

## DESENVOLVIMENTO DE MARCADORES ASSOCIADOS À RESISTÊNCIA AO BICHO-MINEIRO: SELEÇÃO E VALIDAÇÃO DE SNPs<sup>1</sup>

Jaqueline Moraes Bazioli<sup>2</sup>, Paula de Sousa Guimarães<sup>3</sup>, Bruna Aparecida Strazza<sup>4</sup>, Juliana Camargo Martinati Schenk<sup>5</sup>, Lilian Padilha<sup>6</sup>, Oliveiro Guerreiro-Filho<sup>7</sup>, Mirian Perez Maluf<sup>8</sup>

<sup>1</sup> Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café

<sup>2</sup> Bolsista Consórcio Pesquisa Café, BS, jaqueline.bazioli@gmail.com

<sup>3</sup> Bolsista Consórcio Pesquisa Café, DSc, psquim@yahoo.com.br

<sup>4</sup> Bolsista Consórcio Pesquisa Café, BS, bruna.strazza@yahoo.com

<sup>5</sup> Bolsista INCT/CNPq, DSc, Instituto Agronômico de Campinas, Campinas-SP, julianamartinati@gmail.com

<sup>6</sup> Pesquisadora, DSc, Embrapa Café, Brasília-DF, lilian.padilha@embrapa.br

<sup>7</sup> Pesquisador, DSc, Instituto Agronômico de Campinas, Campinas-SP, oliveiro@iac.sp.gov.br

<sup>8</sup> Pesquisadora, PhD, Embrapa Café, Brasília-DF, mirian.maluf@embrapa.br

**RESUMO:** O objetivo deste estudo é o desenvolvimento de marcadores do tipo SNPs associados com a resistência ao bicho-mineiro em café. A estratégia principal utilizou as análises genômicas de microarranjo, para caracterização de expressão diferencial de genes-candidatos em plantas resistentes infestadas com o bicho-mineiro. Inicialmente, análises *in silico* de 2137 destes genes identificaram aqueles potencialmente relacionados com a especificidade da resposta de defesa da planta. Uma vez selecionados, o perfil de expressão de 22 genes foi confirmado por qRT-PCR em experimentos utilizando plantas resistentes e suscetíveis, infectadas pelo bicho-mineiro. Para busca de polimorfismos do tipo SNPs, que poderão ser utilizados como marcadores para seleção-assistida, selecionamos 4 genes, cujas regiões genômicas foram clonadas e sequenciadas em genótipos parentais da população em estudo. Após análises *in silico* de sequências genômicas, os polimorfismos do tipo SNPs identificados serviram de base para construção de sondas alelo-específicas, utilizadas na genotipagem de plantas parentais e progênie em seleção. A análise de segregação de SNPs identificados indicou que nenhum dos polimorfismos avaliados apresenta correlação com a característica de resistência ao bicho-mineiro. Estes polimorfismos estão associados com alguma outra característica agrônômica não avaliada neste estudo. A genotipagem de outros SNPs identificados está em andamento.

**PALAVRAS-CHAVE:** seleção-assistida, estresse biótico, melhoramento molecular

## DEVELOPMENT OF MOLECULAR MARKERS ASSOCIATED WITH LEAF-MINER RESISTANCE: SELECTION AND VALIDATION OF SNPs

**ABSTRACT:** The aim of this study is the development of SNPs-markers associated with leaf miner resistance in coffee. The main strategy consisted of microarray analyses, which characterized gene differential expression in resistant and susceptible plants infested with the leaf miner. Initially, *in silico* analysis of 2137 of these genes identified those involved in the specific plant defense response. Once selected, the profile of expression of 22 genes was confirmed by qRT-PCR experiments using resistant and susceptible plants, infected with leaf miner. To search for SNPs, and potential markers for selection-assisted, genomic regions from 4 selected genes were cloned and sequenced in parental genotypes of the population under study. After *in silico* analysis, SNPs identified were the basis for construction of allele-specific probes, used in genotyping of parent plants and progenies in selection. Segregation analysis of identified SNPs indicated that none of the evaluated polymorphisms has correlation with the characteristic of resistance to leaf miner. These polymorphisms are probably associated with other agronomic characteristic not evaluated in this study. The genotyping of additional identified SNPs is in progress.

**KEY WORDS:** assisted-selection, biotic stress, molecular breeding

## INTRODUÇÃO

O uso de marcadores moleculares para seleção-assistida é atualmente uma realidade no melhoramento de várias espécies de interesse agrônômico (Heffner et al., 2010). Atualmente, com a crescente disponibilização de dados genômicos obtidos por análises genômicas em larga-escala, a gama de sequências que podem potencialmente ser utilizadas como marcadores moleculares aumentou. Uma estratégia para identificação de genes-marcadores é a caracterização da expressão de genes em larga-escala a partir de hibridizações em microarranjos. Em café, um estudo realizou a primeira hibridização em larga-escala utilizando sequências do Banco de Dados do Genoma Café. Para este estudo foram selecionados dois genótipos de cafeeiros, resistentes e suscetíveis ao bicho mineiro (*Leucoptera coffeella*), a fim de avaliar alterações na expressão dos genes ao longo do processo de infestação. A análise de microarranjo foi feita utilizando o modelo 12 X 135K e contou com um total de 22.500 sequências de EST derivadas no banco de dados do genoma do café (CAFEST). As análises de hibridização indicaram que em todos os tempos de infestação avaliados houve uma pequena parcela do total de genes que apresentou expressão significativa. Nas plantas não infectadas houve

expressão diferencial de 2137 genes (4,09% super-expresso e 5,94% reprimidos em plantas suscetíveis). No tempo correspondente à oviposição, 5,29% do total de genes foram super-expressos e 5,7% foram reprimidos em plantas suscetíveis. Durante a eclosão e alimentação da larva, 4,17% e 6,42% do total de genes foram super-expressos e reprimidos respectivamente em plantas suscetíveis (Martinati et al, 2011; Cardoso et al., 2014). A partir destes perfis de expressão diferencial, selecionamos para validação e sequenciamento 22 genes que potencialmente podem distinguir uma interação compatível de uma incompatível e assim, atuarem como marcadores para seleção-assistida de cafeeiros resistentes ao bicho-mineiro.

## MATERIAL E MÉTODOS

**Validação de genes diferencialmente expressos** - A partir do *screening* de genes diferenciais identificados em análises de microarranjo (Cardoso et al., 2014) selecionamos 22 genes (12 superexpressos e 10 reprimidos em resposta de defesa), para avaliação da expressão por PCR em tempo real e análises da correlação entre os resultados obtidos pelos dois métodos (microarranjo e PCR quantitativo). Os genes que apresentaram valores numéricos de “*FoldChange*” positivos (superexpressão) e negativos (repressão) maiores que +2 ou menores que -2 nas interações de microarranjo, tiveram suas sequências identificadas e anotadas tendo como fonte de dados o banco de ESTs gerado pelo projeto Genoma Brasileiro do Café-CAFEST (Vieira et al., 2006), por meio da interface “gene annotation, new assembly”. Para a seleção de *primers* gene-específicos, as sequências de nucleotídeos dos reads/contigs de interesse foram comparadas com as sequências de ESTs de banco de dados públicos (NCBI), utilizando-se, para isso, o algoritmo tBLASTx. Para classificar a parte funcional da proteína, bem como sua matriz de leitura, foi utilizada a ferramenta Open Reading Frame Finder (ORF FINDER) e ainda se realizou a busca dos domínios conservados das proteínas, a fim de definir as regiões para anelamento dos *primers*.

**Extração de RNA e síntese de cDNA** - Para extração do RNA foram coletadas folhas infestadas com bicho-mineiro, em quatro estágios de desenvolvimento da praga: oviposição (T1 = 1 dia após instalação do experimento), eclosão da lagarta (T2 = 7 dias após instalação do experimento), com lesão em desenvolvimento intermediário (T3 = 11 dias após instalação do experimento), e com lesão em fase final de desenvolvimento (T4 = 18 dias após instalação do experimento). O RNA total foi extraído pelo método citado por Chang et al. (1993). A quantificação do RNA foi realizada medindo-se a absorbância a 260nm, 280nm e 230 nm. As amostras de RNA total foram tratadas com DNase I (Promega™). A partir dos RNAs tratados com DNase I foi sintetizada a primeira fita de cDNA utilizando o kit SuperScript™ II Reverse Transcriptase (Thermo). Em seguida, as amostras foram armazenadas a -20°C.

**Análises de expressão gênica** - A quantificação da expressão dos genes selecionados foi realizada por qRT-PCR, em equipamento ABI7300. As reações foram realizadas em volume final de 15 µL contendo 400 ng de cDNA, 200mM de cada primer e 1X de SYBR Green master mix (Promega). Todas as amostras foram amplificadas em triplicatas. Foram utilizadas também amostras sem adição de cDNA para detectar qualquer sinal de contaminação (controle negativo). Para confirmar a presença de amplicons únicos, os produtos do PCR foram submetidos à curva de dissociação com temperatura variando entre 60°C e 95°C. Para todos os genes testados utilizou-se como calibrador os valores obtidos para as plantas suscetíveis, em T0.

**Identificação de polimorfismos do tipo SNPs** - Para identificação de polimorfismos gênicos selecionamos 4 genes cuja expressão diferencial foi validada. Os *contigs* foram avaliados e *primers* gene-específicos incluindo toda a região codificadora foram obtidos. Os fragmentos contendo regiões genômicas foram obtidos a partir de metodologia de PCR genômico, utilizando-se o DNA extraído de folhas de café da cultivar H14 (tolerante) e IAC 62 (suscetível). O DNA genômico foi extraído segundo o protocolo com CTAB (Paillard et al, 1996) e sua quantificação foi feita em espectrofotômetro (Abs230nm, Abs260nm, Abs280nm). As sequências dos genes selecionados foram obtidas do banco de dados do Genoma Café (CAFEST). O programa Primer3 plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) foi utilizado para o desenho dos “primers” e o Oligo analyzer (<http://eu.idtdna.com/analyzer/applications/oligoanalyzer/>) para avaliar as condições (%GC, Tm, etc.) e parâmetros desejados. Os fragmentos genômicos amplificados e purificados foram clonados em vetor do tipo PGEM-T (Promega), inseridos na bactéria *E. coli* (DH5α) para multiplicação. Colônias de bactérias contendo o vetor desejado foram crescidas e submetidas à extração do vetor. O vetor foi isolado e purificado utilizando-se kits comerciais e os fragmentos inseridos foram observados no gel de agarose (1,0%). O sequenciamento dos clones contendo inserto de interesse foi por empresa terceirizada. Para as análises das sequências e dos polimorfismos foi utilizado o programa BioEdit (Hall,1999). As sequências dos genes em plantas de cafeeiros suscetíveis e resistentes ao bicho mineiro foram analisadas e comparadas quanto à composição de bases e os polimorfismos identificados e caracterizados. Para validação dos SNPs como marcadores, sondas alelo-específicas do tipo TaqMan (Life Technologies) foram utilizadas para genotipagem de progênies selecionadas.

## RESULTADOS E DISCUSSÃO

A estratégia para obtenção de marcadores associados à resistência ao bicho-mineiro em café foi buscar polimorfismos do tipo SNPs em genes-candidatos, relacionados com a resposta de defesa e identificados pela análise de expressão diferencial em larga-escala em cafeeiros resistentes e suscetíveis (Cardoso et al., 2014). Inicialmente, selecionamos

genes diferencialmente expressos, tanto ativados quanto reprimidos, entre progênies resistentes e suscetíveis na ausência do inseto (T0). Baseados nestas análises selecionamos 22 genes para dar continuidade ao proposto e testar a estratégia. Para validação do perfil diferencial a expressão destes genes foi quantificada por qRT-PCR, nos tempos T0 e T2. A lista destes genes, com anotação e valores comparativos das análises de expressão estão apresentados na Tabela I e na Figura 1. Observa-se que a expressão da maioria dos genes apresentou diferenças no nível de expressão (maior ou menor quantidade), porém manteve o mesmo perfil de ativação ou repressão em plantas resistentes e suscetíveis (Figura 1). No entanto, para alguns genes além de variações no nível de expressão foi também observada uma inversão no perfil. Assim, por exemplo os genes contigs 3 e 4 que pela análise de microarranjo se apresentavam reprimidos nas plantas resistentes em T0, não tiveram seus valores de expressão diferencial validados pelas análises de qRT-PCR. A situação oposta também foi observada: análises de microarranjos indicaram uma super expressão, como é o caso dos contigs 17 e 19, e as análises de qRT-PCR mostraram repressão da expressão gênica. Análises de dados genômicos em larga-escala requerem o uso de algoritmos e protocolos que convertem os dados brutos em valores relativos e normalizados. Este tratamento dos dados pode ocasionalmente resultar em desvios do perfil real de expressão dos genes. Assim, a validação dos resultados obtidos por análises *in silico* é fundamental para efetivamente identificar genes diretamente relacionados com a resposta em estudo. Neste caso, as diferenças observadas entre perfil de expressão de genes categorizados pelas análises *in silico* e o da quantificação em tempo-real demonstram a necessidade desta validação, não só para definir genes ligados à resistência, mas também para identificar aqueles potencialmente associados com uma infestação bem-sucedida pelo bicho-mineiro.

Tabela 1. Genes ativados e reprimidos selecionados a partir de análises *in silico* das interações entre T0R e T0S, T0S e T2R, com anotação funcional, valores de *foldchange* e validação em tempo-real.

Gene	Anotação	Microarranjo (T0S x T0R)	qRT-PCR (T0S x T0R)	qRT-PCR (T0S x T2R)
Contig 1	Sem homologia	-43,72708262	-14,53008741	-2,571
Contig 2	negative cofactor 2 transcriptional	1,118970506	1,697400103	-
Contig 3	proteína desconhecida	-115,7701332	2,669667372	-
Contig 4	proteína desconhecida	-97,12980896	4,336887431	-
Contig 5	60S ribosomal protein l37	1,269400249	1,019124283	-
Contig 6	organ-specific protein	-155,9213346	-1,040062337	-
Contig 7	caffeine synthase	-148,7526086	-8,633805944	-
Contig 8	polygalacturonase-1 non-catalytic subunit	-236,3703428	-1,747141755	-
Contig 9	proteína desconhecida	235,4565896	2,063366349	413,586
Contig 10	organ-specific protein	-41,68995503	1,643373032	-
Contig 11	extensin-like protein	14,88901672	1,77768528	245,933
Contig 12	proteína desconhecida	445,8766536	5,413869079	379,114
Contig 13	isocitrate lyase	8,84438282	3,204286914	1,807
Contig 14	hypothetical protein VITISV_000181	62,54369404	1,021005049	-
Contig 15	polygalacturonase-1 non-catalytic subunit	-152,1769324	3,249008835	-8,329
Contig 16	glycerol-3-phosphate acyltransferase 6	17,93674884	3,410539567	109,833
Contig 17	gdsl-motif lipase hydrolase-like	16,98350616	-3,010486539	33,578
Contig 18	cellulose synthase	-7,762211719	1,438933912	-
Contig 19	proteína desconhecida	24,73317491	-1,584107234	34,746
Contig 20	acid phosphatase	-236,5339843	1,148961137	1,775
Contig 21	zinc finger protein	126,6590875	1,427340956	30805,701
Contig 22	acidic endochitinase-2	-447,4557263	-11,20965522	-234,002

Para identificação de polimorfismos gênicos associados com a característica de resistência ao bicho-mineiro selecionamos 4 genes cuja expressão diferencial foi validada (Figura 1). Os contigs foram avaliados e *primers* gene-específicos incluindo toda a região codificadora foram obtidos. A região genômica foi clonada a partir dos genótipos parentais da população em análise. Análises *in silico* das sequências genômicas obtidas resultaram na identificação de um número reduzido de SNPs entre os genótipos parentais: dois SNPs no gene *isocitrate lyase*; um SNP nos genes

*glycerol-3-phosphatase acyltransferase 6, sem homologia*; e nenhum SNP no gene *caffeine synthase*. Esses resultados demonstram a reduzida diversidade genética presente na população em estudo, apesar do cruzamento inicial interespecífico (*C. arabica* X *C. racemosa*) implicar em uma maior variabilidade. A partir destas análises, desenhamos conjuntos de sondas e primers para amplificação de regiões genômicas contendo SNPs. A genotipagem de acessos parentais indicou que os polimorfismos selecionados não segregam com a característica de resistência.

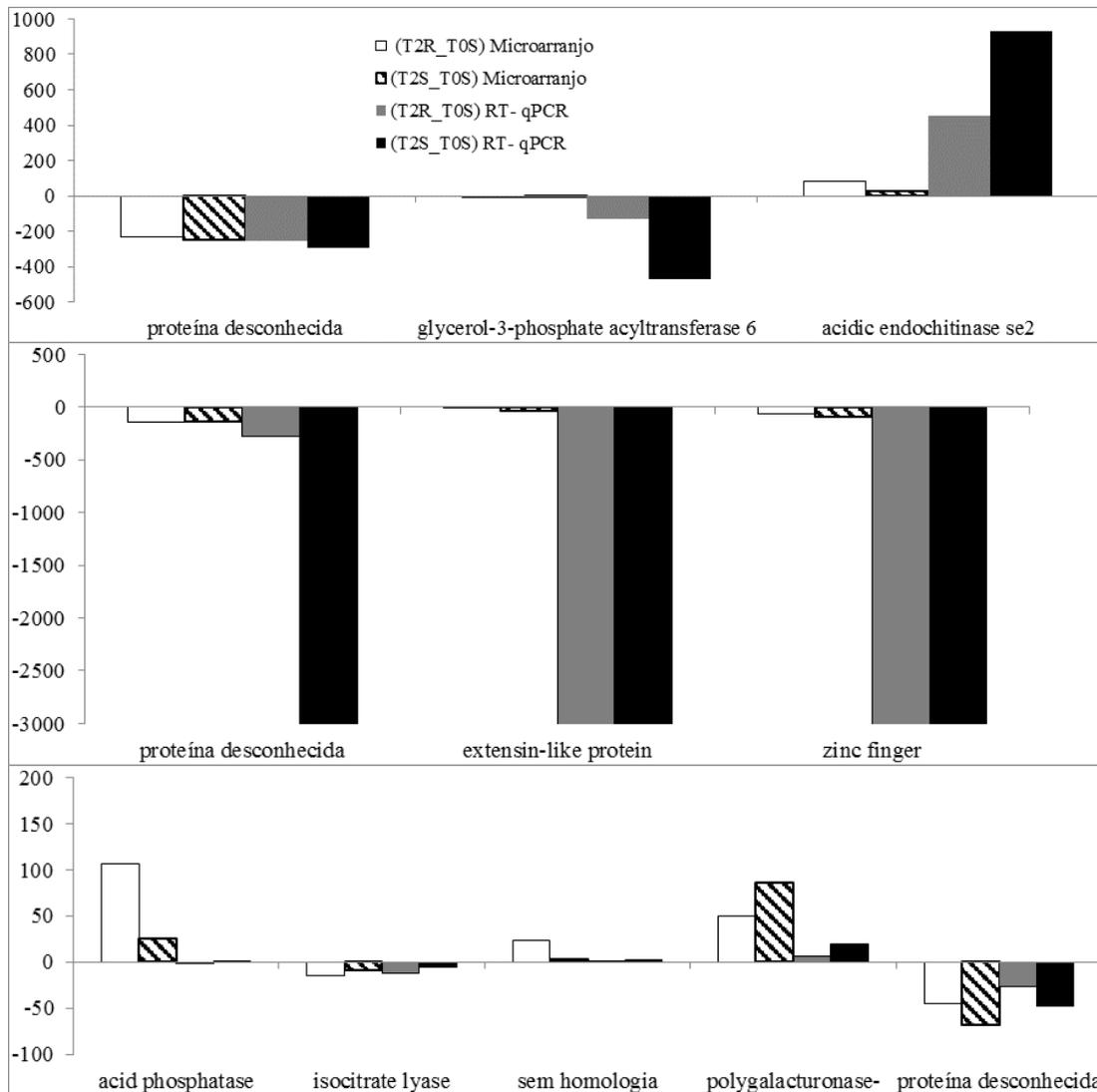


Figura 1. Comparação da expressão de genes selecionados entre plantas resistentes e suscetíveis ao bicho-mineiro analisada por técnicas de microarranjo e RT-qPCR no tempo 2

## CONCLUSÃO

1 - Até o momento não foi possível identificar polimorfismos do tipo SNPs associados com a característica de resistência ao bicho-mineiro. Porém já foram identificados genes expressos somente em plantas resistentes ou suscetíveis, o que representa também um tipo de marcador importante para seleção de progênies contendo genes de defesa. Este fato em si é bastante animador, uma vez que nenhuma das estratégias avaliadas anteriormente permitiu a identificação de marcadores que separassem progênies segregando para diferentes características em uma mesma população.

## **AGRADECIMENTOS**

Os autores agradecem o apoio financeiro do Consórcio Pesquisa Café e do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

- CARDOSO, D. C ; MARTINATI, J. C. ; GIACHETTO, P F ; VIDAL, R O ;CARAZZOLLE, M F ; PADILHA, L. ; GUERREIRO-FILHO, O. ; MALUF, M. P. Large-scale analysis of differential gene expression in coffee genotypes resistant and susceptible to leaf miner toward the identification of candidate genes for marker-assisted selection. *BMC Genomics* , v. 15, p. 66, 2014.
- CHANG, S.; PURYERAR, J.; CAIRNEY, J. A single and efficient method for isolating RNA from pine trees. *Plant Molecular Biology*. v.11, p.113-116, 1993.
- HALL, T.A. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, p. 95-98, 1999.
- HEFFNER, E.L., LORENZ, A.J., JANNINK, J.L., SORRELLS, M.E. Plant breeding with genomic selection: gain per unit time and cost. *Crops Science*, 50, p.1681-1690, 2010.
- MARTINATI, J. C., CARDOSO, D. C., GUERREIRO FILHO, O., VIDAL, R., CARAZZOLLE, M., MALUF, M. P. Large-scale analysis of differential gene expression in coffee genotypes resistant and susceptible to leaf miner In: III Simpósio Brasileiro de Genética Molecular de Plantas, 2011, Ilhéus.
- PAILLARD, M., LASHERMES P., PÉTIARD V. Construction of a molecular linkage map in coffee. *Theor Appl Genet*, 93, p. 41-47, 1996.