

NOVOS MARCADORES PARA FINS DE MAPEAMENTO E LOCALIZAÇÃO DE QTLs A PARTIR DE PCR-RFLP DE GENES DO GENOMA CAFÉ BRASILEIRO¹

Carlos Augusto Colombo^{2*}, Paula Yuri Yamamoto^{3a}, Luis Carlos S. Ramos², Paulo Mazzafera⁴, Paulo Boller Gallo⁵, Otávia T. Vilella^{3b}, Sergio D. Lannes^{3c}, David Pol⁶, Lucia P. Ferreira^{3c}, Luiz Gonzaga E. Vieira⁷, Luiz Filipe P. Pereira^{7,8}

¹ Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – CBP&D/Café

² IAC (Instituto Agronômico), Centro de Recursos Genéticos Vegetais, Campinas-SP

^{3a} Bolsista CBP&D/Café, IAC, Campinas, SP

^{3b} Bolsista Iniciação Científica FAPESP, IAC (Instituto Agronômico), Campinas, SP

^{3c} Bolsista CBP&D/Café, IAPR, Londrina, PR

⁴ UNICAMP - Fisiologia Vegetal, Campinas, SP

⁵ APTA-Pólo Regional Nordeste do Estado de São Paulo

⁶ CIRAD, UMR DAP, Montpellier, França

⁷ IAPAR (Instituto Agronômico do Paraná), Londrina, PR

⁸ Embrapa Café, LBI/IAPAR Londrina PR

* autor para correspondência: ccolombo@iac.sp.gov.br

RESUMO: O mapeamento genético é uma das estratégias mais visadas para fins de melhoramento genético e ganha maior importância a partir do surgimento de marcadores do tipo SNPs ou INDELS, facilitado pelos projetos genomas, sobretudo de ESTs. Assim, análises *in silico* de busca de polimorfismo SNPs em sequências relacionadas com qualidade de bebida e derivadas de *C.arabica* e *C.canephora* foram analisadas e o polimorfismo entre as duas espécies validado para seis genes (quatro proteases e dois de sacarose) a partir de estudos de laboratório utilizando a técnica PCR-RFLP. Para tanto, após confirmação do polimorfismo nas duas espécies parentais, foram genotipadas 90 plantas F2 derivadas da autofecundação de um híbrido interespecífico de *Coffea arabica* e *C. canephora* 4x. Após a obtenção dos amplificadores, estes foram digeridos com diversas enzimas de restrição de quatro bases (*Alu I*, *Dde I*, *Eco RI*, *Hae III*, *Mse I* e *Msp*, *Fnu DII*, *Taq I*; *Scr FI*). Dos seis genes analisados, quatro deles apresentaram segregação do tipo 3:1 na população F2 (Cisteína 8, Cisteína 5, β -Fructosidase e Sacarose Fosfato Sintase), demonstrando a utilização dos mesmos para fins de mapeamento.

Palavras-chave: *Coffea spp.*, qualidade de bebida, seleção assistida, enzimas de restrição, ESTs, genes candidatos

NEW MARKERS TO MAPPING AND LOCATION OF QTLs FROM THE BRAZILIAN COFFEE GENOME BASED ON PCR-RFLP

ABSTRACT: The genetic mapping strategy is very targeted for breeding and its importance is growing with the emergence of SNPs or INDELS markers, facilitated by the genome project, especially for ESTs. Thus, *in silico* analysis of polymorphic SNPs in sequences related to cup quality and derive from *C.arabica* and *C.canephora* were analyzed and the polymorphisms between the two species validated for six genes (four proteases and two sucrose) from laboratory studies using PCR-RFLP technique. Thus, after confirmation of polymorphism in both parental species, 90 F2 plants derived from self an interspecific hybrid of *Coffea arabica* and *C. canephora* 4x were genotyped. After obtaining the amplified, they were digested with various restriction enzymes of four bases (*Alu I*, *DDE I*, *Eco RI*, *Hae III*, *Mse I* and *Msp*, *Fnu DII*, *Taq I*, *Scr FI*). Of the six genes analyzed, four of them showed 3:1 segregation in the F2 population (Cysteine8, Cysteine 5, β -Fructosidase and sucrose phosphate synthase), demonstrating the use of them for mapping.

Key words: *Coffea spp.*, cup quality, assisted selection, restriction enzymes, ESTs, candidate genes

INTRODUÇÃO

Mapas genéticos oferecem grande potencial de uso em programas de melhoramento, sobretudo de plantas perenes, cujo ciclo de obtenção de novos cultivares pode chegar a 20 anos, como é o caso do café. O sucesso da construção de mapas depende de diversos fatores, fundamentalmente, do tamanho efetivo e da variabilidade genética da população de estudo, da qualidade e ou quantidade dados fenotípicos e ou das marcas genéticas disponíveis para fins de mapeamento. No caso da construção de mapas genéticos em *Coffea arabica*, uma das principais dificuldades é a baixa diversidade genética presente na espécie (Maluf et al., 2005). No Brasil, esta dificuldade pode e vem sendo contornada por meio da utilização de populações representadas por híbridos interespecíficos, como são os casos de estudo do grupo de pesquisa da UFV e do IAC-IAPAR, com híbridos de Timor e F2 derivados da autofecundação de F1 entre *C.arabica* e *C.canephora*, respectivamente. Em ambos os casos, grande número de marcadores moleculares (RAPD, AFLP e SSR) vem sendo mapeados nas populações de estudo. Mais recentemente, o surgimento dos projetos genoma, sobretudo de ESTs, como é o caso do café (Vieira et al., 2006) tem oferecido o aproveitamento de novos marcadores a partir de

seqüências expressas do genoma. Em café, o aroma pode formado por mais de 800 compostos voláteis e os precursores relacionados com a formação dos mesmos são: açúcares, aminoácidos, peptídeos, trigonelina, ácidos clorogênicos, ácidos orgânicos, lipídeos e carotenóides (Clifford, 1985). Interações entre açúcares redutores e o grupamento amino de aminoácidos e peptídeos, ou mesmo de proteínas (reação de Maillard) é tida como reação essencial para o desenvolvimento de aroma. O termo genérico protease inclui enzimas que degradam proteínas intactas, que devem ser chamadas de proteinases ou endopeptidases, e aquelas que degradam ligações peptídicas específicas, ou seja, que necessitam de uma posição terminal para gerar a especificidade, que devem ser denominadas exopeptidases (Barret, 1986). Estas últimas poderiam ser divididas ainda em carboxi ou amino peptidases, dependendo da posição terminal gerando a especificidade (Brzin & Kidric, 1995). Pesquisas preliminares têm revelado que cafés colhidos de plantas crescendo em regiões com acentuada diferença na temperatura média anual apresentam diferenças marcantes na qualidade da bebida e que, dentro os diversos compostos químicos do grão que revelaram variação, a atividade de proteases explicam grande parte da variância observada. Assim, o propósito do presente trabalho é apresentar marcadores do tipo SNPs e ou INDELS para fins de mapeamento em café, derivados de genes envolvidos em vias metabólicas associadas à qualidade de bebida, em especial, genes do metabolismo de açúcares (Geromel et al., 2006) e proteases (Shimizu et al., 2000). SNPs (Single Nucleotide Polymorphisms) e INDELS (INsertion / DEletion) são marcadores moleculares com capacidade de diferenciar indivíduos por meio de variações em apenas um nucleotídeo de seqüências de DNA que podem ou não codificar genes. Marcadores de variação nucleotídica têm sido considerados uma das melhores opções para utilização em mapeamento genético, genotipagem, tendo também um grande potencial para o melhoramento por seleção assistida (MAS).

MATERIAL E MÉTODOS

Os genes escolhidos para estudo foram obtidos da base de dados ESTs do genoma café (Vieira et al., 2006) a partir de análises *in silico*. Foram desenhados *primers* para amplificação de três genes que codificam para Cisteína Proteinase (EC 3.4.22 - *blast hits*: gil2944446; gil18414611; gil17979125), um Aspártico (EC 3.4.23 - *blast hit*: gil82623417), uma β -Fructosidase/Invertase (EC 3.2.1.26 - *blast hit*: gil551259), uma Sacarose Fosfato Sintase (EC 2.4.1.14 - *blast hit*: gil77176831) e uma Sacarose Sintase (EC 2.4.1.13 - *blast hit*: gil115310620). Para a análise de segregação do polimorfismo de restrição desses genes, foram coletadas folhas na população mapa (F_2) no município de Mococa (SP) bem como dos genitores e híbrido F_1 (Campinas). As folhas foram liofilizadas e o DNA total extraído, quantificado, amplificado por PCR e digerido por enzimas de restrição. Os fragmentos foram analisados em gel de agarose (1,5%), após eletroforese e coloração com brometo de etídeo. Os indivíduos foram genotipados na forma de presença ou ausência (1 e 0) de fragmentos, permitindo a geração de uma matriz de dados para posteriores análises estatísticas de mapeamento genético, que se encontram em execução no presente momento.

RESULTADOS E DISCUSSÃO

Foram detectados polimorfismos *in silico* entre *C. arabica* e *C. canephora* para quatro genes de cisteína e um aspártico. Sítios de restrição dentro dos locos foram identificados *in silico*, incluindo sítios reconhecidos por enzimas que revelam polimorfismo entre as duas espécies genitoras do mapa genético, *C. arabica* e *C. canephora* (Tabela 1). Com exceção da enzima *EcoRI*, que apresenta apenas uma região de corte no gene cisteína 5, as demais enzimas possuem vários sítios de restrição em todos os genes. A enzima *Alu I*, por exemplo, apresenta oito sítios de reconhecimento na espécie *C. arabica* contra sete na espécie *C. canephora*, tanto para o gene cisteína 5 quanto para o aspártico 1. Da mesma forma, o número de fragmentos esperados a partir da digestão com a enzima *DdeI* é de cinco em *C. arabica* contra quatro em *C. canephora*. Os polimorfismos destes locos foram testados em laboratório e podem ser vistos na Figura 1. Os produtos amplificados correspondem ao tamanho esperado nos casos em que a amplificação foi realizada a partir de cDNA, porém, a partir de DNA total os fragmentos apresentam tamanho superior ao esperado, configurando a presença de introns nesses genes e revelando uma correspondência possivelmente não exata entre os sítios de restrição esperados e os obtidos. Para cada loco de protease, foi determinada a enzima que de forma mais evidente revelou polimorfismo de restrição entre os parentais para fins de genotipagem da população mapa (F_2). As buscas de polimorfismos dentro dos genes de açúcares foram feitas com cinco a oito enzimas de restrição, nos genitores e no híbrido F_1 , revelando sítios polimórficos com pelo menos uma delas em cada loco e permitindo, assim, a genotipagem de toda população mapa (Figura 2):

Cisteína 8 – Enzimas testadas *Alu I*, *Dde I*, *Eco RI*, *Hae III*, *Mse I* e *Msp I*; Genotipagem com *Hae III*; Fragmentos polimórficos 750 e 1000 pb; Segregação possível 9:1 e 3:1 (respectivamente).

Cisteína 22 – Enzimas testadas *Alu I*, *Dde I*, *Eco RI*, *Hae III*, *Mse I* e *Msp I*; Genotipagem com *Msp I*; Fragmento Polimórfico 600 pb (ou 1750); Segregação possível 9:1.

Cisteína 5 - Enzimas testadas *Alu I*, *Dde I*, *Eco RI*, *Hae III*, *Mse I* e *Msp I*; Genotipagem com *Mse I*; Fragmento Polimórfico 300 pb; Segregação possível 3:1.

Aspártico - Enzimas testadas: *Fnu* DII, *Hae* III, *Eco* RI, *Msp* I e *Taq* I; Genotipagem com *Msp* I; Fragmento Polimórfico 3000 pb; Segregação possível 9:1.

β-Fructosidase - Enzimas testadas: *Alu* I, *Scr* FI, *Fnu* DII, *Hae* III, *Eco* RI e *Msp* I; Genotipagem com *Msp* I; Fragmento Polimórfico 600 pb; Segregação possível 3:1.

Sacarose Fosfato Sintase - Enzimas testadas: *Alu* I, *Scr* FI, *Fnu* DII, *Hae* III, *Eco* RI, *Msp* I, *Mse* I e *Taq* I; Genotipagem com *Taq* I;

Fragmento Polimórfico 400 pb; Segregação possível 3:1.

Sacarose Sintase - Enzimas testadas: *Alu* I, *Scr* FI, *Fnu* DII, *Hae* III, *Eco* RI, *Msp* I, *Mse* I e *Taq* I; Genotipagem com *Msp* I; Fragmento Polimórfico 500 pb; Segregação possível 3:1.

A matriz de dados gerada a partir da genotipagem será utilizada para testar a hipótese de segregação através do Teste do Qui-Quadrado. Assim, os marcadores que apresentarem segregação de acordo com o esperado, serão inseridos no mapa parcial já obtido por marcadores SSR e AFLP, assim como pela inserção de marcas SSR em desenvolvimento por pesquisadores do IAPAR na mesma população mapa. Dessa forma, será possível posicionar os genes analisados nos grupos de ligação que serão obtidos pela reunião de todos os marcadores disponíveis, bem como serão estudados outros locos da via da sacarose e correspondentes a ácidos clorogênicos e diterpenos.

CONCLUSÕES

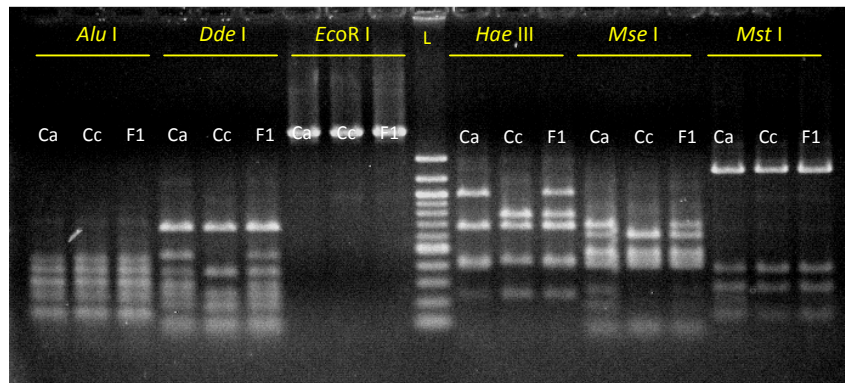
Marcadores SNPs derivados da digestão de ESTs amplificados (PCR-RFLP) a partir de análises *in silico* do banco de genes do genoma café são potencialmente informativos para fins de mapeamento genético e deverão ser buscados com maior intensidade. Estes marcadores serão disponibilizados no banco de dados do Genoma Café para utilização de outros grupos de pesquisa em melhoramento genético do Brasil.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Barret, A. J. The classes of proteolytic enzymes. In Plant Proteolytic Enzymes. Vol. I.; M. J. Dalling, Ed.; CRC Press, Inc.: Boca Raton, 1986; pp 1-16.
2. Brzin, J.; Kidric, M. Proeinases and their inhibitors in plants: role in normal growth and in responses to various stress conditions. Biotechnol. Gen. Engin. Rev. 1995, 13, 421-467.
3. Clifford MN (1985) Chemical and physical aspects of green coffee and coffee products. In MN Clifford, KC Wilson, eds, Coffee: Botany, Biochemistry and Production of Beans and Beverage. AVI Publishing company, Inc., Westport, Connecticut, pp 305-3744. GEROMEL, Clara ; FERREIRA, Lucia Pires ; CAVALLARI, Ana A ; C, Guerreiro S M ; PEREIRA, Luiz Filipe Protasio ; VIEIRA, Luiz Gonzaga ; LEROY, Thierry ; POT, David ; MAZZAFERA, Paulo ; MARRACCINI, Pierre R . Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. Journal of Experimental Botany, v. 57, p. 3243-3258, 2006.
5. VIEIRA, Luiz Gonzaga ; ANDRADE, Alan Carvalho ; COLOMBO, Carlos Augusto ; PEREIRA, Luiz Filipe Protasio ; OUTROS ; CARAZZOLLE, Marcelo Falsarella ; PEREIRA, Gonçalo Amarante Guimarães . The Brazilian coffee genome project: an EST-based genomic source. Brazilian Journal Of Plant Physiology, Campinas SP, v. 18, n. 1, p. 95-108, 2006.
6. Shimizu, M.M., Mazzafera, P., 2000. Compositional changes of proteins and amino acids in germinating coffee seeds. Brazilian Archives of Biology and Technology 43(3): 259-265.

Tabela 1. Sítios de restrição das enzimas *Alu* I, *Dde* I, *EcoR* I, *Hae* III, *Mse* I e *Mst* I avaliados *in silico* para os contigs de *C. arabica* e *C. canephora* nos genes aspártico 1, cisteína 5, cisteína 8, cisteína 22 e cisteína 23.

Enzimas	Aspártico		Cisteína 5		Cisteína 8		Cisteína 22		Cisteína 23	
	Ca	Cc	Ca	Cc	Ca	Cc	Ca	Cc	Ca	Cc
<i>Alu</i> I	250	250	42	42	61	61	105	N	100	100
	438	438	181	181	311	311	336	336	122	122
	658	658	384	384	596	596	490	490	316	316
	766	766	597	597	620	620	502	502	-	-
	847	847	686	686	-	-	525	525	-	-
	912	912	713	713	-	-	549	549	-	-
	960	960	787	787	-	-	-	-	-	-
	1065	N	1012	N	-	-	-	-	-	-
<i>Dde</i> I	10	10	34	34	58	58	530	530	76	76
	502	502	100	100	299	299	536	536	148	148
	793	793	242	242	344	344	720	720	339	339
	930	930	563	563	415	415	N	762	402	402
	1074	1074	1004	1004	607	607	-	-	506	506
	-	-	-	-	-	-	-	-	861	861
<i>EcoR</i> I			229	229						
<i>Hae</i> III	305	305	441	441	585	585	201	201	31	31
	-	-	953	953	874	874	220	220	210	210
	-	-	-	-	-	-	299	299	333	333
	-	-	-	-	-	-	803	803	454	454
	-	-	-	-	-	-	-	-	630	630
	-	-	-	-	-	-	-	-	636	636
	-	-	-	-	-	-	-	-	865	N
	-	-	-	-	-	-	-	-	910	910
<i>Mse</i> I	405	405	118	118	395	395	14	14	65	65
	522	522	538	538	691	691	303	303	-	-
	642	642	967	967			578	578	-	-
<i>Msp</i> I	N	344	257	257	160	N	398	398	135	N
	-	-	-	-	375	375	N	612	298	298
	-	-	-	-	557	557	-	-	710	710
	-	-	-	-	1111	1111	-	-	961	961

Figura 1. Gel de agarose 1,5% evidenciando polimorfismo do gene de protease cisteína_8. Produtos de amplificação por PCR-RFLP de DNA genômico de *C. arabica* (Ca), *C. canephora* (Cc) e de F1, tratados com as enzimas de restrição *Alu* I, *Dde* I, *EcoR* I, *Hae* III, *Mse* I e *Mst* I. L-ladder 100 pb.

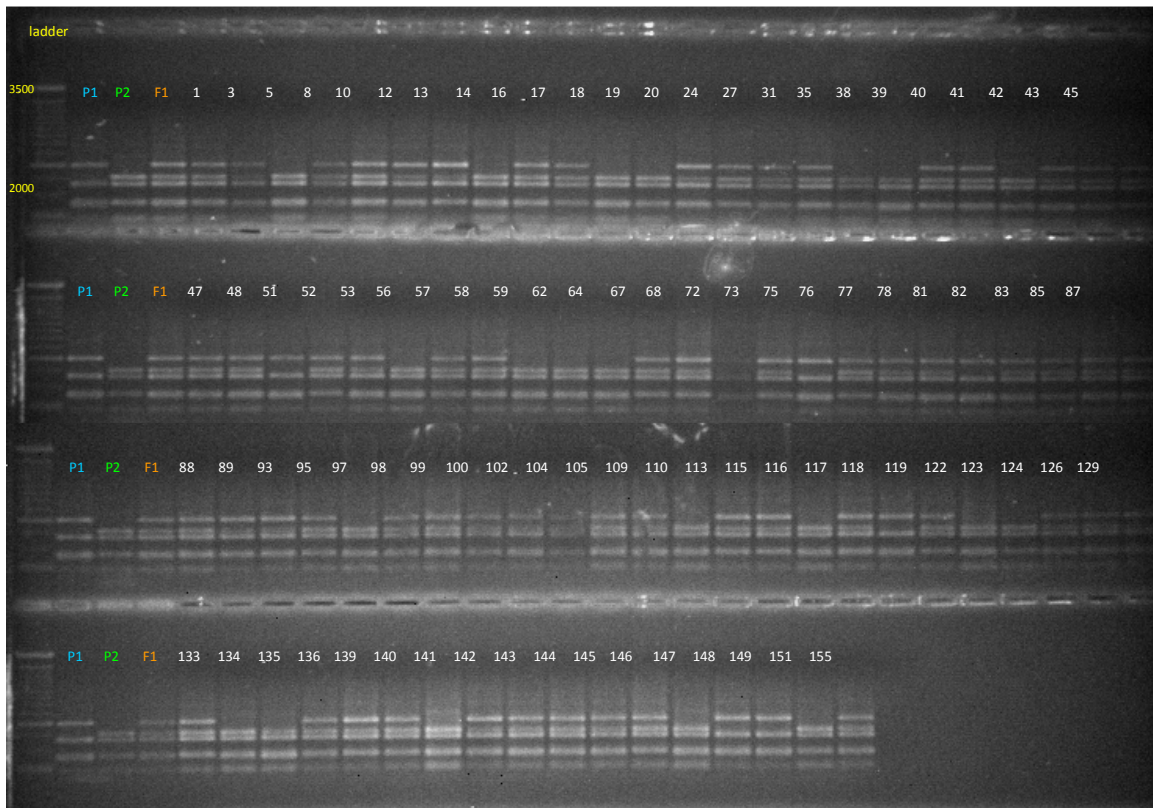


Figura 2. Gel de agarose 1,5% para ilustrar polimorfismo de restrição obtido com a enzima HaeIII após amplificação do gene de protease Cisteína-8 em 90 plantas F_2 de café do híbrido interespecífico *C.arabica* x *C.canephora*, cuja segregação apresentada encontra-se na proporção 3:1.