

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE ESTs PRESENTES NO BANCO DO GENOMA CAFÉ RELACIONADOS A PROTEASES

Guilherme G. BAPTISTA¹, Paulo MAZZAFERA², Regina H. G. PRIOLLI¹, Milton MASSAO SHIMIZU², Natalie R. LEÓZ SCHOCKEN¹ e Carlos COLOMBO¹ E-mail: ccolombo@iac.sp.gov.br

¹ Recursos Genéticos – IAC; ² Departamento de Fisiologia Vegetal – IB-UNICAMP

Resumo:

Evidências recentes indicam que proteases podem ter importante papel na qualidade de bebida de cafeeiros cultivados em regiões com temperaturas médias anuais contrastantes. Assim, o objetivo do presente trabalho foi identificar ESTs apresentando alta similaridade com proteases no banco de dados do Projeto Genoma Café. Cerca de 1800 reads foram identificados e blastados. Contigs contendo em sua maioria seqüências de bibliotecas de frutos foram classificadas nos em serinas, cisteínas, aspárticas e metalo proteinases. Em seguida tais contigs foram submetidos a uma análise filogenética com seqüências similares encontradas em bancos públicos como também identificados os domínios funcionais das mesmas. Primers para amplificação destes domínios foram desenhados e o próximo passo será verificar o nível de expressão destes genes em endospermas de frutos de café provenientes de locais com temperaturas contrastantes.

Palavras-chave: Protease, café, EST, qualidade de bebida, nitrogênio.

Abstract:

Recent evidences indicate that proteases may have an important role in the quality of beverage obtained from coffee beans from plants cultivated in areas contrasting in annual mean temperatures. Therefore, the aim of this study was to identify ESTs displaying high similarity to proteases in the Coffee Genome database. Approximately 1,800 reads were identified and blasted. Then we selected the contigs displaying similarity with proteases from seeds and they were classified as serine, cysteine, aspartate or metallo proteinases. Subsequently, the contigs were subjected to phylogenetic analysis and we also identified their functional domains. Primers based on these domains were designed and in a next step they will be used to study the expression of these genes in the endosperm of coffee fruits collected from plants growing in areas contrasting in temperature.

Keywords: proteases, coffee, EST, beverage quality, nitrogen.

Introdução:

O mercado café vem seguindo uma tendência pela melhoria da qualidade de bebida. Vários fatores estão relacionados com a obtenção de café de boa qualidade, como o processamento pós-colheita de frutos que representa uma etapa fundamental para essa obtenção. Os produtores atentos a esta demanda vem procurando cafeeiros com qualidades fisiológicas específicas, que apresentem um bom desempenho, tanto ao nível de produção quanto de qualidade final do produto, em regiões geográficas definidas. Outro fator que influencia diretamente a qualidade do café é a composição química da semente tanto durante o desenvolvimento do fruto quanto após os tratamentos pós-colheita.

Em café, o aroma pode formado por mais de 800 compostos voláteis e os precursores relacionados com a formação dos mesmos são: açúcares, aminoácidos, peptídeos, trigonelina, ácidos clorogênicos, ácidos orgânicos, lipídeos e carotenóides (Clifford, 1985). Interações entre açúcares redutores e o grupamento amino de aminoácidos e peptídeos, ou mesmo de proteínas (reação de Maillard) são consideradas essenciais para o desenvolvimento de aroma. Além do conteúdo endógeno, outra fonte de açúcares para a ocorrência da reação de Maillard são os açúcares redutores formados durante a torração a partir da quebra sacarose já presente na semente e acumulada durante sua formação.

Sementes de café apresentam atividade de protease e pode ser encontrada em sementes secas (Ludwig et al., 2000), além de aminoácidos. É provável também a presença de fragmentos de polipeptídeos, como mostram os trabalhos de Rogers et al. (1999) e Ludwig et al. (2000). Entretanto, não se sabe ao certo qual a contribuição dos mesmos na qualidade da bebida (Montavon et al., 2003). Selmar et al. (2001) destacam que alanina é um importante aminoácido no desenvolvimento de compostos durante a torração de café, através da reação de Maillard. Junto com asparagina, ácido glutâmico, ácido aspártico e serina, alanina quantificam para aproximadamente 80% dos aminoácidos livres em sementes de café (Arnold and Ludwig, 1996; Shimizu and Mazzafera, 2000). Por outro lado, alanina não está entre os aminoácidos mais abundantes na principal proteína de reserva em café (Rogers et al., 1999), como também não está entre os aminoácidos em maior quantidade quando analisada a constituição de aminoácidos no hidrolisado de proteínas totais da semente de café (Shimizu and Mazzafera, 2000).

Silva (2004) detectou, em plantas de café cultivadas em condições de temperatura média anual contrastantes (mais amena e mais quente), que este fator climático pode ter forte influência na qualidade e quantidade de proteínas das sementes, que por sua vez depende da ação de proteases. Ainda, evidenciou que no período de enchimento de grãos as sementes possuem proteínas com alto peso molecular e pouca atividade de protease. Além dos locais, as plantas foram submetidas a diferentes procedimentos de irrigação, não tendo sido detectado qualquer efeito destes tratamentos nos níveis

de qualidade e quantidade de proteínas das sementes. Seu estudo demonstrou que cafês colhidos em local com temperatura média mais amena produz bebida de melhor qualidade, pois além das diferenças bioquímicas observadas, testes de xicara e com a língua eletrônica.

Sendo assim, o presente estudo tem por objetivo principal identificar e caracterizar seqüências de ESTs depositadas no banco de seqüências do Projeto Genoma Café supostamente envolvidas com a ação de proteases.

Material e métodos

Foram utilizadas ferramentas de bioinformática disponíveis na página web do Projeto Genoma Café para localizar, categorizar e clusterizar, por comparação, os reads com elevada similaridade (e-value menor que 1^{-20}) aos genes de proteases localizadas em bancos genômicos públicos (NCBI - National Center for Biotechnology Information). Os reads encontrados foram categorizados em proteinases serinas, cisteínas, aspárticas e as metalo-proteinases. Após clusterizados, foram selecionados contigs com alta incidência de reads provenientes de bibliotecas de frutos e estes foram submetidos à busca de domínio conservado por meio da ferramenta Pfam (Protein families database of alignments and HMMs). Após análise da qualidade da seqüência, foram desenhados primers flanqueando estes domínios por meio do programa GeneRunner para estudos futuros de expressão. O estudo filogenético foi feito utilizando o programa de alinhamento múltiplo Malign com seqüências similares encontrados em outras espécies vegetais e a partir dos dados obtidos foram feitas análises de parcimônia para a obtenção de árvores filogenéticas utilizando o programa filogenético PAUP.

Resultados e discussão:

Na mineração dos dados do Genoma Café foram encontrados mais de 1800 reads potencialmente envolvidos com proteases e destes as proteinases serinas correspondem a 8% do total de reads, as cisteínas 47%, as aspárticas 16%, as metalo-proteinases 8%, endopeptidases não classificadas 3% e as demais seqüências relacionadas a proteases não foram categorizadas.

Os reads de serinas apresentaram 2 contigs após clusterização, com predominância de reads vindos de bibliotecas de frutos (contigs com pelo o menos 5 reads com no mínimo 40% representados por bibliotecas de frutos). Os reads de cisteínas revelaram 4 contigs, do grupo das aspárticas 1 contig, assim como aqueles representando as metalo-proteinases, conforme mostra a Tabela 1.

Nome do Contig	Cisteína		Aspártica		Metallo		Serina	
	Nº reads	Frutos (%)	Nº reads	Frutos (%)	Nº reads	Frutos (%)	Nº reads	Frutos (%)
C01	210	25 (11,9)	77	5 (6,5)	28	4 (14,3)	28	3 (10,7)
C02	117	17 (14,5)	15	5 (33,3)	6	1 (16,7)	17	4 (23,5)
C03	59	3 (5,1)	11	3 (27,3)	5	3 (60,0)	13	1 (7,7)
C04	46	8 (17,4)	11	1 (9,1)	4	3 (75,0)	9	3 (33,3)
C05	37	5 (13,5)	9	1 (11,1)	3	2 (66,7)	8	8 (100)
C06	35	3 (8,6)	7	3 (42,9)	3	2 (66,7)	8	2 (25,0)
C07	25	4 (16,0)	5	1 (20,0)	3	1 (33,3)	5	4 (80,0)
C08	23	2 (8,7)	4	2 (50,0)	2	2 (100)	4	1 (25,0)
C09	20	1 (5,0)	4	1 (25,0)	2	1 (50,0)	3	3 (100)
C10	18	4 (22,2)	3	1 (33,3)	2	1 (50,0)	3	2 (66,7)
C11	17	1 (5,9)	3	1 (33,3)	2	1 (50,0)		
C12	12	11 (91,7)	2	2 (100)	2	1 (50,0)		
C13	9	9 (100)	2	1 (50,0)	2	1 (50,0)		
C14	8	8 (100)						
C15	8	7 (87,5)						
C16	8	1 (12,5)						
C17	4	4 (100)						
C18	4	1 (25,0)						

Tabela 1 – Principais resultados da clusterização, que continham reads de bibliotecas de frutos, de 4 proteinases encontradas no banco de dados do Genoma Café, com o número total de reads de cada contig e o número e porcentagem de reads de frutos.

As análises filogenética e as árvores genéticas produzidas se encontram na Figura 1. Nas quatro árvores obtidas o contig de cada grupo de proteinase revelou ser mais distante das demais espécies incluídas na análise. No grupo aspártico o café se aproxima mais de batata e tomate, se distanciando das demais espécies, provavelmente menos aparentadas que as primeiras. Nas comparações de metalo e serino proteinases, café aparece mais próxima de arabidopsis, seguido pelo tomate, enquanto que para cisteína proteinase o café se aproxima mais de Helianthus, seguido por arabidopsis.

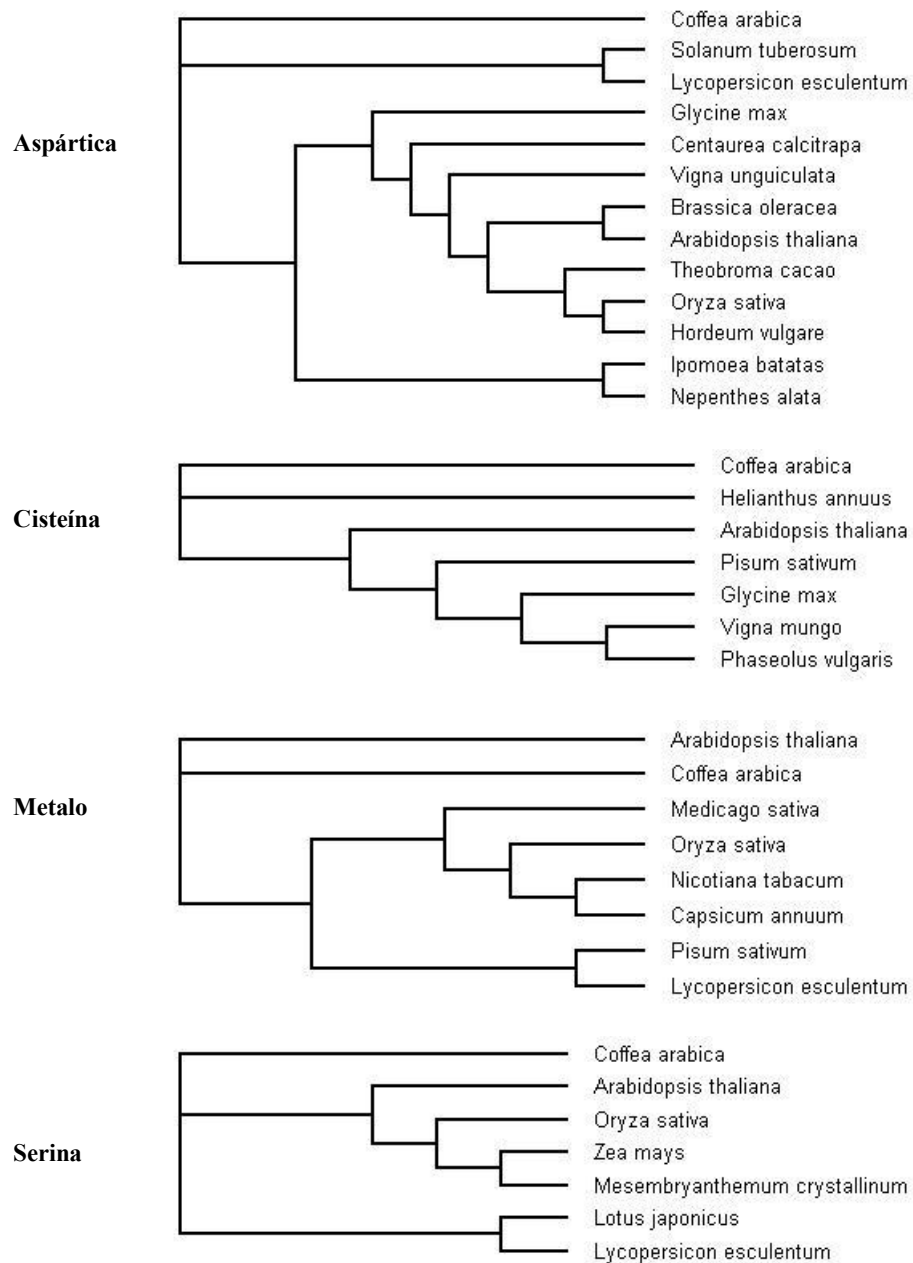


Figura 1 - Árvores filogenéticas construídas a partir de regiões conservadas de genes de Aspártico, Cisteíno, Metalo e Serino proteinases provenientes de frutos de café (*Coffea arabica*) em relação a outras espécies.

Foram selecionados quatro contigs representados por ESTs de bibliotecas de frutos para estudos de expressão. Para tanto, foram realizadas análises para determinar o domínio conservado do gene e desenhados primers para posterior amplificação destas regiões. A seguir, pretende-se obter cDNAs de endospermas de café provenientes de plantas crescendo em ambientes contrastantes quanto à temperatura média anual para início de estudos de expressão, na tentativa de relacionar a presença de proteases e a qualidade de bebida do café.

Bibliografia

- Arnold U, Ludwig E (1996) Analysis of free amino acids in green coffee beans II. changes of the amino acid content in arabica coffees in connection with post-harvest model treatment. *Z. Lebensm Unters Forsch* 203: 379-384
- Clifford MN (1985) Chemical and physical aspects of green coffee and coffee products. *In* MN Clifford, KC Wilson, eds, *Coffee: Botany, Biochemistry and Production of Beans and Beverage*. AVI Publishing company, Inc., Westport, Connecticut, pp 305-374

- Montavon P, Duruz E, Rumo G, Pratz G (2003) Evolution of green coffee protein profiles with maturation and relationship to coffee cup quality. *J. Agric. Food Chem.* 51: 2328-2334
- Ludwig E, Lipke U, Raczek U, Jager A (2000) Investigations of peptides and proteases in green coffee beans. *Eur. Food Res. Technol.* 211: 111-116
- Rogers WJ, Bézard G, Deshayes A, Meyer I, Pétiard V, Marraccini P (1999) Biochemical and molecular characterization and expression of the 11S-type storage protein from *Coffea arabica* endosperm. *Plant Physiol. Biochem.* 37: 261-272
- Shimizu, MM, Mazzafera, P, 2000. Compositional changes of proteins and amino acids in germinating coffee seeds. *Brazilian Archives of Biology and Technology* 43(3): 259-265.
- Silva, EA (2004) Influência de distintas condições edafoclimáticas e do manejo de irrigação no florescimento, produção e qualidade de bebida do café (*Coffea arabica* L.). Tese de Doutorado. Biologia Vegetal. UNICAMP. 81p.