

DESENVOLVIMENTO DE NOVAS METODOLOGIAS PARA A IDENTIFICAÇÃO DE PROTEÍNAS DURANTE O DESENVOLVIMENTO DE FRUTOS DO CAFÉ (*COFFEA ARABICA* L.)¹

Ingrid Gomes Renoldi Heimbeck²; Jorge Alex Taquita³; Maura Vianna Prates⁴; Felipe Vinecky⁵ Gabriel Sérgio Costa Alves⁶; Natália Gomes Vieira⁷. Luciana Pereira Freire²; Carlos Bloch Jr⁸; Pierre Marraccini⁹; Alan Carvalho Andrade¹⁰

¹ Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café com apoio da FINEP e INCT-Café (CNPq/FAPEMIG).

² Bolsista, M.Sc., Embrapa Café, Brasília-DF, ingridheimbeck@gmail.com

³ Analista, Bs., Embrapa Recursos Genéticos e Biotecnologia (LEM-NTBio)

⁴ Pesquisadora, PhD., Embrapa Recursos Genéticos e Biotecnologia (LEM-NTBio)

⁵ Doutorando, M.Sc., UnB – Embrapa Recursos Genéticos e Biotecnologia (LGM-NTBio), Brasília-DF

⁶ Doutorando, M.Sc., Universidade Federal de Lavras-MG

⁷ Mestranda, Bs., Universidade Federal de Lavras, Lavras-MG

⁸ Pesquisador, PhD., Embrapa Recursos Genéticos e Biotecnologia (LEM-NTBio)

⁹ Pesquisador, D.Sc., CIRAD UMR AGAP, Montpellier, FR

¹⁰ Pesquisador, D.Sc., Embrapa Recursos Genéticos e Biotecnologia (LGM-NTBio), alan@cenargen.embrapa.br

RESUMO: O café é uma das bebidas mais consumidas e um dos principais produtos agrícolas, comercializado mundialmente. A espécie *Coffea arabica*, preferível pelos agricultores, fornece grãos cuja bebida é reconhecida como de melhor qualidade pois constitui-se bioquimicamente de elevadas concentrações de carboidratos, lipídeos e trigonelina. O desenvolvimento e crescimento do fruto são caracterizados pela evolução dos tecidos embrionários e alterações bioquímicas, ilustrando a importância do seu estudo para melhor compreender as características finais dos grãos. O objetivo deste trabalho foi desenvolver uma nova metodologia eficaz na identificação do conteúdo protéico expresso no endosperma do fruto do café da espécie *C. arabica* var. IAPAR59 (I59) em diferentes estágios de maturação. A metodologia utilizada neste trabalho baseia-se na estratégia proteômica *bottom-up*, com digestão protéica, análises por cromatografia líquida em fase reversa e espectrometria de massa com interfaces *on-line* e *off-line*. O primeiro permitiu o assinalamento da estrutura primária completa de vinte e sete peptídeos que foram identificados por alinhamento semântico local (BLAST) e apresentaram homologia com quatro proteínas com função ainda desconhecida e nove proteínas com funções já descritas, dentre elas a globulina 11S e a dehidrina, envolvida na proteção osmótica das células vegetais. A interface *on-line* permitiu o assinalamento da estrutura primária de trinta e quatro peptídeos, dos quais treze deles correspondiam a proteínas com funções ainda desconhecidas, quatro com sequências ainda não descritas e quatorze proteínas com funções conhecidas expressas nos diferentes estágios de desenvolvimento. Diante do exposto, as técnicas utilizadas neste trabalho abrem boas perspectivas para uma melhor compreensão das alterações proteômicas que ocorrem durante os diferentes estágios de desenvolvimento e maturação dos grãos cafeeiros. Entretanto, ajustes metodológicos ainda se fazem necessários para permitir a identificação de um número maior de proteínas, especialmente aquelas em menor abundância.

Palavras-Chave: *Coffea arabica*, desenvolvimento do grão, cromatografia líquida, LC-MS e espectrometria de massa.

DEVELOPMENT OF NEW METHODS FOR IDENTIFICATION OF PROTEINS DURING COFFEE FRUIT DEVELOPMENT (*Coffea arabica* L.)

ABSTRACT: The coffee in addition to being one of the most widely consumed beverages is a major agricultural product marketed worldwide. Preferred by farmers, the *Coffea arabica* species provides grain for a best quality beverage once they are composed biochemically of high concentrations of carbohydrates, lipids and trigonelline. The fruit growth and development are characterized by the development of embryonic tissues and biochemical changes, emphasizing the importance of its study to better understand the final characteristics of the grains. The aim of this study was to develop a new methodology effective in identifying the protein expressed in the endosperm of the coffee bean of *C. arabica* var. IAPAR59 (I59) in different stages of maturation. The methodology used was based on the bottom-up proteomics strategy, with protein digestion, and analysis by reverse phase liquid chromatography and by mass spectrometry interfaces online and offline. The first approach allowed the assignment of the complete primary structure of twenty-seven peptides that were identified by local semantic alignment (BLAST) that analysis showed homology with four proteins with yet unknown function and nine proteins with functions already described, among which the 11S globulin and dehydrin, involved in osmotic protection of plant cells. The online interface has allowed the primary structure assignment of thirty-four peptides and within those identified thirteen proteins remains with unknown function, four proteins which sequences has not been described previously and fourteen proteins with known functions expressed in different developmental stages. Given the above, the techniques used in this study offer good prospects for

a better understanding of the proteomic changes that occur during different stages of development and maturation of coffee beans. However, further methodological adjustments are necessary to permit the identification of a much larger number of proteins, especially those of lower abundance.

Key words: *Coffea arabica*, bean development, liquid chromatography, LC-MS mass spectrometry.

INTRODUÇÃO

O café além de ser uma das bebidas mais consumidas é um dos principais produtos agrícolas comercializado mundialmente, sendo considerado o segundo item em importância do comércio internacional de *commodities*. O gênero *Coffea* pertence à família Rubiaceae e é oriundo de regiões da África (Etiópia, Sudão e Quênia). O café comercial é produzido pelas espécies *C. arabica* e *C. canephora*. O café “Arabica” é mais apreciado pelo consumidor, considerado de melhor qualidade, pois apresenta em sua composição bioquímica concentrações mais elevadas de carboidratos, lipídeos e trigonelina que o café “Robusta” (Leroy *et al.*, 2006).

O sabor e o aroma da bebida de café são resultantes da presença combinada de vários constituintes químicos destacando-se os ácidos, aldeídos, cetonas, açúcares, aminoácidos, ácidos graxos, compostos fenólicos, assim como a ação de enzimas em alguns destes constituintes e proteínas (Sarrazin *et al.*, 2000).

Estudos com a finalidade de se determinar o conteúdo protéico dos grãos de café foram realizados, utilizando-se estratégias conjuntas de gel de eletroforese bi-dimensional, biblioteca de cDNA e degradação de Edman para confirmação das proteínas. Alguns trabalhos permitiram caracterizar as proteínas de reserva 11S do grão de café e mostraram que essas proteínas representam cerca de 45% do conteúdo protéico dos grãos (De Castro & Marraccini, 2006).

O desenvolvimento do fruto de café é um processo longo caracterizado por várias mudanças e evoluções nos tecidos, sendo que importantes alterações bioquímicas quantitativas e qualitativas acompanham esse desenvolvimento, ilustrando a importância do seu estudo para uma melhor compreensão das características finais dos grãos de café. O presente trabalho teve por objetivo desenvolver uma ferramenta eficaz na separação, purificação e identificação das proteínas expressas nos diferentes estágios de desenvolvimento em frutos de *C. arabica* var. I59 por meio de digestão protéica, análise por cromatografia líquida de fase reversa e com análises posteriores de espectrometria de massa com interface *online* e *offline*.

MATERIAL E MÉTODOS

Material vegetal:

Neste trabalho, a coleta do material biológico foi realizada no campo experimental da Embrapa Cerrados (Planaltina-DF) no ano de 2008/2009. Frutos da espécie *C. arabica* var. IAPAR59 (I59) foram colhidos em intervalos de 30 dias após o florescimento (DAF) até 237 DAF. Os frutos foram dissecados em pericarpo, perisperma e endosperma. Os tecidos foram pulverizados em nitrogênio líquido até se obter um pó homogêneo.

Extração e quantificação das proteínas

As extrações das proteínas totais foram realizadas utilizando dois protocolos distintos. O protocolo SDS-Fenol modificado (Wang *et al.*, 2003) foi utilizado para análises cromatográficas com interface *offline* no endosperma dos estágios 150, 210 e 237 DAF. O protocolo de extração protéica em solução ácida e precipitação por sulfato de amônio foi utilizado para análises cromatográficas com interface *online* para os estágios 150, 180, 210 e 237 DAF em dois tecidos embrionário (pericarpo e endosperma). A dosagem das proteínas foi feita pelo método de Bradford (1976).

Deteção e identificação de proteínas com cromatografia líquida e espectrometria de massa (off-line)

Os extratos protéicos do endosperma dos estágios 150, 210 e 237 DAF foram fracionados em um sistema de cromatografia líquida ultra rápida usando uma coluna analítica em fase reversa Vydac 214TP510 (C₄) e re-cromatografados uma coluna analítica C₁₈ (Shimpack XR-ODS), ambas equilibradas com 0.1 % TFA com gradiente de 0.1 % TFA em acetonitrila em um fluxo de 1.0 mL.min⁻¹. A absorbância foi monitorada a 216 e 280 nm. Os peptídeos foram analisados por MALDI-TOF MS e MS/MS usando um espectrômetro Autoflex II com óptica TOF-TOF. Foram realizados o seqüenciamento *de novo* dos peptídeos encontrados. Buscas por similaridade foram feitas na base de dados CAFEst utilizando análises por tblastn e as montagens dos bancos café (2004-07).

Deteção e identificação de proteínas com cromatografia líquida (on-line) acoplada à espectrometria de massa

Os extratos protéicos do endosperma e pericarpo dos estágios 150, 180, 210 e 237 DAF foram analisados por cromatografia líquida em fase reversa acoplada a espectrometria de massas (UPLC nanoAcquity / Synapt). Foi utilizado uma coluna c₁₈ nanoAcquity (UPLC 100um x 100mm). em um gradiente de eluição de 0-2 min. 97% B; 2-82 min. 60%; 82-87 min. 15% B; 87-88 min. 97%. As análises foram conduzidas em um espectrômetro de massas de alta resolução híbrido com dois analisadores de massa- um quadrupolo (Q) e um analisador por tempo de voo (TOF - *Time-of-flight*).

Os dados brutos obtidos, resultantes dos experimentos de LCMS/MS, foram processados usando o programa “ProteinLynx” (versão 2.0) da Waters/Micromass, o qual já possui um aplicativo para identificação das proteínas.

RESULTADOS E DISCUSSÃO

O extrato protéico bruto de todas as amostras citadas acima foram digeridas com tripsina antes de serem fracionados. Para as amostras do protocolo de Wang *et al.* (2003), o fracionamento protéico foi realizado em duas etapas. Na primeira utilizou-se uma coluna C₄, onde as frações foram coletadas nas seguintes regiões: fração 1 (F1)-15 a 40% AcN; F2 -40-60% AcN e F3 61%- até o final do gradiente. Com o intuito de se obter melhor resolução, as frações 1, 2 e 3 foram re-cromatografadas, separadamente, utilizando uma coluna C₁₈ e resultou em aproximadamente 20 frações, as quais foram submetidas à análises por espectrometria de massa. A Figura 1 demonstra o perfil cromatográfico referente a amostra do grão cereja (210 DAF). Todas as amostras (150, 210 e 237 DAF) tiveram perfis cromatográficos parecidos (dados não mostrados).

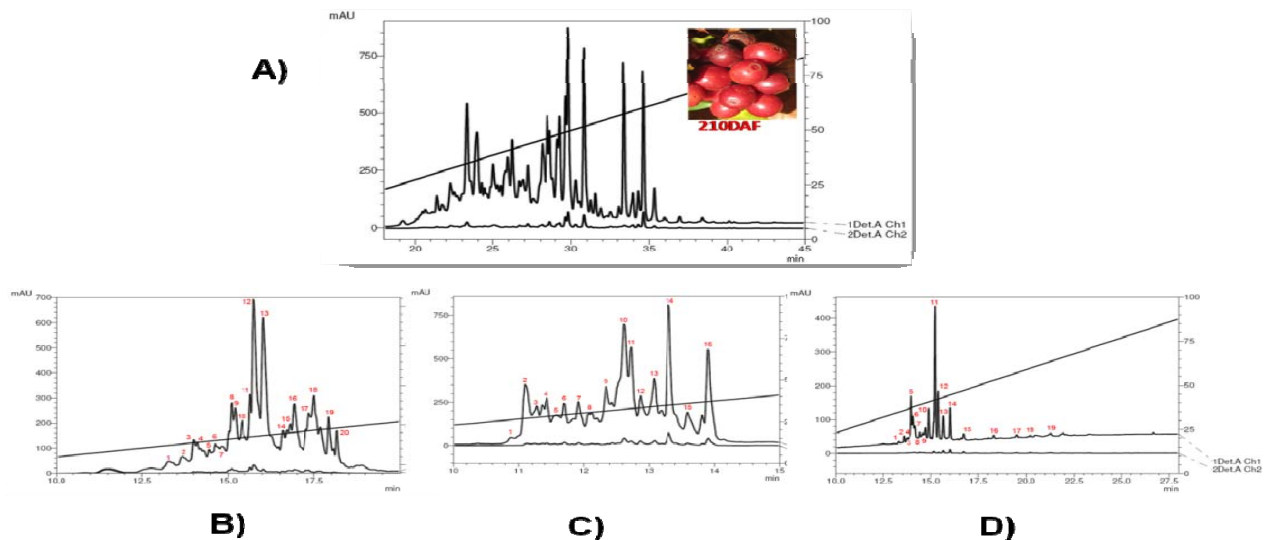


Figura 1. A) Fracionamento do extrato protéico do endosperma 210 DAF da espécie *C. arabica* var. I59 por cromatografia líquida de alta eficiência em modo de fase reversa utilizando uma coluna Vydac C₄. B-C e D) frações na faixas de 15 a 40%, 40-60% e 61% até o final do gradiente de AcN, respectivamente utilizando uma coluna XR-ODS C₁₈.

Vinte e sete peptídeos tiveram sua sequência primária completa assinalada manualmente pelo método de sequenciamento *de novo*. A identificação das proteínas correspondentes foi feita por meio de análises por TBLASTn, utilizando a Base de Dados do genoma Café (CAFEst), como referência. A Tabela 1 demonstra parte dos resultados encontrados, alguns peptídeos tiveram a estrutura primária idêntica à tradução teórica dos contigs presentes na base de dados, enquanto outros divergiram em alguns aminoácidos. Desta forma, a metodologia utilizada foi eficiente em identificar doze proteínas diferentes, a partir dos extratos protéicos de grãos de café analisados

Tabela 1: Correlação dos peptídeos encontrados nos diferentes estágios de desenvolvimento da maturação dos grãos com a sua identificação *in silico* utilizando o banco genômico do café (CENARGEN, Brasília-DF) como referência.

Íon	ID	150 DAF	210 DAF	237 DAF	Seq. Peptídeo	Proteína
1061.80	Contig 1050				LNAQEPSFR 1747 LNAQEPSFR 1721	11S storage protein [<i>Coffea arabica</i>]
1150.00	Contig11904				QTDEYGNPAR 706 QTDEYGNPAR 677	emb 20238.1 putative dehydrin [<i>Helianthus annuus</i>]
1337.00	Contig17766				498 IQEQQQQLWH 469 516 IQEQQQQLWH 487	gb AAK15088.1 2S albumin [<i>Sesamum indicum</i>]

1474.00	Contig13457			SVEILEGDGGVGTIK 231 SVEILEGDGGVGTIK 275	sp O50001 PRU1_PR UAR major allergen protein homolog [<i>Prunus armeniaca</i>]
1619.00	Contig12015			QPIISTEFGEVSGVR 153 QPIISTEFGEVSGVR 197	dbj 78478.1 prepro MP73 [<i>Cucurbita maxima</i>]
1626.00	Contig10670			LTQGGTPSSSCCSGVK 1239 LTQGGNPSSSCCSGVK 1286	dbj BAA82107.1 NtWRKY1 [<i>Nicotiana tabacum</i>]
1626.00	Contig10670			LTQGGTPSSSCCSGVK 472 LTQGGNPSSSCCSGVK 425	dbj 77694.1 lipid transfer protein [<i>Atriplex nummularia</i>]

No segundo método de análise do fracionamento das proteínas do endosperma dos estágios 150, 180, 210 e 237 DAF se deu por detecção e identificação em um cromatógrafo líquido de alta eficiência acoplada “on-line” ao espectrometro de massas-LC-MS/MS. A Figura 2 representa o perfil cromatográfico da amostra do endosperma verde (150 DAF), os espectros MS/MS gerados pelo sequenciamento das amostras foram processados pelo software “Protein Lynx”.

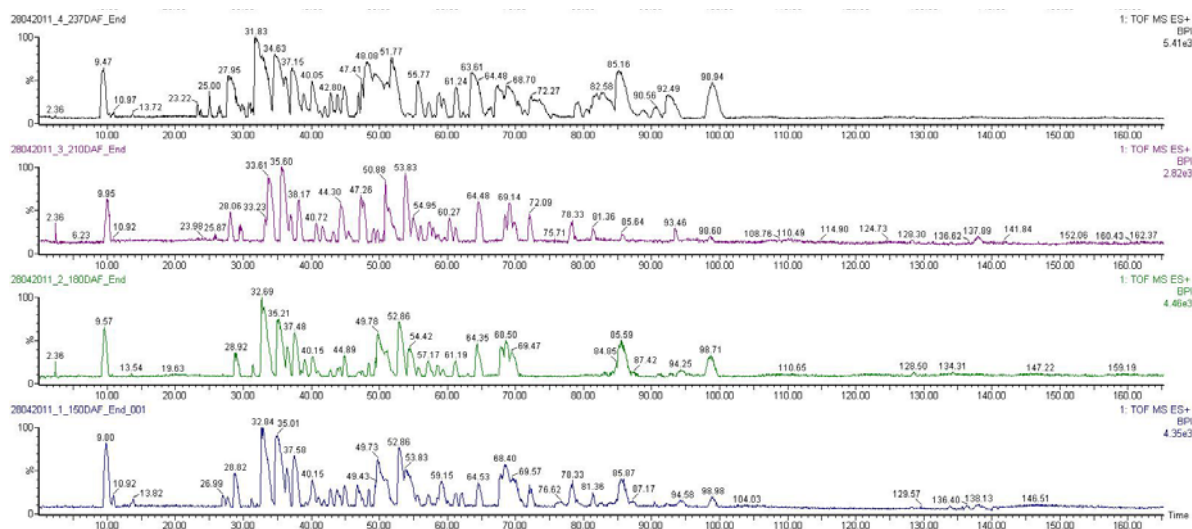


Figura 2. Fracionamento do extrato protéico de endosperma de grãos de café em diferentes estágios de desenvolvimento (150 - 237 DAF) de *C. arabica* var. I59 por cromatografia líquida de alta performance (fase reversa com coluna capilar (NanoAcquity, UPLC-Waters, USA) em gradiente 2 -97% B (AcN + Ácido Fórmico 0.1%).

Trinta e quatro peptídeos puderam ser identificados (Tabela 2), sendo 13 deles correspondentes a proteínas que ainda não tiveram sua função descrita, quatorze proteínas com funções descritas e quatro proteínas que ainda não tiveram sua sequência descrita no banco de dados.

Tabela 2: Correlação dos peptídeos encontrados nas análises realizadas com o “Protein Lynx”, nas amostras de frutos de café em diferentes estágios de desenvolvimento com a Base de Dados CAFest (CENRGEN, Brasília-DF), como referência.

Contig	150 DAF	180 DAF	210 DAF	237 DAF	Resultado Blastp
<u>Contig412</u>					gi 148278099 ABQ53932.1 lipid transfer protein [<i>Sesamum indicum</i>]
<u>Contig14025</u>					gi 269996497 ACZ57583.1 allergenic thaumatin [<i>Olea europaea</i>]
<u>Contig18204</u>					gi 302777019 ADL70200.1 Late embryogenesis abundant protein [<i>Solanum lycopersicum</i>]
<u>Contig1050</u>					gi 2979526 AAC61881.111S storage globulin [<i>Coffea arabica</i>]
<u>Contig1509</u>					gi 224071850 XP_002303583.1 glutathione peroxidase [<i>Populus trichocarpa</i>]
<u>Contig3197</u>					gi 84314116 ABC55670.1 dehydrin DH1a [<i>Coffea canephora</i>]
<u>Contig3616</u>					gi 89275331 ABD66069.1EMZ08 [Elaeis guineensis]
<u>Contig839</u>					gi 300827456 ADK36668.1 cytosolic class I small heat shock protein 3B [<i>Nicotiana tabacum</i>]
<u>Contig17091</u>					gi 156891145 ABU96710.1 glutaredoxin [<i>Solanum tuberosum</i>]

A eficácia dos métodos utilizados permitiu a identificação de proteína abundantes em grãos de café já relatadas em estudos anteriores. Dentre as proteínas mais encontradas destacou-se a globulina 11S, que perfaz cerca de 45% do conteúdo protéico dos grãos de café (Rogers *et al.*, 1999). Outro grupo de proteínas encontrado foi o das dehidrinas, que estão envolvidas na proteção das células vegetais contra a desidratação (Close, 1996). As proteínas LEA (late embryogenesis abundant), embora não tenham uma função ainda bem estabelecida, acredita-se que estejam também associadas a importância a tolerância a dessecação (Blackman *et al.*, 1991). A proteína MP73 está envolvida na formação de vacúolos de armazenamento protéico. Algumas proteínas reconhecidas como alergênicas também foram encontradas. Várias funções têm sido sugeridas a proteína de transferência de lipídeos (LTP), como por exemplo, participação na formação da cutícula e embriogênese, no estabelecimento de relações simbióticas, entre outras (Terras, *et. al.*, 1992)

CONCLUSÕES

Os resultados apresentados mostram que as técnicas utilizadas são adequadas para entender as alterações proteômicas que ocorrem durante os diferentes estágios de desenvolvimento dos grãos de café. Foi demonstrado que proteínas abundantes previamente descritas em grãos de café, puderam ser identificadas além de outras ainda não descritas. Entretanto, ajustes metodológicos ainda se fazem necessário para permitir a identificação de um número maior de proteínas, especialmente aquelas em menor abundância. Neste sentido, pode-se sugerir tentativas de remoção das albuminas por imunoafinidade ou outros métodos de precipitação e preparo dos extratos. Da mesma forma, novos parâmetros de corrida nas análises com a cromatografia (*on-line*) acoplada à espectrometria de massa, podem ser testados, visando-se uma melhor separação dos peptídeos tripticos e conseqüentemente a obtenção de melhores resultados.

REFERÊNCIAS BIBLIOGRÁFICAS

- BLACKMAN, S.A.; WETTLAUFER, S.H.; OBENDORF, R.L.; LEOPOLD, A.C. Maturation proteins associated with desiccation tolerance in soybean seeds. **Plant Physiology**, v.96, n.3, p.868-874, 1991.
- BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, v.72, p 248-254, 1976.
- CLOSE, T.J. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. **Physiologia Plantarum**, v.97, p.795-803, 1996.
- DE CASTRO, R.; MARRACCINI, P. Cytology, biochemistry and molecular changes during coffee fruit development. **Brazilian Journal Plant Physiology**, v.18, p.175-199, 2006.
- LEROY T., RIBEYRE F., BERTRAND B., CHARMETANT P., DUFOUR M., MONTAGNON C., MARRACCINI P., POT D. Genetics of coffee quality. **Brazilian Journal Plant Physiology**, v.18, n.1, p.229-242, 2006.
- SARRAZIN, C.; LE QUÉRE, J.L.; GRETSCH, C.; LIARDON, R. Representativeness of coffee aroma extracts: a comparison of different extraction methods. **Food Chemistry**, v.70, p.99-106, 2000.
- ROGERS, W.J.; BÉZARD, G.; DESHAYES, A.; MEYER, I.; PÉTIARD, V.; MARRACCINI, P. Biochemical and molecular characterization and expression of the 11S-type storage protein from *Coffea arabica* endosperm. **Plant Physiology**, v.37, n.4, p.261-272, 1999.
- TERRAS, F. R.; SCHOOF, H.M.; DE BOLLE, M.F.; VAN LEUVEN, F.; REES, S.B.; VANDERLEYDEN, J.; CAMMUE, B. P.; BROEKAERT, W.F. Analysis two novel classes of plant antifungal proteins from radish (*Raphanus sativus L.*) seeds. **The Journal of Biological Chemistry**, v.267, p.15301-15309, 1992.
- WANG, W.; SCALI, M.; VIGNANI, R.; SPADAFORA, A.; SENSI, E.; MAZZUCA, S. Protein extraction for twodimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. **Electrophoresis**, v.24, n.14, p.2369-75, 2003.