

VALIDAÇÃO DE MARCADORES MICROSSATÉLITES E DIVERSIDADE GENÉTICA DE GENÓTIPOS DE *Coffea arabica* PROVENIENTES DA ETIÓPIA¹

Bruna Silvestre Rodrigues da Silva²; Sandra Bellodi Cação³; Rafaelle Vecchia Ferreira⁴; Maria Aparecida Santos⁵; Suzana Tiemi Ivamoto⁶; Juliana Costa Silva⁷; Leonardo Godoy Androcioli⁸; Douglas Silva Domingues⁹; Thierry Leroy¹⁰; Pierre Charmetant¹¹; Gustavo Sera¹²; Luiz Filipe Protasio Pereira¹³

¹Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café

²Bolsista Consórcio Pesquisa Café, Londrina – PR, brunasilvestrerodrigues@hotmail.com

³Pesquisadora, Dr, TMG, Londrina – PR, sandracacao@sercomtel.com.br

⁴Doutoranda de Genética e Biologia Molecular – UEL, Londrina – PR, rafaelle.ferreira@gmail.com

⁵Pesquisadora, Dr, Pioneer, Planaltina–GO, mariaaparecida.ciapsan@gmail.com

⁶Doutoranda de Genética e Biologia Molecular – UEL, Londrina – PR, suzanatiemi@yahoo.com.br

⁷Mestranda em Bioinformática, UTFPR, Cornélio Procópio – PR, jujucostasilvaanalise@gmail.com

⁸Mestre em Agricultura Conservacionista, IAPAR, Londrina – PR, leonardoandrocioli@hotmail.com

⁹Pesquisador, Dr, Unesp, Rio Claro – SP, doug@rc.unesp.br

¹⁰Pesquisador, Dr, CIRAD, Montpellier – FRA, thierry.leroy@cirad.fr

¹¹Pesquisador, Dr, CIRAD, Montpellier – FRA, pierre.charmetant@cirad.fr

¹²Pesquisador, Dr, IAPAR, Londrina – PR, gustavosera@iapar.br

¹³Pesquisador, Dr, Embrapa Café, Brasília – DF, filipe.pereira@embrapa.br

RESUMO: O café é uma valiosa *commodity* agrícola, sendo sua bebida uma das mais populares do mundo. Apesar do sucesso dos programas de melhoramento em *Coffea arabica*, o desenvolvimento de novas cultivares é demorado, podendo levar mais de 20 anos. Além disso, a estreita base genética de *C. arabica* pode dificultar a obtenção de cultivares resistentes a pragas e a doenças ou tolerantes a estresses abióticos. Os acessos selvagens de *C. arabica* geneticamente distantes das variedades cultivadas proporcionam novos alelos para enriquecer sua variabilidade genética. O objetivo deste trabalho foi validar marcadores moleculares do tipo SSR em um painel de 24 acessos de *C. arabica* da coleção da Etiópia, realizar um estudo da diversidade, estrutura e seleção de genótipos para futuros trabalhos de introgressão no melhoramento. Neste trabalho foram utilizados 58 marcadores microsatélites (SSR) obtidos de análises *in silico* a partir de dados de RNA-seq da cultivar Iapar 59 e de sequências de BACs de Híbrido de Timor 832/2, e 77 SSR obtidos de uma extensa revisão bibliográfica em *C. arabica* e/ou *C. canephora*. Dos 135 SSR utilizados, 122 resultaram em produtos de amplificação e 31 foram polimórficos no painel. Desses, 20 SSR foram altamente informativos com valores de PIC acima de 0,7, e sete são descritos pela primeira vez. Um total de 165 alelos foram obtidos com média de 5,3 alelos por loco. O estudo de diversidade e estrutura genética dos acessos separou os genótipos em três principais grupos. A utilização de marcadores SSR informativos assim como o estudo de diversidade e estrutura genética destes e dos demais acessos permitirá a criação de uma coleção nuclear, facilitando a conservação, acessibilidade e uso de recursos genéticos de cafeeiros da Etiópia.

PALAVRAS-CHAVE: Café, Marcadores, SSR, RNA-seq, BAC.

MICROSATELLITE MARKERS VALIDATION AND GENETIC DIVERSITY OF *Coffea arabica* GENOTYPES FROM ETHIOPIA

ABSTRACT: The Coffee is a valuable agricultural *commodity*, being your beverage one of the most popular in the world. Despite the success of breeding programs in *Coffea arabica*, the development of new cultivars is time consuming, may take more than 20 years. Furthermore, the narrow genetic base of *C. arabica* can difficult the obtaining of cultivars resistant to pests and diseases or tolerant to abiotic stresses. The wild accessions of *C. arabica* genetically distant of the cultivated varieties can provide new alleles to enrich your genetic variability. The objective of this study was to validate molecular markers SSR type on a panel of 24 *C. arabica* access of the collection of Ethiopia, to conduct a study of the diversity, structure and selection of genotypes for future introgression work on improving. In this work were used 58 microsatellite markers (SSR) obtained *in silico* analysis from RNA-seq data cultivar Iapar 59 and BACs sequences of Timor Hybrid 832/2, and 77 SSR obtained from an extensive literature review in *C. arabica* and / or *C. canephora*. Of the 135 SSR used, 122 resulted in amplification products and 31 were polymorphic in the panel. Of these, 20 SSR were highly informative with values PIC above 0.7, and seven are first described in the literature. A total of 165 alleles were obtained with an average of 5.3 alleles by locus. The study and genetic diversity of the structure of the access separated the genotypes into three main groups. The use of informative SSR markers and the study of genetic

diversity and structure of these and other access will allow the creation of a core collection, facilitating conservation, accessibility and use of genetic resources of coffee in Ethiopia.

KEYWORDS: Coffee, markers, SSR, RNA-seq, BAC.

INTRODUÇÃO

O café é um dos principais produtos agrícolas mundiais, com cultivo em mais de 80 países e uma importante fonte de renda na Ásia, África, e América Latina. O cultivo em sua maior parte é de *Coffea arabica*, representando 65% a 70% da produção mundial, sendo considerado como um café de qualidade superior, aroma agradável e sabores diversificados. O sucesso da produção agrícola depende fortemente do uso de cultivares com qualidades diferenciadas e adaptadas a diferentes ambientes de cultivo. O melhoramento de plantas tem um papel fundamental no desenvolvimento da agricultura ao gerar novas variedades em espécies de interesse agrônomo, tolerantes a doenças, pragas e a estresses abióticos que são fatores restritivos ao alcance do máximo potencial produtivo de uma cultivar. Para tanto, a utilização de ferramentas biotecnológicas como os marcadores moleculares permite a caracterização genética de espécies de interesse agrônomo e a escolha precoce de características com menor tempo e custo. Dentre os marcadores atualmente disponíveis, *Simple Sequence Repeats* (SSR) ou microssatélites são sequências de DNA compostas por repetições curtas em tandem em um dado loco. Essa classe de marcadores tem sido amplamente utilizada devido à simplicidade da técnica utilizada, rapidez, grande poder de resolução e altos níveis de polimorfismo. Além disso, são amplamente dispersos nos genomas de plantas possibilitando uma discriminação precisa mesmo entre indivíduos geneticamente relacionados. Estudos sobre a caracterização do germoplasma de *Coffea spp*, incluindo análises com marcadores moleculares, relataram a baixa diversidade genética de *C. arabica*. As justificativas para a estreita base genética e diversidade molecular da espécie incluem a sua recente origem, sua reprodução preferencialmente autógama e ao seu histórico de dispersão limitado a poucos indivíduos. Hoje a maioria das cultivares resultam de apenas duas bases populacionais, *Typica* e *Bourbon*. Assim, mesmo com as vantagens dos marcadores SSR, poucos foram descritos até o momento para *C. arabica*, sendo importante expandir em número esta classe de marcador. As coleções de acessos de *C. arabica* da Etiópia (Centro de origem da espécie) possuem maior variabilidade fenotípica para tamanho de folhas, altura, concentração de cafeína, açúcares e diterpenos. Devido à importância dos acessos do centro primário de origem da espécie, o Instituto Agrônomo do Paraná (IAPAR) possui atualmente 132 acessos de *C. arabica* provenientes da Etiópia, sendo a maioria da região Oeste do Vale do Rift. Este trabalho teve como objetivo a validação de marcadores SSR obtidos de revisão bibliográfica e de análise *in silico* em dados de RNA-seq de Iapar 59 e de sequências de BACs de Híbrido de Timor 832/2 em um painel de genótipos representativos da coleção da Etiópia presente no IAPAR, assim como realizar o estudo de diversidade e estrutura genética desse painel.

MATERIAL E MÉTODOS

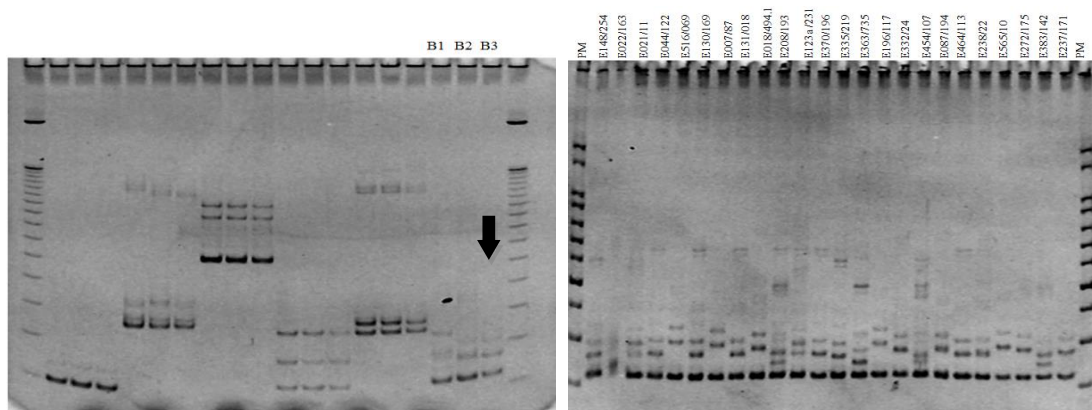
Para a validação dos SSR, foram utilizados 58 marcadores SSR obtidos a partir da análise *in silico* de dados de sequenciamento de *C. arabica* cv. Iapar 59 (RNA-seq) e Híbrido de Timor 832/2 (sequenciamento de clones BACs) juntamente com 77 marcadores provenientes de revisão bibliográfica. Amostras de sementes de plantas foram enviadas inicialmente ao Catie na Costa Rica, que repassou parte da coleção ao Instituto Agrônomo de Campinas (IAC). Posteriormente o IAC forneceu 132 amostras de sementes ao IAPAR. A validação foi realizada em um painel de 24 acessos representativos da diversidade existente na coleção de 132 acessos de *C. arabica*. Os 24 acessos selecionados no estudo foram coletados em várias regiões da Etiópia, sendo nativos das províncias de Kaffa, Illubador, Gojjam do lado Oeste do Vale do Rift, e Sidamo e Harar do lado Leste do Vale do Rift. O DNA genômico foi isolado a partir de folhas de plantas adultas pelo método CTAB. Após a extração a qualidade do DNA foi verificada em gel de agarose 0,8%, corado com brometo de etídeo e a quantificação estimada por espectrofotometria em NanoDrop com absorvância de 260 e 280 nm. Para diminuir os custos e o tempo das análises, foi utilizada a estratégia de *bulks* ou *pools* de DNA, uma adaptação da técnica de BSA (*Bulk Segregant Analysis*), e 15 amostras de DNA foram divididas em três *bulks* com base nos dados de origem geográfica. Para cada marcador SSR foram realizadas três reações de amplificação, onde quantidades equimolares de DNA dos acessos de cada grupo foram combinados a uma concentração final de 100 ng/ μ L. Os marcadores que apresentaram padrão polimórfico entre os grupos foram avaliados no painel de 24 genótipos a fim de genotipar esta população. Os acessos individuais foram diluídos para uma concentração final de 5 ng/ μ L. Os produtos de PCR foram submetidos à eletroforese em gel de poliacrilamida 10%, corados por brometo de etídeo, e os géis capturados pelo sistema digital Kodak KS 120. O tamanho molecular dos produtos amplificados foi estimado utilizando uma escala de DNA de 50 pb como um padrão (*Ludwig Biotec*). A informação genética de cada loco SSR foi avaliada pelo número de alelos por loco, frequência alélica por loco e PIC (*Polymorphism Information Content* - PIC). Os dados alélicos dos marcadores foram utilizados para determinar a diversidade genética e as relações filogenéticas entre os 24 acessos testados. Apesar dos microssatélites serem codominantes, a análise molecular de cada acesso foi baseado na presença/ausência de cada fragmento amplificado devido à natureza tetraplóide de *C. arabica*, sendo os dados utilizados para a construção de uma matriz binária que demonstra o perfil dos SSR em cada genótipo. Para

realizar a análise de diversidade genética, uma matriz de dissimilaridade foi utilizada para a construção de um dendograma de diversidade global utilizando o método de agrupamento (UPGMA) implementado no *software* DARwin 5. A Análise de Coordenadas Principais (PCO) baseada nas distâncias genéticas de Nei. (1978) foi realizada pelo programa FAMD 1.31. Foi utilizada também uma abordagem de agrupamento *bayesiano* descrito por Pritchard et al. (2000), implementado no *software* STRUCTURE 2.3.4. O STRUCTURE foi executado utilizando o método estatístico ΔK , inferindo $K=5$ para a análise.

RESULTADOS E DISCUSSÃO

1. SELEÇÃO DOS *PRIMERS* POLIMÓRFICOS

Os 58 pares de *primers* obtidos da análise *in silico* mais os 77 da revisão de literatura foram analisados em 15 acessos da coleção da Etiópia divididos em três *bulks*. Quando ocorria a visualização de alelos polimórficos entre os *bulks*, os SSR eram posteriormente avaliados no painel de 24 acessos selecionados no estudo (Figura 1).



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

Figura 1: Gel de Poliacrilamida revelando o polimorfismo dentro de *Coffea arabica*. A - Identificação de marcadores polimórficos pela estratégia de *bulks* de DNA (B1, B2 e B3 = grupos). A seta indica o polimorfismo revelado pelo marcador **Ca61**. B - Perfil de amplificação do marcador **Ca61** em genótipos individuais.

Dos 135 *primers* SSR desenhados, 31 *primers* foram polimórficos com padrões de bandas reprodutíveis. Dos 77 *primers* SSR desenhados a partir da revisão bibliográfica, 47 foram monomórficos e 24 foram polimórficos no painel de 24 acessos da Etiópia, na análise pela estratégia de *bulks* de DNA. Para os 37 *primers* SSR desenhados para CHT 832/2, 22 *primers* foram monomórficos e 6 foram polimórficos. E dos 21 SSR obtidos e testados em *bulks* na cultivar Iapar 59 apenas o *primer* (Clapar 14) demonstrou-se polimórfico no painel (Tabela 1).

Tabela 1. Número de *primers* SSR polimórficos e alelos nos acessos da Etiópia.

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

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

O sucesso desta estratégia é observado em outras culturas como a soja, onde Rabel et al. (2010) demonstraram que as análises em “*bulks*” reduziram em cinco vezes o número das reações necessárias para a realização da genotipagem com

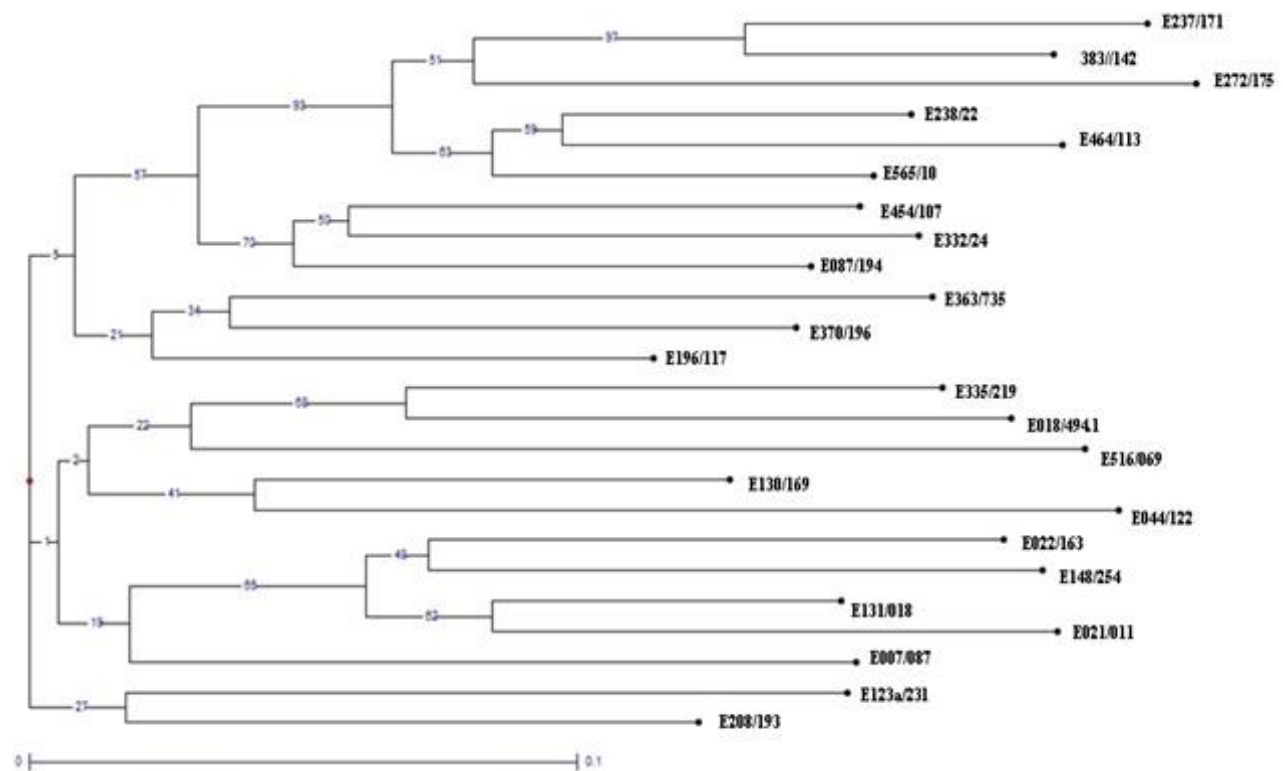
uma conseqüente diminuição de custos, tempo e reagentes. Ferreira, 2010 em um estudo anterior em nosso laboratório utilizou 292 *primers* em um painel de acessos da Etiópia, e destes, 48 marcadores microssatélites polimórficos foram obtidos por esta estratégia. Quatorze SSR apresentaram 2 a 5 alelos em genótipos que representam parte da diversidade da coleção da Etiópia.

1.1 ANÁLISE DA DIVERSIDADE GENÉTICA DOS 24 ACESSOS DA ETIÓPIA

Após a análise dos *bulks*, os 31 SSR polimórficos foram analisados no painel de 24 genótipos da coleção da Etiópia. Foi observado um total de 165 alelos com 2 a 11 alelos por loco detectado, e uma média de 5,3 alelos SSR. As frequências alélicas variaram de 0,01 a 0,8 com 78 dos alelos (47,3%) com frequências entre 0,2 a 0,8, e 87 (52,7%) dos alelos considerados raros pela baixa frequência (0,2). Do total dos alelos identificados, um alelo obteve alta frequência (0,8). Os valores de PIC variaram de 0,3 (CHT25) a 0,8 (M24), com média de 0,7. Dos 31 *primers*, 20 apresentaram valores de PIC superior a 0,7 demonstrando ser informativos para estudos de diversidade genética. Essas análises permitem identificar entre os marcadores testados, aqueles geneticamente mais informativos para testes de identificação genética. Um ponto importante para os marcadores polimórficos obtidos da análise *in silico* foi a origem dos dados, onde os SSR desenvolvidos para o genótipo CHT 832/2 foram de sequenciamento de BACs genômicos, enquanto que para Iapar 59 foram analisados dados de transcriptoma (RNA-seq). Assim, neste estudo seis marcadores SSR de CHT 832/2 foram validados no painel e apenas um marcador SSR obtido de Iapar 59 apresentou-se polimórfico. Portanto um nível inferior de polimorfismo observado nos SSR obtidos de dados de transcriptoma pode ocorrer devido a alta conservação de seqüências de DNA em regiões transcritas no genoma, o que não ocorre com os SSR genômicos, que são amplamente distribuídos, principalmente em regiões intergênicas.

1.2 ESTRUTURA GENÉTICA DOS ACESSOS

A ancestrabilidade entre os acessos observados através do dendograma de diversidade global (Figura 2) gerou três principais grupos que se distinguiram entre os acessos tetraplóides, fato corroborado pela análise da estrutura populacional ($K=5$) (Figura 3) e pela PCO (Figura 4).



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

Figura 2. Dendrograma de diversidade global entre os 24 acessos de *C. arabica* com base no coeficiente de Dice utilizando o método UPGMA.

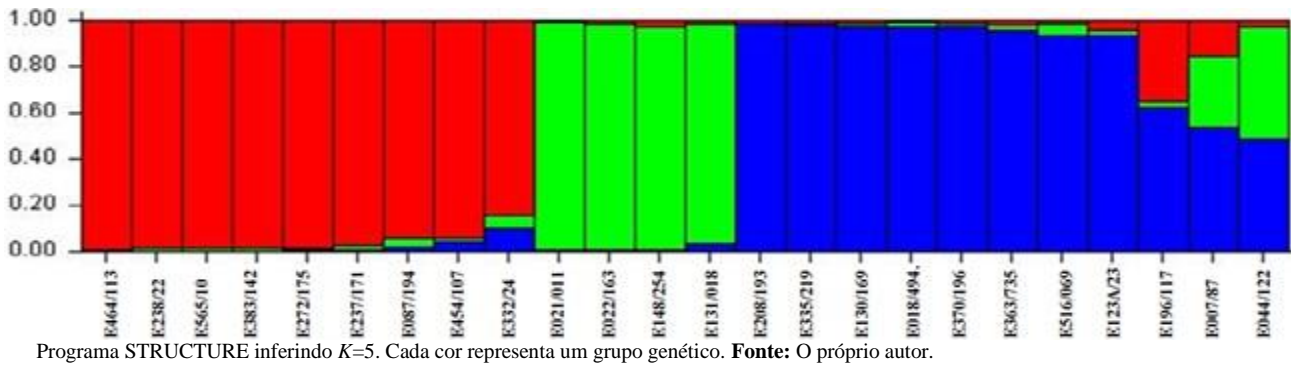


Figura 3. Gráfico de estrutura genética a partir dos dados dos 31 marcadores SSR.

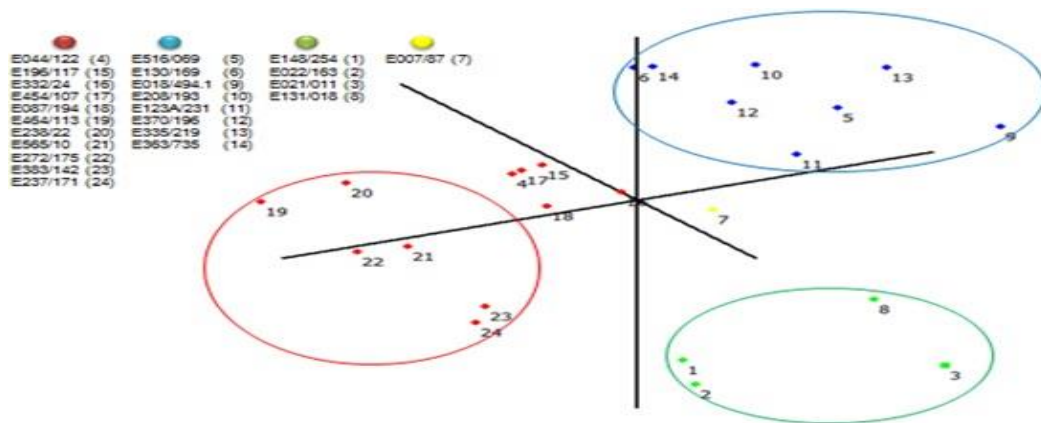


Figura 4. Análise da coordenada principal entre 24 acessos de *C. arabica* provenientes da Etiópia.

O dendograma de diversidade global gerado a partir dos acessos de *C. arabica* do centro de origem reforça a sua considerável diversidade genética, demonstrando que alguns dos acessos da coleção da Etiópia não constituem um grupo geneticamente homogêneo. Silvestrini et al. (2007), em um estudo de diversidade genética em genótipos comerciais e selvagens de *C. arabica*, demonstraram que a diversidade dos acessos da Etiópia foi maior do que as cultivares demonstrando a sua importância como fonte de variabilidade genética para os programas de melhoramento do café. Nesse estudo, a maioria dos acessos do lado Leste da Etiópia ficaram no grupo 2 (E018/494.1, E022/163, E021/11, E007/87), porém não no mesmo subgrupo e o genótipo E238/22 e 237/171 dentro do grupo 3, não havendo uma partição clara entre acessos de ambos os lados do Vale do Rift. Do total dos acessos, 20 possuem ancestralidade semelhante, porém os acessos E007/087, E196/117, E044/122 e E332/24 demonstraram assim como na PCO possuir mistura de genótipos ancestrais de ambos os três agrupamentos. Já o acesso E007/087, o único da província de Harar, demonstrou ser o mais diverso entre os estudados e o que mais compartilha similaridade genética entre os três grupos em todas as análises, sendo um genótipo alvo para posteriores estudos de melhoramento da espécie. Nossos resultados corroboram com os trabalhos de Anthony et al., 2001, Chaparro et al., 2004 e Lopes-Gartner et al., 2009, no qual os acessos originários de províncias próximas foram mais estreitamente relacionados do que os coletados na mesma província, não existindo uma relação muito clara entre os sítios de origem geográfica dos acessos da Etiópia e os agrupamentos resultantes.

CONCLUSÕES

1. Dos 135 loci SSR testados, validou-se 31 *primers* polimórficos pela estratégia de *bulks*, e destes, sete são descritos pela primeira vez na literatura;
2. Com a estratégia de *bulks* de DNA houve uma grande economia de materiais e tempo otimizando as análises dos resultados;
3. A análise da estrutura genética separou os acessos em três principais grupos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANTHONY, F.; BERTRAND, B.; QUIROS, O.; WILCHES, A.; LASHERMES, P.; BERTHAUD, J.; CHARRIER, A. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. *Euphytica*, v.118, p.53-65, 2001.
- CHAPARRO, A. P.; CRISTANCHO, M. A.; CORTINA, H. A AND GAITAN, A. L. Genetic variability of *Coffea arabica* L. accessions from Ethiopia evaluated with RAPDs. *Genetic Resources and Crop Evolution*, 51:291–297, 2004.
- FERREIRA, R.V. Seleção de marcadores microssatélites polimórficos em *Coffea arabica*. [Trabalho de Conclusão de Curso]. Londrina: Universidade Estadual de Londrina, Centro de Ciências Biológicas, 2010.
- LÓPEZ-GARTNER, G.; CORTINA, H.; MC COUCH, SUSAN.R.; MONCADA, M.D.P. Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. *Tree Genetics and Genomes*, v.5, p.435-446, 2009.
- NEI, M. 1978. Estimation of average heterozygosity and genetic distance from small number of individuals. *Genetics* v. 89: p.583-590. 1978.
- POT, D.; SCHOLZ, M. B. S.; LANNES, S. D.; DEL GROSSI, L.; PEREIRA, L. F. P.; VIEIRA, L. G.; SERA, T. Phenotypic analysis of *Coffea arabica* accessions from Ethiopia: contribution to the understanding of *Coffea arabica* diversity. In: 22nd INTERNATIONAL CONFERENCE ON COFFEE SCIENCE, Campinas. *Anais...* Campinas, p.165, 2010.
- PRITCHARD, J. K.; STEPHANS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. *Genetics*, v. 155, p. 945–959, 2000.
- RABEL, M. et al. Marcadores Moleculares Microssatélites na Avaliação de Sementes de Soja com Variação na Coloração do Hilo. *Revista Brasileira de Sementes*, 32, 2, 019-025, 2010.
- SILVESTRINI, M.; JUNQUEIRA, M.G.; FAVARIN, A.C.; GUERREIRO-FILHO, O.; MALUF, M.P.; SILVAROLLA, M.B AND COLOMBO, C.A. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. *Genetic Resources and Crop Evolution*, 54: 1367-1379, 2007.