

EXPLORAÇÃO DO BANCO DE DADOS CafEST EM BUSCA DE PROVÁVEIS GENES DE *Coffea arabica* ENVOLVIDOS NA BIOSÍNTESE DE FENILPROPANÓIDES

Magnólia A. CAMPOS¹, E-mail: camposma@ufla.br; Fernanda C.L. MEDEIROS¹; Lívia M. PEREIRA¹; Mário Lúcio V. RESENDE¹; Marília S. SILVA²; Angela MEHTA³; Maria Fátima GROSSI-DE-SA³

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

Resumo:

Compostos naturais denominados fenilpropanóides apresentam funções na defesa vegetal, desempenhando papéis tanto na defesa pré-existente quanto na defesa induzida local e sistêmica em resposta ao ataque de patógenos. Prováveis genes que codificam para cada uma das 21 enzimas-chave envolvidas na biossíntese de fenilpropanóides em *C. arabica* foram identificados dentro do banco de dados brasileiro de genoma funcional de café (CafEST). De um total de 3559 ESTs selecionadas, 101 ESTs provavelmente representam a classe fenilalanina-amônia-liase (PAL). PAL catalisa a reação de desaminação de fenilalanina para gerar ácido cinâmico, dando início à rota dos fenilpropanóides. Os maiores números de ESTs, 521 e 490, foram encontrados representando as classes das enzimas cinamato 4-hidroxilase (C4H) e isoflavona O-metil-transferase (IOMT), respectivamente. Os menores números de ESTs encontrados, 21 e 36, representam as classes das enzimas chalcona isomerase (CHI) and cafeoil coenzima A O-metil-transferase (CCOMT). Análise detalhada dos 11 clusters de ESTs do tipo PAL revelou que seqüências de aminoácidos deduzidas compartilham alta similaridade (84-100%) com proteínas PAL isoladas das plantas *Coffea canephora*, *Ipomoea nil*, *Catharanthus roseus*, *Jatropha curcas* e *Ulmus pumila*. Além disso, a presença de uma possível ORF completa foi revelada. Ainda foi possível observar que das 101 ESTs identificadas para a classe PAL, 38 foram expressas em folhas de *C. arabica* sob diferentes condições, sendo a maioria encontrada em folhas de ramos plagiotrópicas de plantas adultas não tratadas com Bion. As ESTs do tipo PAL expressas em folhas estão presentes em 5 dos 6 contigs e em 2 singlets. A análise de 'neighbour-joining' gerou um filograma que agrupou cinco seqüências oriundas de contigs do tipo PAL de *C. arabica* em dois grupos maiores. A presença de possíveis membros de todas as enzimas-chave envolvidas na biossíntese de fenilpropanóides no genoma de *C. arabica* ainda não tinha sido relatada. Os prováveis genes identificados neste trabalho são candidatos para análises *in silico* e experimentais sobre o perfil de expressão. A elucidação da via de biossíntese de fenilpropanóides no metabolismo de café pode gerar impactos sobre o desfecho da resistência a doenças nesta cultura economicamente importante, levando a ganhos positivos sobre o agronegócio cafeeiro.

Palavras-chave: Genoma, resistência induzida, DNA, PAL, parasitos.

EXPLOITING THE CafEST DATABASE TO SEARCH *Coffea arabica* PUTATIVE GENES INVOLVED IN PHENYLPROPANOID BIOSYNTHESIS

Abstract:

Phenylpropanoid natural compounds encompass functions in plant defense, playing roles in preformed or inducible as well as local and systemic defense in response to pathogen attack. Putative genes encoding all of the 21 key enzymes involved in phenylpropanoid biosynthesis in *Coffea arabica* were identified within the Brazilian coffee functional genome database (CafEST). From a total of 3559 ESTs-reads present in *C. arabica*, 101 ESTs were found representing the phenylalanine ammonia-lyase (PAL) enzyme. PAL catalyzes the deamination reaction of phenylalanine to give cinnamic acid, which is the first step of the phenylpropanoid pathway. The higher numbers of ESTs, 521 and 490, were found representing the Cinnamate 4-hydroxylase (C4H) and Isoflavone O-methyl-transferase (IOMT) enzymes, respectively. The lower numbers of ESTs, 21 and 36, were found representing the Chalcone isomerase (CHI) and Caffeoyl coenzyme A O-methyl-transferase (CCOMT) enzymes. Detailed analysis of the 11 PAL-like-EST-clusters revealed that deduced amino acid sequences share high similarities (84-100%) with Pal proteins from *Coffea canephora*, *Ipomoea nil*, *Catharanthus roseus*, *Jatropha curcas* e *Ulmus pumila* plants. In addition, the presence of a putative complete PAL-like ORF. It was also possible to observe that of the 101 identified PAL-ESTs 38 were expressed under different conditions, being that the majority was found in plagiotropic leaves from adult plants untreated with Bion. The PAL-ESTs that were expressed in leaves were grouped within 5 of the 6 contigs and 2 singlets. A neighbour-joining analyses of aligned amino acid sequences generated a phylogram which grouped 5 PAL-like sequences from *C. arabica* contigs into two major groups. The presence of all members of key enzymes involved in phenylpropanoid biosynthesis in *C. arabica* genome has not yet been reported. The putative genes here identified are candidate to *in silico* and experimental expression profiles analyses. Elucidation of the phenylpropanoid pathway in coffee metabolism may strongly impact the outcome of the disease resistance in this economically important crop, leading to many positive gains on the coffee agribusiness.

Keywords: Genome, induced resistance, DNA, PAL, parasites.

Introdução

Fenilpropanóides são compostos naturais produzidos por plantas e que apresentam funções importantes na defesa vegetal contra pragas, predadores e patógenos. A via de fenilpropanóides é específica de plantas e gera uma grande gama de moléculas importantes na fisiologia vegetal, tais como fitoalexinas (stilbenos, coumarinas e isoflavonóides), flavonóides (protetores UV e antocianinas), moléculas de sinalização (AS, ácido salicílico) e monolignóis, que são as unidades do biopolímero lignina, o principal componente da parede celular de plantas (La Camera et al., 2004). Essa via é iniciada pela desaminação de fenilalanina para gerar ácido cinâmico, reação catalisada pela fenilalanina-amônia-liase (*Phenylalanine Ammonia-Lyase*, PAL, EC 4.3.1.5). Através de uma série de reações de hidroxilação, metilação e desidratação, vários fenilpropanóides simples são produzidos a partir de cinamato, incluindo ácidos p-coumárico, caféico, felúrico, sinápico. Lignina e suberina são formadas da mistura de fenilpropanóides simples, variando em composição conforme a espécie (Lewis e Yamamoto, 1990; Whetten e Sederoff, 1995). Seja exibindo atividade antimicrobiana de amplo espectro ou agindo em vias de sinalização de defesa, fenilpropanóides atuam direta ou indiretamente na resistência de plantas, desempenhando funções na defesa pré-formada ou induzida, tanto local quanto sistêmica, em resposta ao ataque de patógenos. Portanto, a elucidação e manipulação dos genes que codificam as enzimas-chave da biossíntese de fenilpropanóides no metabolismo secundário de cafeeiro poderia gerar grandes impactos no resultado da resistência a doenças nesta cultura agronomicamente importante, levando a muitos ganhos positivos no agronegócio cafeeiro. Um método eficiente para a descoberta de genes e que ainda pode levar a inferências sobre a expressão gênica é extrair informações de coleções de seqüências expressas (ESTs) presentes em bancos de dados específicos. O seqüenciamento em larga escala de ESTs de cafeeiro foi realizado como uma iniciativa do Consórcio Brasileiro de Pesquisas e Desenvolvimento de Café em parceria com a rede AEG-FAPESP e Embrapa Recursos Genéticos e Biotecnologia. Usando a estratégia de contagem de ESTs de genes anotados, as facilidades do banco de dados brasileiro do genoma funcional de café (CafEST) foram usadas para buscar por similaridade enzimas-chave da biossíntese de fenilpropanóides presentes no genoma funcional de *C. arabica*. Prováveis genes que codificam membros de todas as enzimas envolvidas nesta via foram encontrados nesta espécie de café, os quais são discutidos neste trabalho.

Material e Métodos

Para encontrar seqüências de *C. arabica* homólogas a genes que codificam cada uma das enzimas da via de biossíntese de fenilpropanóides dentro do banco de dados do genoma café (CafEST, <http://www.lge.ibi.unicamp.br/cafe/>), análises de busca por comparação foram feitas utilizando o programa tBLASTn (Altschul et al., 1997) e seqüências de aminoácidos bem descritas no GenBank (NCBI, www.ncbi.nlm.nih.gov) para cada uma das enzimas dessa via (Dixon et al., 2002). Usando os critérios de *E-value* $e-10^{-04}$ e matriz BLOSUM62, um total de 3559 ESTs foi selecionado do banco de dados CafEST (Tabela 1), o qual cobre um número de 173967 reads úteis.

Clusters de ESTs foram construídos para cada classe de enzima separadamente, usando o programa CAP3 (*Contig Assembly Program*) (Huang e Madan, 1999). Em seguida, a seqüência consenso de cada contig usado em análise filogenética foi comparada com seqüências de aminoácidos de proteínas homólogas depositadas no Genbank, usando o programa BLASTx (Altschul et al., 1997). Alinhamento final das seqüências de aminoácidos foi obtido no Clustal X 1.83 (Thompson et al., 1997), o qual foi posteriormente usado na construção de árvore filogenética no programa MEGA 3 (Kumar et al., 2004), através do método de Neighbor joining e análise de *bootstrap* com 1.000 replicatas.

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

Prováveis genes que codificam para cada uma das 21 enzimas-chave envolvidas na biossíntese de fenilpropanóides em *C. arabica* foram identificados dentro do banco de dados brasileiro do genoma café (CafEST). Com base em buscas por similaridades usando o programa tBLASTn (Altschul et al., 1997), um total de 3559 ESTs relacionadas a esta via biossintética foi selecionado (Tabela 1). Entre estas, 101 ESTs provavelmente representam a classe fenilalanina-amônia-liase (PAL), as quais foram agrupadas em em 11 clusters. PAL catalisa a reação de desaminação de fenilalanina para gerar ácido cinâmico que dá início à rota dos fenilpropanóides. Os maiores números de ESTs, 521 e 490, foram encontrados representando as classes das enzimas cinamato 4-hidroxilase (C4H) e isoflavona O-metil-transferase (IOMT), respectivamente, os quais foram distribuídos em em 139 e 37 clusters. Os menores números de ESTs encontrados, 21 e 36, representam as classes de enzimas chalcona isomerase (CHI) and cafeoil coenzima A O-metil-transferase (CCOMT), respectivamente, os quais foram ambos agrupados em 6 clusters.

Tabela 1. Distribuição de ESTs presentes no CafEST que provavelmente codificam enzimas da biossíntese de fenilpropanóides em *C. arabica*.

Enzima de Biossíntese de Fenilpropanóides		Nº Total de ESTs no CafEST	Nº de Clusters no CafEST	
Nome	Sigla		Contigs	Singlets
<i>Phenylalanine ammonia-lyase</i>	PAL	101	6	5
<i>Cinnamate 4-hydroxylase</i>	C4H	521	80	59
<i>4-Coumarate:coenzyme A ligase</i>	4CL	158	5	0
<i>Caffeic acid O-methyl-transferase</i>	COMT	55	7	9
<i>Caffeoyl coenzyme A O-methyl-transferase</i>	CCOMT	36	3	3
<i>Ferulate 5-hydroxylase</i>	F5H	230	31	32
<i>Cinnamyl coenzyme A reductase</i>	CCR	154	21	22
<i>Cinnamyl alcohol dehydrogenase</i>	CAD	153	22	22
<i>Laccase</i>	LAC	59	14	9
<i>Chalcone synthase</i>	CHS	171	9	5
<i>Chalcone reductase</i>	CHR	205	7	0
<i>Chalcone isomerase</i>	CHI	21	5	1
<i>Flavanone 3-B-hydroxylase</i>	F3BH	141	25	13
<i>Flavonoid 3-B-hydroxylase</i>	F3'H	153	29	25
<i>Flavonoid 3', 5'-B-hydroxylase</i>	F3'5'H	235	44	44
<i>Dihydroflavonol reductase</i>	DFR	102	18	14
<i>Anthocyanidin synthase</i>	ANS	117	19	18
<i>Isoflavone synthase</i>	IFS	203	34	40
<i>Isoflavone O-methyl-transferase</i>	IOMT	490	19	18
<i>Isoflavone 2'-hydroxylase</i>	I2'H	119	27	26
<i>Isoflavone reductase</i>	IFR	135	19	11
Nº Total		3559		

Os clusters de ESTs do tipo PAL foram analisados em detalhes. Uma busca por proteínas com seqüências de aminoácidos similar aquelas deduzidas a partir dos contigs foi realizada e as melhores similaridades foram encontradas via Blast com seqüências de PAL isoladas das plantas *Coffea canephora*, *Ipomoea nil*, *Catharanthus roseus*, *Jatropha curcas* e *Ulmus pumila* (Tabela 2). De um total de 32 bibliotecas usando cDNAs da espécie *C. arabica*, ESTs que codificam PAL não foram encontradas em 09 bibliotecas, as quais foram construídas sob as seguintes condições: plântulas e folhas tratadas com ácido araquidônico, raízes tratadas e não-tratadas com Bion, células de calos, calos primários de linhagem embriogênica, sementes e embriões de frutos e sementes germinando. Nas demais 23 bibliotecas foram encontradas ESTs relacionadas a PAL, as quais foram agrupadas em contigs quiméricos, formados por ESTs expressas em diferentes órgãos/tecidos sob diferentes condições, sem padrão aparente para a expressão gênica. Entretanto, as 11 seqüências de aminoácidos preditas a partir de contigs e singlets compartilham alta similaridade com proteínas PAL disponíveis no GenBank (NCBI). Já foi demonstrado que a atividade da enzima PAL é dificilmente detectada em cotilédones de café em desenvolvimento, enquanto que em folíolos é 900 vezes maior que nos cotilédones (Aerts e Baumann, 1994). Das 101 ESTs identificadas para a classe PAL, 38 foram expressas em folhas de *C. arabica* sob diferentes condições, sendo a maioria encontrada em folhas de ramos plagiotrópicas de plantas adultas não-tratadas com Bion. Com exceção do Contig C3, ESTs do tipo PAL expressas em folhas estão presentes nos demais contigs e em 2 singlets. Três prováveis genes do tipo PAL foram expressos em ramos de *C. arabica* infectados com *Xylella fastidiosa*, os quais foram agrupados dentro dos contigs C1, C4 e C5. Se transcritos para estes 3 membros de genes do tipo PAL apresentam um padrão de expressão constitutivo e/ou induzido por patógenos precisa ser demonstrado experimentalmente.

Tabela 2. Distribuição de ESTs de *C. arabica* do tipo PAL em contigs e similaridades da sequência de aminoácidos deduzida com proteínas do GenBank.

Contig	Nº de ESTs	Tamanho (nt)	Seqüência de melhor similaridade via Blast			
			Organismo	Nº de acesso	E-value	Similaridade
C1	6	1510	<i>Catharanthus roseus</i>	dbj BAA95629.1	0	330/355 (92 %)
C2	25	2031	<i>Ipomoea nil</i>	gb AAG49585.1	0	531/557 (93 %)
C3	4	1539	<i>Jatropha curcas</i>	gb ABI33979.1	1 e ⁻¹⁷⁹	366/417 (84 %)
C4	30	2652	<i>Coffea canephora</i>	gb AAN32866.1	0	717/717 (100 %)
C5	26	2699	<i>Jatropha curcas</i>	gb ABI33979.1	0	656/710 (90 %)
C6	4	375	<i>Ulmus pumila</i>	gb AAV82486.1	1 e ⁻²¹	51/56 (91 %)

As seqüências de aminoácidos deduzidas do tipo PAL de *C. arabica* e seqüências similares foram posteriormente usadas em análises filogenéticas, com exceção do C6 devido ao menor comprimento da seqüência. Uma árvore baseada na análise de ‘neighbour-joining’ foi gerada a partir de alinhamento múltiplo entre seqüências de aminoácido preditas e similares (Figura 6). As cinco seqüências oriundas de contigs do tipo PAL de *C. arabica* foram agrupadas em dois grupos maiores. O primeiro grupo é constituído pelo Contig C1 e seqüência homóloga de *Catharanthus roseus*, além dos contigs C3 e C5 que parecem mais similares entre si do que com suas seqüências homólogas. O segundo grupo é composto por uma seqüência de *Jatropha curcas*, pelo contig C2 e homólogo de *Ipomoea nil* e pelo contig4 e homólogo de *Coffea canephora*. No contig C4 foi encontrada uma seqüência de 2151 nucleotídeos de comprimento que parece ser uma ORF (*open reading frame*) completa, cuja seqüência de 717 aminoácidos preditos compartilha 99% de identidade (100% de similaridade) com uma PAL de *C. canephora* (citada na Tabela 2).

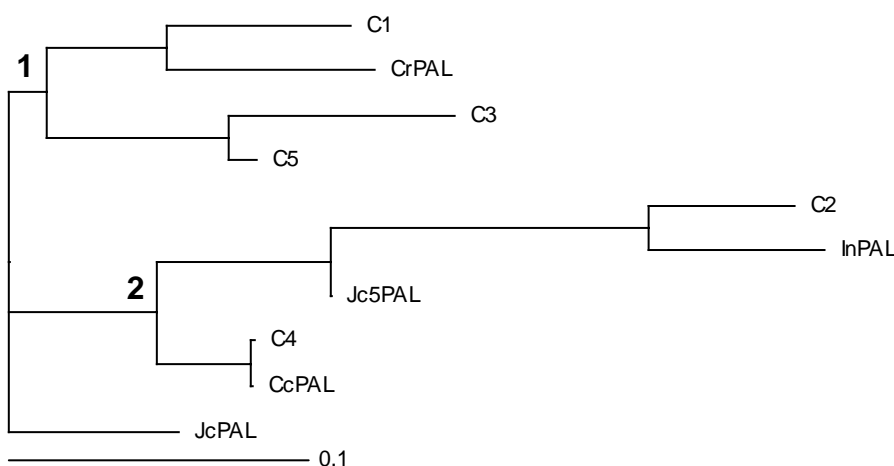


Figura 1. Filograma da análise de *neighbour-joining* de seqüências de aminoácidos de PAL de *C. arabica* e similares, construído e visualizado no Programa MEGA a partir de alinhamento múltiplo gerado no ClustalX. C1 a C5 indicam contigs do tipo PAL de *C. arabica* e as seqüências similares obtidas no GenBank EMBL estão indicadas por: CcPAL, *Coffea canephora*; JcPAL, *Jatropha curcas*; CrPAL, *Catharanthus roseus*; InPAL, *Ipomoea nil*, conforme detalhes na Tabela 2.

Conclusão e Perspectivas

A presença de possíveis membros de todas as enzimas-chave envolvidas na biossíntese de fenilpropanóides no genoma de *C. arabica* ainda não tinha sido relatada. Os prováveis genes identificados neste trabalho são candidatos para análises *in silico* e experimentais sobre o perfil de expressão para validação funcional. Isto poderia trazer informações adicionais sobre a regulação positiva e negativa da expressão gênica sob diferentes condições da planta, “cross-talk” entre vias de transdução de sinais de salicilato e jasmonato e canalização de intermediários da via. Ainda, o entendimento da regulação da biossíntese de fenilpropanóides abre perspectivas também para a manipulação genética dessa via (engenharia metabólica) visando controle genético de doenças do cafeeiro e alteração de sabor da bebida, ambas características afetadas por compostos fenilpropanóides. Portanto, a elucidação da via de biossíntese de fenilpropanóides no metabolismo

de café pode gerar impactos sobre o desfecho da resistência a doenças nesta cultura economicamente importante, levando a ganhos positivos sobre o agronegócio cafeeiro.

Agradecimentos

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

Referências Bibliográficas

Aerts, R. J.; Baumann, T. W. (1994) Distribution and utilization of chlorogenic acid in *Coffea* seedlings. *Journal of Experimental Botany* 45: 497-503.

Altschul, S.F.; Madden, T.L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25:3389-3402.

Dixon, R. A.; Achnine, L.; Kota, P.; Liu, C.; Srinivasa Reddy, M. S.; Wang, L. (2002). The phenylpropanoid pathway and plant defence — a genomics perspective. *Molecular Plant Pathology*, 5: 371-390.

Huang, X.; Madan, A. (1999). CAP3: A DNA Sequence Assembly Program. *Genome Research*, 9: 868-877

Kumar, S.; Tamura, K.; Nei, M. (2004). MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Briefings in Bioinformatics*, 5: 150-163.

La Camera, S.; Gouzerh, G.; Dhondt, S.; Hoffmann, L.; Fritig, B.; Legrand, M.; Heitz T. (2004). Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways *Immunological Reviews*, 198: 267.

Lewis, N.; Yamamoto, E. (1990). Lignin: occurrence, biogenesis and biodegradation. *Physiology and Molecular Plant Pathology*, 41: 455-496.

Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 24: 4876-4882.

Vieira, L.G. E.; Andrade, A. C.; Colombo, C. A.; Araújo, A. H.; Mehta, A. (2006). Brazilian coffee genome project: an EST-based genomic resource. *Brazilian Journal of Plant Physiology*, 18:95-108

Whetten R.; Sederoff R. (1995). Lignin biosynthesis. *Plant Cell*. 7: 1001-1013.