

ANÁLISE *IN SILICO* DE GENES POTENCIALMENTE ENVOLVIDOS NA RESPOSTA AOS ESTRESSES ABIÓTICOS, PRESENTES NA BASE DE DADOS DO GENOMA CAFÉ

Felipe VINECKY, Kelly M. de BRITO, Felipe R. DA SILVA e Alan C. ANDRADE¹ E-mail: alan@cenargen.embrapa.br

¹Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF.

Resumo:

Os danos potenciais dos estresses abióticos tais como seca, salinidade e temperaturas extremas na produção agrícola mundial é alarmante, tendo em vista as constantes alterações climáticas observadas nas últimas décadas ao redor do globo. Desta forma, a ampliação do conhecimento científico acerca dos fatores genéticos envolvidos na tolerância dos vegetais a esses estresses, com vistas ao melhoramento genético destas características, é prioritário. Foi concluído recentemente, com o apoio do Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (CBP&D-Café), o projeto Genoma Café. Este projeto consistiu no seqüenciamento em larga escala de genes expressos (ESTs) de café e teve como meta, identificar e anotar de 20 a 30.000 genes de *Coffea* spp. a partir do seqüenciamento de 200.000 clones de ESTs obtidos de diversas bibliotecas de cDNA, inclusive aquelas oriundas de material vegetal submetido a estresses abióticos. Com a conclusão do projeto, está disponível para pesquisa um banco de dados contendo uma parcela significativa dos genes expressos em café (transcriptoma), associados às diversas condições de estresse submetidas ou tecidos utilizados para a construção das bibliotecas. O objetivo do presente trabalho, foi realizar uma análise prospectiva dos fatores genéticos potencialmente associados à resposta aos estresses abióticos, presentes na Base de Dados do Genoma Café. Essa análise baseou-se na identificação de genes de café com alta similaridade aos previamente descritos, caracterizados como genes envolvidos na resposta aos estresses abióticos (EREA), em outras espécies vegetais. Através das análises de tBlastn, genes de café com alto nível de similaridade (e-value $\leq 10^{-20}$) aos 355 genes EREA pesquisados, foram identificados, em mais de 91% dos casos. Apenas 5% dos genes EREA pesquisados, não resultaram em genes análogos de café, com nível significativo de similaridade (e-value $\geq 10^{-5}$). A observação de que a maioria desses genes não encontrados na base de dados do café, são genes relacionados às respostas ao frio, pode ser justificada pela ausência de ESTs provenientes de bibliotecas de cDNA, com esse tipo de estresse. Por outro lado, essas diferenças poderiam estar relacionadas às diferenças fisiológicas marcantes entre arábido e café, com relação à tolerância ao frio, uma vez que grande parte dos genes EREA utilizados nesta pesquisa, são provenientes de arábido. Pode-se concluir com os resultados obtidos, que a utilização da Base de Dados do Genoma Café é uma ferramenta poderosa na rápida identificação e seleção de potenciais genes de interesse agrônomo de café, para a realização de ensaios experimentais de caracterização funcional.

Palavras-chave: tolerância a seca, estresses abióticos, transcriptoma, bioinformática, melhoramento genético, biotecnologia, *Coffea*

***IN SILICO* ANALYSIS OF GENES POTENTIALLY INVOLVED IN ABIOTIC STRESS RESPONSES, IDENTIFIED IN THE COFFEE-EST DATABASE**

Abstract:

Abiotic stresses, such as drought, salinity and extreme temperatures, are serious threats to agriculture due to severe climate changes observed in the last decades, around the globe. Therefore, the identification of the molecular mechanisms involved in abiotic stress responses aiming at improving breeding for abiotic-stress tolerance in crop plants is considered priority in plant biotechnology programs. Supported by the Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (CBP&D-Café), it was recently concluded the Brazilian Coffee EST-project. The project aimed at identifying 20 to 30 thousand genes from a large-scale sequencing effort (200.000 EST), obtained from several coffee-cDNA libraries, including those made from plants submitted to abiotic stresses. The conclusion of this project provided to coffee researchers a database containing a large set of transcribed genes, associated with the diverse stress conditions applied before the construction of the cDNA libraries. The aim of the present work was to identify in the Coffee-EST database, genes potentially involved in the response to abiotic stresses. The analysis consisted initially on the identification of coffee genes displaying high degree of sequence similarity to well characterized genes, described as to be involved in the abiotic stress responses (IASR), from other plant species. Results of tBlastn analysis indicated that coffee ESTs displaying high degree of primary-sequence similarity (e-value $\leq 10^{-20}$) to 355 IASR genes, could be identified in more than 90% of the cases. The results also show that only 5% of the IASR genes searched for, could not be identified at a significant level of similarity (e-value $\geq 10^{-5}$). The observation that the majority of these "non-identified" ESTs correspond to genes involved in response to cold stresses, could be explained by the fact that there was no cDNA library made under cold stress. Another explanation could be related to the obvious differences between arábido and coffee, regarding cold-tolerance, in view of the fact that most of the IASR sequences used for screening the database are genes from arábido. The results of this work indicate that the use of the Coffee-EST database provided a fast and efficient way to identify and select candidate genes involved in abiotic-stress responses in coffee, that will be further studied and characterized.

Key words: Drought tolerance, abiotic stresses, transcriptome, bioinformatics, crop breeding, biotechnology, *Coffea*

Introdução

Apesar do alto nível tecnológico já adquirido na agricultura mundial, as perdas ocorridas nas principais culturas por estresses são muito elevadas. Estresses abióticos são a causa principal destas perdas, na agricultura mundial, reduzindo a produtividade média das principais culturas em mais de 50% (BOYER, 1982; BRAY, 1993). Levando em conta o potencial de produtividade das principais culturas em relação à média real obtida nas lavouras nos Estados Unidos, KRIEG (1993) apresentou dados das principais culturas que indicam perdas de até 87% por fatores bióticos e abióticos. As perdas por fatores de estresses abióticos são as mais significativas para a maioria das culturas, representando no mínimo quatro vezes mais em relação aos fatores bióticos.

As plantas superiores, bem como todos os organismos vivos, possuem a necessidade de estarem “sintonizadas” com as constantes alterações ambientais durante o seu crescimento e desenvolvimento. As plantas utilizam vários estímulos ambientais tais como luz, temperatura e disponibilidade de água, para regular os seus processos normais de crescimento, germinação e florescimento. Entretanto, condições climáticas severas tais como seca, temperaturas extremas (baixas e altas), salinidade alta ou inundação, causam efeitos adversos no crescimento e desenvolvimento vegetal. Na busca de se tolerar esses estresses e se garantir a sobrevivência, as plantas respondem a esses fatores ambientais através de alterações fisiológicas, aos níveis celular e molecular. Grande parte da regulação biológica responsável pelas alterações metabólicas e fisiológicas em resposta às mudanças ambientais ocorre na regulação da expressão dos genes envolvidos nestes processos (SHINOZAKI & DENNIS, 2003). Desta forma, o conhecimento dos genes que possuem expressão alterada em resposta a determinado fator ambiental poderá revelar grande parte das bases genéticas determinantes ou envolvidas neste processo biológico.

O desenvolvimento de tecnologias aplicadas à biologia e genética molecular vem propiciando uma produção enorme de informações na área genômica, através do seqüenciamento de diferentes organismos. Recentemente, as seqüências genômicas da planta modelo *Arabidopsis*, foram totalmente determinadas (THE ARABIDOPSIS GENOME INITIATIVE, 2000). A utilização da tecnologia de arranjos de DNA para se analisar simultaneamente a expressão de vários genes tem proporcionado uma gama enorme de informações acerca do perfil da expressão gênica em plantas submetidas a condições de estresses abióticos. Os resultados destas análises, começam a revelar a complexidade dos mecanismos moleculares envolvidos nas respostas dos vegetais aos estresses abióticos (SHINOZAKI et al, 2003; ZHU, et al, 2003). Esses mecanismos moleculares são baseados na ativação e regulação da expressão de vários genes, cujos produtos protéicos podem estar envolvidos tanto na resposta (percepção, tradução do sinal e controle da expressão gênica/fatores de transcrição) como na tolerância ao estresse (antioxidantes, protetores osmóticos, transporte iônico, etc) (SHINOZAKI et al, 2003). O controle da expressão dos genes induzidos por estresses abióticos pode ser dependente ou independente de ácido abscísico (ABA) e pode ainda ser comum aos diferentes tipos de estresse. Por exemplo, mais da metade dos 299 genes induzidos por estresse hídrico são também induzidos por estresses de alta salinidade e tratamento com ABA. Entretanto, somente 10% destes são também induzidos por baixas temperaturas, indicando haver certa especialização (SEKI et al, 2002 a,b).

O melhoramento genético convencional do cafeeiro tem obtido sucesso no atendimento do setor produtivo, e é realizado através dos métodos tradicionais de cruzamentos e seleção de progênes favoráveis. Este processo de seleção envolve testes exaustivos para avaliar as características desejáveis da progênie seguidos de propagação dos indivíduos selecionados e avaliação em campo da planta adulta. Atualmente, as técnicas de biologia avançada podem oferecer alternativas para a redução do tempo de obtenção de novos genótipos e atuar como uma ferramenta auxiliar nos programas de melhoramento genético através da identificação de marcadores em programas de seleção assistida. Além disso, a possibilidade de produção de plantas transgênicas abre novas perspectivas ao melhoramento convencional, permitindo a rápida incorporação de características desejáveis às espécies perenes. No entanto, as alterações genéticas experimentadas até hoje, com o objetivo de se aprimorar a tolerância das plantas aos estresses abióticos, têm sido baseadas na manipulação de genes envolvidos na proteção e manutenção da função e estrutura dos componentes celulares. Ao contrário da maioria das heranças monogênicas, tais como resistência a pragas e doenças, a complexidade e multiplicidade dos componentes genéticos envolvidos nas respostas aos estresses abióticos tornam a manipulação e o melhoramento genético para essas características muito mais difícil (WANG et al, 2003).

Enquanto a adaptação das plantas aos estresses abióticos nas condições naturais propicia vantagens ecológicas, o custo metabólico e energético pode, em alguns casos, limitar os benefícios agrônômicos e resultar em baixa produtividade. Desta forma, o melhoramento genético direcionado para a obtenção de plantas de interesse agrícola com maior tolerância aos estresses abióticos, somente poderá ser obtido na prática, através de uma ação combinada do melhoramento clássico e o melhoramento avançado com técnicas de biologia molecular (KASUGA et al., 1999; DUNWELL, 2000; WANG et al., 2001). Segundo WANG et al. (2003), um programa racional e estratégico de melhoramento para tolerância aos estresses abióticos deve incluir as seguintes etapas: (1) Seleção de germoplasma de espécies selvagens de relevância; (2) Elucidação dos mecanismos moleculares específicos nos genótipos tolerantes e sensíveis; (3) Seleção e melhoramento utilizando técnicas de biologia avançada, através de genoma funcional, uso de sondas e marcadores moleculares para a seleção entre as populações naturais e cruzamentos, bem como transformação com genes específicos; (4) Adaptação das práticas agrônômicas em uso. Desta forma, um dos requisitos primordiais para um programa estratégico de melhoramento genético direcionado para tolerância a estresses abióticos, é a identificação dos genes envolvidos no controle destas características, conforme descrito acima.

O presente trabalho teve como objetivo principal prospectar e identificar na Base de Dados resultante do Projeto Genoma Café, os genes com alta similaridade aos envolvidos na resposta aos estresses abióticos, previamente descritos e caracterizados em outras espécies vegetais.

Material e Métodos

Inicialmente, uma coleção de genes envolvidos na resposta aos estresses abióticos (EREA), já caracterizada em outras espécies vegetais, foi construída, a partir de relatos de literatura. Esses genes foram estratificados em três grandes grupos, de acordo com a função desenvolvida na resposta aos estresses abióticos: 1. Percepção e transferência do sinal; 2. Controle da transcrição e 3. Mecanismos de resposta ao sinal (Fig. 1). Utilizando-se a ferramenta tBlastn (e-value $\leq 10^{-20}$) da Base de Dados do Genoma Café (<http://genoma.cenargen.embrapa.br/>) procedeu-se a identificação dos genes de café com alta similaridade aos genes do conjunto EREA.

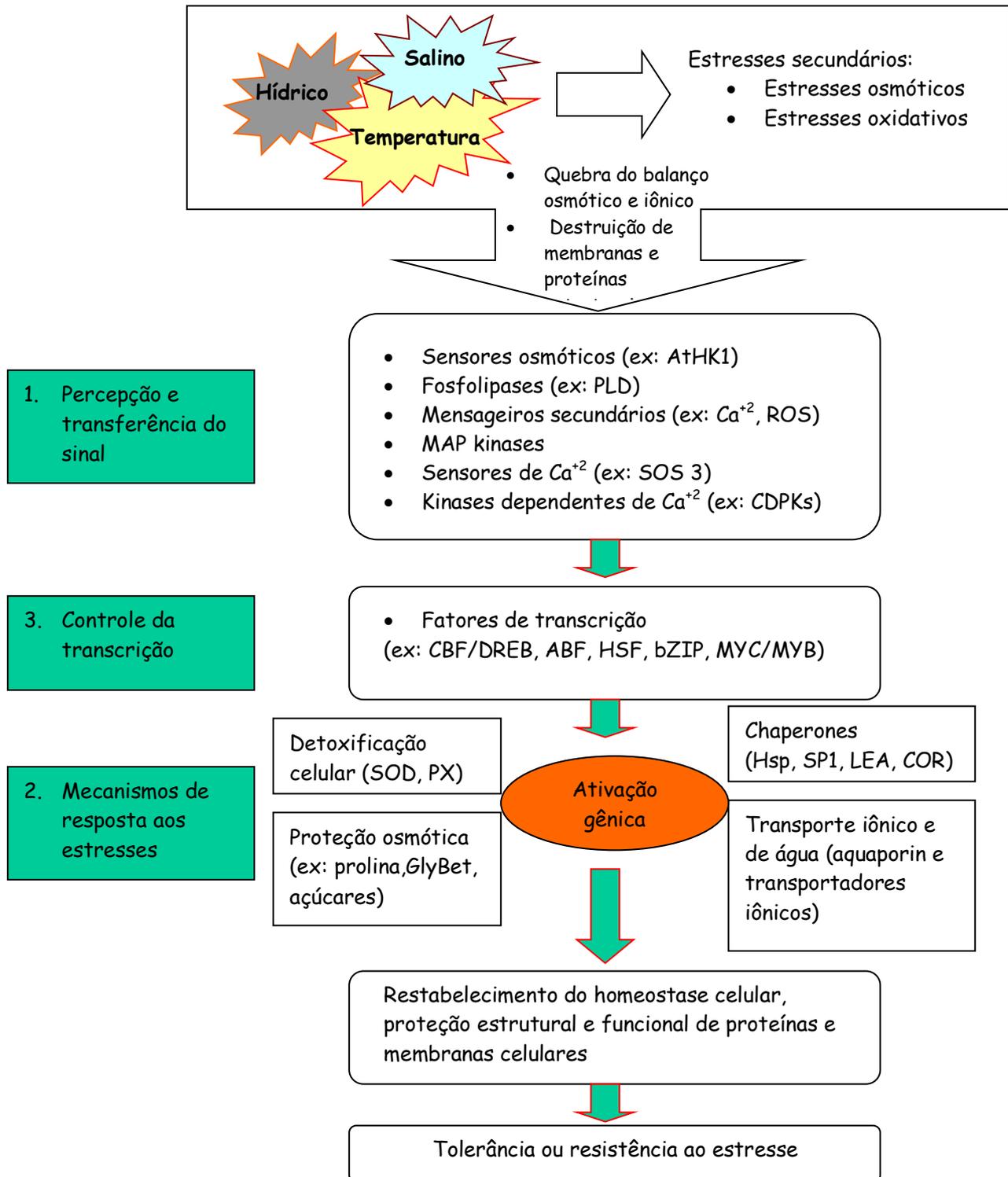


Figura 1: Cadeia de eventos moleculares na resposta das plantas aos estresses abióticos. Classificação dos eventos em três estratos: 1. Percepção e transferência do sinal, 2. Mecanismos de resposta aos estresses e 3. Controle da transcrição (Wang et al, 2003).

Resultados e Discussão

Os resultados da análise tBlastn utilizando cerca de 355 seqüências protéicas, presentes no conjunto EREA, indicam que 324 destas seqüências apresentam alta similaridade ($e \leq -20$) com inúmeros ESTs presentes no conjunto Unigene-CAFEEST, disponível na Base de Dados do Genoma Café da Embrapa Recursos Genéticos e Biotecnologia. O grande número de seqüências encontradas pode ser devido a alta conservação de seqüência primária entre os diferentes parálogos das proteínas pesquisadas e/ou à presença de vários alelos no genoma do café. Esta possibilidade representaria um impedimento na identificação dos potenciais ortólogos dos genes EREA em café. No entanto, uma análise da distribuição de freqüência dos “Best Hits”, extratificados em 7 classes de e-value, indicam que mais de 70% dos ESTs identificados, apresentam e-value menores que e-50 sugerindo alta similaridade. Esta distribuição de freqüência é diferente da extratificação resultante da análise do conjunto Unigene (Fig. 2). Isto sugere especificidade e maior confiabilidade nos dados da análise realizada com os genes EREA, onde 47 % dos EST identificados apresentam-se agrupados nas classes com e-value \leq e-100. Já no caso do Unigene somente 27 % dos “Best Hits” estão agrupados nessas classes.

Os dados apresentados na Fig. 3A, ilustram a identificação de potenciais ortólogos, em café, de várias proteínas envolvidas no controle transcricional da resposta aos estresses abióticos. Os resultados indicam que das 40 seqüências dos fatores de transcrição EREA, utilizadas na pesquisa do banco de dados do café, o EST análogo com nível de similaridade significativo, não foi indentificado para somente uma delas. A Fig. 3B, apresenta uma distribuição do resultado da análise de tBlastn ($e\text{-value} \leq e\text{-}20$) com essas 40 seqüências EREA, apresentando todos os ESTs identificados, classificados por família.

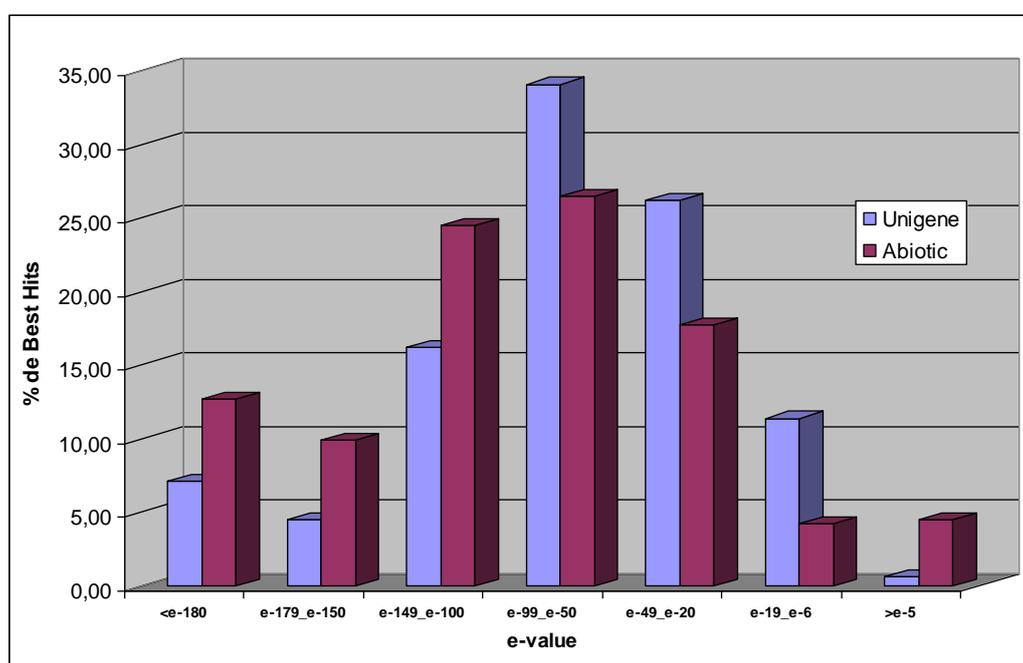


Fig. 2. Distribuição de freqüência dos “best hits” encontrados na base de dados do genoma café, resultante das análises de tBlastn e Blastx, realizados com as seqüências EREA (Abiotic) e Unigene (NCBI-Non redundant), respectivamente.

Pode-se concluir com os resultados obtidos, que a utilização da Base de Dados do Genoma Café é uma ferramenta poderosa na rápida identificação e seleção de potenciais genes de interesse agrônomo de café, para a realização de ensaios experimentais de caracterização funcional. Os dados apresentados mostram que através das análises de tBlastn, genes de café com alto nível de similaridade ($e\text{-value} \leq 10^{-20}$) aos 355 genes EREA pesquisados, foram identificados, em mais de 91% dos casos. Apenas 5% dos genes EREA pesquisados, não resultaram em genes análogos de café, com nível significativo de similaridade ($e\text{-value} \geq 10^{-5}$). A observação de que a maioria desses genes não encontrados na base de dados do café, são genes relacionados às respostas ao frio, pode ser justificada pela ausência de ESTs provenientes de bibliotecas de cDNA, com esse tipo de estresse. Por outro lado, essas diferenças poderiam estar relacionadas às diferenças fisiológicas marcantes entre arábido e café, com relação à tolerância ao frio, uma vez que grande parte dos genes EREA utilizados nesta pesquisa, são provenientes de arábido.

Finalizando, a realização de ensaios experimentais de caracterização funcional para a validação das observações *in silico* realizadas neste trabalho é o foco atual de nossas pesquisas. Neste sentido, a utilização do conjunto Unigene arranjado em membranas de nylon (Macroarranjos), para a realização de experimentos de hibridização com sondas marcadas radiotivamente, sintetizadas a partir de mRNA obtido de genótipos contrastantes e/ou submetidos a tratamentos de estresses abióticos, fornecerá sem dúvida, dados importantes na discriminação dos genes envolvidos nas respostas do café aos estresses abióticos. Esse conhecimento poderá trazer grandes benefícios aos programas de melhoramento genético do café, acelerando o processo de seleção.

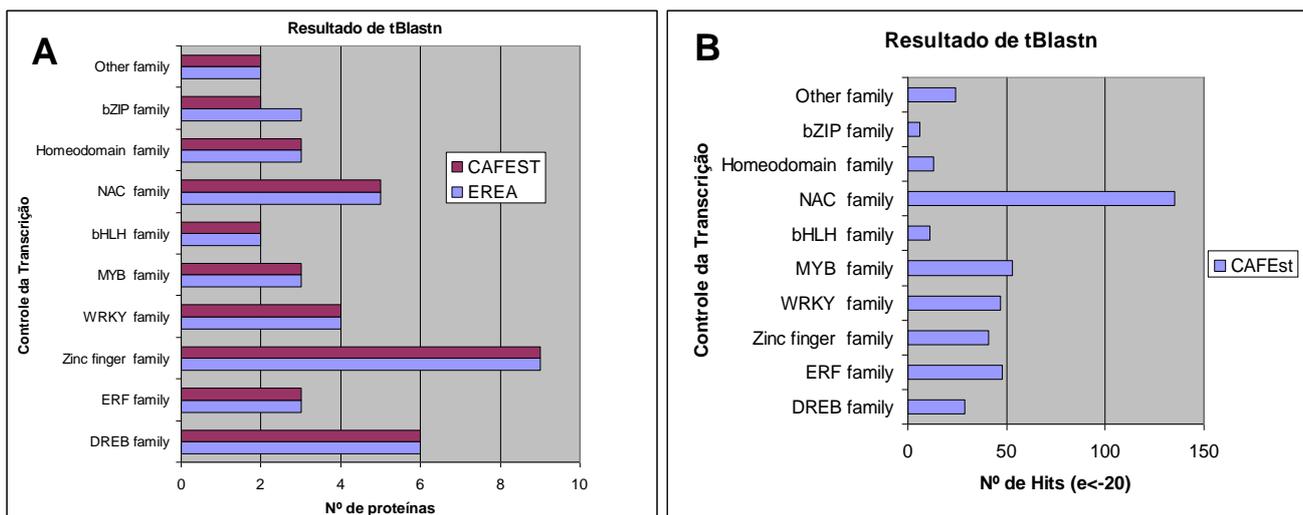


Fig. 3. Resultado das análises de tBlastn realizadas com 40 seqüências de fatores de transcrição EREA **A.** Número de ESTs identificados, com alta similaridade às seqüências EREA utilizadas na pesquisa. **B.** Resultado total dos ESTs identificados (e-value $\leq e-20$), agrupados por famílias proteicas dos fatores de transcrição EREA.

Referências bibliográficas

- Boyer, J. S. (1992) Plant productivity and environment *Science*, 218:443-448.
- Boyer, J. S. (1996) Advances in drought tolerance in plants. *Advances in Agronomy*, 56:187-218.
- Bray, E.A. (1993) Molecular Responses to Water Deficit. *Plant Physiol*, 103(4):1035-1040.
- Dunwell Jm. (2000) Transgenic approaches to crop improvement. *J Exp Bot*, 51:487-496.
- Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K. (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat Biotechnol*, 17(3):287-291.
- Krieg, D. R. Stress tolerance mechanisms in above ground organs. In: A Workshop on .Adaptation to Soil Stress, August, 1-4, 1993. Proceedings ..., Lincoln, Nebraska, University of Nebraska, 1993. pg. 65-79.
- Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K. Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J*. v. 31, n. 3, p.279-292, 2002.
- Seki M, Ishida J, Narusaka M, Fujita M, Nanjo T, Umezawa T, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K. Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. *Funct Integr Genomics*. v. 2, n. 6, 282-291, 2002 .
- Shannon, M. C. Adaptation of plants to salinity. *Advances in Agronomy*, vol. 75, pg. 75-120, 1997.
- Shinozaki K, Dennis Es. Cell signalling and gene regulation: global analyses of signal transduction and gene expression profiles. *Curr Opin Plant Biol*. v.6, n.5, p. 405-409, 2003.
- Shinozaki K, Yamaguchi-Shinozaki K, Seki M. Regulatory network of gene expression in the drought and cold stress responses. *Curr Opin Plant Biol*. v.6, n. 5, p. 410-417, 2003
- THE ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* v 408, n 6814, p. 796-815, 2000.
- Zhu T. Global analysis of gene expression using GeneChip microarrays. *Curr Opin Plant Biol* v. 6, n.5, p.418-425, 2003
- Wang W, Vinocur B, Shoseyov, O., Altman A. Biotechnology of plant osmotic stress tolerance: physiological and molecular considerations. *Acta Hort* v. 560, p. 285-292, 2001.
- Wang W, Vinocur B, Altman A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta*. Sep 26 , 2003.