

**IDENTIFICAÇÃO E ANÁLISE DA EXPRESSÃO *IN SILICO* DOS FATORES DE
TRANSCRIÇÃO CBF/DREB EM CAFÉ (*Coffea arabica*)**

ANDRÉ ALMEIDA LIMA¹, SOLANGE APARECIDA SÁGIO²; ANTONIO CHALFUN-JÚNIOR³

RESUMO

As plantas têm seu crescimento e produtividade altamente influenciados por estresses abióticos tais como seca, salinidade e mudanças de temperatura. O estudo de fatores de transcrição envolvidos na percepção e resposta a estes estresses, como os fatores de transcrição CBF/DREB, se mostra de grande importância no desenvolvimento de culturas mais tolerantes. Assim, este trabalho teve como objetivo a identificação e caracterização *in silico* dos prováveis fatores de transcrição CBF/DREB em café (*Coffea arabica*) no banco de dados CAFEST. Foram realizadas buscas por palavra-chave e similaridade no banco CAFEST, e as sequências selecionadas foram clusterizadas, anotadas e analisadas quanto à presença de domínios conservados. Posteriormente realizaram-se alinhamentos múltiplos envolvendo as sequências encontradas no banco CAFEST e sequências de outras espécies, sendo a filogenia avaliada por meio de árvores filogenéticas geradas pelo programa MEGA4. A análise de expressão foi avaliada por *Northern*s Eletrônicos, construídos pelos programas Cluster e Treeview. A análise filogenética permitiu que as sequências aqui encontradas fossem alocadas em quatro dos seis grupos presentes nesta família de fatores de transcrição e o *Northern* eletrônico mostrou que estes prováveis fatores de transcrição são expressos em diferentes tecidos, estádios de desenvolvimento e condições ambientais, com destaque para as condições de estresse e tecidos reprodutivos.

Palavras-chaves: Bioinformática, estresse, CAFEST.

INTRODUÇÃO

As plantas têm seu crescimento e produtividade altamente influenciados por estresses abióticos como seca, salinidade e mudanças de temperatura. Estas condições adversas representam um grande desafio para a produção de alimentos, podendo reduzir seu potencial em até 70%, com o estresse hídrico atuando como o tipo de estresse abiótico mais predominante. As plantas respondem e se adaptam a estas condições por meio de uma série de mudanças bioquímicas e fisiológicas.

A superfamília de fatores de transcrição AP2/ERF, a qual é definida pelo domínio AP2/ERF, está envolvida na regulação transcricional de uma variedade de processos biológicos relacionados ao crescimento e desenvolvimento vegetal, assim como em várias respostas a estímulos ambientais. Esta superfamília é dividida em três famílias principais, AP2 (*APETALA2*) e RAV (*Related to ABI3/VPI*) e ERF (*Ethylene Response Factors*) (Riechmann et al., 2000), de acordo com o número e a similaridade da sequência de seus domínios AP2/ERF (Nakano et al., 2006). A família ERF é caracterizada por possuir um único domínio AP2/ERF, sendo subdividida em duas subfamílias: a subfamília ERF (*Ethylene Response Factors*), que está principalmente envolvida na resposta a estresses bióticos como ataque de patógenos, e a subfamília CBF/DREB (*C-repeat/DRE-Binding Factor / Dehydration-Responsive Element Binding proteins*), cujo os membros estão principalmente envolvidos nas respostas das plantas a estresses abióticos (Sakuma et al., 2002).

Os fatores de transcrição DREBs/CBFs interagem especificamente com os elementos cis DRE/CRT (*dehydration-responsive element/C-repeat*) e regulam a expressão gênica em respostas a estresses como seca, salinidade e baixas temperaturas (Yamaguchi-Shinozaki and Shinozaki, 1994; Dubouzet et al., 2003; Wang & Dong, 2009; Agarwal et al., 2010). Considerando que a tolerância a estresses abióticos envolve uma série de mudanças bioquímicas e fisiológicas, se colocando-se como uma característica multigênica, o estudo dos fatores de transcrição envolvidos na percepção e resposta

¹ Mestrando em Agronomia/Fisiologia Vegetal, DBI/ UFLA, andrelima4@yahoo.com.br

² Doutoranda em Agronomia/Fisiologia Vegetal, DBI/ UFLA, solsagio@gmail.com

³ Professor Adjunto, DBI/UFLA, chalfunjunior@dbi.ufla.br

ao estresse se mostra de grande importância no desenvolvimento de culturas mais tolerantes ao estresse.

Com a criação do banco de ESTs (*Expressed Sequence Tags*) CAFEST, gerado pelo sequenciamento do transcriptoma do cafeeiro (Vieira et al., 2006), tornou-se possível a realização de buscas por sequências formadoras de prováveis genes relacionados às mais diversas características de interesse, permitindo também a determinação dos possíveis tecidos em que estes genes são expressos, além de uma estimativa desta expressão.

Desta forma, este trabalho teve como objetivo realizar a caracterização *in silico* dos prováveis fatores de transcrição CBF/DREB em café (*Coffea arabica*).

MATERIAIS E MÉTODOS

Busca por sequências

Primeiramente o banco de dados CAFEST foi investigado através dos critérios de busca por palavra chave e similaridade [tblastn utilizando sequências homólogas publicadas no banco de dados públicos *GenBank* (www.ncbi.nlm.nih.gov)], para identificação dos prováveis receptores do etileno. As sequências apresentando similaridade confiável ($e\text{-value} > 10^{-4}$) foram depositadas no sistema de gerenciamento e manipulação de sequências, o *GeneProject*, e clusterizadas, utilizando o programa CAP3 (Huang & Madan, 1999). Após a clusterização, os EST-contigs e EST-singlets, foram anotados e comparados contra bancos de dados públicos de proteínas para se obter maiores informações sobre as prováveis proteínas codificadas por estas sequências e eliminação de sequências falsas. Após esta etapa, as sequências selecionadas foram ainda utilizadas como moldes para uma nova procura no próprio banco de dados CAFEST, processo denominado de saturação, com intuito de estender ou completar estas sequências, ou mesmo encontrar novas sequências.

Análise filogenética

A caracterização dos prováveis fatores de transcrição CBF/DREB foi realizada através de um alinhamento múltiplo entre as sequências selecionadas no CAFEST e sequências homólogas publicadas. Os alinhamentos foram realizados através do programa ClustalW (Thompson et al., 1994) com os parâmetros padrões (*default*) e utilizando-se de sequências de nucleotídeos traduzidas em aminoácidos as quais foram deduzidas através da ferramenta *translate*, encontrada no site do banco de proteínas ExPASy (www.expasy.ch), sendo suas ORFs (*Open Reading Frames*) obtidas através da ferramenta ORF finder, encontrada no site do NCBI (www.ncbi.nlm.nih.gov). A partir destes alinhamentos, árvores filogenéticas foram geradas através do programa MEGA4 (Tamura et al., 2007), com o modelo de comparação *Neighbor-joining* (Saitou & Nei, 1987), método de distância *p* e supressão *pair-wise*. A validade das árvores quanto à distância filogenética dos clusters pôde ser medida pelo teste probabilístico de *bootstraps* (Sitnikova et al., 1995).

Análise da expressão *in silico*

As análises de expressão *in silico* foram determinadas pela construção de *Northern* Eletrônicos, gerados a partir da construção de uma tabela contendo o número de vezes que cada *read* formador de um EST-contig apareceu expresso em cada biblioteca. Esses dados foram então normalizados para obtenção do grau de expressão exata dos prováveis genes em cada biblioteca. Posteriormente, foi feita uma matriz, relacionando os prováveis genes e as bibliotecas em que estes foram expressos, a qual foi submetida a um agrupamento hierárquico pelo programa Cluster (Eisen et al., 1998) e os resultados obtidos foram então visualizados pelo programa TreeView (Eisen et al., 1998), utilizando uma escala do preto ao vermelho, onde quanto mais próximo à cor vermelha maior o grau de expressão.

RESULTADOS E DISCUSSÃO

Com relação aos fatores de transcrição da família ERF foram encontrados um total de 706 *reads* os quais, após clusterização, formaram 59 EST-contigs e 453 EST-singlets. A análise destas sequências verificou que somente 26 contigs e 15 singlets demonstraram codificar para proteínas

possuindo o domínio AP2/ERF, característico destes fatores de transcrição. De acordo com a identidade dos aminoácidos presentes dentro do domínio AP2/ERF, estas sequências puderam ser separadas entre as duas subfamílias formadoras da família ERF, onde 18 contigs e 9 singlets demonstraram pertencer à subfamília ERF, sendo descartados das análises posteriores, e 8 contigs e 6 singlets demonstraram pertencer à subfamília CBF/DREB, sendo somente estas sequências avaliadas nas análises de filogenia e de expressão *in silico*.

A distinção entre os fatores de transcrição das subfamílias CBF/DREB e ERF foi feita pela análise dos aminoácidos 14 e 19 do domínio AP2/ERF, que de acordo com Sakuma et al. (2002), correspondem à valina e ácido glutâmico nos fatores de transcrição CBF/DREB e alanina e ácido aspártico. Estes aminoácidos apresentam um papel essencial na ligação às sequências alvo destes fatores de transcrição, no entanto, o resíduo de ácido glutâmico pode não se apresentar conservado como observado por Dubouze et al. (2003). Todos os EST-contigs e EST-singlets pertencentes à subfamília CBF/DREB encontrados no banco de dados CAFEST apresentaram estes resíduos conservados, à exceção dos contigs C26 e C28 e dos singlets CA00-XX-CS1-024-D10-CC.F e CA00-XX-IC1-018-F06-EC.F que não apresentaram o resíduo de ácido glutâmico na posição 19 do domínio AP2/ERF.

A análise filogenética das sequências das prováveis proteínas codificadas pelas sequências encontradas no banco de dados CAFEST e de sequências de fatores de transcrição CBF/DREB de outras espécies, mostrou que as sequências encontradas neste trabalho puderam ser alocadas em quatro dos seis subgrupos da subfamília CBF/DREB (Figura 1), os quais se distinguem de acordo com a similaridade de seus domínios AP2/ERF (Sakuma et al.,2002).

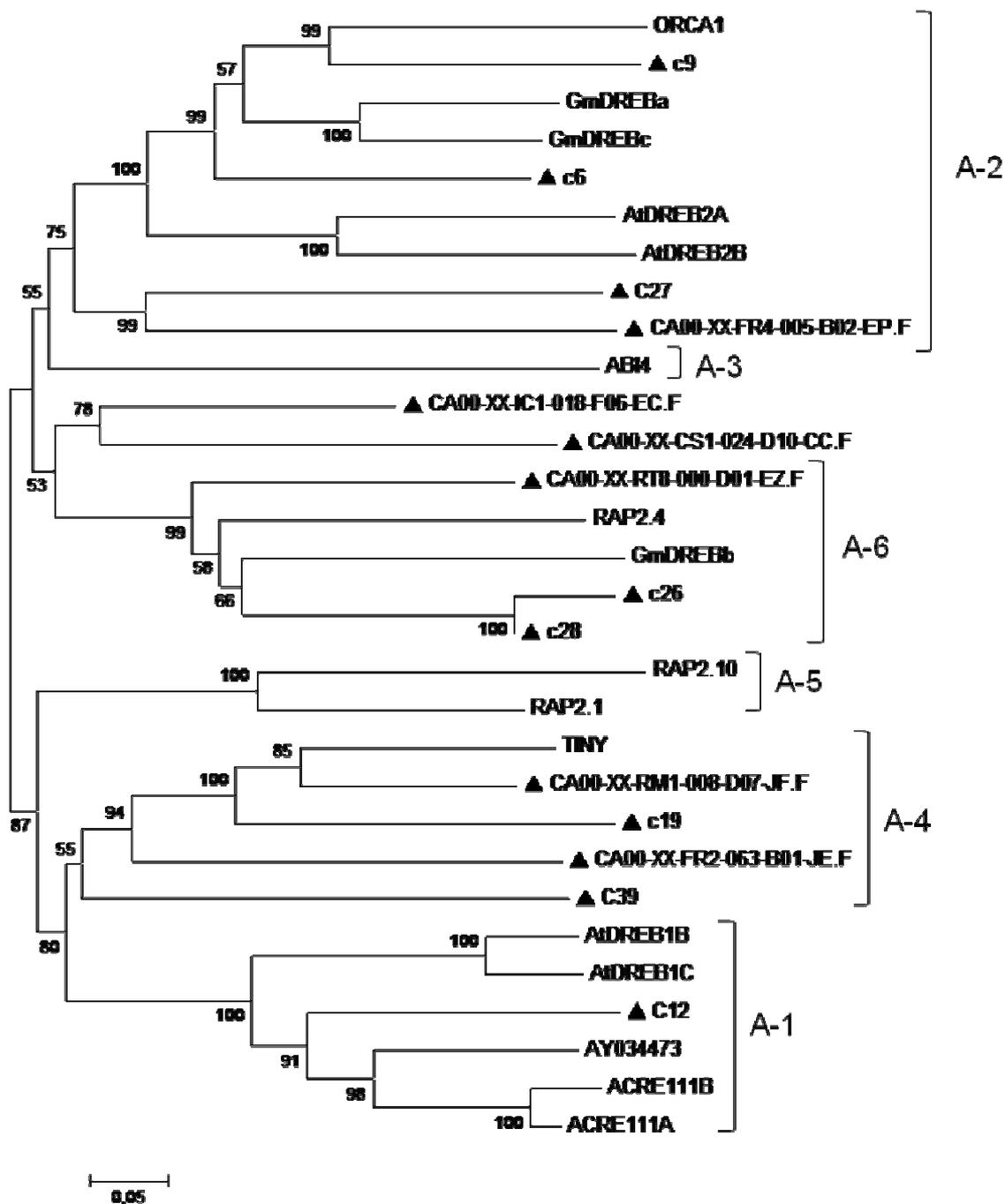


Figura 1 - Árvore filogenética envolvendo as sequências de aminoácidos dos prováveis fatores de transcrição CBF/DREB encontrados no banco de dados CAFEST (▲), e sequências de *Arabidopsis thaliana* [AtDREB1B(AAC49662), AtDREB1C(BAA33793), AtDREB2A(BAA33794), AtDREB2B(BAA33795), AIB4(AAC39489.1), RAP2.4(AAC49770.1), RAP2.1(AAM63886.1), RAP2.10(AAC49776.1), TINY(CAA64359.1)], *Catharanthus roseus* [(ORCA1(CAB93939.1)], *Glycine Max* [GmDREBa(AY542886), GmDREBb(AY296651), GmDREBc(AY244760)], *Lycopersicon esculentum* (AY034473), *Nicotiana tabacum* [ACRE111A(AF211530_1), ACRE111B(AF211531_1)]. Valores de *bootstrap* menores que 50% foram omitidos.

Quanto à expressão dos possíveis genes encontrados no banco de dados CAFEST pertencentes à subfamília CBF/DREB, o *Northern* Eletrônico demonstrou que juntos, estes genes se mostraram expressos em 21 bibliotecas, envolvendo tanto tecidos vegetativos como tecidos reprodutivos, diferentes estádios de desenvolvimento e diferentes condições, com destaque para aquelas envolvendo

estresses abióticos como nas bibliotecas SH2 e RT8 (Figura 2). Os fatores de transcrição CBF/DREB dos grupos A-1 e A-2 têm a sua expressão induzida pelo frio e seca, respectivamente (Dubouzet et al., 2003; Wang & Mao, 2009; Moreira et al., 2010), sendo que um dos contigs encontrados (C6) também foi agrupado em A-2 e se mostrou expresso na biblioteca de estresse hídrico (SH2) (Figura 2), corroborando com estes estudos anteriores. Poucos trabalhos analisaram a expressão destes fatores de transcrição em tecidos reprodutivos (Li et al., 2005), no entanto, os resultados aqui encontrados sugerem que estes fatores de transcrição podem apresentar um papel diferente nestes tecidos (Figura 2).

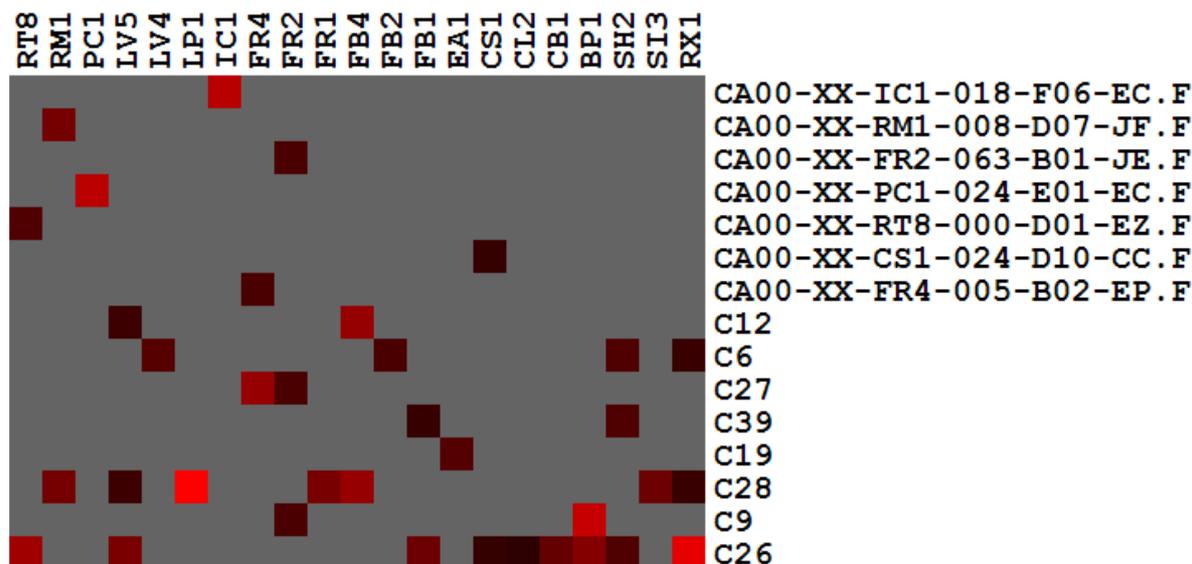


Figura 2 - Northern Eletrônico representando os níveis de expressão dos contigs e singlets dos possíveis genes CBF/DREB nas diferentes bibliotecas, em uma escala de cor do preto ao vermelho, onde quanto mais próximo ao vermelho maior o nível de expressão. Bibliotecas (Vieira et al., 2006): Plântulas e folhas tratadas com ácido aracdônico (LP1); Calos não embriogênicos com e sem 2,4D (IC1, PC1); Células em suspensão tratadas com acibenzolar-S-metil (BP1); Suspensão de células tratadas com acibenzolar-S-methyl e brassinoesteróides (CB1); Hipocótilos tratados com acibenzolar-S-metil (CL2); Células em suspensão tratadas com NaCl (CS1); Calo embriogênico (EA1); Botões florais em diferentes estádios de desenvolvimento (FB1, FB2, FB4); Botões florais + "pinhead" frutos + frutos em diferentes estádios (FR1, FR2); Fruto (*Coffea racemosa*) (FR4); Folhas jovens de ramos ortotrópicos (LV4, LV5); Folhas infectadas com lagartas e ferrugem (RM1); Células em suspensão em estresse com alumínio (RT8); Plantas estressadas por déficit hídrico (*Pool* de tecidos) (SH2); Sementes em germinação (Sementes inteiras e embriões zigóticos) (SI3), Caules infectados com *Xylella spp.*(RX1).

CONCLUSÃO

O banco de dados CAFEST se mostrou bastante representativo quanto aos possíveis fatores de transcrição CBF/DREB, os quais puderam ser alocados em quatro dos seis grupos presentes nesta subfamília de fatores de transcrição. A análise de expressão *in silico* mostrou que estes fatores de transcrição são expressos em diferentes tecidos, estádios de desenvolvimento e condições, com destaque para as condições de estresse e tecidos reprodutivos.

REFERÊNCIAS BIBLIOGRÁFICAS

AGARWAL, P.; AGARWAL, P. K.; JOSHI, A. J.; SOPORY, S K.;REDDY, M. K. Overexpression of PgDREB2A transcription factor enhances abiotic stress tolerance and activates downstream stress-responsive genes. **Molecular Biology reports**, v. 37, p.1125-1135, 2010.

DUBOUZET, J.G.; SAKUMA, Y.; ITO, Y.; KASUGA, M.; DUBOUZET, E.D; MIURA, S.; SEKI, M.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. Os-DREB genes in rice, *Oryza sativa* L., encoded transcription activators that function in drought-, high-salt- and cold responsive gene expression. **Plant Journal**, v. 33, p.751–763, 2003.

EISEN, M. B.; SPELLMAN, P. T.; BROWN, P. O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences**, (USA), 95: 14863-8, 1998.

HUANG, X, MADAN, A. CAP3: A DNA assembly program. **Genome Research**, v. 9 p. 868-877, 1999.

LI, X. P.; TIAN, A. G.; LUO, G. Z.; GONG, Z. Z; ZHANG, J. S.; CHEN, S. Y. Soybean DRE-binding transcription factors that are responsive to abiotic stresses. **Theoretical Applied Genetics**, v. 110, p. 1355–1362, 2005.

MOREIRA, R. S.; MEDRI, M.E.; NEUMAIER, N.; LEMOS, N. G.; BROGIN, R. L.; MARCELINO, F. C.; OLIVEIRA, M. C. N.; FARIAS, J. R. B.; ABDELNOOR, R. B.; NEPOMUCENO, A. L. Cloning and quantitative expression analysis of drought-induced genes in soybean. **Genetics and Molecular Research**, v. 9, n. 2, p. 858-867, 2010.

NAKANO, T.; SUZUKI, K; FUJIMURA, T.; SHINSHI, H. Genomewide analysis of the ERF gene family in Arabidopsis and rice. **Plant Physiology**, v. 140, p. 411–432, 2006.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, v. 4, p. 406– 425, 1987.

SAKUMA, Y.; LIU, Q.; DUBOUZET, J.G.; ABE, H.; SHINOZAKI, K.; YAMAGUCHI-SHINOZAKI, K. DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. **Biochemical and Biophysical Research Communications**, v. 290, p. 998–1009. 2002.

SITNIKOVA, T.; RZHETSKY, A.; NEI, M. Interior-branch and bootstrap tests of phylogenetics trees. **Molecular Biology and Evolution**, v. 12, p. 319-33, 1995.

TAMURA, K.; DUDLEY, J.; NEI, M.; KUMAR, S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. **Molecular Biology and Evolution**, v. 24 , p. 1596-1599, 2007.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, p. 4673–4680, 1994.

VIEIRA, L. G. E. ; ANDRADE, A. C. ; MONTE, D. C. ; ALMEIDA, E. R. P. ; SA, M. F. G. Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, v. 18, n. 01, p. 95-108, 2006.

WANG, C. T.; DONG, Y. M. Overexpression of maize ZmDBP3 enhances tolerance to drought and cold stress in transgenic Arabidopsis plants. **Biologia**, v. 64, n. 6, p. 1108-1114, 2009.

YAMAGUCHI-SHINOZAKI, K. & SHINOZAKI, K. A novel cisacting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature or high-salt stress. **Plant Cell**, v. 6, p. 251–264, 1994.