTTGTTGTTTCC	236	2/0.31	1/0.00	4/0.26	1/0.00	-	2/0.14	4/0.38
SSRCa 078		(TCC) ₅	F: AGCCTCCCTTAGTTTGTCTC R: GGAAAGTCGTCAGATTGGTT	210	2/0.27	1/0.00	1/0.00	1/0.00	-	1/0.00	2/0.10
SSRCa 079		(CCCT) ₂ N ₅ (GAAAA) ₃	F: AAGTGGAGGATTTTGTGGA R: CCAAGTGGATAGGTGTGAGAG	287	2/0.37	1/0.00	1/0.00	1/0.00	2	1/0.00	3/0.19
SSRCa 080		(CA) ₉ N ₈ (CT) ₃₀	F: GTTCTTTCCGCGTCAAT R: GAGAAGAGAGAGGAAGGGAAA	250	4/0.72	1/0.00	4/0.61	2/0.44	-	4/0.71	10/0.79
SSRCa 081		(CT) ₃₈	F: ACCGTTGTTGGATATCTTTG R: GGTTGAACCTAGACCTTATTT	229	3/0.59	1/0.00	2/0.24	1/0.00	-	1/0.00	4/0.28
SSRCa 082		(CT) ₁₇ CG(CT) ₆	F: GCTTGTTCATCGCTAAA R: TTACACGTCAACCCACAAAC	178	3/0.50	2/0.24	3/0.54	2/0.35	1	3/0.36	7/0.62
SSRCa 083		(TC) ₃₂	F: TCCAACAACATTAAGCGTATTC R: GACAAACCTGAGGGAAAAGA	223	3/0.59	1/0.00	2/0.35	1/0.00	-	1/0.00	5/0.32

Continuação...

Nº de acesso GenBank	Primer	Repetição	Sequência do primer (5' → 3')	Tam. frag. (pb)	Número de alelos/PIC						Número total de alelos/PIC (n=24)
					C.c (n=5)	C.a (n=6)	HT (n=3)	T (n=3)	C.r ¹	V (n=6)	
SSRCa 084		(CCA) ₄ / (CAC) ₅	F: ATCGGAAAGATGTCAACCAT R: CAAATTGAAGCCAGTGGTG	157	+	+	+	+	+	+	+
SSRCa 085		(TC) ₂₄	F: ATGTGAAAATGGGAAGGATG R: CACAGGAAAGTGACACGAAG	105	-	-	-	-	-	-	-
SSRCa 086		(AC) ₁₁	F: AGAGAGAAGCCATGATTTGA R: TCAGTCCCAGAGAATAAGGA	105	-	-	-	-	-	-	-
SSRCa 087		(TC) ₂₂	F: TCACTCTCGCAGACACACTAC R: GCAGAGATGATCACAAGTCC	143	4/0.57	3/0.54	4/0.67	3/0.13	1	4/0.53	8/0.65
SSRCa 088		(TTTTCT) ₃	F: TACCTCTCCTCCTCTCTCT R: ATTTCTATGGACCGCAAC	180	3/0.49	3/0.25	3/0.13	2/0.38	2	2/0.38	5/0.47
SSRCa 090		(GA) ₂₁	F: TGACTCGATTACATCCCTAATG R: GTATTTTGGTTCCCCATGTT	120	-	-	-	-	-	-	-
SSRCa 091		(GT) ₈ (GA) ₁₀	F: CGTCTCGTATCACGCTCTC R: TGTTCTCGTTCCTCTCTCT	110	8/0.82	3/0.48	2/0.35	1/0.00	-	3/0.46	10/0.60
SSRCa 092		(CCA) ₇ CT(TCCACC) ₅	F: ATAGCCTGAGCCGTAACCA R: GGGTAATTATGACGAGGGACA	142	4/0.64	2/0.38	3/0.15	2/0.38	2	2/0.38	6/0.56
SSRCa 093		(CT) ₃₇	F: TTGCCTACAATACCTGTCTCC R: CCCAATTCCTCTCCATTCT	196	-	-	-	-	-	-	-
SSRCa 094		(TC) ₄ (TTCT) ₃ / (TTTCCT) ₃ (TTTC) ₅	F: GTGTCCTAGGGAAGGGTAAG R: GAGTGCTAGGAGAGGGAGAG	195	2/0.30	3/0.59	2/0.38	2/0.38	1	2/0.24	4/0.49
SSRCa 095		(TG) ₁₁	F: GAGAGAGCCGAGTGAAGAGA R: GAGAGAGAAGCCATGATTTGA	185	4/0.45	1/0.00	1/0.00	1/0.00	-	1/0.00	4/0.30
SSRCa 096		(CT) ₁₈	F: GAAATGGTGAACCTCTCTTTGG R: ATTTGCATGGCTTTGGTG	183	+	+	+	+	+	+	+
Média					3.3/0.46	1.8/0.22	2.2/0.27	1.7/0.22	1.5	2.0/0.26	5.1/0.43

C.a: *Coffea arabica*; C.c: *Coffea canephora*; HT: Híbrido de Timor; T: Triploide (*C. arabica* x *C. racemosa*); C.r: *Coffea racemosa*; V: variedades comerciais resistentes à ferrugem; ¹ somente um acesso; (-) nenhum produto de amplificação; (+) primers que amplificaram multi-bandas; *primers monomórficos.

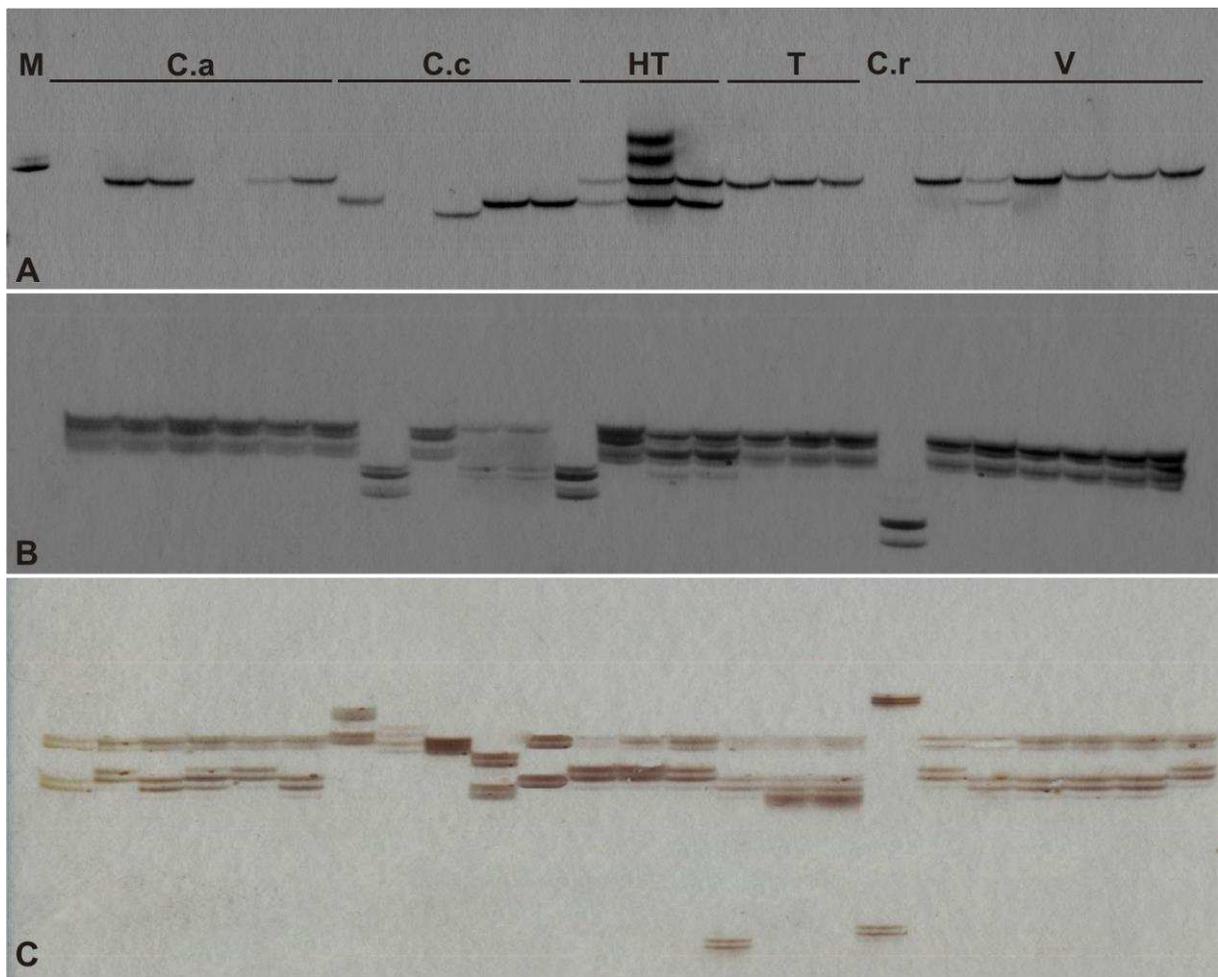


Figura 1. Variação alélica apresentada pelos *primers* SSRCa 068 (A), SSRCa 092 (B) e SSRCa 018 (C). M: Marcador de peso molecular; C.a: *Coffea arabica*; C.c: *Coffea canephora*; HT: Híbrido de Timor (*C. arabica* x *C. canephora*); T: Triplóides (*C. arabica* x *Coffea racemosa*); C.r: *Coffea racemosa*; V: Variedades comerciais resistentes à ferrugem.

CAPÍTULO III

Comparação da eficiência de marcadores EST-SSR e SSR genômico para estudo da diversidade genética em *Coffea*

Resumo

Neste trabalho foi comparada a eficiência dos marcadores EST-SSR e SSR genômico para o estudo da diversidade genética entre 24 acessos do gênero *Coffea*. Um total de 17 *primers* EST-SSR oriundos do Projeto Brasileiro do Genoma Café e 18 *primers* SSR genômicos foram avaliados. A análise genética foi realizada baseada no número médio de alelos/loco, nível de polimorfismo, número de alelos exclusivos, coeficientes de dissimilaridade genética e agrupamento genético. O número médio de alelos/loco foi de 5,1 e 5,3 para os marcadores EST-SSR e SSR genômico, respectivamente. Nenhum marcador EST-SSR apresentou alelos nulos, enquanto que nos SSR genômico eles foram 22%. Os marcadores SSR genômicos apresentaram maior número de alelos exclusivos e foram em média 43% mais polimórficos dentro e 35% entre os grupos de acessos analisados, quando comparados com os EST-SSR. Os SSR genômicos apresentaram maiores coeficientes de dissimilaridade genética e número de alelos discriminantes do que os EST-SSR, e foram eficientes em distinguir todos os acessos estudados. Além dessas análises, o número ótimo de marcas SSR, estimado neste trabalho, para produzir o mesmo agrupamento e diversidade genética foi de 120 marcas. Portanto, comparado aos marcadores EST-SSR, os SSR genômicos apresentaram maior nível de polimorfismo, maior número de alelos exclusivos, maiores coeficientes de dissimilaridade genética e maior poder de discriminação. Essas características tornam os SSR genômicos mais adequados para estudos de diversidade genética em *Coffea*, principalmente em *C. arabica* que apresenta baixa variabilidade genética.

Palavras-chave: marcadores microssatélites, diversidade genética, café, análise discriminante, número ótimo de marcadores.

Introdução

Os microssatélites (SSR) são marcadores moleculares adequados para os estudos genéticos, principalmente por apresentarem elevado polimorfismo, alta reprodutibilidade, multialelismo, abundância e distribuição por todo o genoma. Além disso, dependem de pequena quantidade de DNA para serem detectados, são baseados em PCR e apresentam a possibilidade de automação. Em razão disso, os marcadores SSR têm sido utilizados em análises genéticas com as diversas finalidades, como identificação de clones, linhagens, híbridos e cultivares, testes de paternidade, diversidade genética, fluxo gênico, taxa de cruzamento, mapeamento genético e seleção assistida (Buso et al., 2003). A maior desvantagem dos marcadores SSR é o alto custo e longo tempo necessário para o desenvolvimento dos *primers*, uma vez que os protocolos envolvem a construção de bibliotecas genômicas, hibridização, clonagem e seqüenciamento dos clones positivos (Zane et al., 2002; Squirrell et al., 2003). Alternativamente, a obtenção de marcadores SSR pode ser realizada por meio de buscas em bancos de dados, públicos ou privados, utilizando recursos computacionais apropriados. A maior limitação para o uso desta metodologia é o número limitado de seqüências de DNA disponíveis nos bancos de genes. Entretanto, com o aumento dos projetos de seqüenciamento em andamento e a preferência por seqüenciar apenas as regiões transcritas do genoma (Expressed Sequence Tag – ESTs), uma nova classe de marcadores, denominada EST-SSR, tem sido desenvolvida. Os EST-SSR são marcadores SSR obtidos a partir de ESTs geradas e disponibilizadas pelos projetos de seqüenciamento genômico. Por meio de programas computacionais, as ESTs contendo os SSRs são identificadas e a partir delas os *primers* SSR podem ser desenhados rapidamente e a um relativamente baixo custo.

Para algumas espécies de plantas, os marcadores EST-SSR apresentam um menor polimorfismo genético quando comparados a SSR derivados de biblioteca genômica, principalmente pela natureza conservada das seqüências codantes (Varshney et al., 2005). Porém, os EST-SSRs apresentam as vantagens de elevada transferência de *primers* entre espécies do mesmo gênero e a grande probabilidade de estarem associados com regiões funcionais do genoma (Poncet et al., 2006). Esta classe de marcadores tem sido utilizada para o estudo da diversidade genética em videira (Scott et al., 2000), cana-de-açúcar (Cordeiro et al., 2003) e café (Aggarwal et al., 2007; Poncet et al., 2006), na integração de mapa genético da soja (Song et al., 2004), no mapeamento genético em trigo (Gao et al., 2004), batata (Feingold et al., 2005) e algodão (Han et al., 2006) e na identificação de genes ligados a resistência do cafeeiro ao bicho mineiro (Pinto et al., 2007).

Para a análise de cafeeiros da espécie *C. arabica* e seus derivados, o elevado nível de polimorfismo de marcadores é uma característica de grande importância, pois genótipos dessa espécie apresentam baixa variabilidade genética, principalmente pela biologia reprodutiva (espécie autógama) e processos evolutivos (Lashermes et al., 1999). Devido a isso, é essencial identificar os marcadores moleculares mais adequados para acessar, com o máximo de eficiência, a variabilidade genética da espécie. Neste contexto, o objetivo deste trabalho foi comparar a eficiência de marcadores EST-SSR e SSR genômico no estudo da diversidade e distinção entre acessos de *Coffea* do banco de germoplasma da Universidade Federal de Viçosa (UFV) e da Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG).

Material e Métodos

Material genético e extração de DNA

Foram analisados 18 acessos de cafeeiros do banco de germoplasma do Programa de Melhoramento Genético da UFV/EPAMIG (Tabela 1), incluindo seis acessos de *C. arabica*, cinco de *C. canephora*, três Híbridos de Timor (*C. arabica* x *C. canephora*), três Triplóides (*C. arabica* x *C. racemos*) e um de *C. racemosa*. Seis variedades comerciais de *C. arabica* resistentes à ferrugem também foram incluídas neste estudo (Tabela 1). O DNA de cada um dos 24 acessos foi extraído de folhas jovens, seguindo o protocolo de Diniz et al. (2005).

Marcadores microssatélites

Foram utilizados dois conjuntos de marcadores SSR que apresetaram produto de amplificação bem definidos (Tabela 2). O primeiro contendo 17 *primers* foi obtido das ESTs de *C. arabica* do Projeto Brasileiro do Genoma Café. O segundo de 18 *primers* foi obtido de biblioteca genômica enriquecida do genótipo de *C. arabica*, Bourbon amarelo UFV 570.

A reação de PCR foi realizada em um volume total de 20µL contendo 50ng do DNA genômico, 0,6 unidades de *Taq* DNA polimerase e tampão 1x (Promega), 1mM de MgCl₂, 150µM de cada dNTP e 0,1µM de cada *primer*. A amplificação foi efetuada em termociclador PTC 200 (MJ Research) utilizando-se o procedimento *touchdown*-PCR, que consistiu na desnaturação inicial a 94°C/2min, seguido de 13 ciclos a 94°C/30s, 67°C a 55°C/30s, reduzindo 1°C a cada ciclo, e 72°C/30s. Os 13 ciclos foram seguidos de mais 30 ciclos a 94°C/30s, 55°C/30s e 72°C/30s e extensão final a 72°C/8min. O polimorfismo foi visualizado em gel de poliacrilamida desnaturante 6% e corado com prata (Creste et al., 2001).

Análise estatística

Os géis foram avaliados visualmente, onde a presença da banda ou alelo foi designado por 1 e a ausência por 0 (zero) para cada um dos 17 *primers* EST-SSR e 18 SSR genômico. A distância genética foi calculada utilizando-se o índice de dissimilaridade de Jaccard (Jaccard, 1908). A representação gráfica das distâncias genéticas foi efetuada por meio de dendrograma obtido pelo método UPGMA, com o auxílio do *software* GENES (Cruz, 2006). Foram avaliados o número médio de alelos por *primer*, a proporção de locos polimórficos e o número de alelos exclusivos para as duas classes de SSR (Silvestrini et al., 2007).

Os acessos foram também agrupados de acordo com a origem genética: (1) acessos de *C. arabica*; (2) *C. canephora*; (3) Híbridos de Timor; (4) Triplóides e (5) Variedades comerciais de *C. arabica*, resistentes à ferrugem, com diferentes níveis de introgressão de *C. canephora*. A matriz de distância genética entre os grupos diferentes grupos genéticos foi obtida pelo coeficiente de Jacard (1908), com o auxílio do *software* GENES (Cruz, 2006). Para estudar a diversidade genética entre e dentro das populações foi realizada uma análise de variância molecular (Amova), segundo Excoffier et al. (1992), utilizando-se o *software* GENES (Cruz, 2006). O procedimento PROC STEPDISC do programa SAS (SAS Institute 1996) foi utilizado para verificar quais as melhores marcas (alelos) para discriminar as grupos genéticos estudados (Fahima et al., 1999; Beharav e Nevo 2003). Este procedimento foi aplicado para as duas classes de marcadores (EST-SSR e SSR genômico).

O número ótimo de marcadores necessários para acessar o mesmo nível de diversidade genética foi estimado a partir de correlações entre matrizes de dissimilaridade, utilizando-se o *software* GENES (Cruz, 2006). Para isso, foram obtidas várias matrizes de dissimilaridade, as quais foram correlacionadas com a matriz obtida com todas as marcas. O procedimento foi iniciado com cinco marcas (bandas ou alelos), sendo repetido com incrementos de cinco marcas, até o total de 183 marcas. Para cada tamanho amostral foram realizadas 30 repetições aleatórias e os resultados foram apresentados graficamente, por meio de uma correlação entre as matrizes com m_i marcas amostradas e a matriz original com m marcas ($m_i < m$).

Resultados

Número de alelos

Os 17 marcadores EST-SSR e 18 SSR genômicos, quando analisados nos 24 genótipos de cafeeiros, geraram um total de 87 e 96 alelos, com um número médio de alelos/*primer* de 5,1 e 5,3, respectivamente (Tabela 2). O número de alelos/*primer* variou de 3 a 11 para os marcadores EST-SSR e de 3 a 13 para os SSR genômicos. Alelos nulos, considerados neste trabalho como alelos que não apresentaram produto de amplificação em

um ou mais acessos, foram observados para aproximadamente 22% dos marcadores SSR genômicos, enquanto que nenhum marcador EST-SSR apresentou alelos nulos. Observou-se também que 14 (82%) marcadores EST-SSR e 7 (39%) SSR genômico apresentaram mais do que dois alelos por indivíduo (Tabela 2), sendo esperado até quatro alelos/indivíduo para os acessos alotetraplóides de *C. arabica*.

Os marcadores SSR genômicos apresentaram 29 alelos exclusivos para os acessos de *C. canephora*, enquanto somente 19 foram encontrados para os marcadores EST-SSR (Tabela 3). Para os acessos de *C. arabica* e Híbridos de Timor apenas os marcadores SSR genômicos apresentaram alelos exclusivo, 1 e 3, respectivamente. Para as variedades resistentes à ferrugem foram observados alelos exclusivos apenas com os EST-SSRs (1). Para *C. racemosa*, um maior número de alelos exclusivos (9) foi encontrado utilizando os marcadores SSR genômicos.

Nível de polimorfismo e diversidade genética entre e dentro de populações de *Coffea*

Todos os *primers* avaliados apresentaram polimorfismo entre os 24 acessos de *Coffea* analisados. Os marcadores SSR obtidos de biblioteca genômica foram mais eficientes em detectar polimorfismo entre e dentro dos grupos genéticos estudados (Tabela 3). Dentro dos grupos eles foram, em média, 42,9% mais polimórficos que os SSR obtidos das ESTs do Projeto Brasileiro do Genoma Café. A superioridade dos SSR genômicos foi menor entre os grupos (em média 35,0%). As duas classes de marcadores SSR avaliadas foram mais eficientes em detectar diferenças genéticas entre do que dentro dos grupos genéticos. Os marcadores SSR genômicos foram aproximadamente 170% e 158% mais polimórficos que os EST-SSR dentro dos grupos de Híbridos de Timor e *C. arabica*, respectivamente (Tabela 3).

Analisando dentro dos grupos genéticos, observou-se que os maiores níveis de polimorfismo foram encontrados com os SSR genômicos para os acessos de *C. canephora* (89,2%) e variedades comerciais de *C. arabica* resistentes à ferrugem (39,5%). Por outro lado, o menor grau de polimorfismo (10,6%) foi encontrado com os EST-SSR dentro do grupo de Híbrido de Timor (Tabela 3).

Os maiores níveis de polimorfismos entre os grupos genéticos foram detectados envolvendo o grupo de *C. canephora* com os demais, tanto com os EST-SSR quanto com os SSR genômicos. O maior polimorfismo (97,5%) encontrado foi entre os acessos de *C. canephora* e Triplóides, considerando os SSR genômicos. Os menores níveis de polimorfismos foram verificados entre os acessos de *C. arabica* e Híbridos de Timor (22%), com os marcadores EST-SSR. Para os SSR genômicos os menores níveis de polimorfismos

foram entre os acessos de *C. arabica* e variedades comerciais (43,6%) e os acessos Triplóides (Tabela 3).

Os resultados da amostra mostraram que os marcadores SSR (EST-SSR e SSR genômico) foram mais eficientes em detectar a diversidade genética entre os grupos genéticos (Tabela 4). Considerando os EST-SSR, aproximadamente 64% da variação genética ocorreram entre grupos, enquanto que para os SSR genômico esta foi de 54%. Os marcadores SSR genômico foram mais eficientes para detectar a diversidade genética dentro dos grupos (46%) do que os EST-SSR (36%).

As maiores distâncias genética foram encontradas entre os grupos de *C. canephora* e *C. racemosa*, para as duas classes de marcadores SSR (Tabela 5). Tanto os marcadores EST-SSR quanto os SSR genômicos, indicaram os grupos de *C. arabica* e as variedades comerciais como as mais próximas geneticamente (Tabela 5).

Para as análises dentro de cada grupo genético, embora as medidas de dissimilaridade não sejam independentes entre si (medidas altamente correlacionadas), o teste *t* (Student) foi aplicado para comparar os coeficientes de dissimilaridade genética considerando as duas classes de marcadores SSR (Figura 1). Os maiores coeficientes de dissimilaridade genética foram encontrados dentro do grupo de *C. canephora*, seguidos por Híbrido de Timor e variedades comerciais resistentes à ferrugem. Por outro lado, dentro dos grupos de *C. arabica* e Triplóides foram encontrados os menores coeficientes de dissimilaridade genética (Figura 1). Os marcadores SSR genômico apresentaram os maiores coeficientes de dissimilaridade genética em todos os grupos estudados, sendo estatisticamente superiores aos EST-SSR para acessar a diversidade dentro dos grupos de *C. arabica* ($t = 4.48$, $p < 0,01$), *C. canephora* ($t = 3.62$, $p < 0,05$) e Híbridos de Timor ($t = 3.25$, $p < 0,05$). Para o grupo dos acessos Triplóides e variedades comerciais resistentes à ferrugem, apesar dos marcadores SSR genômicos apresentarem os maiores coeficientes de dissimilaridades, não houve diferença estatística entre eles.

Análise de agrupamento e distância genética

O potencial dos marcadores EST-SSR e SSR genômico foi avaliado individualmente e em conjunto nos estudos da diversidade genética nos acessos de *Coffea* do banco de germoplasma da UFV/EPAMIG.

O dendrograma gerado pela matriz de dissimilaridade genética de Jaccard, considerando todos os acessos, possibilitou a formação de grupos distintos de acordo com a constituição genética das espécies, tanto para os EST-SSR e SSR genômico, quanto para o agrupamento conjunto (Figura 2, 3 e 4). Os acessos de *C. canephora*, *C. racemosa* e *C.*

arabica e seus derivados foram separadamente agrupados utilizando as duas classes de marcadores SSR. Entretanto, muitos acessos de *C. arabica* e variedades comerciais resistentes à ferrugem, assim como dois Híbridos de Timor (HT CIFC 832/2 e HT CIFC 4106) e dois acessos Triplóides (UFV 557-2 e 557-4), não foram distinguidos pelos 17 *primers* EST-SSR (Figura 2). Por outro lado, os 18 *primers* SSR genômico foram capazes de diferenciar todos os acessos estudados (Figura 3). As duas classes de marcadores SSR permitiram separar todos os acessos de *C. canephora* e *C. racemosa* dos demais.

Os maiores coeficientes de dissimilaridade genética foram obtidos com os marcadores SSR genômicos, os quais variaram de 0,03 a 0,80 entre os 24 acessos analisados (Tabela 6). Para os marcadores EST-SSR os coeficientes de dissimilaridade variaram de 0,00 a 0,77, sendo que muitos acessos de *C. arabica* e variedades comerciais resistentes à ferrugem não foram distinguidas por estes marcadores. Os maiores coeficientes de dissimilaridade foram encontrados entre os acessos de *C. canephora* e *C. racemosa* para as duas classes de marcadores.

Número de marcas e análise discriminante

O número mínimo de marcadores SSR necessário para obter o mesmo padrão de agrupamento e diversidade genética entre os 24 acessos de *Coffea* foi de 120 marcas (Figura 6). A matriz de dissimilaridade gerada utilizando as 120 marcas SSR apresentou uma correlação de 0,90 ($t = 109,11$; $p < 0,001$) com a matriz usando todas as 183 marcas SSR. O dendrograma com as 120 marcas foi o mesmo que o obtido com todas as marcas SSR (dados não mostrados). Além disso, os acessos mantiveram a mesma ordem dentro dos grupos de populações. Pelo resultado obtido observou-se que números menores que 120 marcas, também apresentaram matrizes de alta correlação com a obtida usando o total de marcas (Figura 5), mas modificou a ordem e número de grupos dos dendrogramas.

A análise discriminante, baseada na presença e ausência de alelos, mostrou os melhores locos para diferenciar os grupos genéticos estudados com as duas classes de marcadores SSR (Tabela 7). Dez *primers* e dezessete alelos dos marcadores SSR genômico foram os que melhor discriminaram os grupos avaliados. Para os EST-SSR, oito alelos de cinco *primers* discriminaram todas os grupos. Alguns marcadores contribuíram com mais de um alelo para a análise discriminante, como o SSRCa 087 e EST-SSR 007 (5 e 4 alelos, respectivamente) que foram fundamentais para a discriminação dos grupos. Aproximadamente 17,7% dos alelos dos marcadores SSR genômicos e 9,2% dos EST-SSR foram fundamentais para distinção entre os grupos. Os marcadores SSR genômicos foram aproximadamente 52,9% mais eficientes que os EST-SSR para a análise discriminante dos grupos.

Discussão

A variação do número de alelos em um loco de microssatélite pode ser devido a diferenças na taxa de mutação e na pressão de seleção exercida em cada loco (Métais et al., 2002), que por sua vez são influenciadas pela estrutura, tipo e comprimento da repetição e também pelo contexto genômico do loco de microssatélite. Em nosso trabalho o número médio de alelos por loco foi praticamente igual para as duas classes de marcadores SSR, diferentemente do encontrado para arroz (Cho et al., 2000), feijão (Blair et al., 2006) e cana-de-açúcar (Cordeiro et al., 2001), onde os SSR genômico apresentaram maior número médio de alelos que os marcadores EST-SSR. O número total de alelos de uma população é altamente dependente do seu tamanho, constituição genética e número de *primers* ou loco avaliado. O nível de polimorfismo de um conjunto de 25 *primers* EST-SSR foi estudado por Poncet et al. (2006), em diferentes espécies do gênero *Coffea*, encontrando uma média de 10,5 alelos/loco. Maluf et al. (2005), ao estudar a diversidade genética entre 28 acessos de *C. arabica* e *C. canephora* com 23 *primers* SSR, encontraram uma média de 2,87 alelos por loco num total de 66 alelos amplificados. Moncada e McCouch (2004), utilizando 34 *primers* SSR genômico para estudar a diversidade genética em *Coffea*, encontraram uma média de 3,0 alelos/loco em espécies tetraplóides selvagens e 2,0 alelos/loco em espécies cultivadas.

Neste trabalho, os *primers* SSR genômico SSRCa 003, 052, 068 e 081 apresentaram alelos nulos, o que não foi observado para o conjunto de *primers* EST-SSR (Tabela 2). Alelos nulos resultantes de amplificação de marcadores microssatélites também foram encontrados em algumas espécies como trigo (Gupta e Varshney 2000 e Leigh et al., 2003), arroz (Cho et al., 2000) e café (Poncet et al., 2004 e Poncet et al., 2006). Alelos nulos é possivelmente o resultado de mutações de pontos, tais como os SNP (*Single Nucleotide Polymorphism*), ou outros eventos de mutação que ocorrem na região onde os *primers* foram desenhados, ou ainda, podem ser originados de rearranjos genômicos de alta escala, incluindo eventos de inserção-deleção dentro da região amplificada (Leigh et al., 2003). A principal explicação para a maior proporção de alelos nulos encontrada nos SSR genômicos está no fato de que mutações são menos toleradas em regiões transcritas do genoma, como é o caso dos EST-SSR. Isto afeta a taxa com que as regiões que flanqueiam os SSRs, onde os *primers* são desenhados, estejam envolvidas. Muitos SSRs não são parte constitutiva do gene, por isso nessas regiões são toleradas mutações, porém, grandes inserções ou deleções podem ser evolutivamente deletérias ou prejudiciais para a espécie, não sendo, portanto, selecionadas.

Aproximadamente 82% dos *primers* EST-SSR e 39% dos SSR genômicos apresentaram mais de dois alelos por indivíduo (Tabela 2). Este fato é esperado em espécies

com genoma poliplóide, como os acessos alotetraplóides de *C. arabica* em que até quatro alelos podem ser observados no mesmo indivíduo. Para as espécies diplóides (*C. canephora* e *C. racemosa*) nenhum *primer* SSR apresentou mais de dois alelos por indivíduo, diferente do relatado por Silvestrini et al. (2007) para o loco E12-3CTG (SSR9) em acessos das espécies de *C. canephora* e *C. eugenioides*.

Os marcadores SSR genômicos avaliados neste trabalho, apresentaram um maior número de alelos exclusivos para grupo genético, quando comparados com os EST-SSRs (Tabela 3). Alelos exclusivos são úteis para identificação de espécies, distinção de acessos e *fingerprint* de variedades (Donini et al., 1998; Eujayl et al., 2001; He et al., 2005). Portanto, o elevado número de alelos exclusivos encontrados pelos *primers* SSR genômicos demonstram que estes marcadores são úteis para *fingerprint* genético, estudos de introgressão e orientações de cruzamentos para formação de novas populações de mapeamento genético.

Os *primers* SSR genômico foram mais polimórficos do que os EST-SSR nos acessos do gênero *Coffea* avaliados. Um polimorfismo aproximadamente 42,8% maior dentro e 35,0% entre os grupos foi obtido pelos SSR genômicos (Tabela 3). Maiores níveis de polimorfismo dessa classe de marcadores também foram encontrados para outras espécies como videira (Scott et al., 2000), arroz (Cho et al., 2000), trigo (Eujayl et al., 2001), cana-de-açúcar (Cordeiro et al., 2001) e cevada (Chabane et al., 2006). Segundo Varshney et al. (2005), a superioridade de polimorfismo dos SSR genômicos comparada com a dos EST-SSR é devido a elevada conservação das seqüências de DNA em regiões transcritas do genoma. A posição dos *primers* EST-SSR dentro da seqüência pode influenciar o nível de polimorfismo dos *primers*, sendo os derivados de regiões 3'ESTs mais polimórficos do que aqueles derivados de regiões 5'ESTs em muitas espécies de plantas (Scott et al., 2000; Gao et al., 2003; Gupta et al., 2003). Outro fator importante entre as duas classes de SSR avaliadas neste trabalho é a porcentagem de *primers* com repetições de di- e trinucleotídeos. Aproximadamente 56% dos SSR genômicos correspondem a marcadores com repetições dinucleotídeos, contra 47% dos EST-SSR (Tabela 2). Marcadores SSR com repetições de dinucleotídeos foram mais polimórficos do que os trinucleotídeos em trigo (Nicot et al., 2004 e Gadaleta et al., 2007) e cevada (Baek et al., 2003). Em espécies do gênero *Coffea* este tipo de estudo ainda não foi realizado, mas este fato pode também ser uma das explicações para o maior polimorfismo apresentado pelos *primers* SSR genômicos.

O conjunto de 17 *primers* EST-SSR não foi suficiente para diferenciar todos os 24 acessos de *Coffea* avaliados. Não foi possível discriminar, com esses marcadores, alguns acessos de *C. arabica*, variedades comerciais, Híbridos de Timor e Triplóides (Figura 2 e Tabela 6). Variedades comerciais derivadas de *C. arabica*, também não foram separadas com

o conjunto de *primers* SSR utilizados por Silvestrini et al. (2007) para estudos de diversidade genética em *Coffea*. Por outro lado, os 18 *primers* SSR genômicos, utilizados neste trabalho, possibilitaram diferenciar todos os acessos (Figura 3), com uma maior distância genética entre (Tabela 5) e dentro das populações (Figura 1). Os *primers* SSR Ca 018 e 091 permitiram distinguir todos os acessos de *C. canephora* (T 3751; T 3580; Guarini UFV 514 e Apoatã IAC 2258, do grupo Robusta; e UFV 513, do grupo Conillon). Estes *primers* podem ser úteis para *fingerprint* genético e certificação de variedades comerciais clonais derivadas de *C. canephora*. Com as duas classes de marcadores SSR foi possível identificar os Híbridos de Timor mais próximos e mais distantes dos genótipos de *C. arabica*. Esse resultado é importante para os programas de melhoramento que visam incorporar um ou alguns genes de *C. canephora* e manter as demais características de *C. arabica*. Para isso, deve-se escolher o acesso do Híbrido de Timor mais próximo geneticamente de *C. arabica*, como o Híbrido CIFIC 1343/269. Entretanto, quando a prioridade for aumentar a base genética do programa de melhoramento, o uso de genitores divergentes geneticamente torna-se mais apropriado. Para obtenção de novas populações de mapeamento, por exemplo, recomenda-se cruzar os genótipos de *C. arabica* mais divergentes possível dos Híbridos de Timor, como os Híbridos CIFIC 832/2 e CIFIC 4106 (Tabela 6, Figuras 2, 3 e 4). As duas classes de marcadores SSR foram também eficientes em agrupar os diferentes acessos de acordo com seus respectivos grupos de origem (Figuras 2, 3 e 4).

Na Figura 1 estão apresentados os coeficientes médios de dissimilaridade genética para cada grupo genético. Apesar do baixo polimorfismo dentro dos grupos de *C. arabica*, Híbrido de Timor, Triplóides e variedades comerciais, as duas classes de marcadores SSR utilizados neste trabalho foram eficientes para diferenciar estes grupos (Tabela 5). Os SSR genômicos foram significativamente mais eficientes que os EST-SSR para detectar diferenças genéticas dentro do grupo de *C. arabica* ($t = 4,48$; $p < 0,01$), *C. canephora* ($t = 3,62$; $p < 0,05$) e Híbridos de Timor ($t = 3,25$; $p < 0,05$). Este resultado demonstra que em *Coffea*, assim como em outras espécies de plantas (Varshney et al., 2005), os marcadores SSR genômicos se destacam pelo maior polimorfismo comparado com os EST-SSR.

Elevados níveis de polimorfismos foram encontrados para os acessos de *C. canephora*. Este resultado está em concordância com trabalhos anteriores (Baruah et al., 2003; Poncet et al., 2004; Poncet et al., 2006; Aggarwal et al., 2007; Silvestrini et al., 2007; Hendre et al., 2008) e sustenta a hipótese de elevada diversidade desta espécie, resultante da composição genética e sistema de cruzamento alógama. Por outro lado, as duas classes de marcadores SSR apresentaram um menor polimorfismo em acessos de *C. arabica* e variedades comerciais resistentes à ferrugem, assim como encontrado por Silvestrini et al. (2007) para algumas

variedades do IAC (Instituto Agronômico de Campinas, Brasil). Esta baixa variabilidade genética em *C. arabica* e seus derivados tem sido atribuída aos processos evolutivos da origem da espécie, incluindo sua biologia reprodutiva (autogamia).

O número de marcas necessárias para produzir o mesmo agrupamento e diversidade genética, para este estudo, foi de 120 marcas (Figura 6). Isto evidencia que muitas marcas SSR utilizadas neste trabalho, não são importantes para discriminar entre os 24 acessos de *Coffea* estudados, assim como também foi observado pela análise discriminante (Tabela 7). Resultados similares foram também obtidos por Fahima et al. (1999) em que apenas 15 dos 48 locos RAPD foram importantes para discriminar 11 populações de trigo. No estudo da diversidade genética de 136 acessos de cajueiro (*Anacardium* spp), utilizando marcadores ISSR, Pessoni (2007) encontrou um número ótimo entre 115 a 155 marcas necessárias para produzir o mesmo nível de diversidade que o total de 223 marcas, indicando também a presença de marcas pouco informativas. Este trabalho pode servir de base para futuros estudos de diversidade genética utilizando marcadores SSR em *Coffea*.

Todos os *primers* SSR genômicos avaliados neste trabalho, embora desenvolvidos de *C. arabica*, amplificaram fragmentos nos acessos de *C. canephora*, Híbridos de Timor e Triplóides. Treze (59%) *primers* SSR genômicos e 16 (94%) EST-SSR foram validados também para *C. racemosa*. Devido a natureza conservada das seqüências EST utilizadas para o desenho dos *primers* avaliados neste trabalho, os mesmos apresentaram uma maior taxa de funcionalidade em outras espécies, principalmente em *C. racemosa*. Esses resultados demonstram o potencial dos marcadores SSR para estudos genéticos em espécies relacionadas do gênero *Coffea*. A elevada transferência de *primers* entre espécies de *Coffea* foi anteriormente observada para marcadores EST-SSRs (Bhat et al., 2005; Poncet et al., 2006 e Aggarwal et al., 2007).

Este é o primeiro trabalho que compara os marcadores EST-SSR e SSR genômico para o estudo da diversidade genética em *Coffea*. Os marcadores SSR utilizados neste trabalho elucidaram claramente as relações genéticas previamente apresentadas por outros grupos de marcadores de DNA para o gênero *Coffea* (Antony et al., 2001; Baruah et al., 2003; Poncet et al., 2004; Poncet et al., 2006; Aggarwal et al., 2007; Silvestrini et al., 2007; Hendre et al., 2008). A elevada superioridade em detectar o nível de polimorfismo, o maior número de alelos exclusivos, os maiores coeficientes de dissimilaridade genética e o elevado poder de discriminação dos acessos tornam os marcadores SSR genômicos preferíveis para estudos de diversidade genética em *Coffea*, principalmente em cafeeiros arábicas que apresentam baixa variabilidade genética. Estes marcadores são de grande utilidade para identificação de variedades derivadas de *C. arabica* e *C. canephora*, uma vez que foram 41% e 15% mais

polimórficos que os EST-SSR, respectivamente. Os *primers* SSR genômicos SSRCa 018, 068, 082, 087, 091 e 094 e os EST-SSR 007 e 029, devem ser os preferidos para estudar o acessos de *C. arabica*, pois apresentaram os maiores níveis de polimorfismo entre os acessos (Tabela 8). Os SSR genômicos podem também ser eficientemente utilizados para identificação dos Híbridos de Timor do banco de germoplasma da UFV/EPAMIG, principalmente pela superioridade em relação aos EST-SSR (~170%) no nível de polimorfismo genético. Embora as duas classes de marcadores possuam suas próprias vantagens e desvantagens, ficou claramente demonstrada a importância dos SSR genômicos para estudos da diversidade genética em *Coffea*. O elevado custo, mão-de-obra e tempo para a obtenção dos *primers* SSR genômicos são recompensados pelo maior nível de polimorfismo desta classe de marcadores, principalmente para *C. arabica*, a espécie de maior importância sócio-econômica mundial, que apresenta estreita base genética.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), a Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (CBP&D) pelo apoio financeiro.

Referências

- Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. **Theor Appl Genet** 114:359-372
- Anthony F, Bertrand B, Quiros O, Wilches A, Lashermes P, Berthaud J, Charrier A (2001) Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. **Euphytica** 118:53-65
- Baek HJ, Beharav A, Nevo E (2003) Ecological-genomic diversity of microsatellites in wild barley *Hordeum spontaneum*, population in Jordan. **Theor Appl Genet** 106:397-410
- Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P, Aggarwal RK (2003) Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. **Mol Ecol Notes** 3:647-650
- Beharav A, Nevo E (2003) Predictive validity of discriminant analysis for genetic data. **Genetica** 119: 259-267
- Blair MW, Giraldo MC, Buendía HF, Tovar E, Duque MC, Beebe SE (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). **Theor Appl Genet** 113: 100-109
- Buso GSC, Ciampi AY, Moretzsohn MC, Amaral ZPS, Brondani RV (2003) Desenvolvimento e caracterização de marcadores microssatélites em espécies vegetais tropicais. **Revista Biotecnologia e Desenvolvimento** 30:46-50
- Chabane K Ablett GA, Cordeiro GM, Valkoun J, Henry RJ (2006) EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. **Genet Resour Crop Evol** 52:903-909
- Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). **Theor Appl Genet** 100:713-722
- Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. **Plant Sci** 160:1115-1123
- Cordeiro GM, Pan YB, Henry RJ (2003) Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. **Plant Science** 165:181-189
- Creste S, Tulmann-Neto A, Figueira A (2001) Detection of simple sequence repeat polymorphisms in denaturing polyacrilamide sequencing gels by silver staining. **Plant Molecular Biology Report** 4: 299-306

- Cruz CD (2006) **Genes versão 2006.4.1: Programa GENES versão Windows**. Editora UFV, Viçosa, MG, Brazil
- Diniz LEC, Sakiyama NS, Lashermes P, Caixeta ET, Oliveira ACB (2005) Analysis of AFLP marker associated to the Mex-1 resistance locus in Icatu progenies. **Crop Breed Appl Biotech** 5:387-393
- Donini P, Stephenson P, Bryan GJ, Koebner RMD (1998) The potential of microsatellite for high throughput genetic diversity assessment in wheat and barley. **Genetic resources and Crop evolution** 45: 415-421
- Eujayl I, Sorrells M, Baum M, Wolters P, Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRS and genomic SSRS. **Euphytica** 119: 39-43
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. **Genetics** 131:479-491
- Fahima T, Sun GL, Beharav A, Krugman T, Beiles A, Nevo E (1999) RAPD polymorphism of wild emmer wheat populations, *Triticum dicoccoides*, in Israel. **Theor Appl Genet** 98: 434-447
- Feingold S, Lloyd J, Norero N, Bonierbale M, Lorenzen J (2005) Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). **Theor Appl Genet** 111:456-466
- Gadaleta A, Mangini G, Mulè G, Blanco A (2007) Characterization of dinucleotide and trinucleotide EST-derived microsatellites in the wheat genome. **Euphytica** 153: 73-85
- Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. **Theoretical and Applied Genetics** 108:1392-400
- Gao LF, Tang J, Li H, Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. **Mol Breed** 12:245-261
- Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. **Euphytica** 113:163-185
- Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. **Molecular Genetics and Genomics** 270: 315-323.

- Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. **Theor Appl Genet** 112:430-439
- He G, Meng R, Gao H, Guo B, Gao G, Newman M, Pittman RN, Prakash CS (2005) Simple sequence repeat markers for botanical varieties of cultivated peanut (*Arachis hypogaea* L.). **Euphytica** 142:131-136
- Hendre PS, Phanindranath R, Annapurna V, Lalremruata A, Aggarwal RK (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 Biology** 8: 51(doi:10.1186/1471-2229-8-51)
- Jaccard P (1908) Nouvelles recherches sur la distribution florale. **Bulletin de la Société Vandoise des Sciences Natureles** 44:223-270
- Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A (1999) Molecular characterization and origin of the *Coffea arabica* L. genome. **Mol Gen Genet** 261: 259-266
- Leigh F, Lea V, Law J, Wolteret P, Powell W, Donini P (2003) Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. **Euphytica** 133:359-366
- Maluf MP, Silvestrini M, Ruggiero LMC, Guerreiro Filho O, Colombo CA (2005) Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. **Scientia Agricola** 62:366-373
- Métais I, Hamon B, Jalouzot R, Peltier D (2002) Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. **Theor Appl Genet** 104:1346-1352
- Moncada P, McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. **Genome** 47:501-509
- Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). **Theor Appl Genet** 109:800-805
- Pessoni LA (2007) **Estratégias de análise da diversidade em germoplasmas de cajueiro (*Anacardium spp.* L.)**. Tese de doutorado, Universidade Federal de Viçosa, 159p
- Pinto FO, Maluf MP, Guerreiro-Filho O (2007) Study of simple sequence repeat markers from coffee expressed sequences associated to leaf miner resistance. **Pesq Agropec Bras** 42:377-384

- Poncet V, Hamon P, Minier J, Carasco C, Hamon S, Noirot M (2004) SSR cross-amplification and variation within coffee trees (*Coffea* spp.). **Genome** 47:1071-1081
- Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, Kochko A, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. **Mol Gen Genomics** 276:436-449
- Rohlf FJ (2000) **NTSYSpc: numerical taxonomy and multivariate analysis system, version 2.1**. Exeter Software, Setauket, NY
- SAS Institute (1996) **SAS User Guide: Statistics, Version 6.09**. SAS Institute Inc., Cary, NC
- Scott KD, Egger P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. **Theor Appl Genet** 100:723-726
- Silvestrini M, Junqueira MG, Favarin AC, Guerreiro-Filho O, Maluf MP, Silvarolla MB, Colombo CA (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. **Genet Resour Crop Evol** 54:1367-1379
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. **Theor Appl Genet** 109:122-128
- Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? **Molecular Ecology** 12:1339-1348
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. **TRENDS in Biotechnology** 23: 48-55
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. **Molecular Ecology** 11:1-16

Tabela 1. Acessos de cafeeiros utilizados neste trabalho.

Acessos	Espécie (ploídia)
Acessos do Banco de germoplasma UFV/EPAMIG	
1 - UFV 2144 (Catuaí Vermelho IAC 44)	<i>Coffea arabica</i> (2n = 4x = 44)
2 - Típica UFV 2945	<i>Coffea arabica</i> (2n = 4x = 44)
3 - Bourbon UFV 2952	<i>Coffea arabica</i> (2n = 4x = 44)
4 - Bourbon Amarelo UFV 535-1	<i>Coffea arabica</i> (2n = 4x = 44)
5 - Arábica UFV 10832	<i>Coffea arabica</i> (2n = 4x = 44)
6 - Bourbon Amarelo UFV 10745	<i>Coffea arabica</i> (2n = 4x = 44)
7 - T 3751 (Robusta)	<i>Coffea canephora</i> (2n = 2x = 22)
8 - T 3580 (Robusta)	<i>Coffea canephora</i> (2n = 2x = 22)
9 - Conillon UFV 513 (Conillon)	<i>Coffea canephora</i> (2n = 2x = 22)
10 - Guarini UFV 514 (Robusta)	<i>Coffea canephora</i> (2n = 2x = 22)
11 - Apatã IAC 2258 (Robusta)	<i>Coffea canephora</i> (2n = 2x = 22)
12 - Híbrido de Timor CIFC 832/2	<i>C. arabica</i> x <i>C. canephora</i> (2n = 4x = 44)
13 - Híbrido de Timor CIFC 4106	<i>C. arabica</i> x <i>C. canephora</i> (2n = 4x = 44)
14 - Híbrido de Timor CIFC 1343/269	<i>C. arabica</i> x <i>C. canephora</i> (2n = 4x = 44)
15 - UFV 557-2	Triplóide (<i>C. arabica</i> x <i>C. racemosa</i>) (2n = 3x = 33)
16 - UFV 557-3	Triplóide (<i>C. arabica</i> x <i>C. racemosa</i>) (2n = 3x = 33)
17 - UFV 557-4	Triplóide (<i>C. arabica</i> x <i>C. racemosa</i>) (2n = 3x = 33)
18 - <i>Coffea racemosa</i>	<i>Coffea racemosa</i> (2n = 2x = 22)
Variedades comerciais resistentes à ferrugem	
19 - Catiguá MG2	Variedade comercial (<i>C. arabica</i> x HT) (2n = 4x = 44)
20 - IAPAR 59	Variedade comercial (<i>C. arabica</i> x HT) (2n = 4x = 44)
21 - Oeiras MG6851	Variedade comercial (<i>C. arabica</i> x HT) (2n = 4x = 44)
22 - Sacramento MG1	Variedade comercial (<i>C. arabica</i> x HT) (2n = 4x = 44)
23 - Catucaí Amarelo 2SL	Variedade comercial (<i>C. arabica</i> x Icatu vermelho) (2n = 4x = 44)
24 - Obatã Amarelo IAC 4932	Variedade comercial (<i>C. arabica</i> x HT) (2n = 4x = 44)

Tabela 2. Descrição e comparação dos marcadores EST-SSR e SSR genômico.

Marcador	Repetição	Tamanho do fragmento (pb)	Tm (°C)	Nº de alelos ¹	Alelos nulos ²	Nº máximo de alelos/indivíduo ³
EST-SSR						
EST-SSR 005	(CA) ₇	129	58	3	N	3
EST-SSR 007	(GTC) ₅	117	58	10	N	3
EST-SSR 010	(TG) ₇ (TA) ₄ // (GT) ₉	152	51	11	N	3
EST-SSR 012	(CA) ₇ (AC) ₃	118	53	4	N	3
EST-SSR 013	(CA) ₉	141	52	3	N	3
EST-SSR 023	(TA) ₁₅	207	58	5	N	3
EST-SSR 025	(AT) ₇	108	53	3	N	3
EST-SSR 027	(AT) ₃ (AC) ₄ (AT) ₉	259	53	6	N	4
EST-SSR 029	(AT) ₁₀	169	48	6	N	4
EST-SSR 047	(CAA) ₅ N ₃ (CTG) ₆	178	52	7	N	4
EST-SSR 048	(AAG) ₆ N ₉ (TGA) ₄	126	54	4	N	2
EST-SSR 054	(TAT) ₇	184	53	4	N	2
EST-SSR 055	(TAT) ₅	167	52	4	N	2
EST-SSR 057	(TTA) ₆	90	51	4	N	3
EST-SSR 058	(AAT) ₄	201	58	4	N	3
EST-SSR 069	(GCG) ₆	101	53	3	N	3
EST-SSR 073	(CGC) ₂ (GGC) ₅	160	54	6	N	4
Subtotal				87		
Média				5,1		
SSR genômico						
SSRCa 003	(GT) ₁₂	196	57	4	S	2
SSRCa 016	(GAA) ₃ // (GGAAAG) ₃	172	56	3	N	2
SSRCa 018	(GT) ₁₈ (GA) ₁₀	115	57	13	N	3
SSRCa 019	(GA) ₁₁	329	55	4	N	2
SSRCa 020	(AGA)G(AGA) ₃ // (TG) ₄ // (ATT) ₆	264	57	3	N	2
SSRCa 026	(T) ₁₆ N ₁₂ (TC) ₇ // (CAC) ₄	289	57	5	N	2
SSRCa 052	(TTG) ₇	129	57	4	S	3
SSRCa 062	(CAA) ₂ G(AGAA) ₂ // (AG) ₄ N ₈ (GA) ₄	275	57	3	N	3
SSRCa 068	(AGG) ₇ // (GAA) ₄	236	57	4	S	4
SSRCa 081	(CT) ₃₈	229	53	4	S	2
SSRCa 082	(CT) ₁₇ CG(CT) ₆	178	56	7	N	2
SSRCa 083	(TC) ₃₂	223	56	5	N	2
SSRCa 087	(TC) ₂₂	143	56	8	N	3
SSRCa 088	(TTTTCT) ₃	180	57	5	N	3
SSRCa 091	(GT) ₈ (GA) ₁₀	110	56	10	N	2
SSRCa 092	(CCA) ₇ CT(TCCACC) ₅	142	58	6	N	3
SSRCa 094	(TC)4(TTCT) ₃ // (TTTCCT) ₃ (TTTC) ₅	195	55	4	N	2
SSRCa 095	(TG) ₁₁	185	57	4	N	2
Subtotal				96		
Média				5,3		
Total				183		

¹Número de alelos encontrados nos 24 acessos de *Coffea* estudados; ²Presença de alelos nulos: sim/não (alelos que não apresentaram produto de amplificação em um ou mais acessos de *C. arabica*); ³Número máximo de alelos em cada indivíduo.

Tabela 3. Porcentagem de polimorfismo entre e dentro dos grupos de *Coffea*: comparação da eficiência dos marcadores EST-SSR e SSR genômico.

Grupos	Polimorfismo (%)*		Superioridade (%)**	Número médio de alelos (alelos exclusivos)	
	EST-SSR	SSR genômico		EST-SSR	SSR genômico
Dentro					
<i>C. arabica</i> (Ca)	11,1	28,6	157,7	4,0 (0)	1,9 (1)
<i>C. canephora</i> (Cc)	77,8	89,2	14,7	3,3 (19)	3,6 (29)
Híbrido de Timor (HT)	10,6	28,6	169,8	3,2 (0)	1,9 (3)
Triplóides (T)	12,3	14,3	16,3	3,7 (4)	1,6 (5)
Variedades (V)	28,0	39,5	41,1	3,7 (1)	2,0 (0)
<i>C. racemosa</i> (Cr)	-	-	-	2,0 (1)	1,3 (9)
Média	28,0	40,0	42,9		
Entre					
Ca x Cc	84,5	94,0	11,2		
Ca x HT	22,0	47,8	116,4		
Ca x T	30,4	43,6	43,42		
Ca x V	33,3	43,6	30,9		
Ca x Cr	30,8	91,1	195,8		
Cc x HT	83,1	94,0	13,1		
Cc x T	86,1	97,5	13,2		
Cc x V	86,5	96,3	11,3		
Cc x Cr	90,1	97,3	8,0		
HT x T	36,7	48,0	30,8		
HT x V	37,7	41,3	9,5		
HT x Cr	55,4	88,5	59,7		
T x V	36,1	50,0	38,5		
T x Cr	44,8	85,1	90,0		
V x Cr	61,4	87,8	43,0		
Média	54,6	73,7	35,0		

- não avaliado devido conter somente um acesso; *o nível de polimorfismo foi calculado dividindo o número de bandas polimórficas pelo número total de bandas amplificadas para cada população; **Superioridade = porcentagem de superioridade dos marcadores SSR genômico em relação aos EST-SSR (Exemplo: $[(28,6-11,1) \times 100] / 11,1 = 157,7\%$).

Tabela 4. Análise de variância molecular (Amova) entre e dentro de grupos genéticos de *Coffea*.

Fonte de variação	GL	Soma de quadrados		Componente de variação		Variação (%)	
		EST-SSR	SSR genômico	EST-SSR	SSR genômico	EST-SSR	SSR genômico
Entre grupos	5	151,04	148,03	6,88	6,33	64,22	54,34
Dentro de grupos	18	60,00	95,08	3,83	5,32	35,78	45,66
Total	23	220,04	243,83	10,71	11,66		

Tabela 5. Distância genética entre populações obtida pelos marcadores EST-SSR (baixo da diagonal) e SSR genômico (acima da diagonal).

Populações	<i>C. arabica</i>	<i>C. canephora</i>	Híbrido de Timor	Triplóides	<i>C. racemosa</i>	Variedades
<i>C. arabica</i>	-	0,819	0,291	0,292	0,840	0,190
<i>C. canephora</i>	0,626	-	0,772	0,831	0,867	0,813
Híbrido de Timor	0,136	0,604	-	0,346	0,850	0,290
Triplóides	0,236	0,658	0,294	-	0,862	0,261
<i>C. racemosa</i>	0,476	0,711	0,522	0,397	-	0,828
Variedades	0,124	0,626	0,163	0,253	0,502	-

Tabela 6. Coeficientes de dissimilaridade genética (Jacard, 1908) entre os 24 acessos de *Coffea* do banco de germoplasma da UFV/EPAMIG (valores abaixo e acima da diagonal correspondem as distâncias obtidas com os EST-SSR e SSR genômico, respectivamente).

Acessos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1-Catuai vermelho UFV2144	-	0,29	0,08	0,15	0,18	0,11	0,75	0,71	0,69	0,75	0,69	0,25	0,26	0,17	0,21	0,20	0,24	0,75	0,25	0,13	0,15	0,11	0,13	0,15
2-Típica UFV2945	0,00	-	0,23	0,21	0,18	0,25	0,76	0,73	0,71	0,77	0,75	0,21	0,33	0,26	0,38	0,33	0,36	0,77	0,21	0,31	0,25	0,29	0,23	0,25
3-Bourbon UFV2952	0,00	0,00	-	0,08	0,21	0,03	0,74	0,71	0,71	0,77	0,70	0,27	0,31	0,20	0,24	0,23	0,26	0,74	0,23	0,11	0,08	0,13	0,16	0,18
4-Bourbon amarelo UFV535-1	0,00	0,00	0,00	-	0,14	0,06	0,76	0,73	0,73	0,79	0,75	0,26	0,30	0,18	0,26	0,21	0,24	0,74	0,16	0,13	0,11	0,16	0,18	0,11
5-Arabica UFV10832	0,10	0,10	0,10	0,10	-	0,18	0,78	0,75	0,70	0,76	0,74	0,28	0,32	0,20	0,29	0,23	0,27	0,76	0,19	0,21	0,18	0,14	0,16	0,14
6-Bourbon amarelo UFV10745	0,10	0,10	0,10	0,10	0,00	-	0,76	0,73	0,73	0,79	0,72	0,30	0,33	0,22	0,22	0,21	0,24	0,74	0,21	0,08	0,06	0,11	0,14	0,16
7-Robusta T37511	0,55	0,55	0,55	0,55	0,51	0,51	-	0,65	0,69	0,64	0,65	0,74	0,72	0,70	0,77	0,76	0,77	0,75	0,77	0,72	0,74	0,76	0,75	0,76
8-Robusta T3580	0,54	0,54	0,54	0,54	0,53	0,53	0,18	-	0,64	0,59	0,60	0,71	0,71	0,72	0,75	0,73	0,74	0,80	0,74	0,71	0,70	0,73	0,72	0,70
9-Conillon UFV 513	0,68	0,68	0,68	0,68	0,67	0,67	0,62	0,60	-	0,53	0,61	0,71	0,71	0,67	0,75	0,73	0,75	0,75	0,72	0,74	0,73	0,71	0,70	0,71
10-Robusta Guarini UFV514	0,56	0,56	0,56	0,56	0,53	0,53	0,33	0,36	0,57	-	0,53	0,70	0,70	0,71	0,80	0,79	0,80	0,73	0,78	0,77	0,79	0,75	0,76	0,75
11-Robusta Apoatã IAC2258	0,62	0,62	0,62	0,62	0,61	0,61	0,25	0,34	0,58	0,39	-	0,68	0,66	0,67	0,71	0,72	0,71	0,79	0,74	0,70	0,72	0,70	0,69	0,70
12-HT CIFC 832/2	0,15	0,15	0,15	0,15	0,13	0,13	0,47	0,51	0,67	0,56	0,56	-	0,19	0,19	0,34	0,30	0,33	0,76	0,22	0,27	0,33	0,26	0,24	0,17
13-HT CIFC 4106	0,15	0,15	0,15	0,15	0,13	0,13	0,47	0,51	0,67	0,56	0,56	0,00	-	0,16	0,37	0,33	0,36	0,79	0,30	0,31	0,37	0,30	0,35	0,22
14-HT CIFC 1343/269	0,10	0,10	0,10	0,10	0,08	0,08	0,51	0,53	0,71	0,56	0,61	0,10	0,10	-	0,27	0,22	0,26	0,76	0,23	0,20	0,26	0,22	0,24	0,13
15-Triploide UFV 557-2	0,22	0,22	0,22	0,22	0,18	0,18	0,54	0,57	0,71	0,57	0,66	0,27	0,27	0,24	-	0,13	0,13	0,78	0,31	0,20	0,22	0,22	0,24	0,22
16-Triploide UFV 557-3	0,26	0,26	0,26	0,26	0,22	0,22	0,57	0,60	0,73	0,59	0,69	0,31	0,31	0,28	0,12	-	0,05	0,77	0,26	0,18	0,21	0,21	0,23	0,16
17-Triploide UFV 557-4	0,22	0,22	0,22	0,22	0,18	0,18	0,54	0,57	0,71	0,57	0,66	0,27	0,27	0,24	0,00	0,12	-	0,78	0,29	0,22	0,24	0,24	0,27	0,20
18- <i>Coffea racemosa</i>	0,45	0,45	0,45	0,45	0,42	0,42	0,61	0,60	0,77	0,60	0,72	0,49	0,49	0,47	0,38	0,36	0,38	-	0,73	0,74	0,74	0,74	0,73	0,74
19-Catiguá MG2	0,31	0,31	0,31	0,31	0,23	0,23	0,58	0,60	0,67	0,58	0,64	0,33	0,33	0,30	0,33	0,34	0,33	0,60	-	0,23	0,21	0,21	0,24	0,16
20-IAPAR 59	0,18	0,18	0,18	0,18	0,08	0,08	0,53	0,53	0,67	0,55	0,61	0,21	0,21	0,16	0,25	0,29	0,25	0,49	0,17	-	0,08	0,13	0,16	0,18
21-Oeiras MG6851	0,10	0,10	0,10	0,10	0,00	0,00	0,51	0,53	0,67	0,53	0,61	0,13	0,13	0,08	0,18	0,22	0,18	0,42	0,23	0,08	-	0,11	0,14	0,21
22-Sacramento MG1	0,10	0,10	0,10	0,10	0,04	0,04	0,50	0,55	0,66	0,55	0,60	0,09	0,09	0,08	0,20	0,25	0,20	0,44	0,26	0,12	0,04	-	0,08	0,16
23-Catucai Amarelo 2SL	0,08	0,08	0,08	0,08	0,02	0,02	0,52	0,54	0,68	0,54	0,62	0,11	0,11	0,06	0,19	0,23	0,19	0,43	0,25	0,10	0,02	0,02	-	0,18
24-Obatã Amarelo IAC 4932	0,08	0,08	0,08	0,08	0,02	0,02	0,52	0,54	0,68	0,54	0,62	0,11	0,11	0,06	0,19	0,23	0,19	0,43	0,25	0,10	0,02	0,02	0,00	-

Tabela 7. Sumário dos melhores locos diferenciadores obtidos pela análise discriminante stepwise, para as duas classes de marcadores SSR entre as populações estudadas.

Ordem	Loco	Alelo	R ²	Prob > F
SSR genômico				
1	SSRCa 003	a2	1,00	<0,0001
2	SSRCa 003	a1	1,00	<0,0001
3	SSRCa 020	a2	1,00	<0,0001
4	SSRCa 083	a2	0,62	0,0002
5	SSRCa 087	a5	0,51	0,0024
6	SSRCa 088	a3	0,44	0,0096
7	SSRCa 087	a1	0,40	0,0089
8	SSRCa 016	a1	1,00	<0,0001
9	SSRCa 087	a7	0,20	0,0324
10	SSRCa 087	a4	0,33	0,0242
11	SSRCa 082	a2	0,30	0,0305
12	SSRCa 094	a4	0,40	0,0204
13	SSRCa 091	a9	0,40	0,0273
14	SSRCa 094	a2	0,39	0,0393
15	SSRCa 087	a3	0,73	0,0018
16	SSRCa 068	a2	0,40	0,0200
17	SSRCa 091	a8	1,00	<0,0001
EST-SSR				
1	EST-SSR 007	a5	1,00	<0,0001
2	EST-SSR 005	a2	1,00	<0,0001
3	EST-SSR 007	a2	1,00	<0,0001
4	EST-SSR 025	a1	0,67	<0,0001
5	EST-SSR 047	a6	1,00	<0,0001
6	EST-SSR 007	a8	0,50	0,0007
7	EST-SSR 007	a9	0,30	0,0450
8	EST-SSR 010	a3	0,44	0,0035

Tabela 8. Marcadores SSR polimórficos para cada população.

População	SSR genômico (SSRCa)*	Número (%)**	EST-SSR	Número (%)**
Ca	018, 068, 082, 087, 091, 094	6 (33%)	029, 007	2 (12%)
Cc	003, 016, 018, 019, 026, 052, 068, 081, 082, 083, 087, 088, 091, 092, 094, 095	16 (89%)	025, 027, 029, 023, 047, 048, 054, 055, 057, 058, 010, 012, 013, 069, 073	15 (88%)
HT	019, 052, 068, 081, 082, 083, 091, 092	8 (44%)	025, 027, 029, 058	4 (23%)
T	018, 020, 088	3 (17%)	027, 054, 010, 073	4 (23%)
V	018, 082, 087, 091, 094	5 (28%)	029, 023, 007, 010, 012, 013	6 (35%)

**primers* que permitiram distinguir pelo menos um acesso dos demais; **valores entre parêntese correspondem a porcentagem do total de *primers*; Ca: *C. arabica*; Cc: *C. canephora*; HT: Híbrido de Timor; T: Triplóides; V: Variedades resistentes à ferrugem.

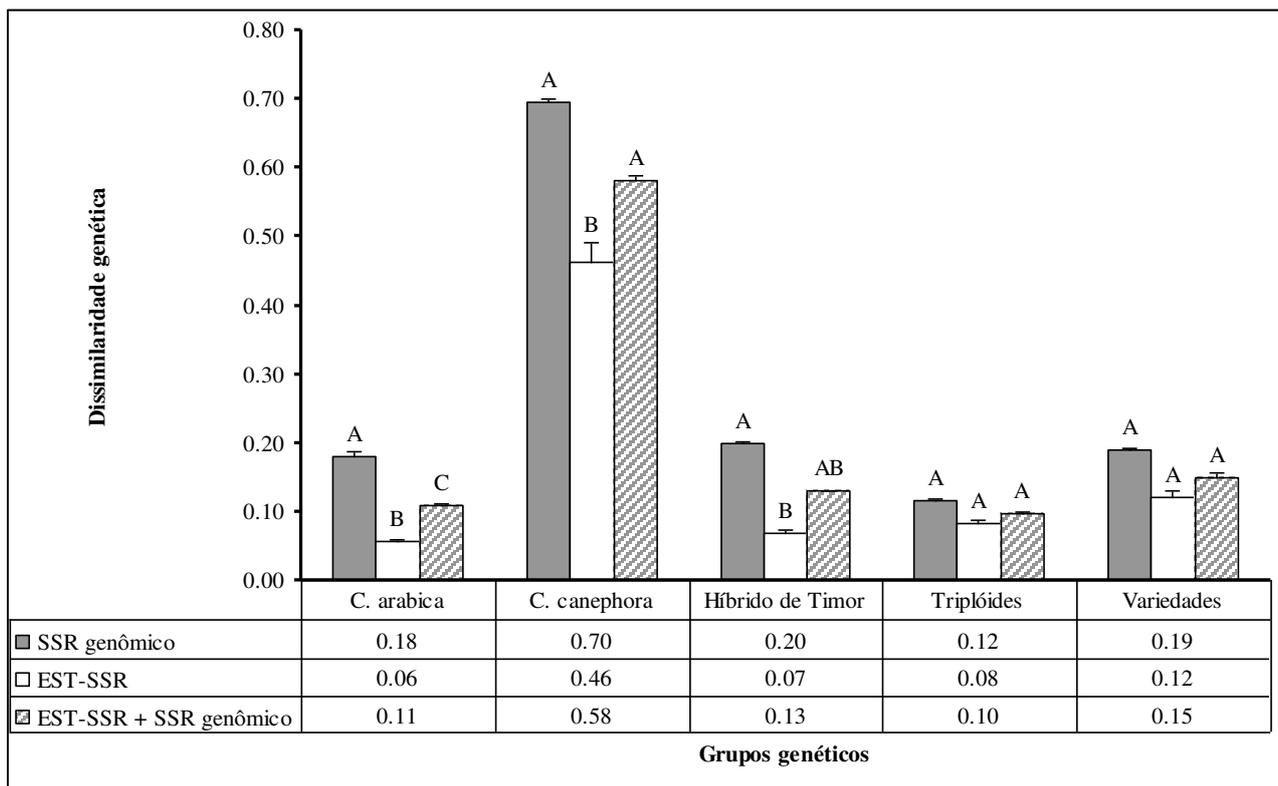


Figura 1. Coeficiente médio de dissimilaridade genética dentro de cada grupo genético (barras verticais acima das colunas correspondem aos valores do desvio padrão).

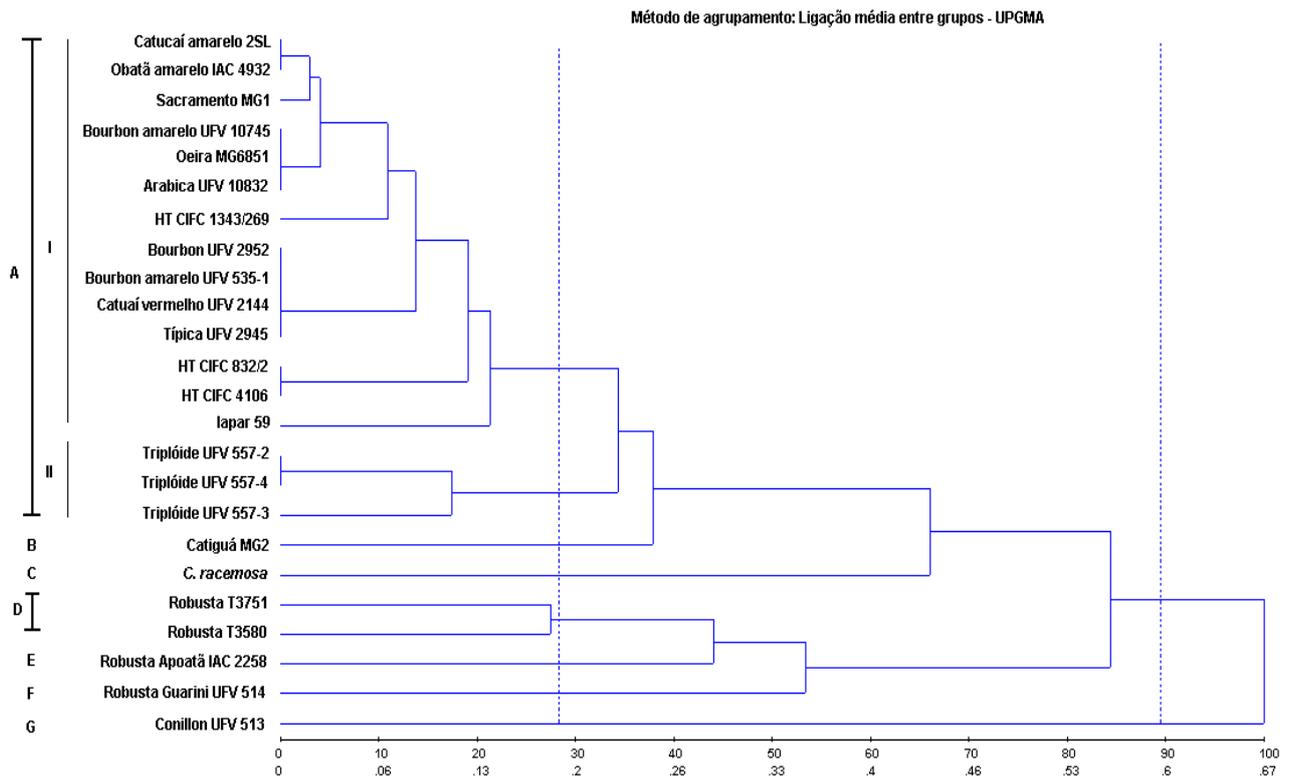


Figura 2. Dendrograma obtido por meio do método UPGMA a partir das distâncias genéticas expressas em complementos de Jaccard, estimadas entre os 24 acessos de cafeeiro, baseadas em 17 marcadores EST-SSR (correlação cofenética = 0,98).

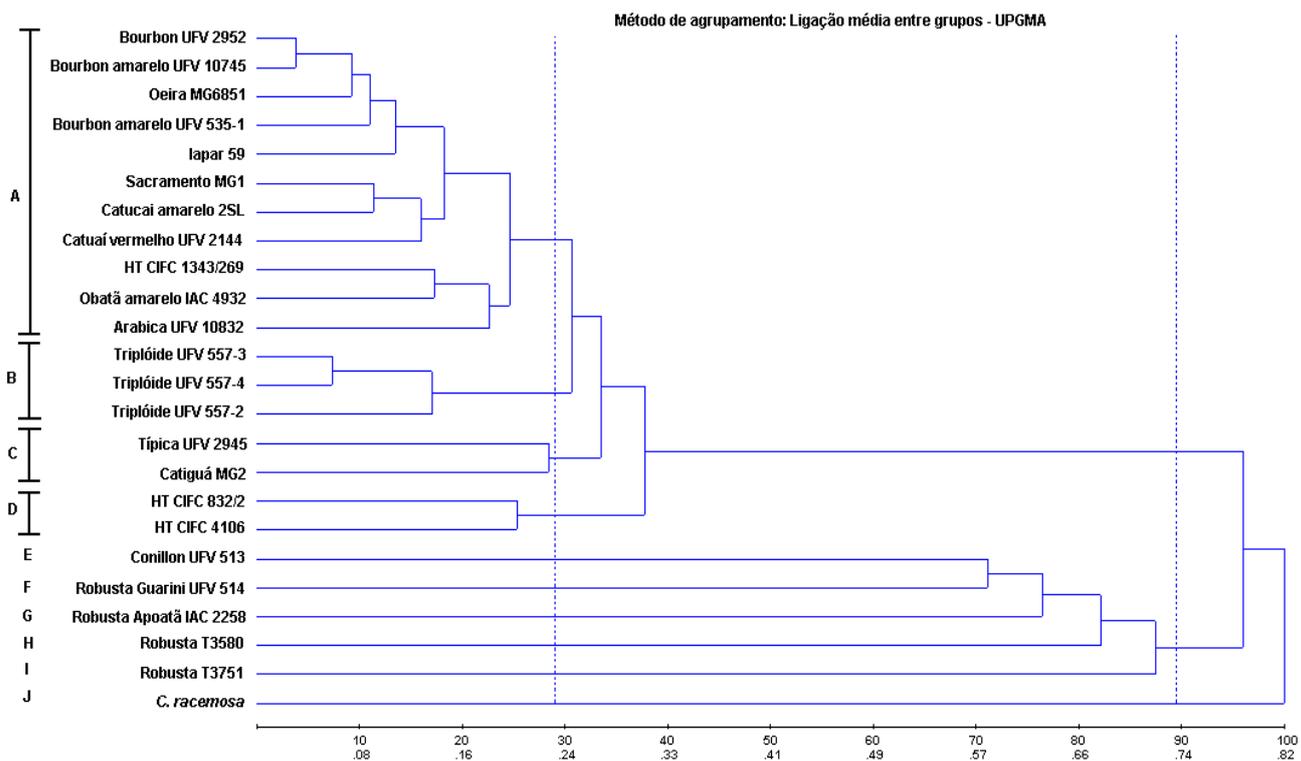


Figura 3. Dendrograma obtido por meio do método UPGMA a partir das distâncias genéticas expressas em complementos de Jaccard, estimadas entre os 24 acessos de cafeeiro, baseadas em 18 marcadores SSR genômico (correlação cofenética = 0,99).

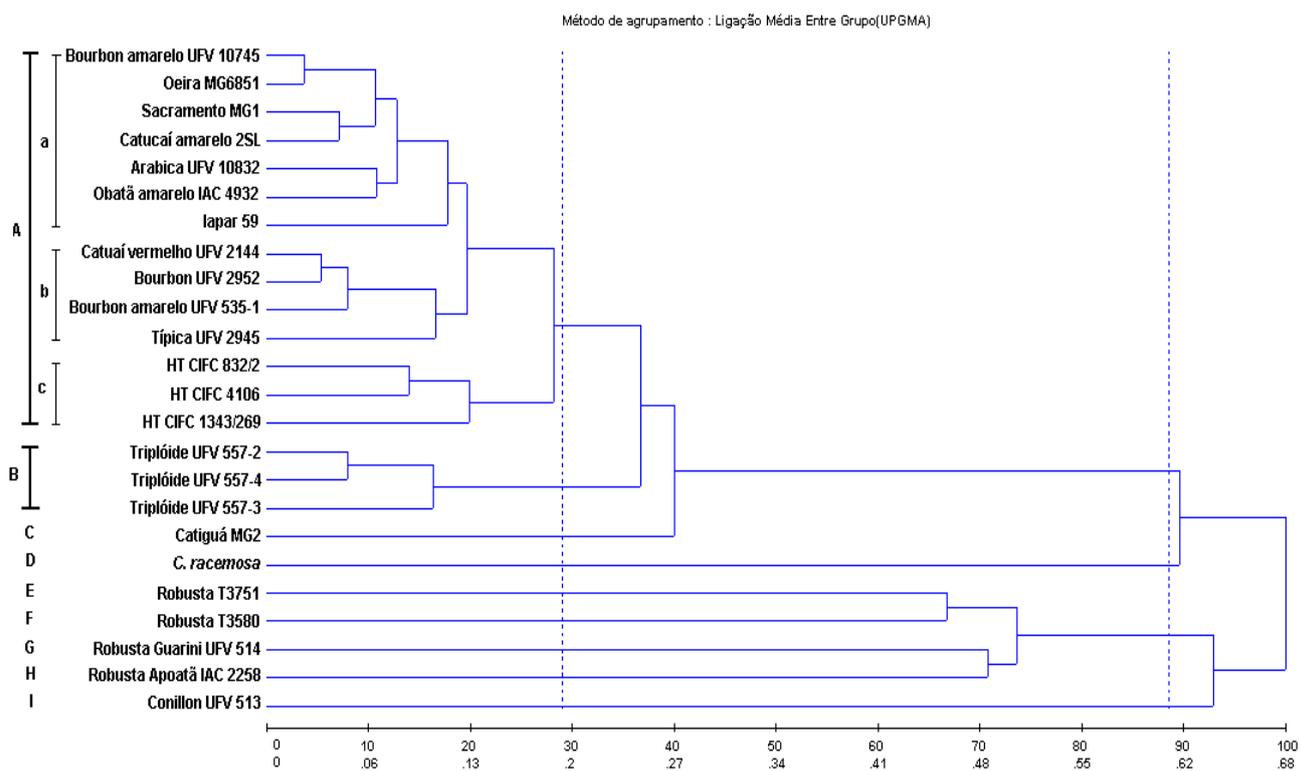


Figura 4. Dendrograma obtido por meio do método UPGMA a partir das distâncias genéticas expressas em complementos de Jaccard, estimadas entre os 24 acessos de cafeeiro, baseadas em 17 marcadores EST-SSR e 18 SSR genômico (correlação cofenética = 0,99).

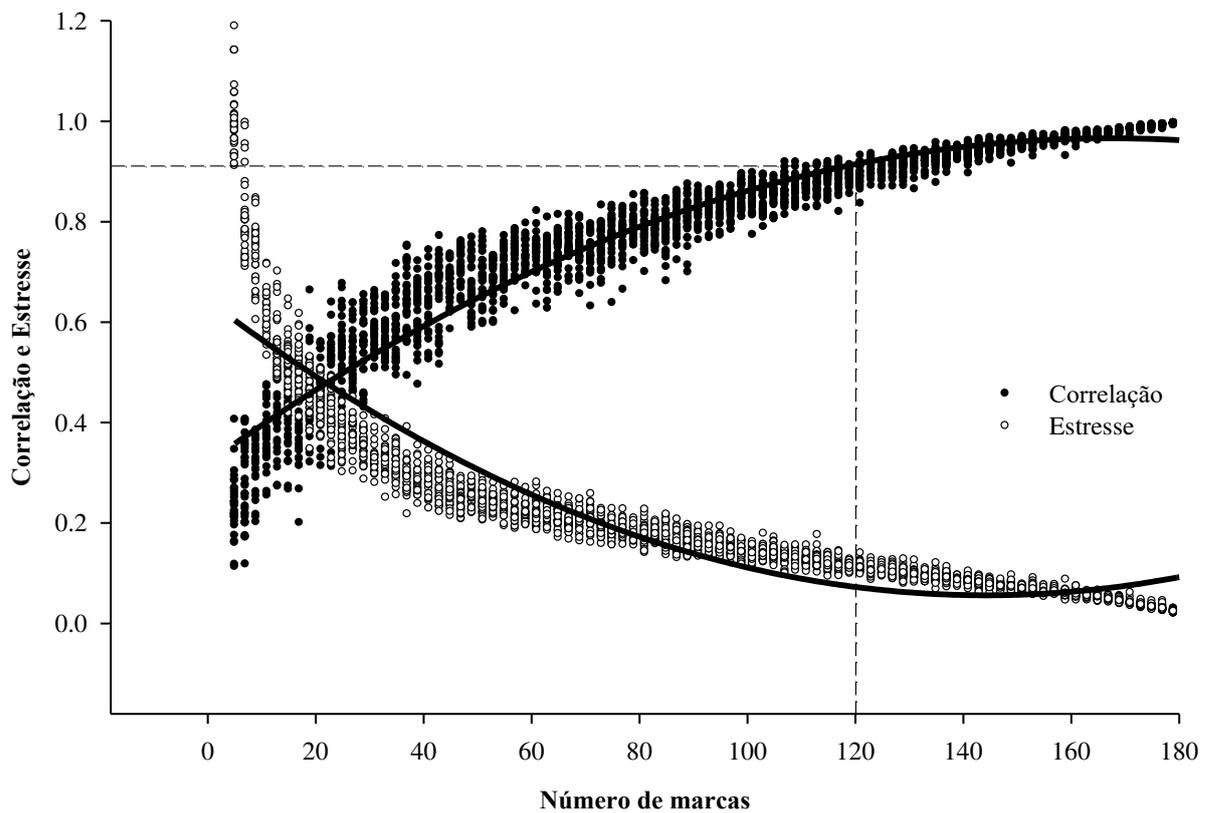


Figura 5. Número ótimo de marcas (alelos ou bandas) SSR para acessar a diversidade genética entre os 24 acessos de *Coffea*. Os pontos referem-se a projeção da correlação e estresse estimados entre as M_i marcas amostradas ($5 \leq M_i \leq 183$) e a matriz original com M marcas ($M = 183$).