

IDENTIFICAÇÃO DE ESTS TECIDO-ESPECÍFICAS NO BANCO DE DADOS DO GENOMA FUNCIONAL DE *Coffea* spp.

Daiene B. M. SANTOS¹, E-mail: daienebms@pop.com.br; Juliana D. ALMEIDA¹; Leila M. G. BARROS¹; Mauro CARNEIRO¹; Felipe R. DA SILVA¹

¹Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF

Resumo:

O projeto Brasileiro Genoma Café foi elaborado e executado com o objetivo de fornecer informações sobre o genoma do cafeeiro aos pesquisadores que desenvolvem variedades melhoradas, em busca da produtividade e qualidade de grãos. Nesse projeto foram seqüenciados 218.150 clones distintos de ESTs (Expressed Sequenced Tags) escolhidos aleatoriamente em 49 bibliotecas de cDNA de *C. arabica*, *C. canephora* e *C. racemosa*, as quais representam diferentes tecidos em estágios específicos de desenvolvimento e tecidos submetidos a estresses biótico e abiótico. Após a remoção de contaminações e seqüências de baixa qualidade, as 154.770 ESTs restantes foram agrupadas em 45.366 *clusters* (UniGenes), dos quais cerca de 27% não apresentam similaridade significativa com seqüências protéicas já descritas. A fim de selecionar genes preferencialmente expressos na raiz, no fruto, na folha e no botão floral, realizamos uma análise *in silico*, visando a prospecção dos seus respectivos promotores em um próximo passo. Para tal, foram feitas comparações por meio do Teste Exato de Fisher, entre grupos formados por bibliotecas de ESTs obtidas a partir de um único tipo de tecido e grupos formados pelas bibliotecas restantes, tendo-se como resultado a seleção de UniGenes que possuem grande chance de serem tecido-específicos. Dessa forma, foram selecionados 103 UniGenes que apresentam níveis significativamente distintos de transcritos oriundos de cada um dos tecidos, sendo 18 de folhas, 40 de frutos, 14 de raiz e 31 de botões florais. Dentre os Unigenes selecionados, foram escolhidos três de cada um dos tecidos e um constitutivo para as validações experimentais. O critério de escolha desses Unigenes se baseou no grau de ineditismo, na especificidade dos unigenes e no nível de expressão. Para confirmar o caráter de tecido especificidade dos genes escolhidos, serão realizadas análises de *Real Time RT-PCR* e *Northern blot*.

Palavras-chave: Café, ESTs, promotores.

IDENTIFICATION OF TISSUE-SPECIFIC ESTS IN THE *Coffea* spp. FUNCTIONAL GENOMIC DATABASE

Abstract:

The Brazilian Coffee Genome Project generated a total of 218,150 EST (Expressed Sequenced Tags) sequences of randomly chosen clones coming from 49 cDNA libraries made from *C. arabica*, *C. canephora* or *C. racemosa*. Those libraries were made of mRNA extracted from several distinct tissues, developmental stages and forms of biotic and abiotic stresses. Contamination and low quality sequences were removed and the 154,770 valid ESTs were grouped in 45,366 clusters (UniGenes), of which about 27% have unknown function. In order to locate candidate tissue-specific promoters, an *in silico* analysis was performed to identify genes preferentially expressed on roots, fruits, leaves and flowers. We've constructed groups of EST libraries, each coming from one tissue type, and have used the Fischer's Exact Test to compare each one to a group formed by the remaining libraries (i.e., the libraries not included on that particular group) and identified 103 UniGenes with a high chance of being tissue-specific: 18 leaf-, 40 fruit-, 14 root- and 31 flower-specific. The expression level, specificity and uniqueness of the previously identified UniGenes were used to select 3 from each tissue to experimental validation. Real Time RT-PCR and Northern blot analysis will be employed to verify those 12 genes tissue-specificity.

Key words: Coffee, ESTs, promoters

Introdução

A produção comercial do café é baseada em duas espécies: *C. arabica* L e *C. canephora* Pierre, as quais representam 70% e 30% do mercado, respectivamente. O Brasil é o maior produtor e exportador mundial de café, respondendo por cerca de um quarto da produção mundial. Contudo, a cafeicultura apresenta uma série de problemas que podem limitar a produção nacional e que requisitam cultivares mais adaptados. A obtenção de melhores cultivares por métodos clássicos de melhoramento é dificultada por aspectos intrínsecos como: a perenidade, a baixa variabilidade no gênero *Coffea* e as dificuldades de cruzamento de *Coffea arabica* devido a diferenças de ploidia. Nesse caso, as técnicas de biotecnologia, entre elas a transgenia, se apresentam como ferramentas bastante úteis na aceleração da obtenção dos resultados esperados. O projeto

Brasileiro Genoma Café foi elaborado e executado na tentativa de auxiliar os pesquisadores na obtenção de variedades mais adequadas ao cenário atual como, por exemplo, tolerantes a estresses bióticos e abióticos ou que produzam grãos de melhor qualidade. Nesse projeto foram seqüenciados 218.150 clones de ESTs (Expressed Sequenced Tags) escolhidos aleatoriamente de 49 bibliotecas de cDNA de *C. arabica*, *C. canephora* e *C. racemosa*, representando estágios específicos de desenvolvimento de tecidos e tecidos submetidos a estresse biótico e abiótico. Após a remoção de contaminações e seqüências de baixa qualidade, as 154.770 ESTs restantes foram agrupadas em 45.366 *clusters* (UniGenes), dos quais 26.242 são formados por uma única seqüência (Sales *et al.*, 2005). Cerca de 27% dos UniGenes não apresentam similaridade significativa (blastx com *e-value* melhor que 10^{-4}) com seqüências protéicas já descritas. De posse de um determinado UniGene, é possível detectar seu promotor. Assim, promotores inéditos poderão ser isolados, caracterizados e potencialmente utilizados para agregação de valor nas mais diversas culturas e no caso específico do café visando resistência a patógenos, tolerância à seca, uniformidade de maturação de frutos e qualidade dos frutos, trazendo benefícios no manejo das culturas e para as características pós-colheita. A importância da obtenção desses promotores reside no fato de que uma das barreiras a utilização da transgenia é o fato de que as plantas transgênicas comercializadas hoje em dia possuem transgenes comandados por promotores constitutivos que fazem com que sua expressão ocorra em todos os tecidos da planta ao invés de somente no local onde é necessária. Uma forma de atingir a expressão pontual do transgene de interesse é utilizar promotores tecido específicos em sua confecção.

O objetivo deste trabalho foi identificar, no banco UniGene do cafeeiro, ESTs abundantes e inéditas preferencialmente expressas em raiz, folha, flor ou fruto, que no futuro serão utilizadas na prospecção de seus respectivos promotores. Esses promotores, além de promover a especificidade de expressão dos genes por eles comandados poderão ser patenteados e utilizados em programas de melhoramento sem onerá-los com o pagamento de royalties de promotores protegidos.

Material e Métodos

Foram realizados Testes Exatos de Fisher, *in silico* (Sales & Silva, 2005), contrastando bibliotecas de ESTs de um único tecido contra bibliotecas dos demais tecidos. Primeiramente, os grupos a serem contrastados eram criados para então serem selecionados para o Teste de Fisher. O primeiro grupo de ESTs de um único tecido a ser criado foi de raiz. Em seguida, foi o grupo formado por ESTs de folhas, sendo o mesmo feito com os grupos das ESTs de fruto e de botão floral. Cada um desses grupos foi contrastado com um grupo formado por todas as bibliotecas dos demais tecidos, que não o dele, presentes no banco de dados do Genoma Café. Como resultado, os Testes selecionaram os UniGenes que apresentam expressão preferencial ou exclusiva nos grupos de cada um dos tecidos. O Teste Exato de Fisher já apresenta em seus resultados o Blast de cada unigene, indicando quais são homólogos à seqüências já conhecidas e quais são inéditos (Figura 1).

Após esse *screening* virtual foram escolhidos entre os unigenes resultantes dos Testes Exatos de Fisher, 3 representantes de cada tecido. Um unigene constitutivo também foi selecionado para servir de controle positivo aos experimentos. Para cada um dos unigenes foram desenhados, para um ensaio preliminar de Reação da Polimerase em Cadeia utilizando Transcriptase Reversa (RT-PCR), pares de *primers* com 20 nucleotídeos, contendo os sítios de restrição *Spe I* e *Not I*, de forma que amplificassem fragmentos entre 400 e 500 pb. Um clone (EST) contendo a região do fragmento amplificado foi escolhido em cada um dos unigenes, para, a partir deles, serem sintetizadas por PCR ou restrição as sondas que serão usadas no ensaio de *Northern blot*.

Para uma outra abordagem do estudo de expressão dos unigenes selecionados virtualmente será realizado um ensaio de Real Time-PCR. Para tal, foi desenhado para cada unigene escolhido, um par de *primers* de 20 nucleotídeos de forma que amplificassem fragmentos entre 90 e 170 pb e tivessem uma temperatura de anelamento (T_m) entre 59 e 61 °C.

Resultados e Discussão

Os contrastes entre bibliotecas de diferentes tecidos e o conjunto das bibliotecas restantes apontaram 103 UniGenes preferencialmente expressos, sendo 18 de folhas, 40 de frutos, 14 de raiz e 31 de botões florais (Tabela 1). A comparação destas seqüências com seqüências protéicas já descritas, utilizando-se BlastX, mostrou que 84% dos *clusters* apresentam similaridade a seqüências conhecidas e 16% são inéditos.

Devido a impossibilidade de trabalhar com todos os unigenes (Unigenes) selecionados, foram escolhidos três representantes de cada tecido levando em consideração se eram formados apenas por bibliotecas do tecido em questão, se sua seqüência era inédita e se apresentavam uma boa expressão relativa em relação ao total de *reads* das bibliotecas que o compunham. Como unigene constitutivo foi escolhido aquele formado pelo maior número possível de bibliotecas de todos os tecidos. Para produzir a sonda a ser usada no *Northern blot* por PCR, foram desenhados *primers* específicos para cada um dos unigenes escolhidos para cada tecido e mais o constitutivo. Para o ensaio de Real Time PCR foram desenhados outros pares de *primers* adequados as exigências da técnica. Dessa forma, a próxima etapa será validar os UniGenes selecionados mediante ensaios de *Northern blot* e RT-PCR e subseqüentemente usar os genes validados na prospecção de seus respectivos promotores.

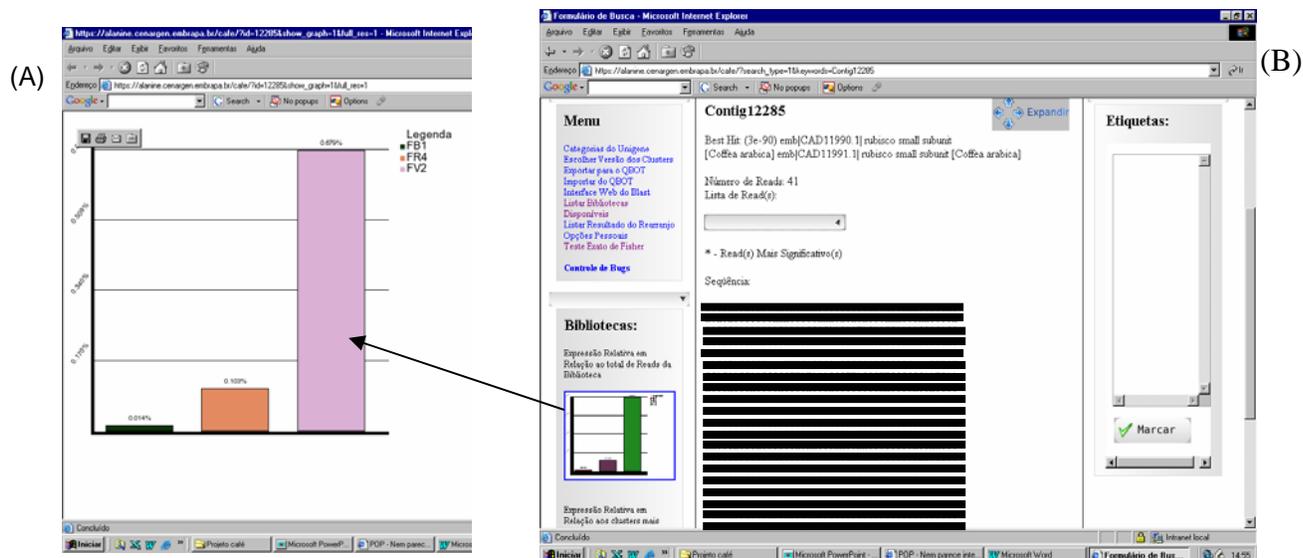


Figura 1 - Demonstração de alguns resultados proporcionados pelo Teste Exato de Fisher, como a expressão relativa em relação ao total de *Reads* da Biblioteca (A) e o Blast (B).

Conclusões

No Banco de Dados do Genoma Funcional de *Coffea* spp existem 18 genes preferencialmente expressos em folha, 40 em frutos, 14 em raiz e 31 em botões florais. Desses, 84% apresentam similaridade a seqüências conhecidas e 16% são inéditos.

Referências Bibliográficas

Sales RMOB, Andrade AC, da Silva FR (2005) Determinação do Unigene do Projeto Genoma Café. In: Anais do IV Simpósio de Pesquisa dos Cafés do Brasil. Londrina, Brasil. Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café. CD-ROM.

Sales RMOB, Silva FR (2005) Desenvolvimento de Ferramenta Web para Análise de Dados de Projetos Genoma do tipo EST. In: Anais do X Encontro do Talento Estudantil da Embrapa Recursos Genéticos e Biotecnologia. Brasília-DF.

Tabela 1 - Relação das bibliotecas que compunham os grupos contrastados, as comparações realizadas e seus respectivos resultados.

Nome dos Grupos	Raiz	Não Raiz	Folha	Não Folha	Flor	Não Flor	Fruto	Não Fruto
Bibliotecas	RT3	AR1	LV4	AR1	FB1	AR1	FV2	AR1
	RT4	FB1	LV5	BP1	FB2	BP1		BP1
	RT5	FB2	LV8	BP2	FB4	BP2		BP2
		FB3	LV9	CA1	FR1	CA1		CA1
		FB4		CB1	FR2	CB1		CB1
		FR1		CD1		CD1		CD1
		FR2		CL1		CL1		CL1
		FR4		CL2		CL2		CL2
		FV2		CM1		CM1		CM1
		IA2		CS1		CS1		CS1
		IC1		EA1		EA1		EA1
		LV4		EC1		EC1		EC1
		LV5		FB1		FP2		FB1
		LV8		FB2		FR4		FB2
		LV9		FB3		FV2		FB3
		PC1		FB4		IA2		FB4
		RM1		FP2		IC1		FP2
		RX1		FR1		LM3		FR1
		SH3		FR2		LP1		FR2
				FR4		LV4		FR4
				FV2		LV5		IA2
				IA2		LV8		IC1
				IC1		LV9		LM3
				LM3		PA1		LP1
				LP1		PC1		LV4
				PA1		RM1		LV5
				PC1		RT3		LV8
				RM1		RT5		LV9
				RT3		RT7		PA1
				RT5		RT8		PC1
				RT7		RX1		RM1
				RT8		SH2		RT3
				RX1		SH3		RT5
				SH2		SI3		RT7
				SH3				RT8
				SI3				RX1
								SH2
								SH3
								SI3
	Raiz X Não Raiz		Folha X Não Folha		Flor X Não Flor		Fruto X Não Fruto	
Clusters Total	39		117		94		57	
Clusters Seleccionados	14		18		31		40	
Clusters Inéditos	2		0		6		8	