

# ESTs DE *Coffea arabica* DO TIPO GENES R CLASSE 3 IDENTIFICADAS NO CafEST

Magnólia A. CAMPOS<sup>1</sup>, E-mail: camposma@ufla.br; Flávia B. SILVA<sup>1</sup>; Marília S. SILVA<sup>2</sup>; Érika E. V. S. ALBUQUERQUE<sup>3</sup>; Cristiane C. TEIXEIRA<sup>3</sup>; Ângela MEHTA<sup>3</sup>; Maria Fátima GROSSI DE SÁ<sup>3</sup>

<sup>1</sup>Universidade Federal de Lavras, Lavras-MG; <sup>2</sup>Embrapa Cerrados, Planaltina-DF; <sup>3</sup>Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF.

## Resumo:

Café é uma importante *commodity* e um dos produtos agrícolas mais comercializados e consumidos no mundo. Por isso, o seqüenciamento em larga escala de seqüências expressas (ESTs) em órgãos de espécies de cafeeiro foi realizado e resultou na formação do banco de dados brasileiro do genoma funcional de café (CafEST). Apesar da produção e consumo de *Coffea arabica* corresponder à cerca de 70% do mercado de café, essa espécie é altamente suscetível a nematóides, pragas e patógenos. Portanto, há um crescente interesse de programas de melhoramento genético de cafeeiro em desenvolver variedades de *C. arabica* com resistência a esses agentes fitopatológicos. Inúmeros genes de resistência (genes *R*) já foram isolados de várias plantas e foram agrupados em classes de 1 a 6. Com o objetivo de elucidar se subclasses das 6 classes de genes *R* estão representadas nos genomas de espécies de café, uma estratégia de identificação de genes baseada em mineração de dados genômicos está sendo empregada para selecionar ESTs do banco CafEST que representem cada classe e subclasse já descrita. O presente trabalho relata a presença de um total de 293 ESTs relacionadas com genes *R* classe 3 (genes do tipo *LRR/NBS/TIR*) no genoma de *C. arabica*, as quais foram identificadas dentro do banco CafEST por homologia no BLAST. Dentre essas seqüências, 101 representam a subclasse RPP4 e foram agrupadas em 56 clusters. Adicionalmente, 93 ESTs representando a subclasse RPP5 foram encontradas e agrupadas em 46 clusters. Finalmente, as últimas 99 ESTs encontrados representam a subclasse RPS4 e foram agrupadas em 54 clusters. No entanto, nenhuma EST foi encontrada representando as outras subclasses de genes *R* classe 3 (L, M, N, P e RPP1) no banco CafEST com base nos critérios empregados. A estratégia usada para selecionar ESTs de *C. arabica* que potencialmente codificam genes *R* classe 3 dentro do CafEST permitiu a identificação de vários prováveis genes.

Palavras-chave: *Coffea arabica*, genes *R* classe 3, CafEST, resistência.

## In silico analyses OF *Coffea arabica* PUTATIVE CLASS 3 R GENES WITHIN THE cafEST DATABASE

### Abstract:

Coffee is a very important commodity and one of the main commercialized and consumed crops worldwide. For this reason, the sequencing in large scale of expressed sequence tags (ESTs) from coffee organs was performed and resulted in the formation of the Brazilian Coffee Genome EST database (CafEST database). Although *C. arabica* production and consumption corresponds to approximately 70 % of the coffee market, it is highly susceptible to nematodes, pests and diseases. Therefore there is a raising interest of genetic breeding programs in developing *C. arabica* varieties with resistance to these phytopathological agents. A large number of plant resistance genes (*R* genes) have already been isolated and were classified into six categories denoted class1-class 6. Herein it is presented a screening of the coffee transcriptome database for class 3 LLR/NBS/TIR-like *R* gene related sequences within the *Coffea arabica* ESTs from the CafEST data bank. Based on sequence similarities in homologue searches we selected a total of 293 ESTs encoding for class 3 *R* proteins putative related to disease resistance in *C. arabica*. Among these reads, 101 ESTs representing the RPP4 subclass were grouped into 56 clusters. We found 93 reads representing the RPP5 subclass and which were grouped into 46 clusters. In addition, we also found 99 reads representing the RPS4 subclass and which were grouped into 54 clusters. However, no EST was found representing other subclasses of *R* genes (L, M, N, P and RPP1) in CafEST database based on adopted criteria. The used approach to select ESTs from *C. arabica* that putatively encode potencialmente *R* genes classe 3 dentro do CafEST led to identification of several putative genes.

Key words: *Coffea arabica*, class 3 *R* genes, CafEST, resistance.

### Introdução

Café é uma importante *commodity* e um dos produtos agrícolas mais comercializados e consumidos no mundo. O Brasil produz aproximadamente 40% do café comercializado no mercado internacional, além de ser o segundo maior consumidor mundial desse produto, particularmente da espécie *Coffea arabica*. Apesar da produção e consumo de *C. arabica* corresponder à cerca de 70 % do mercado de café, essa espécie é altamente suscetível a nematóides, pragas e doenças. Portanto, há um crescente interesse de programas de melhoramento genético de cafeeiro em desenvolver variedades de *C. arabica* com resistência a nematóides, pragas e doenças.

Durante a evolução, as plantas desenvolveram vários mecanismos contra o ataque de patógenos. Além de muitas barreiras pré-formadas, também ativam resistência ao nível de espécie (resistência de não-hospedeira), resistência raça-específica, resistência raça não-específica e mecanismos de resistência basal. Plantas reconhecem a presença de patógenos quando ocorre um contato físico entre eles na superfície externa vegetal ou dentro de seus tecidos (Vidhyasekaran, 1997).

O reconhecimento do patógeno pela planta, o qual é mediado por receptores de resistência (proteínas R), é um pré-requisito essencial para a indução de respostas de defesa no local de penetração. Inúmeros genes de resistência de planta (genes *R*) já foram isolados e agrupados em classes de 1 a 6. Genes *R* classe 3 bem caracterizados codificam proteínas R receptoras do tipo TIR-NBS-LRR, tais como L, M, N, P, RPP1, RPP5 e RPS4, as quais apresentam um domínio N-terminal TIR, que é similar ao domínio de sinalização citoplasmática de receptores transmembrana *Toll* e *Interleukin-1*, um domínio NBS (*nucleotide binding site*) e um domínio LRR (*leucine-rich repeat*) (Meyers et al., 1999; Dangl e Jones, 2001). Receptores Toll de drosófila e do tipo Toll (*Toll like receptors*, TLRs) de mamíferos reconhecem moléculas PAMPs (*pathogen-associated molecular pattern*) através do domínio LRR extracelular e transduzem o sinal de PAMPs para o interior da célula através do domínio TIR. Similarmente, a resistência do tipo não-hospedeira induzida em plantas é comparável à imunidade inata de animais, a qual atua a resistência a patógenos através do reconhecimento de moléculas elicitoras gerais tipo PAMPs (Jones e Takemoto, 2004).

Recentemente, o seqüenciamento em larga escala de seqüências expressas (*Expressed Sequence Tags*, ESTs) de cafeeiro foi concluído, resultando na formação do Banco Brasileiro do Genoma Funcional de Café (Banco CafEST), trabalho este que envolveu a rede formada entre o Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café, a AEG (*Agronomical and Environmental Genomes*)-FAPESP e a Embrapa Recursos Genéticos e Biotecnologia. Uma estratégia de identificação de genes baseada em mineração de dados genômicos está sendo empregada para selecionar ESTs do banco CafEST que representem cada uma das 6 classes de genes *R* já descrita. Mais de 2400 reads relacionadas a genes *R* já foram selecionadas até o momento. Neste trabalho, nós relatamos a presença de ESTs relacionadas com genes *R* classe 3 (genes do tipo *LLR/NBS/TIR*) no genoma de *C. arabica*, as quais foram identificadas dentro do banco CafEST por homologia no BLAST.

## Material e Métodos

Para buscar genes homólogos no genoma de *C. arabica* que codificam proteínas R classe 3 do tipo TIR-NBS-LRR, análises de busca por comparação foram realizadas dentro do banco de dados CafEST (<http://www.lge.ibi.unicamp.br/cafe>), usando o programa *Basic Local Alignment Tool* (tBLASTn) (Altschul et al., 1997) e seqüências de aminoácido de cada uma das subclasses (L, M, N, P, RPS4, RPP1, RPP4 e RPP5) bem descritas e disponíveis em bancos públicos ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). O uso dos critérios E-value  $1,00 \times 10^{-04}$  e matrix BLOSUM62, levou a seleção de 293 reads do banco CafEST (Tabela 1), o qual possui 153.739 reads úteis. Clusters de ESTs foram construídos para cada classe de possíveis genes *R* classe 3, usando o programa CAP3 (*Contig Assembly Program*) (Huang e Madan, 1999). Em seguida, a seqüência consenso de cada contig da subclasse RPS4 foi comparada com seqüências de aminoácidos de proteínas homólogas depositadas no Genbank, usando o programa BLASTx (Altschul et al., 1997).

Todas as seqüências de *C. arabica* usadas neste trabalho, correspondentes às seqüências do tipo ESTs (reads) e seqüências consenso de clusters, foram obtidas no banco de dados brasileiro do genoma funcional de café (CafEST), as quais são derivadas de 32 bibliotecas de cDNA construídas de diferentes genótipos e órgãos (folha, ramo, fruto, flor e raiz) sob condições de desenvolvimento ou estresse (Vieira et al, 2006).

## Resultados e Discussão

Este trabalho relata a presença de possíveis genes *R* classe 3 que codificam proteínas do tipo TIR-NBS-LRR no genoma de *C. arabica* encontradas no banco de dados CafEST, as quais estão envolvidas na percepção de sinais de moléculas efetoras de patógenos. Com base na similaridade de seqüência encontrada via tBLASTn, um total de 293 ESTs foi identificado no CafEST, as quais provavelmente codificam proteínas R classe 3 relacionadas com a resistência a doenças em *C. arabica* (Tabela 1). Dentre essas seqüências, 101 representantes da subclasse RPP4 foram agrupadas em 56 clusters. Além dessas, 93 ESTs representando a subclasse RPP5 foram encontradas e agrupadas em 46 clusters. O gene *RPP4* de *Arabidopsis* é membro da família multigênica *RPP5* de genes *TIR-NBS-LRR* e confere resistência a míldio através de componentes de sinalização (van der Biezen et al., 2002). Finalmente, as últimas 99 ESTs encontradas representam a subclasse RPS4 e foram agrupadas em 54 clusters. O gene *RPS4* de *Arabidopsis* é um gene *R* classe 3 de resistência à bactéria, cuja proteína reconhece o elicitor de avirulência avrRps4, produto do gene de *Pseudomonas syringae* pv. *ptsi* (Hinsch and Staskawicz, 1996; Gassmann et al., 1999). No entanto, nenhuma EST de *C. arabica* foi encontrada representando as outras subclasses de genes *R* classe 3 (L, M, N, P e RPP1) no banco CafEST, com base nos critérios adotados. Adicionalmente a estes dados, uma estratégia utilizando programas de análises de domínios presentes em proteínas R de cada classe de gene *R*, está sendo empregada em nova varredura do banco CafEST, de modo a confirmar ou adicionar novas seqüências de interesse.

Tabela 1. Número de ESTs do tipo genes *R* classe 3 de *C. arabica* encontrados no Banco CafEST.

Subclasse de gene <i>R</i>	Nº de reads	Nº de clusters	
		Contigs	Singlets
RPP4	101	19	37
RPP5	93	17	29
RPS4	99	18	36

Interessantemente, análise detalhada de clusters do tipo RPP4 (Table 2) revelou que os contigs C1, C3, C5, C13, C16 e C18 são formados por ESTs oriundas exclusivamente de bibliotecas únicas de café (dado não mostrado): O contig C1 é formado por 2 ESTs expressas exclusivamente na biblioteca de folhas tratadas com ácido araquidônico. O contig C3 é formado por 2 ESTs isoladas de folhas não-tratadas com Bion. O contig C5 é composto por 5 ESTs oriundas da biblioteca de ramos infectados por *Xylella fastidiosa*, o que sugere envolvimento desse contig em resistência basal de *C. arabica* a essa bactéria patogênica. O contig C13 é composto por 3 ESTs expressas na biblioteca de células de café mantidas em meio salino. Enquanto que o contig C16 é formado por 2 reads exclusivas de biblioteca de tecidos de raiz e suspensão de células na presença de alumínio. Finalmente, o contig C18 é composto por 3 reads originárias exclusivamente de bibliotecas de sementes em germinação.

**Tabela 2.** Distribuição de ESTs de *C. arabica* do tipo do tipo genes *R* classe 3 subclasse *RPS4* em contigs e similaridades da seqüência de aminoácidos deduzida com proteínas do GenBank.

Contig	Tamanho (nt)	Nº de ESTs	Seqüência de melhor similaridade via Blast			
			Organismo	Nº de acesso	E-value	Similaridade
C1	831	2	<i>Solanum tuberosum</i>	gb AAP44392.1	2e-56	131/153 (84%)
C2	1579	9	<i>Oryza sativa</i>	ref NP_915900.1	2e-16	156/378 (41%)
C3	736	2	<i>Arabidopsis thaliana</i>	ref NP_200956.1	5e-68	161/245 (65%)
C4	1556	7	<i>Oryza sativa</i>	ref NP_001067154.1	1e-141	312/408 (76%)
C5	779	3	<i>Populus trichocarpa</i>	gb ABF81421.1	1e-20	119/249 (47%)
C6	769	3	<i>Cucumis melo</i>	gb AAT77096.1	2e-15	88/182 (48%)
C7	1282	3	<i>Populus trichocarpa</i>	gb ABF81421.1	1e-27	158/334 (46%)
C8	734	3	<i>Cucumis melo</i>	gb AAT77098.1	6e-08	74/176 (42%)
C9	999	2	<i>Arabidopsis thaliana</i>	gb ABG00804.1	1e-58	171/281 (59%)
C10	900	2	<i>Medicago truncatula</i>	gb ABE84400.1	3e-29	95/164 (57%)
C11	918	7	<i>Capsicum annuum</i>	gb AAN62015.2	2e-68	161/197 (81%)
C12	1552	4	<i>Glycine max</i>	gb AAR19098.1	4e-29	175/397 (44%)
C13	862	3	<i>Capsicum annuum</i>	gb AAN62015.2	5e-52	151/179 (84%)
C14	922	2	<i>Arabidopsis thaliana</i>	ref NP_176532.2	9e-76	182/257 (70%)
C15	1003	2	<i>Arabidopsis thaliana</i>	gb AAM13028.1	7e-67	167/256 (64%)
C16	685	2	<i>Solanum tarijense</i>	gb AAR29076.1	6e-28	122/215 (56%)
C17	798	3	<i>Solanum tuberosum</i>	gb AAW48301.1	6e-09	59/113 (52%)
C18	784	3	<i>Nicotiana benthamiana</i>	gb AAY54606.1	6e-53	145/214 (67%)

## Conclusões

Neste trabalho, a estratégia usada para selecionar ESTs de *C. arabica* que potencialmente codificam genes *R* classe 3 dentro do CafEST permitiu a identificação de vários prováveis genes homólogos aos genes *RPP4*, *RPP5* e *RPS4*. Por ser o cafeeiro uma cultura perene, os dados aqui apresentados provêm informações relevantes que poderão ser aplicadas em programas de melhoramento tradicional e no desenvolvimento de novas estratégias para a obtenção de resistência durável de cafeeiro contra patógenos, resultando em impactos positivos no agronegócio do café.

## Agradecimentos

Os autores agradecem o apoio recebido da Embrapa Recursos Genéticos e Biotecnologia, da Embrapa Café e do Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café. O primeiro autor (Magnólia A. Campos) agradece ainda o apoio financeiro dado por CAPES/PRODOC/UFLA.

## Referências Bibliográficas

- Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Mille W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
- Dangl JL, Jones JD. 2001. Plant pathogens and integrated defence responses to infection. *Nature* 411:826–33
- Gassmann W, Hinsch ME, Staskawicz BJ. 1999. The *Arabidopsis RPS4* bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease resistance genes. *Plant J.* 20:265–77.
- Hinsch M, Staskawicz B. 1996. Identification of a new *Arabidopsis* disease resistance locus, *RPS4*, and cloning of the corresponding avirulence gene, *avrRps4*, from *Pseudomonas syringae* pv. *ptsi*. *Mol. Plant-Microbe Interact.* 9:55–6177.

- Jones DA, Takemoto D. 2004. Plant innate immunity - direct and indirect recognition of general and specific pathogen-associated molecules. *Curr Opin Immun.* 16:48-62.
- Meyers BC, Dieckman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J* 20:317-332.
- van der Biezen EA, Freddie CT, Kahn K, Parker JE, Jones JD. 2002. *Arabidopsis RPP4* is a member of the *RPP5* multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signalling components. *Plant J.* 29:439-51.
- Vieira LAA, Colombo C, Moraes A, Mehta A. *et al.* (2006) Brazilian coffee genome project: an EST-based genomic resource. *Braz J Plant Physiol.* 18(1):95-108.
- Videhyasekaran P. 1997. *Fungal pathogenesis in plants and crops. Molecular biology and host defense mechanisms.* Marcel Dekker, Inc. New York, 553 págs.