

# LOCALIZAÇÃO DE MARCADOR MOLECULAR DERIVADO DE EST POTENCIALMENTE ASSOCIADO AO DOMÍNIO LRR EM MAPA GENÉTICO DE *Coffea arabica*

Samuel Mazzinghy ALVARENGA<sup>1</sup>, E-mail: [biocafe@ufv.br](mailto:biocafe@ufv.br); Eveline Teixeira CAIXETA<sup>1,2</sup>; Flávia THIEBAUT Andrade<sup>1</sup>; Bárbara HUFNAGEL Maciel<sup>1</sup>; Eunize Maciel ZAMBOLIM<sup>1,3</sup>; Kátia Viana XAVIER<sup>1</sup>; Ney Sussumu SAKIYAMA<sup>1,4</sup>

<sup>1</sup> Universidade Federal de Viçosa (UFV)/ BIOAGRO, Laboratório de Biotecnologia do Cafeeiro (Biocafé), 36570-000, Viçosa-MG; <sup>2</sup>Embrapa Café, <sup>3</sup> UFV/ Departamento de Fitopatologia; <sup>4</sup> UFV/ Departamento de Fitotecnia. Apoio Financeiro: CBP&D/Café, CAPES e FINEP

## Resumo:

O Projeto Brasileiro do Genoma Café (PBGC) gerou um banco de dados de 200.000 ESTs (*Expressed Sequence Tags*). O banco de dados foi minerado por meio de análise *in silico* e 12009 seqüências potencialmente envolvidas na resistência do cafeeiro a doenças foram identificadas. A partir dessas seqüências foram desenhados 40 oligonucleotídeos, iniciadores específicos para amplificar seqüências do DNA genômico do cafeeiro. As condições de reação e amplificação dos iniciadores sintetizados foram ajustadas. Vinte e nove iniciadores amplificaram fragmentos de aproximadamente 400 pb, exibindo bandas únicas e bem definidas, e apenas um deles foi polimórfico entre os indivíduos resistentes e suscetíveis. O iniciador CARF 005 amplificou um fragmento nos 154 indivíduos F<sub>2</sub> da população H464-2. Essa análise permitiu verificar o posicionamento desse marcador no grupo de ligação 9, a 36,2 cM do marcador K13c no mapa genético de *Coffea arabica* previamente desenvolvido com marcadores RAPD.

Palavras-chave: Genoma Café, *Coffea*, oligonucleotídeos iniciadores específicos.

## LOCALIZATION OF MOLECULAR MARKER FROM EST POTENTIALLY ASSOCIATED WITH THE LRR DOMAIN IN GENETIC MAP OF *Coffea arabica*.

### Abstract:

Twelve thousand and nine sequences potentially involved in the resistance of the coffee tree to diseases were identified in the Projeto Brasileiro do Genoma Café (PBGC) database by *in silico* analysis. Based on those sequences, specific primers were designed in order to amplify sequences of the coffee genomic DNA. Forty specific primers were obtained and their reaction and amplification conditions were tested. Twenty nine primers amplified fragments of approximately 400 bp, showing unique and sharp bands, and only one was polymorphic between the resistant and susceptible genotypes. The primer CARF 005 amplified one fragment in 154 F<sub>2</sub> individuals of the population H464-2. That analysis permitted to verify the position of this marker in the linkage group 9, at 36,2 cM from marker K13c in the genetic map of *Coffea arabica* previously developed with RAPD markers.

Key words: Coffee Genome, *Coffea*, specific primers.

### Introdução

O Projeto Brasileiro do Genoma Café (PBGC) consistiu no sequenciamento em larga escala de seqüências expressas e gerou um banco de dados de 200.000 ESTs (*Expressed Sequence Tags*) (Vieira *et al.*, 2006). Os clones seqüenciados são provenientes de 37 bibliotecas de cDNA de café. Essas ESTs, portanto, representam milhares de genes expressos em seus diferentes órgãos, obtidos em vários estágios de desenvolvimento e submetidos a condições de estresse biótico e abiótico.

O sequenciamento do genoma de plantas tem facilitado e acelerado a identificação de genes desejáveis, possibilitando a sua manipulação subsequente por meio de técnicas de genética molecular. Os genes identificados podem, também, ser úteis para desenvolver novos e eficientes marcadores moleculares e direcionar as manipulações genéticas nos programas de melhoramento.

Os marcadores moleculares obtidos a partir de ESTs são bastante informativos e permitem reconhecer a associação da marca com a sua potencial função, o que os tornam muito valiosos na análise de mapa genético. O objetivo desse trabalho foi utilizar os dados gerados pelo PBGC para o desenvolvimento de marcadores moleculares associados com a resistência do cafeeiro a doenças e posicioná-los no mapa genético.

### Material e Métodos

#### A. Mineração de dados

Os dados do PBGC foram acessados pela plataforma de bioinformática desenvolvida pelo Laboratório de Genômica e Expressão (LGE) e disponível no site (<http://www.lge.ibi.unicamp.br/cafe/>). Para minerar as seqüências do PBGC e identificar as ESTs de interesse foram criadas três estratégias de busca. A primeira estratégia consistiu na busca de palavras-chaves relacionadas na literatura científica com o mecanismo de resistência a doenças. Estas palavras-chave foram

utilizadas como “iscas” para mineração dos dados. Outra estratégia foi a busca por similaridades entre algumas seqüências de cafeeiro já publicadas (Fernandez *et al*, 2004 e Lin *et al*, 2005), e potencialmente envolvidas com a resistência, com as seqüências do PBGC, por meio do algoritmo BLAST. Na terceira estratégia utilizou-se o *Northern* Eletrônico, uma ferramenta desenvolvida pelo LGE que agrupa todas as 200.000 ESTs provenientes das 37 bibliotecas de cDNA. Nesta estratégia, foram selecionados apenas os *contigs* constituídos total ou parcialmente por seqüências da biblioteca RM1 (biblioteca de folhas de *C. arabica* infectadas com bicho mineiro e ferrugem).

As anotações automáticas e os valores de similaridade (*score* e *E-value*) de todas as seqüências encontradas foram analisadas e as mais relevantes para o estudo foram selecionadas.

### B. Desenho de oligonucleotídeos iniciadores

A partir das seqüências selecionadas, foram desenhados iniciadores específicos utilizando o programa computacional *Primer3*. As condições para o desenho foram: (1) produto de PCR com tamanho de 400 a 600pb e 200 a 600pb; (2) tamanho do iniciador de 18 a 24pb; (3) tm de 55 a 60°C; e, (4) percentagem de CG de 45 a 55. A estabilidade destes iniciadores foi verificada utilizando o programa *PrimerSelect*. Desta forma, 40 pares de oligonucleotídeos iniciadores específicos foram obtidos.

### C. Testes dos Iniciadores

Para a obtenção dos marcadores moleculares, diferentes concentrações dos componentes da reação de PCR foram analisadas, incluindo concentrações e marcas comerciais de *taq* DNA polimerase, MgCl<sub>2</sub>, dNTPs e iniciadores. Para a amplificação foram avaliadas diferentes combinações de tempo e temperatura, incluindo *touchdown* PCR. Utilizando as condições de reação e amplificação otimizadas, os 40 iniciadores foram testados em 12 genótipos resistentes e 12 susceptíveis a *Hemileia vastatrix*. Os genótipos susceptíveis foram: UFV-570 (“Bourbon Amarelo”), UFV 2945 (“Típica”), UFV 557-06 (*C. racemosa* x *C. arabica*), UFV 2145-79 (“Catuaí Vermelho”), UFV 2154-345 (“Catuaí Amarelo”), UFV 2143-193 (“Catuaí Amarelo”), UFV 2143-235 (“Catuaí Amarelo”), UFV 2148-57 (“Catuaí Amarelo”), UFV 2164-193 (“Mundo Novo”), UFV 2190-100 (“Mundo Novo”), Caturra 812 (“Caturra Amarelo”) e *C. dewevrei excelsa*. Os genótipos resistentes foram os acessos de “Híbrido de Timor”: UFV 440-22, UFV 443-3, UFV 445-46, CIFC 832-1, CIFC 832-2, UFV 438-52, UFV 446-08, CIFC 33-1, CIFC 1343-269, CIFC 4106, CIFC 420-10 e CIFC 147-1.

### D. Localização do Marcador em Mapa Genético

O iniciador polimórfico foi utilizado para amplificar o DNA de 154 indivíduos F<sub>2</sub> da população H464-2, segregante para a resistência do cafeeiro à ferrugem. Essa população é proveniente do cruzamento de um Híbrido de Timor (H 440-22) resistente à ferrugem e Mundo Novo (2190-100), cultivar susceptível. O programa GQMOL (Cruz e Schuster, 2001) foi utilizado para a análise dos dados moleculares. Essa análise permitiu verificar o posicionamento desse marcador no mapa genético de *C. arabica* previamente desenvolvido com marcadores RAPD (dados não publicados).

## Resultados e Discussão

### A. Mineração de dados

Na estratégia utilizando 15 palavras-chave relacionadas à resistência foram identificadas um total de 11.116 ESTs nos 15 projetos criados (Figura 1). As seqüências presentes em cada um dos projetos foram agrupadas (Tabela 1) e a anotação automática de cada *contig* formado foi analisado. Apenas os *contigs* (1257 no total) foram avaliados, pois possuem um tamanho maior e são mais representativos que os *singlets* (1381 no total).

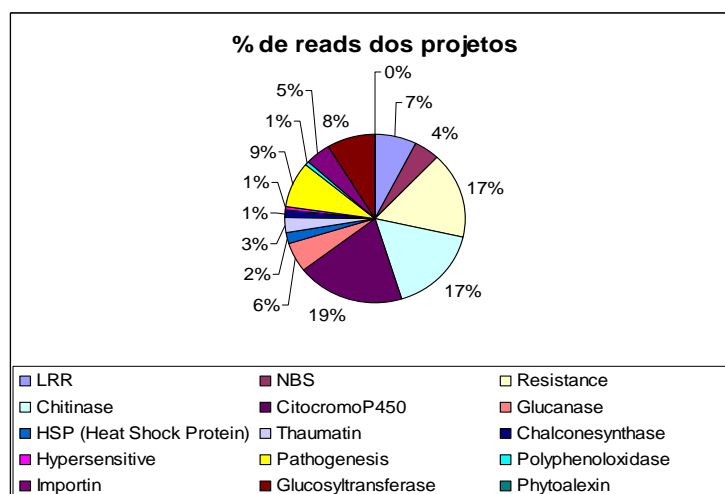


Figura 1– Porcentagem relativa do número de *reads* dos 15 projetos criados pela estratégia Palavras-Chave.

Tabela 1–Resultado do agrupamento das seqüências dos projetos criados a partir de palavras-chave.

<b>Projeto</b>	<b>ESTs</b>	<b>Contigs</b>	<b>Singlets</b>
LRR	825	160	243
NBS	494	105	142
Resistance	1864	347	416
Chitinase	1855	47	48
CitocromoP450	2120	235	202
Glucanase	642	92	68
HSP (Heat Shock Protein)	240	31	27
Thaumatin	317	16	7
Chalconesynthase	153	5	8
Hypersensitive	86	8	4
Pathogenesis	979	63	37
Polyphenoloxidase	67	4	3
Importin	532	11	14
Glucosyltransferase	930	130	160
Phytoalexin	12	3	2
<b>TOTAL</b>	<b>11116</b>	<b>1257</b>	<b>1381</b>

Na outra estratégia, foram identificadas ESTs do PBGC que possuam similaridade com as 57 seqüências relacionadas à resistência, publicadas em artigos científicos.

Na estratégia *Northern* Eletrônico, foram formados 2.206 *clusters* que continham ESTs da biblioteca RM1 (Tabela 2). Deste total, 602 foram formados apenas por ESTs da biblioteca RM1, 21 por seqüências com 99% a 75% e 213 de 74% a 50% dos ESTs provenientes da RM1.

Tabela 2–Número de seqüências com *reads* na Biblioteca RM1.

<b>Estratégia Eletronic Northern</b>	<b>N. seqüências</b>	<b>% em relação a RM1</b>
Sequências com reads em RM1	2206	100
Sequências 50% a 74% de reads em RM1	213	9.65
Sequências 75% a 99% de reads em RM1	21	0.95
Sequências 100% dos reads em RM1	602	27.29

### B. Teste dos iniciadores

Dos 40 iniciadores testados, 11 não amplificaram nenhuma banda. Os 29 restantes geraram marcadores moleculares, visualizadas como bandas únicas, de aproximadamente 400 pb, e bem definidas, sendo que apenas o iniciador CARF 005 apresentou polimorfismo entre os indivíduos resistentes e susceptíveis (Figura 2).

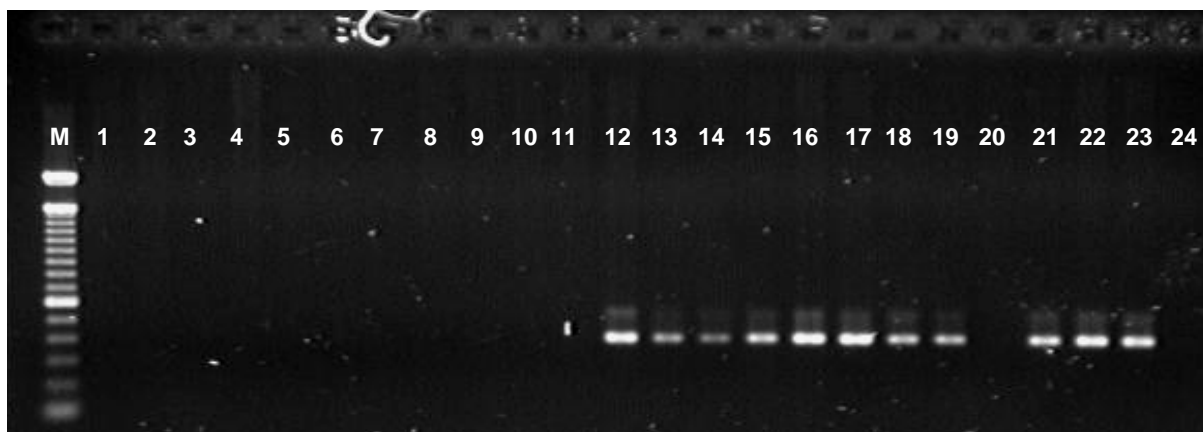


Figura 2– Perfil de Amplificação do iniciador CARF 005. (M) marcador de peso molecular100bp. De 1-12, plantas susceptíveis; de 13-24, plantas resistentes. (1) UFV-570, (2) UFV 2945, (3) UFV 557-06, (4) UFV 2145-79, (5) UFV 2154-345, (6) UFV 2143-193, (7) UFV 2143-235, (8) UFV 2148-57, (9) UFV 2164-193, (10) UFV 2190-100, (11) Caturra 812, (12) *C. dewevrei excelsa*, (13) UFV 440-22, (14) UFV 443-3, (15) UFV 445-46, (16) CIFIC 832-1, (17) CIFIC 832-2, (18) UFV 438-52, (19) UFV 446-08, (20) CIFIC 33-1, (21) CIFIC 1343-269, (22) CIFIC 4106, (23) CIFIC 420-10 e (24) CIFIC 147-1.

O iniciador CARF 005 amplificou um fragmento em *C. dewevrei excelsa*, suscetível a várias raças de *Hemileia vastatrix* conhecidas, diferentemente dos demais indivíduos susceptíveis testados. Para tentar elucidar esse resultado, o fragmento amplificado será seqüenciado e comparado com a seqüência de indivíduos resistentes.

Em relação às ampliações que não apresentaram polimorfismo, é possível que os fragmentos possuam pequenas diferenças dentro da seqüência e, portanto, não puderam ser observadas por eletroforese em gel. Isso será verificado com análises de enzima de restrição e sequenciamento para detecção de SNPs (*Single Nucleotide Polymorphisms*). Existe também a possibilidade da diferença estar no padrão de expressão dos genes de indivíduos resistentes e susceptíveis. Neste caso análises de RT-PCR e ensaios de macroarranjo serão efetuados para demonstrar se existe essa expressão diferencial.

### C. Localização do Marcador em Mapa Genético

O marcador CARF 005, obtido a partir do projeto LRR (*Leucine Rich Repeat*, domínio de proteína de resistência) foi posicionado no mapa genético de café, previamente construído. Essa análise mostrou que esse marcador está presente no grupo de ligação 9, a 36,2 cM do marcador K13c (Figura 3). A análise molecular desse marcador, em conjunto com a análise fenotípica para a resistência, que está sendo efetuada, poderá revelar se o loco onde o marcador se encontra é uma região de genes relacionados com a resistência do cafeeiro. As possibilidades de sucesso da piramidação de genes de resistência à ferrugem dependem do número de genes envolvidos na resistência e a distribuição deles no genoma (ligados ou independentes). Portanto, é importante estudar a distribuição dos possíveis genes de resistência, por exemplo, por meio da localização dos marcadores derivados de ESTs ligados a estes genes no mapa genético.

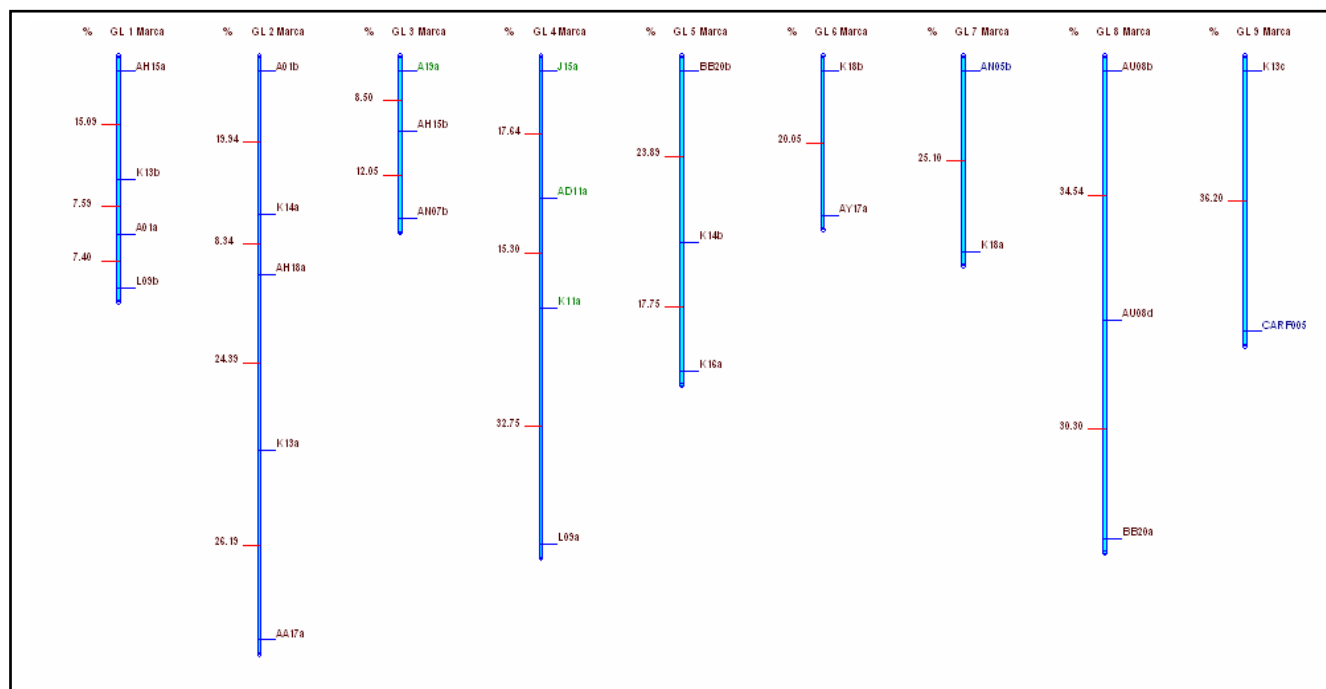


Figura 3– Mapa com a localização do iniciador CARF 005

### Conclusões

Um marcador molecular obtido a partir de EST foi posicionado em mapa genético de *C. arabica* previamente desenvolvido com marcadores RAPD. Esta marca amplifica uma região do DNA que está potencialmente associada ao domínio LRR de proteína de resistência a doenças.

### Referências bibliográficas

Cruz, C. D e Schuster, I. (2001) GQMOL: Programa para análise de dados moleculares. (<http://www.ufv.br/dbg/gqmol/gqmol.htm>). Desenvolvido pelo setor de Genética da Universidade Federal de Viçosa.

Fernandez, D.; Santos, P.; Agostini, C.; Bon, M.-C.; Petitot, A.-S.; Silva, M. C.; Guerra-Guimarães, L.; Ribeiro, A.; Argout, X; Nicole, M. (2004) Coffee (*C. arabica* L.) genes early expressed during infection by the rust fungus (*Hemileia vastatrix*). *Mol. Plant Pathol.*, 5(6):527-536.

Lin, C.; Mueller, L. A.; Carthy, J. M.; Crouzillat, D.; Pétiard, V.; Tanksley, S.D. (2005) Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theor. Appl. Genet.*, 112(1):114-30.

Vieira, L. G. E.; Andrade, A. C.; Colombo, C. A.; Moraes, A. H. de A.; Metha, A., et. al. (2006) Brazilian coffee genome project: an EST-based genomic resource. *Braz. J. Plant Physiol.*, 18(1):95-108.