

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

ANÁLISE IN SILICO DE GENES QUE PARTICIPAM DAS
RESPOSTAS AOS ESTRESSES ABIÓTICOS (SECA/SALINIDADE)
NOS GENOMAS DE PLANTAS SUPERIORES

RECIFE
FEVEREIRO/2012

NINA DA MOTA SOARES CAVALCANTI

**ANÁLISE IN SILICO DE GENES QUE PARTICIPAM DAS
RESPOSTAS AOS ESTRESSES ABIÓTICOS (SECA/SALINIDADE)
NOS GENOMAS DE PLANTAS SUPERIORES**

Tese apresentada ao curso de Doutorado do Programa de Pós Graduação em Ciências Biológicas, da Universidade Federal de Pernambuco, como parte dos requisitos para obtenção do título de Doutor em Ciências Biológicas, na área de concentração Biotecnologia/Biologia Celular e Molecular.

Orientador(a): Ana Maria Benko Iseppon

RECIFE

FEVEREIRO/2012

Catálogo na Fonte:
Bibliotecário Bruno Márcio Gouveia, CRB-4/1788

N244c Cavalcanti, Nina da Mota Soares

Análise in silico de genes que participam das respostas aos stresses abióticos (seca/salinidade) nos genomas de plantas superiores / Nina da Mota Soares Cavalcanti. – Recife: O Autor, 2012.

207 folhas: il., fig., tab.

Orientador: Ana Maria Benko Iseppon

Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Ciências Biológicas. Pós-graduação em Ciências Biológicas, 2012.

Inclui bibliografia e anexos

- 1. Genética vegetal 2. Plantas – Efeito da seca 3. Angiosperma 4. Bioinformática I Iseppon, Ana Maria Benko (orientador) II. Título.**

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

Parecer da comissão examinadora da defesa de tese de Doutorado de **NINA DA MOTA SOARES CAVALCANTI**, intitulada: **ANÁLISE IN SILICO DE GENES QUE PARTICIPAM DAS RESPOSTAS AOS ESTRESSES ABIÓTICOS (SECA/SALINIDADE) NOS GENOMAS DE PLANTAS SUPERIORES**

A comissão examinadora, composta pelos professores abaixo, sob a presidência do primeiro, considera a candidata **APROVADA**.

Recife, 10 de Fevereiro de 2012

Prof^ª. Dr^ª. Ana Maria Benko Iseppon - Orientadora/Membro Interno - UFPE

Prof^ª. Dr^ª. Valesca Pandolfi - Membro Interno - UFPE

Prof^º. Dr^º. Adriano Barbosa da Silva - Membro Externo - UFPR

Prof^º. Dr^º. Valdir Queiroz de Balbino - Membro Externo – UFPE

Prof^º. Dr^º. Tercílio Calsa Junior - Membro Externo - UFPE

Dedido

*À Ana Elizabete Mota,
super-heroína de minha história.*

*À Clara Mota,
co-autora da minha história.*

AGRADECIMENTOS

A minha mãe, Ana Elizabete Mota, por perseguir os meus sonhos ao meu lado, respeitando meus pontos de vista e compreendendo minhas inquietudes;

A meu pai, Antonio Fernando, por seu apoio incondicional, muitas vezes sem saber o porquê, e por sua disponibilidade quase ilimitada - sempre alerta para meus intermináveis pedidos de favor;

A filhota Clarinha por admirar e valorizar minha profissão, de “cuidadora de plantas”;

Ao Leonardo Raffa, pelo companheirismo irrestrito, acompanhado de banhos de realidade (necessários) e declarações de amor (indispensáveis);

À Prof^a Ana Maria Benko Iseppon, que me acompanha desde a graduação, contribuindo para minha formação profissional. Mais de 10 anos de construção de laços de trabalho e de amizade;

Aos amigos Ana Carolina e Luis Carlos, união de anos que ultrapassa as grades da universidade;

Aos amigos da Bioinfo, tanto os que ainda estão na luta, quanto os que traçaram outras rotas. Um obrigada especial a João Pacífico sempre à disposição para ajudar nos trabalhos;

A todos - estudantes, mestrandos, doutorandos, técnicos e professores - que fazem do LGBV um verdadeiro time;

Ao Prof^o Ederson Kido e à Valesca Pandolfi pela contribuição e parceria em vários trabalhos;

Aos amigos extra-UFPE que acompanharam e torceram por mim em todas as etapas do doutoramento;

À Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) pela concessão da bolsa de doutorado;

À Universidade Federal de Pernambuco (UFPE), por meio do Programa de Pós-Graduação em Ciências Biológicas (PPGCB), pela oportunidade e suporte durante todo o período do doutoramento.

*“A verdade é que as primeiras
mudanças são tão lentas que mal se
notam, e a gente continua se vendo por
dentro como sempre foi, mas de fora os
outros reparam”*

Gabriel Garcia Marquez

LISTA DE FIGURAS	ix
LISTA DE TABELAS	xi
LISTA DE ABREVIATURAS.....	xii
RESUMO	16
ABSTRACT	17
INTRODUÇÃO	18
REVISÃO BIBLIOGRÁFICA.....	20
1. As mudanças ambientais e os estresses abióticos	20
2. Seca e salinidade: consequências para as plantas	22
3. Genes estresse-induzidos: respostas moleculares na tolerância aos estresses abióticos	25
4. Abordagens biotecnológicas da tolerância à seca e à salinidade	29
5. Bioinformática: desvendando as “ômicas”	33
6. Organismos Modelo	37
7. Culturas de interesse agrícola no Brasil e seus projetos de genômica	38
7.1. Feijão-caupi e soja: genomas de leguminosas.....	39
7.2. O eucalipto e o FORESTs	43
7.3. A cana-de-açúcar e o SUCEST.....	44
REFERÊNCIAS BIBLIOGRÁFICAS.....	46
CAPÍTULO I. Comparative in silico evaluation of MYB transcription factors in eucalyptus, sugarcane and rice transcriptomes	57
CAPÍTULO II. <i>In silico</i> identification of known osmotic stress responsive genes from <i>Arabidopsis</i> in soybean and <i>Medicago</i>	70
CAPÍTULO III. Overall picture of expressed Heat Shock Factors in <i>Glycine max</i> , <i>Lotus japonicus</i> and <i>Medicago truncatula</i>	84
CAPÍTULO IV. Differential gene expression in the prospection of tolerance genes against drought and salinity in native and cultivated plants.....	98
CAPÍTULO V. Plant-specific transcription factors in <i>Vigna unguiculata</i> transcriptome: sequence conservation and DNA-binding domain integrity	114

CAPÍTULO VI. Deciphering the Transcription Factors World	139
CONCLUSÕES	154
ANEXOS.....	155
Anexo I. Instrução para autores: Revista Genetics and Molecular Reserach	155
Anexo II. Instrução para autores: Revista PLoS Computational Biology	160
Anexo III. Instrução para autores: Revista Current Protein & Peptide Science	173
Anexo IV. Súmula curricular	183

REVISÃO BIBLIOGRÁFICA

Figura 1. Representação esquemática da cascata de eventos moleculares, desde a percepção do estresse até as respostas envolvidas na aquisição da tolerância.....	26
--	----

CAPÍTULO I

Figure 1. Dendrogram generated after Neighbor Joining analysis, showing relationships among the MYB seed sequences of arabidopsis and orthologs of eucalyptus, sugarcane, rice and other organisms.....	62
Figure 2. Prevalence of MYB factors (upper) and <i>in silico</i> expression pattern (bottom) in FORESTs (A) and SUCEST (B) libraries	63
Figure 3. Comparative analysis of MYB ORFs from arabidopsis and rice genomic sequences with eucalyptus and sugarcane clusters.....	64

CAPÍTULO II

Figure 1. Main categories of group I stress-inducible genes (protective molecules), indicating the number of orthologs identified in <i>Glycine max</i> and <i>Medicago truncatula</i>	73
Figure 2. Percentage of stress-inducible group II genes (signaling molecules), including cell signaling factors.....	73
Figure 3. Graphic representation of transcription factors identified in <i>Glycine max</i> and <i>Medicago truncatula</i>	74
Figure 4. Number of gene candidates from group I for <i>Medicago truncatula</i> and <i>Glycine max</i>	74
Figure 5. Categorization of soybean and medicago ‘unknown category’ candidates based on COG (Cluster of Orthologous Groups) functional database.....	75

CAPÍTULO III

Figure 1. Graphic representation of SuperSAGE tags distribution per cluster and compared libraries. A) Quantitative analysis (number of unitags per category). B) Qualitative analysis of tag prevalence (in %)......	90
Figure 2. Dendrogram showing relationships among HSF proteins of <i>A. thaliana</i> , <i>G. max</i> , <i>L. japonicus</i> and <i>M. truncatula</i>	92
Figure 3. Hierarchical clustering (Cluster3.0) of up-regulated (red), down-regulated (green) and not-regulated (black) soybean EST clusters ($p < 0.05$) related to HS response.....	93
Figure 4. <i>In silico</i> hybridization of HSF sequences against the SOYBASE database. Schematic representation of clusters that were anchored in the soybean based on BLAST similarity results.....	95

CAPÍTULO IV

Figure 1. Diversity and abundance stress-inducible genes, considering 1,010 candidates from different categories identified in cowpea.....	111
Figure 2. <i>Vigna unguiculata</i> transcription factor repertory; percentage of transcripts for each genic family.....	112
Figure 3. Graphic representation of transcription factor families expression in cowpea generated in the NordEST Transcriptome Project.....	113

CAPÍTULO V

Figure 1. Graphic representation of the SuperSAGE tags diversity for each transcription factor family. Numbers refer to absolute values.....	132
Figure 2. Quantitative distribution of SuperSAGE tags.....	133
Figure 3. Graphic representation showing the distribution of the differentially expressed transcripts (red circles), considering the libraries comparatively for DREB/ERF (A), NAC (B) and WRKY (C) families.....	134
Figure 4. Dendrogram generated after Neighbor-Joining analysis, showing relationships among WRKY orthologous and paralogous sequences.....	135
Figure 5. Graphic representation of the virtual 2D electrophoresis of WRKY transcription factor.....	136

CAPÍTULO VI

Figure 1. Online integrative script for transcription factor identification, characterization and annotation.....	147
---	-----

CAPÍTULO I

Table 1. Data mining resources and results using arabidopsis seed sequences, matches of eucalyptus, sugarcane and rice with respective domains.....	61
---	----

CAPÍTULO II

Table 1. Functional categories procured and respective seed-sequence amount.....	72
Table 2. Sequence description annotated according to the COG (Cluster of Orthologous Groups) functional category in <i>Glycine max</i> and <i>Medicago truncatula</i>	75
Table 3. Description of sequences with unknown function after AutoFACT analysis	75
Table S1. Identified candidates among abiotic stress responsive genes categories in soybean and medicago genomes based on selected arabidopsis seed sequences, as well as number of other hits, e-value and score, against the respective database of <i>Medicago truncatula</i> (Mt) and <i>Glycine max</i> (Gm).....	78

CAPÍTULO III

Table 1. <i>Arabidopsis thaliana</i> heat shock factors used as seed sequences and respective matches from <i>Glycine max</i> , <i>Lotus japonica</i> and <i>Medicago truncatula</i> , with corresponding domain information.....	88
Table 2. Distribution of unique SuperSAGE tags in the compared libraries, as compared with the soybean EST contigs identified.....	90
Table 3. Correspondence among identified GENOSOJA expressed sequence tags (ESTs) and characterized genes of <i>Glycine max</i>	94

CAPÍTULO IV

Table 1. SOS pathway orthologous genes in cowpea and soybean genomes.....	111
---	-----

CAPÍTULO V

Table 1. Data mining resources and results using <i>Arabidopsis thaliana</i> seed sequences, matches of vigna with respective informations about the alignments against NORDEST and GenBank databases, ORF and domains.....	137
Table 2. Information about <i>Arabidopsis thaliana</i> seed sequences, including gene names and locus.....	138

LISTA DE ABREVIATURAS

ABA	<i>Abscisic Acid</i> ; Ácido Abiscísico
ABI	<i>ABA-Insensitive</i> ; Insensível ao ABA
AD	<i>Activation Domain</i> ; Domínio de Ativação
AKT	<i>Potassium Channel</i> ; Canal de Potássio
AP2	APETALA 2
ATP	<i>Adenosine Triphosphate</i> ; Adenosina Trifosfato
BAC	<i>Bacterial Artificial Chromosome</i> ; Cromossomo Artificial de Bactéria
BLAST	<i>Basic Local Alignment Search Tool</i> ; Ferramenta Básica de Busca por Alinhamento Local
bHLH	<i>Basic Helix-Loop-Helix</i> ; Hélice-Alça-Hélice Básica
BRACELPA	Associação Nacional dos Fabricantes de Celulose e Papel
bZIP	<i>Basic Leucine Zipper</i> ; Ziper de Leucina Básico
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CARE	<i>Cis-Acting Regulatory Element</i> ; Elemento Regulatório Atuando em Cis
CAX	<i>Cation/H⁺ Exchanger</i> ; Trocador Cátion/H⁺
CBF	<i>C-repeat Binding Factor</i> ; Fator de Ligação de Repetição C
CD	<i>Conserved Domain</i> ; Domínio Conservado
cDNA	<i>Complementary DNA</i> ; DNA Complementar
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CONAB	Companhia Nacional de Abastecimento
CRT	<i>C-Repeat</i> ; Repetição C
CTD	<i>C-Terminal Domains</i> ; Domínios C-Terminais
DBD	<i>DNA-Binding Domain</i> ; Domínio de Ligação ao DNA
DDBJ	<i>DNA DataBank of Japan</i> ; Banco de Dados de DNA do Japão

DNA	<i>Desoxyribonucleic Acid; Ácido Desoxirribonucleico</i>
DREB	<i>Drehydration Responsive Element Binding Factor; Fator de Ligação Responsivo à Desidratação</i>
EMBL-Bank	<i>European Molecular Biology Laboratory Databank; Banco de Dado do Laboratório de Biologia Molecular Europeu</i>
EMBRAPA	<i>Empresa Brasileira de Pesquisa Agropecuária</i>
EREBP	<i>Ethylene-Responsive-Element-Binding Protein; Proteína de Ligação ao Elemento Responsivo ao Etileno</i>
ERF	<i>Ethylene Responsive Factor; Fator Responsivo ao Etileno</i>
EST	<i>Expressed Sequence Tag; Etiqueta de Sequência Expressa</i>
EUA	<i>Estados Unidos da América</i>
FACEPE	<i>Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco</i>
FAO	<i>Food and Agriculture Organization; Organização das Nações Unidas para Agricultura e Alimentação</i>
FAPESP	<i>Fundação de Amparo à Pesquisa do Estado de São Paulo</i>
FORESTs	<i>Projeto do Transcriptoma do Eucalipto (https://forests.esalq.usp.br/)</i>
GEE	<i>Gases do Efeito Estufa</i>
GenBank	<i>GeneBank (http://www.ncbi.nlm.nih.gov/Genbank/); Banco de Genes</i>
GENOLYPTUS	<i>Rede Brasileira de Pesquisa do Genoma do <i>Eucalyptus</i></i>
GENOSOJA	<i>Soybean Genome Project; Projeto do Genoma da Soja</i>
HKT	<i>High-Affinity Potassium Transport; Transportador de Alta Afinidade ao Potássio</i>
HS	<i>Heat Shock; Choque Térmico</i>
HSE	<i>Heat Shock Element; Elemento de Choque Térmico</i>
HSF	<i>Heat Shock Factor; Fator de Choque Térmico</i>
HSP	<i>Heat Shock Protein; Proteína de Choque Térmico</i>
HSR	<i>Heat Shock Response; Respostas ao Choque Térmico</i>

HTH	<i>Helix-Turn-Helix; Hélice-Volta-Hélice</i>
IPCC	<i>Intergovernmental Panel on Climate Change; Painel Intergovernamental de Mudanças Climáticas</i>
LEA	<i>Late Embryogenesis Abundant; Abundante na Embriogênese Tardia</i>
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MEGA	<i>Molecular Evolutionary Genetics Analysis; Análises Genéticas da Evolução Molecular</i>
MPSS	<i>Massively Parallel Signature Sequencing; Sequenciamento Paralelo Massivo de Assinaturas</i>
mRNA	<i>Messenger RNA; RNA Mensageiro</i>
MTF	<i>Membrane-bound Transcription Factors; Fator de Transcrição Ligado à Membrana</i>
MW	<i>Molecular Weight; Peso Molecular</i>
NCBI	<i>National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/); Centro Nacional para Informação Biotecnológica</i>
NES	<i>Nuclear Export Signals; Sinal de Exportação Nuclear</i>
NHX	<i>Na⁺/H⁺ exchanger; Trocador de Na⁺/H⁺</i>
NJ	<i>Neighbor-Joining; Agrupamento por Vizinhança</i>
NLS	<i>Nuclear Localization Signal; Sinal de Localização Nuclear</i>
NR	<i>Non-Redundant; Não Redundante</i>
ONSA	<i>Organization for Nucleotide Sequencing and Analysis; Organização para Sequenciamento e Análise de Nucleotídeos</i>
ORF	<i>Open Reading Frame; Quadro de Leitura Aberto</i>
PCR	<i>Polymerase Chain Reaction; Reação em Cadeia da Polimerase</i>
pI	<i>Isoelectric Point; Ponto Isoelétrico</i>
QTL	<i>Quantitative Trait Loci; Loci de Características Quantitativas</i>
RACE	<i>Rapid Amplification of cDNA Ends; Amplificação Rápida das Extremidades do cDNA</i>

RD	<i>Repression Domain; Domínio de Repressão</i>
RNA	<i>Ribonucleic Acid; Ácido Ribonucleico</i>
ROS	<i>Reactive Oxygen Species; Espécies de Oxigênio Reativa</i>
RT-qPCR	<i>Real Time quantitative PCR; PCR quantitativa em Tempo Real</i>
SAGE	<i>Serial Analysis of Gene Expression; Análise Serial da Expressão Gênica</i>
SNPs	<i>Single Nucleotide Polymorphism; Polimorfismo de Nucleotídeo Único</i>
SOS	<i>Salt Overly-Sensitive; Altamente Sensível ao Sal</i>
SSR	<i>Simple Sequence Repeat; Repetição de Sequência Simples</i>
SUCEST	<i>Sugarcane EST Project (http://sucest.lad.dcc.unicamp.br/en/); Projeto EST da Cana-de-açúcar</i>
TAIR	<i>The Arabidopsis Information Resource; O Recurso de Informação sobre Arabidopsis</i>
TF	<i>Transcription Factor; Fator de Transcrição</i>
TFDB	<i>Transcription Factor Binding Site; Sítio de Ligação de Fator de Transcrição</i>
TGI	<i>The Gene Index Project (http://compbio.dfci.harvard.edu/tgi/); O Projeto de Índice de Genes</i>
TSS	<i>Transcription Start Site; Sítio de Início da Transcrição</i>
UFPE	<i>Universidade Federal de Pernambuco</i>
UPGMA	<i>Unweighted Pair Group Method with Arithmetic Means; Método não polarizado de Agrupamentos aos Pares com Médias Aritméticas</i>
UTR	<i>Untranslated Region; Região Não Traduzida</i>
UV-B	<i>Ultravioleta - B</i>
WGS	<i>Whole Genome Shotgun; Quebra do Genoma Inteiro</i>

As mudanças climáticas globais sinalizam para um rearranjo importante na distribuição geográfica e abundância das espécies, diminuição da biodiversidade e extensão das terras afetadas pela seca e salinidade, dentre outros. Juntos os estresses abióticos são responsáveis por desencadear uma série de respostas nas plantas, que podem ser percebidas através das modificações morfológicas, fisiológicas, metabólicas e/ou moleculares, a fim de tolerar estes estresses. As respostas moleculares envolvidas nestes processos podem ser divididas em dois grupos principais: (I) aquelas que envolvem a ativação ou repressão de genes que atuam diretamente nas respostas, como as aquaporinas e os transportadores iônicos, dentre outros; e (II) aquelas que participam das etapas intermediárias destas respostas, como as proteínas quinases e os fatores de transcrição. Neste contexto, a pesquisa aqui apresentada teve como objetivo principal identificar e caracterizar os genes envolvidos na tolerância ao déficit hídrico e ao excesso de sais solúveis no solo em plantas superiores, partindo das informações disponíveis para a planta modelo *Arabidopsis thaliana* L.. A partir de bancos de dados privados e públicos e com o auxílio de ferramentas e programas *in silico* foi possível observar que a maioria dos genes superexpressos em *arabidopsis* sob condições de estresse salino e hídrico foram identificados nos genomas das plantas superiores estudadas. As diferenças observadas com relação à abundância dos genes, diversidade de isoformas ou nas estruturas gênicas e proteicas parecem ser atribuídas às mutações, incluindo *indels*, que ocorreram durante o processo de evolução e especiação das plantas. Tais modificações nos genomas das plantas são provavelmente resultantes dos processos de adaptação e seleção natural sofridos por estes organismos. Ainda, considerando a importância dos fatores de transcrição na modulação das respostas aos estresses abióticos foi elaborado um *pipeline* que integra vários programas e ferramentas de caracterização destes fatores; a criação deste *workflow* surgiu da necessidade de uma ferramenta única que fosse capaz de avaliar diferentes atributos dos fatores de transcrição, de forma que sua caracterização fosse a mais completa possível. Finalmente, os resultados apresentados serão de grande importância para entender melhor os mecanismos de respostas das plantas superiores aos estresses abióticos. Os dados obtidos neste projeto poderão ser úteis no delineamento de experimentos *in vitro* e *in vivo* visando o melhoramento genético de plantas de interesse econômico.

Palavras-chave: genes estresse-induzidos, estresse abiótico, angiospermas, bioinformática.

The global climate changes point to an important rearrangement in the geographical distribution and abundance of species, decrease of the biodiversity and extension of lands affected by drought and salinity, among others. Together the abiotic stresses are responsible for triggering a series of plant responses, which can be perceived through the morphological, physiological, metabolic and/or molecular modifications, to tolerate these stresses. The molecular responses involved in these processes can be divided into two major groups: (I) those that involve the activation or repression of genes that act directly on the responses, as aquaporins and ion transporters, among others; and (II) those that participate in the intermediary steps of these responses, such as kinase proteins and transcription factors. In this context, the here presented research aimed to identify and characterize genes involved in the tolerance to hydric deficit and excess of soluble salt in the soil in higher plants, based on the available information for the model plant *Arabidopsis thaliana* L.. From private and public databanks and with the aid of *in silico* tools and programs it was possible to observe that most genes over-expressed in arabidopsis under water and salt stress were identified in the studied genomes of higher plants. The observed differences regarding abundance of genes, diversity of isoforms or in the genes and proteins structures may be attributed to mutations, including *indels*, that occurred during the process of evolution and speciation of plants. Such changes in the genomes of plants were the result of adaptation and natural selection processes experienced by these organisms. Also, considering the importance of transcription factors in modulating the responses to abiotic stresses has been designed a *pipeline* that integrates various programs and tools for the characterization of these factors; the creation of this *workflow* arose from the necessity of a unique tool that is able to evaluate different attributes of transcription factors, so that their characterization was most complete as possible. Finally, the presented results will be of great importance to better understand the mechanisms of higher plants responses to abiotic stresses. The data from this project may be useful in the delineation of *in vitro* and *in vivo* experiments with a view to plants genetic improvement of economic interest.

Keywords: stress-induced genes, abiotic stress, flowering plant, bioinformatic.

As mudanças climáticas globais tendem a alterar substancialmente a biodiversidade, causando mudanças na fenologia, fisiologia e composição genética, entre outras, e afetando as interações entre as espécies e os processos do ecossistema. A liberação contínua e crescente dos gases do efeito estufa (GEE) está diretamente relacionada ao aumento da temperatura terrestre, alterando o clima no qual as espécies existentes evoluíram e estão plenamente adaptadas e do qual os seres humanos e os sistemas agrícolas dependem.

Como problemáticas associadas às alterações climáticas globais, a seca e a salinidade constituem uma séria ameaça à agricultura, resultando na deterioração das terras agricultáveis em todo o mundo. A baixa disponibilidade hídrica e a alta concentração de sais solúveis são consideradas os fatores climáticos de maior efeito sobre a produtividade agrícola, sendo também responsáveis pela distribuição das espécies nas diferentes zonas climáticas do globo.

Os efeitos da seca e da salinidade sobre as plantas são bastante variáveis em função de sua intensidade, velocidade de imposição do estresse e estágio de desenvolvimento da planta. Para manter o crescimento e a produtividade, as plantas devem adaptar-se às condições de estresse e ativar mecanismos de tolerância específicos. As modificações das plantas para aumentar esta tolerância são mais baseadas na ativação e regulação de genes que protegem e mantêm as funções e os componentes da estrutura celular. As plantas desenvolveram estratégias únicas para responder a cada mudança ambiental, exaustivamente monitorando o ambiente que as rodeiam e ajustando seus sistemas metabólicos e fisiológicos para manter a homeostase.

A elucidação dos mecanismos de controle molecular da tolerância aos estresses abióticos, os quais resultam na utilização de ferramentas moleculares para a engenharia de plantas mais tolerantes, é baseada na expressão de genes específicos relacionados ao estresse. Estes genes estão divididos em duas categorias: (I) os envolvidos na cascata de sinalização e no controle transcricional, como as quinases e os fatores de transcrição; (II) aqueles que funcionam diretamente na proteção de membranas e proteínas, como proteínas de choque térmico e chaperonas, proteínas LEA, osmoprotetores, aquaporinas, trocadores e transportadores de íons, dentre outros.

As respostas moleculares das plantas ao estresse hídrico e salino são complexas uma vez que envolvem muitos genes com efeitos aditivos. Por esta razão, o melhoramento das culturas na aquisição de tolerância não é um processo simples. Nas últimas décadas muitos esforços foram feitos no intuito de

gerar linhagens de diferentes espécies tolerantes aos estresses abióticos; muitas das linhagens transgênicas foram produzidas para superexpressar características específicas.

Do ponto de vista biotecnológico, uma maneira eficiente e rápida para estudar as respostas e adaptações moleculares das plantas aos estresses abióticos é através da análise do “pool” gênico de cada organismo. Para a identificação e caracterização destas características genéticas é essencial a bioinformática, um ramo multidisciplinar das ciências, que une esforços da informática, biologia, estatística, química e física, entre outras áreas. Com a bioinformática é possível ‘desvendar’ os genomas de plantas de interesse sócio-econômico, como feijão-caupi, soja, eucalipto e cana-de-açúcar; a partir da análise de suas sequências com ferramentas e programas específicos é possível comparar genes de organismos relacionados, traçar perfis de expressão de genes e vias de interesse, fazer predições da estrutura e função de proteínas, entre outros. As informações conseguidas através da bioinformática são essenciais na prospecção de culturas com características específicas de resistência e/ou tolerância aos estresses abióticos.

Uma das maneiras mais usuais de utilizar a bioinformática na prospecção de genes e proteínas e suas funções é através da análise comparativa utilizando como parâmetro as informações genéticas disponíveis para organismos modelo. Como exemplo de organismo modelo é possível citar *Arabidopsis thaliana* para as dicotiledôneas, *Oryza sativa* para as monocotiledôneas e *Medicago truncatula* para as leguminosas.

Este trabalho objetivou identificar, caracterizar e analisar os genes envolvidos nas respostas aos estresses abióticos relacionados à seca e salinidade em plantas superiores, através do reconhecimento de suas regiões codificantes e domínios conservados, avaliando padrões de evolução e expressão, e analisando comparativamente a estrutura destes genes.

Trabalhos como este, que visam identificar e caracterizar algumas famílias gênicas e vias que atuam diretamente nas respostas aos estresses abióticos, identificando potenciais candidatos para o desenvolvimento de cultivares mais tolerantes à seca e às altas concentrações de sais no solo, são essenciais na manutenção da produção agrícola mundial. Os resultados obtidos neste estudo devem auxiliar no conhecimento de genes-alvo relacionados à tolerância/resistência aos estresses abióticos em feijão-caupi, soja, eucalipto e cana-de-açúcar, contribuindo para o maior entendimento sobre seus mecanismos de funcionamento.

1. Mudanças ambientais e estresses abióticos

Atualmente identifica-se nas mudanças climáticas globais uma das maiores preocupações de pesquisadores e estudiosos em todo o mundo. A queima de combustíveis fósseis e a mudança de uso ou cobertura do solo, associados ao crescimento populacional, vêm contribuindo para a intensificação do efeito estufa, o que coloca em risco a biodiversidade do planeta (BUCKERIDGE et al., 2007).

As altas concentrações dos gases do efeito estufa estão diretamente relacionadas com o aumento gradativo da temperatura de equilíbrio da terra, que tem como consequência a ampliação de eventos climáticos extremos, elevação do nível do mar, perda de cobertura do gelo polar, alteração na disponibilidade de recursos hídricos, mudanças nos ecossistemas, extinção de espécies, desertificação, impactos na saúde da população humana e interferências na agricultura (NASCIMENTO, 2009).

A temperatura média global subiu 0,6°C no século 20 e pode ser elevada em mais 1°C até 2030, segundo o Painel Internacional de Alterações Climáticas (2011). A projeção para 2090 indica aumento de até 3°C, caso medidas de prevenção não sejam adotadas (IPCC, 2011).

Pellegrini et al. (2007) concluíram que haverá perdas, em termos nacionais, de área de até 18% para o arroz, 11% para o feijão, 39% para a soja, 58% para o café e 7% para o milho. Já a soja poderá sofrer perdas de produtividade em até 40% em 2070. Estima-se que no Nordeste brasileiro ocorrerá uma redução da área para produção das culturas do milho, arroz, feijão, algodão e girassol (ASSAD E PINTO, 2008).

Como consequências das alterações globais em andamento destacam-se o prolongamento dos períodos de seca e a expansão das áreas salinizadas, que juntos são os principais fatores limitantes para o crescimento e desenvolvimento de culturas de grande importância econômica. Estima-se que as perdas na agricultura causadas pelo baixo rendimento das culturas sejam de bilhões de dólares todos os anos. Baseado nos Relatório do “European Heatwave” de 2003, Cias et al. (2005) especularam que as perdas anuais da produção agrícola sejam de aproximadamente 30%. Segundo previsões, o déficit hídrico e a salinização dos solos afetarão mais de 50% das áreas destinadas à agricultura até 2050 (WANG et al., 2003).

No entanto, é importante ressaltar que embora as previsões sejam globais, as regiões agricultáveis tendem a ser diferencialmente afetadas e impactadas pelas mudanças climáticas e suas consequências (MITTLER e BLUMWALD, 2010). Segundo Bohnert et al. (2001), áreas onde a

produção de alimentos é problemática coincidem geralmente com os lugares onde a água é um artigo precioso, onde secas prolongadas são frequentes, solos salinos e marginais predominam e extremos de temperatura comprometem a produção. Mais de um quinto das regiões tropicais e subtropicais destinadas à produção de milho, por exemplo, são afetadas por seca e salinidade (HAYANO-KANASHIRO et al., 2009); no México, cerca de 80% de todo o milho cultivado está crescendo sob condições de sequeiro, onde as possibilidades de minimizar os efeitos do estresse hídrico são limitadas (HAYANO-KANASHIRO et al., 2009).

A disponibilidade de água impõe forte pressão seletiva, uma vez que a mesma é fundamental em todos os aspectos fisiológicos das plantas (BRAY, 1997). A redução da disponibilidade de água nos solos já afeta sobremaneira o rendimento das colheitas de diferentes grãos em muitos países do mundo; o grão de bico, por exemplo, sofre frequentemente com a seca terminal, que causa atraso no florescimento e perda no rendimento da produção (VARSHNEY et al., 2009).

Embora a alteração no status da água seja um fator em vários estresses abióticos, a mesma tem uma importância mais óbvia na seca (VERSLUES et al., 2006). A seca pode ser definida simplificarmente como um período de precipitações abaixo do normal que limita a produtividade das plantas pela baixa disponibilidade de água nos solos, tanto em sistemas naturais como na agricultura (VERSLUES et al., 2006).

Além da depleção continuada dos níveis de água dos solos, que tem como consequência o acúmulo de sais, também a demanda por irrigação têm levado à salinização das terras agricultáveis. A distribuição geográfica da maioria das plantas altamente sensíveis à salinidade, como algumas leguminosas, é delimitada pelas altas concentrações de NaCl dissolvidos nos solos. Assim, como a água, a salinidade é um fator restritivo ao desenvolvimento de muitas culturas (VARSHNEY et al., 2009).

A deposição de altas concentrações de sais, especialmente o Na^+ , pode alterar a estrutura básica do solo, resultando na diminuição da porosidade do mesmo, provocando redução na aeração e na condutância da água (MAHAJAN e TUTEJA, 2005). Além disso, estas altas concentrações causam diversos tipos de estresse em plantas, incluindo alteração na captura de nutrientes, especialmente os íons K^+ e Ca^{+2} , acumulação de íons tóxicos, principalmente o Na^+ , estresse osmótico e oxidativo (VERSLUES et al., 2006).

A relação entre as plantas e o ambiente influencia todos os processos vitais, seja pela disponibilização do oxigênio seja como fonte de alimento. O equilíbrio desta relação é particularmente crítico para o desenvolvimento agrícola sustentável. Como citado por Yi e Hong-Bo (2008), tornar

harmônica a relação entre as plantas e o ambiente (em particular o solo) é o problema-chave para o desenvolvimento de uma agricultura sustentável moderna frente às rápidas mudanças climáticas globais.

A seca, a salinidade e seus efeitos diretos sobre as plantas estão sendo extensivamente estudados. O interesse na área tem aumentado visto que existe uma necessidade urgente no desenvolvimento de variedades tolerantes ao déficit hídrico e às altas concentrações de sais solúveis no solo, tanto através do melhoramento convencional, quanto por engenharia genética, no intuito de minimizar as injúrias e as perdas causadas por tais estresses (BARTELS e SUNKAR, 2005).

2. Seca e salinidade: consequências para as plantas

Durante os últimos 400 milhões de anos da evolução das plantas, com a pressão seletiva dos ambientes secos e salinos fora dos mares, ocorreram variações no comportamento entre espécies, as quais evoluíram, gerando plantas adaptadas aos ambientes mais áridos (DIETRICH et al., 2001). O processo de seleção em regiões marginais, sujeitas à falta de água e de nutrientes, tornou os materiais genéticos nativos dessas áreas mais adaptados aos estresses ambientais, como é o caso do feijão guandu, do amendoim, do mileto e do sorgo, quando comparados com os genótipos originários de regiões mais úmidas e de alta fertilidade natural, como o feijão comum e o milho (HARLAN, 1992).

Assim, no curso da evolução e da domesticação artificial as plantas evoluíram e adquiriram uma grande variedade de estratégias metabólicas, fisiológicas e de desenvolvimento para otimizar a absorção de água e o controle iônico. O balanço eficiente destes processos é essencial para a manutenção do crescimento vegetativo e da reprodução das plantas (HAYANO-KANASHIRO et al., 2009).

As plantas sobrevivem em um ambiente contínuo solo-planta-atmosfera e precisam coordenar os mais diversos processos fisiológicos, metabólicos e moleculares para responder às mudanças deste ambiente de forma a não comprometer seu crescimento e desenvolvimento (YI E HONG-BO, 2008).

Com relação às plantas, a seca e a salinidade podem ser vistas como estresses multidimensionais, que as afetam em vários níveis de sua organização. Os mecanismos de adaptação e/ou tolerância vegetal são complexos porque refletem a integração dos efeitos dos estresses e das respostas das plantas sob todos os níveis da organização, integrados no espaço e no tempo (PIMENTEL, 2004).

A água é um dos fatores-chave no desenvolvimento vegetal, influenciando efetivamente sua nutrição e processos fisiológicos essenciais, como as reações fotossintéticas; o efeito do déficit hídrico

sobre o crescimento das plantas implica menor disponibilidade de CO₂ para a fotossíntese e limitação dos processos de alongação celular (YI e HONG-BO, 2008).

Existe uma ampla variação no grau de tolerância à falta d'água entre as espécies e mesmo entre variedades. Além disto, o estágio de desenvolvimento em que a planta se encontra no momento da exposição ao estresse também é crítico para a produtividade agrícola. Em grande parte das plantas, a deficiência hídrica causa 25% de diminuição de produção quando aplicada antes da floração, 50% durante a floração e 21% quando ocorre no enchimento de grãos (PIMENTEL, 2004). Na cultura do milho, por exemplo, as semanas que antecedem e que seguem a formação dos órgãos reprodutivos são o período em que a suplementação de água tem maior efeito na sua produtividade, já na cultura do trigo uma deficiência hídrica leve causa precocidade de floração enquanto uma seca severa causa retardo na floração (PIMENTEL, 2004).

Marur et al. (1994) argumentaram que tecidos vegetais mostram diversas alterações nas vias metabólicas em resposta ao déficit hídrico, que ocorrem em três níveis principais: i) perturbações das vias metabólicas levando a um acúmulo ou perdas de metabólitos; ii) alterações das atividades enzimáticas; iii) alterações dos padrões proteicos.

Para suportar os efeitos do déficit hídrico, as plantas podem desenvolver mecanismos tanto para retardar a desidratação, através da manutenção da hidratação do tecido, como para tolerar a desidratação, pela capacidade de funcionar enquanto desidratada (TAIZ e ZEIGER, 2006). Mecanismos para retardar a desidratação podem ser resultantes da diminuição da perda de água (pelo fechamento estomático, desenvolvimento de tricomas, redução da área foliar e senescência de folhas velhas para diminuir a transpiração, além de mudanças no tamanho, orientação e enrolamento das folhas para minimizar a superfície exposta à luz, etc.) ou maximização da absorção de água (aumento do crescimento radicular). Já os mecanismos de tolerância envolvem o ajustamento osmótico e eficiência na remoção de espécies reativas de oxigênio (TAIZ e ZEIGER, 2006).

As respostas das plantas ao déficit hídrico mudam dependendo da espécie e do genótipo, da intensidade de tempo de exposição ao estresse, da idade e do estágio de desenvolvimento, do órgão, bem como do tipo e compartimento celular onde ocorre (BARTELS e SUNKAR, 2005; RAMPINO et al., 2006). Com relação aos mecanismos de adaptação, estes podem ser resultantes de processos em curto prazo, como as mudanças na conformação da parede celular, e longo prazo, como mudanças morfológicas do hábito da planta (BARTELS e SUNKAR, 2005; BECK et al., 2007).

O aumento da concentração de sais no solo afeta primariamente o metabolismo do carbono pela toxicidade iônica, induzindo deficiência nutricional, déficit hídrico e estresse oxidativo, modulando os níveis de metabólitos secundários (DOOKI et al., 2006; WAHID e GHAZANFAR, 2006).

A inibição do crescimento das plantas pela salinidade é causada pelos efeitos osmóticos e iônicos. O ajuste osmótico é considerando um importante mecanismo para a tolerância à salinidade e à seca. A redução no potencial osmótico em plantas estressadas pode ser o resultado da acumulação de íons inorgânicos (Na^+ , Cl^- e K^+) e solutos orgânicos compatíveis. O ajuste osmótico tanto nas raízes como nas folhas contribui para a manutenção da obtenção de água e do turgor celular, permitindo que os processos fisiológicos, como abertura dos estômatos, fotossíntese e expansão celular não sejam alterados (AZEVEDO-NETO et al., 2004). Se os efeitos osmóticos adversos e íon específico da absorção de sais excedem o nível de tolerância funcional da planta, ocorrem distúrbios funcionais e injúrias. A fotossíntese não é limitada apenas devido ao fechamento estomático, mas também pelo efeito do sal sobre os cloroplastos, em particular sobre o transporte eletrolítico e sobre os processos secundários (LARCHER, 2000).

Em suma, a salinidade excessiva reduz o crescimento da planta por causar aumento no consumo de energia para absorver água do solo e para realizar os ajustes bioquímicos necessários para sobreviver em condições de estresse (RHOADES et al., 2000). Culturas que são expostas às condições de alta salinidade por tempo prolongado apresentam sintomas de déficit hídrico, que são percebidos através dos efeitos como paralisação do crescimento, danos nas folhas, necroses ou danos nos tecidos (AYERS e WESTCOT, 1999). Muitas vezes, a planta pode não exibir nenhum sintoma de estresse salino e apenas a comparação com plantas normais pode revelar a magnitude da inibição causada pelo baixo potencial osmótico do meio onde a planta está submetida (BLANCO, 1999). De acordo com Gheyi (2000) os sais – além de afetarem a disponibilidade de água e provocarem toxicidade – podem causar um estado de desequilíbrio nutricional na planta, onde o excesso de determinado íon, na planta ou no solo, afeta a disponibilidade de outro e, como consequência, o crescimento e o desenvolvimento vegetal ficam prejudicados.

A tolerância à seca e salinidade é uma característica multigênica complexa (HAYANO-KANASHIRO et al., 2009). As primeiras respostas das plantas que protegem a planta da perda exaustiva de água e da acumulação de íons tóxicos incluem a redução no crescimento vegetativo, fechamento dos estômatos e diminuição das taxas fotossintéticas. Estudos fisiológicos têm mostrado que açúcares, alcoóis de açúcar, aminoácidos e aminas funcionam como osmólitos, protegendo as funções celulares

dos efeitos da desidratação. O acúmulo de tais compostos está diretamente relacionado às condições estressantes de seca e salinidade às quais diferentes espécies de plantas foram submetidas (HAYANO-KANASHIRO et al., 2009).

Para que todas as respostas e a coordenação de mecanismos que levam à aclimação das plantas às condições de estresse abiótico aconteçam é necessário o envolvimento de centenas de genes. Estas respostas são também afetadas por interações entre diferentes fatores ambientais e o estágio de desenvolvimento da planta e podem encurtar o ciclo de vida da mesma, reduzir ou interromper a produção de sementes ou ainda acelerar a senescência (MILTTLER e BLUMWALD, 2010).

Considerando que a maioria das angiospermas não tolera o déficit hídrico e as altas concentrações de sais, a identificação e compreensão dos mecanismos de tolerância a estes estresses são fundamentais no desenvolvimento de cultivares comerciais mais tolerantes (SOARES-CAVALCANTI, 2004).

3. Genes estresse-induzidos: respostas moleculares na tolerância aos estresses abióticos

Sob condições estressantes, a nível celular, diversos genes com potencial função na adaptação à seca e à salinidade são reprimidos ou ativados, interferindo e modulando as funções metabólicas que serão refletidas nas alterações fisiológicas e de desenvolvimento vegetal.

Os componentes moleculares de respostas aos estresses ambientais incluem o reconhecimento do sinal ambiental (entrada), a transdução do sinal (cascata de reações bioquímicas), a ativação ou repressão de genes estresse-induzidos e a expressão de proteínas que caracterizarão o fenótipo tolerante (saída). Assim, o fenótipo tolerante pode ser traduzido como o resultado de uma rede multidimensional de interações com muitos níveis de expressão e regulação (YI e HONG-BO, 2008) (Figura 1). Genes estresse-induzidos podem ser divididos em dois grupos principais: (i) o primeiro grupo inclui as proteínas que diretamente conferem resistência às células, como as proteínas LEA (*Late Embryogenesis Abundant*; Abundante na Embriogênese Tardia), osmorreguladores, aquaporinas, transportadores de sódio, relacionados ao metabolismo secundário, dentre outros; (ii) o segundo tipo compreende aqueles que desempenham função importante na regulação da expressão de outros genes e na transdução de sinais, como os fatores de transcrição e as proteínas quinase (SEKI et al., 2002).

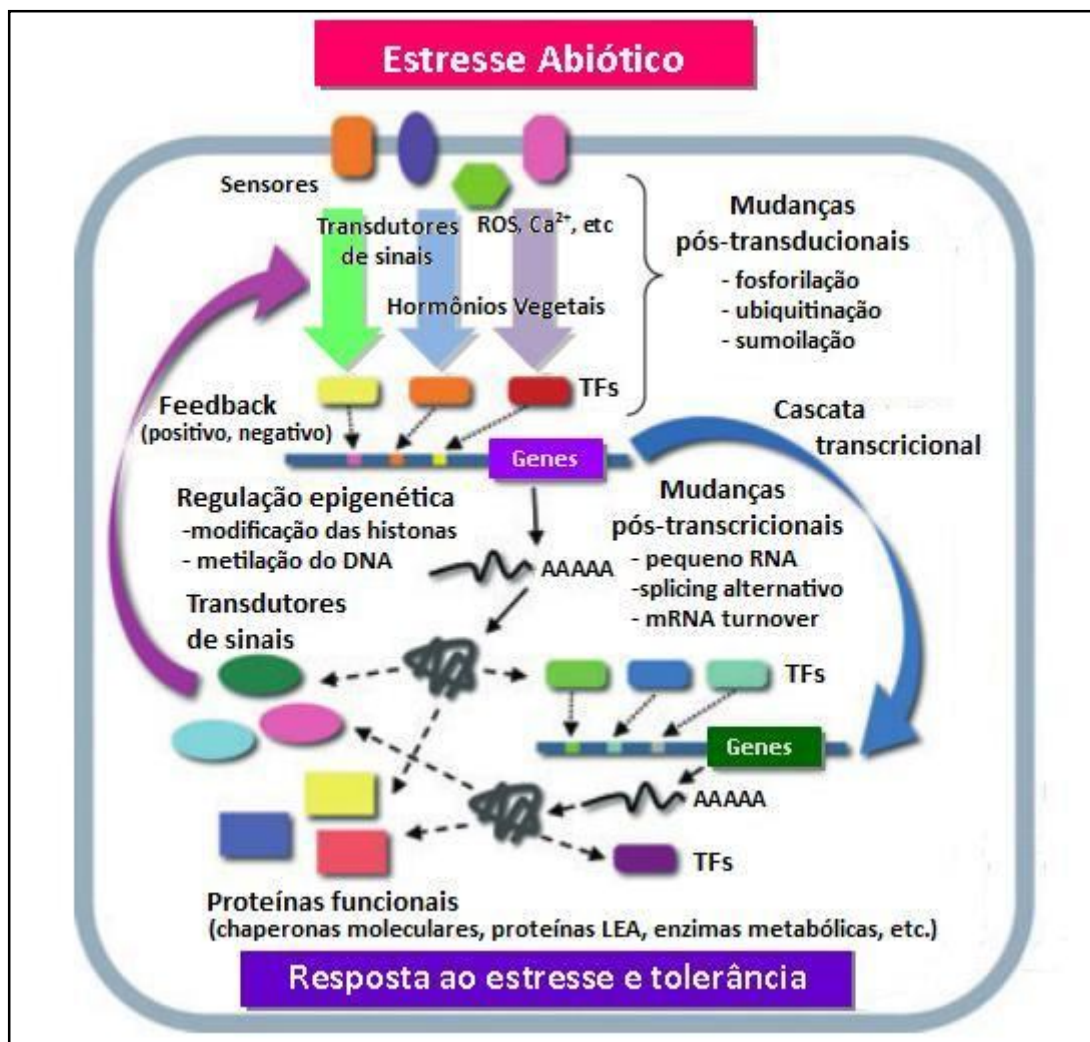


Figura 1. Representação esquemática da cascata de eventos moleculares, desde a percepção do estresse até as respostas envolvidas na aquisição da tolerância (Adaptado de HIRAYAMA e SHINOZAKI, 2010).

Em relação aos mecanismos bioquímicos que as plantas utilizam para sobreviver frente às intempéries ambientais, as principais estratégias envolvem: (i) acúmulo seletivo ou exclusão de íons; (ii) controle da absorção de íons pela raiz e do transporte nas folhas; (iii) compartimentalização de íons em nível celular e na planta como um todo; (IV) síntese de solutos compatíveis; (v) mudanças nas vias fotossintéticas; (vi) alteração da estrutura da membrana; (vii) indução de enzimas antioxidantes e (viii) indução de hormônios vegetais (PARIDA e DAS, 2005).

A habilidade das células em responder aos estímulos externos em tempo hábil e de maneira específica é tema recorrente na biologia molecular. Seki et al. (2002) prepararam microarranjos de ‘cDNA full length’ contendo aproximadamente 7.000 grupos independentes de ORFs (*Open Reading Frames*, Quadros de Leitura Aberto) gênicas completas, para analisar o perfil de expressão dos genes

sob condições de seca, salinidade e baixas temperaturas. Este experimento indicou que os transcritos de 53, 277 e 194 genes, respectivamente, aumentaram cinco vezes ou mais em relação ao grupo controle, depois dos tratamentos com déficit hídrico, alta salinidade e frio, respectivamente. Além disso, foi observado que 22 genes respondiam a todos os três estresses estudados.

Mais recentemente, Matsui et al. (2008) utilizando um subtipo do chip de microarranjo, o *tiling arrays*, para estudar todo o genoma de arábidsopsis sob condições de estresse abiótico observaram a expressão de 16.876 e 17.920 genes responsivos à seca e salinidade, respectivamente; tais genes codificantes encerravam dois grupos principais de proteínas: as funcionais e as regulatórias. Interessante notar que deste total, respectivamente 2.097 e 2.063 genes estavam associados à fotossíntese.

Os mecanismos moleculares de respostas à seca em tomate foram estudados através do perfil de expressão de duas linhagens tolerantes e uma cultivar sensível sob condições de déficit hídrico utilizando experimentos de microarranjo. Dos genes identificados, aproximadamente 400 estavam relacionados à tolerância apenas na linhagem tolerante; dentre estes, grande quantidade de fatores de transcrição e proteínas de sinalização estão incluídos, os quais podem ser reguladores globais da resposta à seca (GONG et al. 2010).

Ademais, Guo et al. (2009) monitorando as mudanças da expressão gênica em nível transcricional nas folhas de dois genótipos tolerantes e um sensível à seca em cevada no estágio reprodutivo, observaram que 17 genes foram expressos exclusivamente no genótipo tolerante, incluindo genes relacionados ao fechamento estomatal, à síntese de osmoprotetores, como a glicina-betaína, e à estabilização da membrana plasmática.

Muitos genes induzidos pelo estresse salino têm sido estudados em plantas superiores. A superexpressão de elementos de ligação responsivos à desidratação (DREB2; *Dehydration Responsive Element Binding*) em soja permitiu o acúmulo de altos níveis de prolina livre, quando comparada às plantas selvagens sob condições de estresse salino; este gene também é um importante ativador transcricional, sendo usualmente empregado no melhoramento de plantas (SOBHANIAN et al., 2010).

Quando sob condições de alta salinidade as plantas mantêm altas concentrações de K^+ e baixas concentrações de Na^+ no citosol através da regulação da expressão e da atividade de transportadores de potássio e sódio e das bombas de H^+ que geram a força motriz para o transporte (ZHU et al., 1998). Evidências sugerem que um complexo proteico formado pela proteína ligadora de cálcio SOS3 e pela proteína quinase de serina/treonina SOS2 possam ser ativados por um sinal de cálcio elicitado pelo estresse salino. Este complexo, então, seria responsável pela fosforilação e ativação de vários

transportadores iônicos, como os trocadores de Na^+/H^+ de membrana (SOS1) e vacúolo (NHX1) (QIU et al., 2002).

Em tomate, Chen e Plant (1999) mostraram uma síntese transitória de proteínas desconhecidas responsivas à salinidade de plantas submetidas a menos de quatro dias de exposição ao estresse salino. Os autores também demonstraram que o ácido abscísico (ABA) não desempenhava papel fundamental na síntese destas proteínas. Ainda utilizando plantas de tomate, enquanto Amini et al. (2007) identificaram apenas cinco proteínas responsivas após 24 h de estresse, Chen et al. (2009) submetendo as plantas a sete dias de condições estressantes identificaram 23 proteínas responsivas; dentre elas foram observadas algumas proteínas anti-oxidantes, proteínas de choque térmico e proteínas associadas ao metabolismo de carboidratos.

Para a aclimação das plantas ao estresse abiótico há o disparo de uma cascata ou rede de eventos que inicia com a percepção do estresse e termina com a expressão de um conjunto de genes-alvo, que fazem uso de rotas e componentes comuns às respostas de diferentes estresses. Este fenômeno, que é conhecido como tolerância cruzada, permite a aclimação das plantas a uma gama de diferentes estresses após exposição a um estresse específico (PASTORI e FOYER, 2002). A tolerância cruzada pode ser explicada pelo alto grau de sobreposição entre conjuntos de genes determinantes da tolerância a um estresse particular (RIZHSKY et al., 2002) ou devido ao fato de que os mecanismos de sinalização e resposta ao estresse envolvem rotas metabólicas comuns (BORISOVA et al., 2001).

Exemplo de tolerância cruzada envolve deficiência hídrica e o excesso de calor (RIZHSKY et al., 2002), o excesso de calor e a salinidade (SONG et al., 2005), bem como a radiação UV-B e deficiência de água (BORISAVA et al., 2001). A deficiência hídrica, por exemplo, muitas vezes está associada à salinidade na biosfera das raízes e ao estresse por calor nas folhas (resultante de diminuição do esfriamento evaporativo devido à baixa transpiração), enquanto o resfriamento e o congelamento levam a reduções na atividade hídrica e ao estresse osmótico (TAIZ e ZEIGER, 2006).

Dados integrados obtidos a partir de abordagens das diferentes ‘ômicas’ e de pesquisas sobre as respostas das plantas aos diferentes estresses abióticos têm possibilitado o aumento do conhecimento sobre os mecanismos de tolerância vegetal no campo e permitem que um panorama mais abrangente das respostas aos estresses abióticos seja criado. Com base neste progresso, o melhoramento das culturas para suportar condições estressantes têm sido conseguido geralmente através da transferência de genes (HIRAYAMA e SHINOZAKI, 2010).

4. Abordagens biotecnológicas da tolerância à seca e à salinidade

As plantas têm se adaptado para sobreviver às condições ambientais adversas e para responder aos estresses em níveis moleculares e celulares, bem como fisiológicos e bioquímicos (YAMAGUCHI-SHINOZAKI e SHINOZAKI, 2006). Enquanto adaptações morfológicas e anatômicas podem ser de vital importância para algumas espécies, estas não acontecem em todas. Em contraste às modificações morfológicas, as respostas celulares parecem ter maior grau de conservação em todo o reino vegetal (RUIZ-LOZANO, 2003). Em nível molecular, as injúrias ambientais engatilham várias mudanças na expressão gênica, que podem definir as respostas de proteção de uma planta. Muitos esforços estão atualmente direcionados na compreensão dos mecanismos de resposta das plantas às condições nas quais a água e altas concentrações salinas limitam seu crescimento e desenvolvimento; tais esforços têm como objetivo identificar produtos gênicos que conferem tolerância a estes estresses (SALINET, 2009).

A engenharia genética é uma ferramenta poderosa na transformação das plantas, que passarão a apresentar maior resistência e/ou tolerância aos diferentes estresses. Muitos estudos têm tentado aumentar a tolerância ao estresse através da transferência de genes específicos. Os genes selecionados para a transformação são geralmente aqueles que codificam proteínas específicas induzidas por estresses abióticos com funções conhecidas; tais proteínas incluem os canais de água, as enzimas-chave para a biossíntese de osmólitos e desintoxificação, além das relacionadas ao transporte iônico. Vários genes individuais parecem ter algum impacto positivo na tolerância ao estresse, no entanto a tolerância à seca e salinidade é uma característica complexa e multigênica, que envolve vários genes e vias, não estando sob controle de um único gene (ATHAR e ASHAF, 2009).

De fato, a transformação vegetal é uma técnica usual na compreensão do funcionamento das plantas sob determinadas condições e para o melhoramento da cultura através da aquisição de características específicas. Para obtenção de atributos desejáveis, modernas ferramentas da biotecnologia têm sido empregadas, apresentando bons resultados. Embora o objetivo do desenvolvimento de novas variedades (que apresentem maior produtividade, vigor, resistência a doenças, qualidade nutricional, etc.) através da biotecnologia seja o mesmo do obtido através do melhoramento convencional, observa-se que as técnicas não convencionais de reprodução oferecem possibilidades adicionais para geração de variabilidade, não considerados anteriormente pelo melhoramento convencional (BORÉM, 1998).

As abordagens atuais indicam a importância dos estudos de genômica funcional e da aplicação da engenharia genética na introdução de características desejáveis em culturas de interesse; vários

relatos descrevem o sucesso da transformação genética em diversas espécies vegetais. Papini-Tersi et al. (2007) realizaram experimentos baseados em microarranjos de cDNA contendo ESTs (*Expressed Sequence Tags*; Etiquetas de Sequências Expressas) de cana-de-açúcar e identificaram dois membros de transportadores da família aquaporina como estreitamente relacionados ao conteúdo de açúcar, indicando o potencial destes genes para caracterizar e distinguir variedades superiores ou para o desenvolvimento de novas variedades com maior conteúdo de açúcar, o que minimizaria os efeitos do estresse osmótico, propiciando um controle eficiente da compartimentalização de solutos.

Também relacionados à tolerância ao estresse abiótico, as tecnologias baseadas nos fatores de transcrição representam uma tendência cada vez maior de uso na biotecnologia vegetal (CENTURY et al., 2008). Agarwal e Jha (2010) discutiram o papel do hormônio ácido abscísico (ABA) na sinalização do estresse e seu envolvimento em diferentes sistemas regulatórios durante condições de estresse abiótico de forma ABA-dependente e independente. Os resultados indicaram que os fatores de transcrição atuam tanto nas vias dependentes quanto independentes do hormônio ABA. No entanto, é provável que estes dois grupos de vias ocorram em paralelo e de forma integrada, possibilitando assim a coordenação adicional entre os sinais elicitados pelo estresse e o ABA na regulação dos genes responsivos a estresses.

Uma estratégia mais ampla e promissora na aquisição de tolerância é o uso de fatores de transcrição, em especial aqueles que funcionam no ‘topo’ da cascata de eventos moleculares, regulando a expressão de vários genes relacionados ao estresse abiótico (BARTELS e SUNKAR, 2005).

Com a finalização do sequenciamento do genoma da soja, Mochida et al. (2009) realizaram uma análise computacional das sequências disponíveis da espécie com o objetivo de identificar todo o repertório de fatores de transcrição. Tal estudo permitiu a identificação dos fatores responsivos aos estresses abióticos de maneira sistêmica. Os resultados apresentados pelos autores deverão acelerar o progresso da “*transcription factoromics*” de soja, a genômica comparativa destas proteínas regulatórias em legumes, facilitando os programas de engenharia genética no sentido de obter ganhos de produtividade da cultura da soja sob condições adversas.

Kasuga et al. (1999) produziram plantas de *arabidopsis* altamente tolerantes à desidratação, pela superexpressão de um único gene, o DREB1A/CBF3, um fator de transcrição induzido por estresse. Também Tran et al. (2004) relataram que a superexpressão das proteínas CBF (codificadas pelos fatores de transcrição *CBF1*, *CBF2* e *CBF3*) em indivíduos transgênicos de *arabidopsis* alterou a expressão de

mais de 40 genes induzidos por estresse, levando a um aumento na tolerância ao congelamento, à seca e à salinidade.

Com relação à expansão dos efeitos das mudanças climáticas globais, principalmente considerando os consequentes impactos dos estresses salino e hídrico, Singh et al. (2010) analisaram os efeitos de múltiplos estresses no crescimento, fisiologia e metabolismo do feijão-caupi com o intuito de identificar genótipos tolerantes. Os autores concluíram que não obstante as altas concentrações de CO₂, o efeito combinado de radiação UV-B e de altas temperaturas possivelmente irão impor sérios problemas para o feijão-caupi e para muitas culturas num futuro próximo. No entanto, a identificação de genótipos tolerantes e de características-chave permitirá a seleção e desenvolvimento de cultivares tolerantes a múltiplos estresses através do melhoramento de plantas baseado nestas características-chave ou em programas de engenharia genética.

Recentemente, muitos genes que codificam antiporters Na⁺/H⁺ vacuolar (*NHX*) foram caracterizados e identificados em diferentes espécies (ADLER et al., 2010; LEIDI et al., 2010; FUKUDA et al., 2011; JHA et al., 2011). Li et al. (2010) clonaram um novo gene de antiporte vacuolar na halófito *Salsola soda* usando a técnica de RACE (*Rapid Amplification of cDNA Ends*; Amplificação Rápida de Extremidades de cDNA) e introduziram tais genes em alfafa (*Medicago sativa*), a qual se apresentou mais tolerante à salinidade e pôde crescer em condições de altos níveis de NaCl. As linhagens transgênicas de alfafa que sobreviveram em solos seriamente comprometidos pelo excesso de sais poderão ser úteis na exploração e no cultivo de áreas degradadas.

De acordo com Yang et al. (2009), apesar do gene *NHX* ser de grande importância no controle osmótico das plantas, para sua melhor atuação é essencial sua expressão em conjunto com os genes da via SOS (*Salt Overly Sensitive*; Altamente Sensível ao Sal). Os autores perceberam que plantas transgênicas co-expressando os genes *SOS1*, *SOS2* e *SOS3* tinham sua tolerância aumentada mais do que quando comparado a plantas transgênicas superexpressando qualquer um dos genes individualmente. Além disso, os autores evidenciaram que a superexpressão do gene *AtNHX1* em *A. thaliana* não aumentou significativamente a tolerância à salinidade quando comparada às plantas controle. Considerando que a intervenção da transformação vegetal pode facilitar a seleção e isolamento de linhagens tolerantes, as estratégias disponíveis têm tido bastante êxito na incorporação de tolerância em diferentes espécies (EL-SAYED et al., 2007).

Além dos exemplos citados, também o conhecimento obtido de estudos de plantas vivendo sob condições ambientais extremas tem fornecido informações importantes sobre os mecanismos de

tolerância ao estresse. Um dos grupos mais bem estudados é o das halófitas, como aquelas que ocorrem em mangues, as quais crescem sob condições de alta salinidade. Genericamente, os mecanismos pelos quais as halófitas toleram alta salinidade diferem entre as espécies, mas características em comum na regulação do conteúdo interno de Na^+ , K^+ e Cl^- vêm sendo observadas (FLOWERS e COLMER, 2008).

A halófito *Thellungiella halophila* é usada como planta modelo por estar intimamente relacionada à planta modelo *arabidopsis*. Como relatado por Inan et al. (2004), análises detalhadas desta halófito revelaram características adicionais na tolerância à salinidade, como a acumulação de mais prolina sob condições de estresse, maior seletividade a íons e maior sensibilidade na regulação estomatal. Dada a grande proximidade com *arabidopsis*, análises comparativas do transcriptoma destas duas espécies foram feitas com o objetivo de revelar tanto similaridades quanto diferenças no perfil de expressão dos transcritos relacionados ao estresse (TAJI et al., 2004; GONG et al., 2005). A expressão diferencial entre glicófitas e halófitas fornecerá maiores informações sobre as bases moleculares da tolerância à salinidade e outros atributos relacionados à tolerância a estresses em halófitas (HIRAYAMA e SHINOZAKI, 2010).

Adicionalmente, as análises comparativas de Loci de Características Quantitativas (QTL; *Quantitative Trait Loci*) entre espécies tolerantes e intolerantes estão recebendo maior atenção nas últimas décadas. Uma das aplicações mais importantes dos mapas genéticos baseados em QTLs é a localização de genes que controlam características de importância econômica e que podem ser aplicados ao melhoramento genético, como produção de grãos, altura da planta, teor de proteína, resistência a doenças, que resultam da ação cumulativa de um conjunto de genes. Assim, um dos grandes benefícios das abordagens baseadas em QTLs é que elas permitem a produção de culturas tolerantes através da combinação de QTLs responsivos a vários estresses abióticos (HIRAYAMA e SHINOZAKI, 2010). Muitos estudos relacionando QTLs a vários estresses abióticos são relatados na literatura; Ren et al. (2005) identificaram o locus *SKC1* codificando transportadores de alta afinidade à K^+ , o *HKT*, através da análise de QTLs para tolerância a salinidade usando variedades tolerantes e susceptíveis de arroz.

A evolução do conhecimento sobre os mecanismos de tolerância à desidratação em várias espécies, através do estudo de genomas funcionais e proteomas, fornece informações preciosas para o desenvolvimento de genótipos capazes de tolerar as condições adversas impostas pelo ambiente sem que a produtividade seja prejudicada substancialmente (NEPOMUCENO et al., 2001).

Estratégias como o uso de chaperonas, que protegem componentes celulares durante a desidratação (ALVIN et al., 2001), ou estratégias que amplificam o sinal molecular da percepção do

estresse hídrico, permitindo à planta antecipar e acelerar os mecanismos de defesa (KASUGA et al., 1999), já são uma realidade. Alcançar esse tipo de resultado e obter plantas cada vez mais eficientes na tolerância aos estresses do ambiente somente está sendo possível devido à compreensão dos mecanismos moleculares de tolerância (NEPOMUCENO et al., 2001).

Também a identificação de novos genes e/ou isoformas de genes conhecidos que desempenhem função importante nas respostas aos estresses abióticos serão potenciais candidatos na aquisição de tolerância à seca e à salinidade para as espécies, mas também para a transferência de característica para outras culturas de interesse agrícola (NEPOMUCENO et al., 2001).

5. Bioinformática: desvendando as ‘ômicas’

De forma simplificada, podemos definir a bioinformática com uma área multidisciplinar da ciência que envolve tecnologia da informação, biologia, matemática, estatística, química, física e outras disciplinas, cuja aplicação principal é nas questões da biologia molecular. Neste sentido, a bioinformática faz uso de programas computacionais para reconhecer, caracterizar e mapear genes, decifrando a estrutura de suas respectivas proteínas (SABBATINI, 1999).

De acordo com Sabbatini (1999), a bioinformática surgiu devido à necessidade de acessar e manusear, de forma eficaz, grande quantidade de dados objetivando a geração de novos conhecimentos. De fato, com o surgimento dos sequenciadores automáticos e após o sequenciamento do genoma humano na década de 90, houve uma explosão na quantidade de sequências gênicas e proteicas, oriundas de diferentes projetos de pesquisa, a ser armazenada e, principalmente, analisada, o que tornava indispensável a utilização de plataformas computacionais eficientes para a interpretação dos resultados obtidos, direcionando os próximos passos e auxiliando as conclusões no final da pesquisa (ROUZÉ et al., 1999; PROSDOCINI et al., 2002).

Assim, com o objetivo, não apenas de armazenar a grande quantidade de dados gerados pelas diferentes ‘ômicas’, mas também de acessar de forma rápida e eficiente esse grande volume de dados, teve início a criação de bancos de dados públicos e privados e de redes de acesso que permitissem a interação entre estes grupos, bem como o acesso e depósito contínuo de dados (PROSDOCINI et al., 2002). Nas últimas décadas vários bancos de dados têm sido construídos visando armazenar e organizar esta grande quantidade de dados, ademais a maioria destas bases de dados disponibilizam ferramentas que permitem a análise das sequências. Embora existam muitos bancos de dados com acesso restrito,

atualmente a tendência é que estes dados sejam disponibilizados e estejam acessíveis a todos os pesquisadores, como é o caso do GenBank (Banco de Genes - <http://www.ncbi.nlm.nih.gov/genbank>), EMBL-Bank (European Molecular Biology Laboratory Databank; Banco de Dado do Laboratório de Biologia Molecular Europeu), DDBJ (DNA DataBank of Japan; Banco de Dados de DNA do Japão), dentre outros. O Genbank faz parte do Centro Nacional para Informação Biotecnológica (NCBI; *National Center for Biotechnology Information* – <http://www.ncbi.nlm.nih.gov>), sendo um banco de dados público integrativo e abrangente, contendo informações/dados de diferentes tipos e origens. Atualmente este banco comporta sequências nucleotídicas de mais 380.000 organismos, obtidas principalmente através da submissão de grupos de pesquisa individuais e também de projetos de sequenciamento em larga escala, que incluem ESTs e WGS (*Whole Genome Shotgun*; Fragmentação do Genoma Inteiro) (BENSON et al., 2011).

No que concerne aos esforços do Brasil no sequenciamento e construção de bancos de dados, alguns projetos de genoma expresso de plantas tiveram grande impacto no mundo científico, com destaque para os projetos: (i) Projeto de Etiquetas de Sequências Expressas de Cana-de-açúcar (*Sugarcane Expressed Sequence Tags*; SUCEST - <http://watson.fapesp.br/sucest.htm>); (ii) Projeto de Etiquetas de Sequências Expressas de Eucalipto (FORESTs - <https://forests.esalq.usp.br>); (iii) Projeto Genoma da Soja (*Soybean Genome Project*; GENOSOJA – <http://www.lge.ibi.unicamp.br/soja>); e (iv) Projeto Transcriptoma do Feijão-Caupi (*Cowpea Transcriptome Project*; NordEST).

De forma geral, a identificação e caracterização dos dados surgidos a partir dos grandes projetos de sequenciamento são feitas através de análises comparativas; a similaridade existente entre genes presentes em diversos organismos é utilizada na anotação de sua função, estrutura, etc. Assim, um gene caracterizado numa levedura pode ajudar na identificação funcional do gene com a mesma função no homem, por exemplo (SANTOS e ORTEGA, 2003).

Com o objetivo de dar sentido à grande quantidade de sequências genômicas e de proteínas disponibilizadas nos diferentes bancos de dados, vários algoritmos foram desenvolvidos para facilitar o processo de anotação, identificação e caracterização em várias etapas; através destes algoritmos, por exemplo, é possível identificar vários tipos de sequências repetitivas (transposons, microssatélites, etc.), sequências estruturais (centrômeros, telômeros, etc.), sequências regulatórias (promotores, *enhancers*, etc.) e regiões transcritas que correspondem aos genes de cada organismo (SANTOS e ORTEGA, 2003).

A ferramenta mais popular de comparação de sequências (alinhamento local) é o BLAST (*Basic Local Alignment Search Tool*; Ferramenta Básica de Busca de Alinhamento Local) (ALTSCHUL et al.,

1990). Através deste algoritmo é possível comparar e alinhar uma sequência de DNA ou proteína com todas as sequências de um banco de dados público ou privado, sendo geralmente utilizado na busca por sequências similares. Esta ferramenta possui várias versões: (I) BLASTp, compara sequências de aminoácidos com o banco de proteínas; (II) BLASTn, compara sequências de nucleotídeos contra o banco de dados de DNA; (III) BLASTx, traduz a sequência de DNA para proteína e a compara com banco de dados proteico; (IV) tBLASTn, compara sequências proteicas com o banco de dados de DNA traduzido; (V) tBLASTx; traduz tanto a sequência nucleotídica quanto o banco de dados de DNA e os compara.

Uma vez selecionadas e/ou obtidas as sequências nucleotídicas de interesse, as mesmas podem ser traduzidas de forma que é possível avaliar suas prováveis proteínas. Várias ferramentas estão disponíveis gratuitamente para esta análise, como o ORF-Finder (*Open Reading Frame Finder*; Identificador de Quadros Abertos de Leitura), que é administrado pelo NCBI e gera uma análise gráfica indicando todos os quadros abertos de leitura de uma sequência de tamanho mínimo, e o *Translate*, disponibilizado através do site do Expasy (<http://web.expasy.org/translate/>), que é um portal de recursos de bioinformática, fornecendo acesso a bases de dados científicos e ferramentas em diferentes áreas, incluindo a proteômica, genômica, transcriptômica, filogenia, biologia de sistemas e genética de populações (SOARES-CAVALCANTI, 2007).

Para a análise comparativa global de sequências nucleotídicas ou proteicas o CLUSTAL (THOMPSON et al., 1997) é um dos softwares mais utilizados para efetuar estes alinhamentos múltiplos; um alinhamento múltiplo é uma hipótese de homologia posicional entre bases nucleotídicas ou aminoácidos de duas ou mais espécies. Além destas comparações, o programa permite a obtenção de árvores filogenéticas utilizando o algoritmo *Neighbor-Joining* (Agrupamento de Vizinhos).

Vários programas para reconstrução filogenética e análise fenética também estão disponíveis *on line* e podem ser obtidos gratuitamente. Dentre eles destaca-se o MEGA (*Molecular Evolutionary Genetics Analysis*) (KUMAR et al., 1993), que pode ser utilizado na análise de sequências, marcadores moleculares, bem como para associá-las à características morfológicas e fisiológicas ou ainda variantes de uma característica. O programa permite a análise da matriz de dados através de métodos utilizados para cálculo de distâncias genéticas, como distância P (NEI, 1991), distância de Jukes-Cantor (JUNKES-CANTOR, 1969), distância de Tajima-Nei (TAJIMA e NEI, 1994), distância de Kimura-2-parâmetros (KIMURA, 1980), distância de Tamura (TAMURA, 1992). Além disso, o programa disponibiliza os algoritmos UPGMA (*Unweighted Pair Group Method with Arithmetic Means*; Método

não Polarizado de Agrupamentos aos Pares com Médias Aritméticas) (SNEATH e SOKAL, 1973), NJ (*Neighbor-Joining*; Agrupamento por Vizinhança) (SAITOU e NEI, 1987) e Máxima Parcimônia (ECK e DAYHOFF, 1966; FITCH, 1971) para construção de dendrogramas. A versão 5.0, a mais recente do programa, além de disponibilizar os algoritmos citados, permite a análise por Máxima Verossimilhança, além de apresentar uma interface mais amigável (TAMURA et al., 2011).

A seleção de genes com dados de expressão variáveis é interessante para a análise de clusterização, a qual pode ser feita com auxílio dos programas integrados CLUSTER (EISEN et al., 1998) e TREEVIEW (PAGE, 1996). Estes programas são capazes de identificar genes com padrão de expressão semelhante para uma determinada situação, apresentando estes resultados na forma de gráficos e dendrogramas. Para cada gene, o perfil de expressão reflete seu comportamento dinâmico sob condições de estresse, estágios de desenvolvimento ou tecidos distintos; genes com perfil de expressão similar são considerados co-expressos.

As ferramentas acima descritas são apenas uma pequena amostra do que está disponibilizado gratuitamente na web. O universo de programas, ferramentas e algoritmos desenvolvidos para executar as mais variadas análises de sequências de DNA, RNA e/ou proteínas tem aumentado sobremaneira nos últimos anos; seja para atender a uma demanda por análises específicas ainda não disponíveis, seja vislumbrando a necessidade de tornar determinadas análises mais refinadas ou específicas. Além disso, esta grande variedade de recursos, muitos podem ser acessados gratuitamente através de suas homepages ou baixados para uso local, enquanto outros são de acesso restrito e podem ser acessados através da compra de licença de uso.

Em conjunto, a grande quantidade de sequências (DNA e proteína) disponíveis nos bancos de dados e a utilização dos diferentes programas e ferramentas de análise viabilizam a construção do painel de genes que participam das respostas aos estresses abióticos. A importância dos dados e informações obtidos através destes programas e ferramentas é percebida no momento do delineamento dos experimentos *in vivo* e *in vitro* visando ao melhoramento genético das plantas. A realização destes experimentos só é possível graças à caracterização dos genes e vias de interesse, assim como do conhecimento das particularidades e atributos intrínsecos do genoma de cada família ou gênero. Como discutido por Santos e Ortega (2003), à medida que o sequenciamento do genoma de espécies adicionais é finalizado, a genômica comparativa assume grande importância, tornando-se essenciais procedimentos computacionais para avaliar a correlação entre organismos em nível molecular. Dentre os procariotos foi demonstrado por genômica comparada que na história evolutiva vários segmentos de DNA foram

trocados entre espécies distintas, num processo de transferência horizontal. Outras aplicações das análises comparativas entre genomas estão emergindo e para cada um desses propósitos novas ferramentas de bioinformática são construídas, sendo muitas delas disponibilizadas via servidores *www* na *internet* (SANTOS e ORTEGA, 2003).

6. Organismos Modelo

A evolução das espécies e a relação entre elas têm sido melhor entendida graças às análises genômica e proteômica comparativas baseadas na crescente quantidade de dados disponíveis de organismos modelo. Estes organismos representam apenas uma pequena fração da biodiversidade existente na terra, no entanto as pesquisas que resultaram de seus estudos constituem a origem do conhecimento biológico. Historicamente, a comunidade científica tem centrado esforços neste organismos modelo para obter um *insight* sobre os princípios gerais que fundamentam as diversas características do organismos, tais como sua fisiologia, metabolismo, crescimento, reprodução, genética, evolução, dentre outros (HEDGES, 2002).

Com a era do sequenciamento em massa, incluindo os genomas completos de muitos organismos modelo, vários campos da ciência foram beneficiados com estes avanços, especialmente a biologia evolutiva; o aumento da quantidade de informações genéticas disponíveis tem sido fundamental para traçar as relações filogenéticas e estimar o tempo de divergência entre táxons (HEDGES, 2002).

No reino vegetal, há mais de 300.000 espécies conhecidas de plantas terrestres, mas modelos genômicos são encontrados em apenas algumas espécies e famílias. Estes incluem o modelo básico de genética e desenvolvimento – *Arabidopsis thaliana* -, de gramíneas - arroz (*Oryza sativa*), milho (*Zea mays*) e trigo (*Triticum aestivum*) -, e outras plantas cultivadas de importância econômica, como algodão (*Gossypium hirsutum*) e soja (*Glycine max*) (HEDGES, 2002).

A. thaliana é uma planta usada como organismo modelo para a compreensão de processos complexos requeridos para o crescimento e desenvolvimento das plantas, sendo o modelo mais bem estudado e elucidado até hoje, seja pelo tamanho de seu genoma (um dos menores no reino vegetal), seja pelo seu curto ciclo de vida. Os esforços de mais 11.000 pesquisadores e 4.000 organizações em todo o mundo foram responsáveis por gerar uma rica e abundante quantidade de informações genéticas desta planta. Tais informações foram compiladas e estão disponíveis no “*The Arabidopsis Information Resource*” (O Recurso de Informação sobre Arabidopsis; TAIR – <http://arabidopsis.org>), que é acessível

através da internet e seus dados podem ser buscados e acessados de inúmeras maneiras (RHEE et al., 2003).

Desde 2001, esforços têm sido concentrados em aumentar o conteúdo e diversidade de dados, anotação funcional dos genes e produtos gênicos, e melhorar as ferramentas de busca, recuperação, análise e visualização de dados. Outras informações incluem polimorfismo de sequências, anotações de ontologia gênica, famílias gênicas, informações de proteínas, vias metabólicas, perfis de expressão, como base em experimentos de microarranjo, dentre outros (RHEE et al., 2003).

Outra planta de grande interesse é *Medicago truncatula*. Com o objetivo de identificar os genes e suas funções por trás dos aspectos-chave da biologia das leguminosas, pesquisadores selecionaram esta planta como sistema modelo. No anos 2000, um conjunto de mais de 17.000 ESTs provenientes de diferentes estágios de desenvolvimento e tecidos foi montando dando origem ao Banco de Dados de *M. truncatula* (*M. truncatula* DataBase – MtDB; <http://www.medicago.org/MtDB>), uma base de dados relacional que integra dados de transcriptoma e permite ao usuário definir a opção de mineração de dados que melhor atenda às suas necessidades (LAMBLIN et al., 2003).

7. Culturas de interesse agrícola no Brasil e seus genomas

No Brasil, por ser a agricultura uma área promissora, maiores investimentos tecnológicos e biotecnológicos visando o aumento da produtividade agrícola são necessários. Assim, as pesquisas que possibilitem compreender funções específicas de genes específicos dos vegetais e seus perfis de expressão em resposta a uma dada condição podem contribuir decisivamente no melhoramento de culturas de interesse (BARBOSA, 2010). Em plantas, embora o número de genomas completos disponíveis ainda seja menor, quando comparado a outros organismos, um crescente número de sequências expressas (como ESTs) têm sido disponibilizadas à comunidade científica, auxiliando no entendimento de processos genéticos em organismos com grandes genomas, como o dos vegetais (BENKO-ISEPPON, 2001).

Neste contexto, os projetos de sequenciamento aumentaram não somente o conhecimento de sequências genômicas para muitos organismos, mas igualmente para sequências de ESTs/cDNA para muitas espécies vegetais, criando novas oportunidades para uso destas informações na compreensão de mecanismos genéticos, de desenvolvimento do controle da planta e suas respostas aos estímulos ambientais (BARBOSA, 2010).

7.1. Feijão caupi e soja: genomas de leguminosas

Segundo estimativas, a produção mundial da cultura de feijão caupi é de aproximadamente 3,7 milhões de toneladas, em uma área cultivada de cerca de 8,7 milhões de hectares. A Nigéria é o maior produtor, com aproximadamente 57% do total da produção mundial, seguida pelo Brasil, que contribui com 17% da produção mundial (PEREIRA et al., 2001).

O feijão caupi (*Vigna unguiculata* (L.) Walp) é a segunda maior cultura agrícola do Nordeste brasileiro, atrás apenas da mandioca (*Manihot esculenta*) em área total cultivada, representando 60% do Nordeste e 32% da área brasileira plantada no Brasil com algum tipo de feijão (Freire-Filho et al., 2008; 2009). O feijão representa aproximadamente 80% da produção total de grãos para alimentação humana, constituindo uma fonte importante de proteínas (23-30%) e carboidratos (56-68%) (HALL et al., 2003).

Com relação ao feijão caupi, seu cultivo é em grande parte praticado por pequenos produtores, desempenhando importante papel econômico-social na região Nordeste, onde constitui o feijão mais consumido (FROTA E PEREIRA, 2000). Segundo Maia et al. (2000), no século XX a cultura do feijão caupi gerou cerca de 2,4 milhões de empregos diretos e abasteceu a mesa de 27,5 milhões de nordestinos.

O sucesso do cultivo desta leguminosa em regiões semi-áridas se deve à capacidade do feijão caupi em crescer sob condições de solo e ambiente estressantes. No Brasil trata-se do único feijão capaz de sobreviver com sucesso na região Norte (alta umidade, muita chuva, solo argiloso) e no Nordeste (seca, solo arenoso, por vezes salino, muito sol), crescendo também em locais menos estressantes e com clima temperado, como no estado de Santa Catarina (FREIRE-FILHO et al., 2008; 2009).

Além disso, devido à sua tolerância a solos com baixa fertilidade, em decorrência da sua capacidade de realizar simbiose efetiva com micorrizas capazes de fixar nitrogênio (MARTINS et al., 1997), bem como habilidade para tolerar solos com grandes variações de pH, o feijão caupi é um dos componentes mais valiosos em sistemas agrícolas, restaurando a fertilidade dos solos para sucessão de outras culturas (TARAWALI et al., 2002; SANGINGA et al., 2003).

Apesar de ser um componente importante nos sistemas de produção no Nordeste brasileiro sua produtividade ainda é relativamente baixa (entre 300 a 400 kg/ha), provavelmente devido aos sistemas de produção usados, onde na maioria não são adotadas práticas adequadas de manejo do solo, de pragas e doenças, além do cultivo em solos com baixa fertilidade (PIO-RIBEIRO, 2005; FREIRE-FILHO et al., 2008; 2009).

No sentido de acessar as informações genéticas da cultura e visando gerar dados genômicos úteis aos programas de melhoramento e de biotecnologia do feijão caupi e de outras espécies de feijão, em 2005 foi estabelecida a rede NordEST, que contou com a participação de 12 laboratórios parceiros de várias regiões do Brasil.

No âmbito do citado projeto, ferramentas de genômica expressa (EST e SAGE - Serial Analysis of Gene Expression; Análise Serial da Expressão Gênica) relacionadas a estresses bióticos (Potyvírus e Mosaico Severo) e abióticos (salinidade) do feijão-caupi também foram utilizadas, incluindo a construção de 10 bibliotecas de EST submetidas a estresse biótico (vírus do mosaico severo) e abiótico (salinidade), nove bibliotecas de SuperSAGE submetidas ao estresse salino e ao vírus do mosaico severo e quatro bibliotecas de LongSAGE submetidas ao estresse por potyvírus (BENKO-ISEPPON et al., 2008).

As principais metas do projeto incluíram: (a) a geração de no mínimo 100 mil transcritos diferencialmente expressos em tecidos e condições importantes para o entendimento de processos de tolerância ou resistência a estresses bióticos e abióticos; (b) o desenvolvimento de um mapa genético de alta resolução; e (c) o desenvolvimento de estratégias eficientes de cultivo, transformação e regeneração *in vitro*, de modo a propiciar a rápida conversão dos dados gerados em benefício da cultura vegetal em questão (BENKO-ISEPPON et al., 2005; AMORIM et al., 2009).

Além da iniciativa que integra a Rede Brasileira de Genômica do Feijão Caupi (Rede NordEST), destaca-se uma iniciativa desenvolvida pelas Universidades de Virgínia e da Califórnia (EUA), com sequências já disponibilizadas em bancos de dados públicos, integrada no projeto HarvEST (<http://harvest.ucr.edu/>), tendo gerado cerca de 180.000 ESTs (CLOSE, 2007; HARVEST, 2011).

Atualmente o projeto envolve atividades referentes à categorização das bibliotecas de ESTs de *V. unguiculata* (sequências do banco NordEST, HarvEST e NCBI) compiladas em um banco local e a análise das bibliotecas controle (injurado e não injurado) de SuperSAGE infectadas ou não pelo mosaico severo, desenvolvidas junto à empresa GenXPro (Frankfurt am Main, Alemanha) no âmbito do projeto. Em setembro de 2009 o referido projeto contava com mais de cinco milhões de transcritos para análise (KIDO et al., 2009), número que atualmente excede 20 milhões de transcritos, incluindo SuperSAGE-tags, Long-SAGE-tags e ESTs, ultrapassando largamente o número de 100 mil transcritos inicialmente planejado (BARBOSA, 2010). Mapas cromossômicos físicos estão sendo desenvolvidos, incluindo as informações de microssatélites, rDNA e BAC, o que permitirá a transferência rápida de dados entre culturas para propósitos de melhoramento (BENKO-ISEPPON et al., 2009).

Outra leguminosa de grande interesse e impacto na agricultura brasileira é a soja. Segundo a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) (2011), a soja cultivada [*Glycine max* (L.) Merr.] é hoje um dos principais produtos do agronegócio brasileiro, correspondendo a 12% do total das exportações deste setor.

Em um contexto global, a soja surge como um dos grãos mais consumidos, sendo fonte de proteínas, minerais, vitaminas e fibras na alimentação humana e animal. Como matéria prima de diversos produtos, esta leguminosa é a base de ingredientes de padaria, massas, produtos de carne, cereais, misturas preparadas, bebidas, alimentação para bebês, alimentos dietéticos e ração animal dentre outros (LIMA, 2010).

Na safra 2009-2010 o Brasil chegou à marca produtiva de 57 milhões de toneladas de soja, ocupando a segunda posição no *ranking* mundial dos maiores produtores do grão, ficando atrás apenas da produção americana, que no mesmo período foi de 76,6 milhões de toneladas. Os rendimentos anuais provenientes da cultura da soja são estimados em torno de R\$ 18,4 bilhões de reais (EMBRAPA, 2011). Segundo a FAO (*Food and Agriculture Organization*; Organização para a Alimentação e a Agricultura) (2011) o Brasil deverá se consolidar como maior produtor e exportador mundial de soja até o ano de 2017, restando 34% de toda a produção, seguido pelos EUA (30%) e Argentina (20%). Com relação ao mercado consumidor, a China é um dos maiores importadores, tendo adquirido 20,7 milhões de toneladas do grão em 2003.

Atualmente, a soja vem crescendo como fonte alternativa de combustível. O biodiesel de soja vem sendo testado por diversas instituições de pesquisa, como a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) em diferentes cidades brasileiras (MISSÃO, 2006; EMBRAPA, 2011).

Em vista de sua importância foi criado o Consórcio Brasileiro do Genoma da Soja (GENOSOJA), estabelecido em 2009 com a finalidade de integrar as diversas instituições que trabalhavam com genômica desta espécie no Brasil, representando uma importante iniciativa para aumentar o conhecimento genético da soja, de forma a subsidiar o aprimoramento do processo de produção deste grão, enfatizando as análises nos fatores limitantes da produção nacional, como a ocorrência de secas, pragas e patógenos (NASCIMENTO, 2009).

Este projeto objetivou investigar vários aspectos da genômica da soja, o que incluiu estudos com genômica funcional e estrutural, transcriptoma e proteoma. A maioria dos experimentos buscou avaliar as sequências extraídas sob estresses bióticos e abióticos, como infecção por ferrugem asiática e nematoide e seca (ABDELNOOR et al., 2009). O GENOSOJA teve como objetivo compreender a

estrutura e função dos genes e proteínas, assim como analisar a expressão destes sob a influência de estresses bióticos e abióticos. Após sua finalização, o projeto pretende depositar os dados gerados em bancos de dados públicos, como o GenBank (ABDELNOOR et al., 2009).

O consórcio GENOSOJA requereu o desenvolvimento de uma plataforma de bioinformática eficiente para integrar os dados públicos mundiais referentes à soja, incluindo ESTs e sequências genômicas, integrando dados disponíveis em bancos públicos com aqueles gerados dentro do projeto GENOSOJA, como tags de SuperSAGE, bibliotecas subtrativas e MicroRNAs. Um conjunto híbrido de 1.276.813 sequências expressas (ESTs) disponíveis no NCBI foi processado, resultando em 30.809 contigs e 29.938 singlets. A aplicação da técnica de SuperSAGE para o estudo de situações fisiológicas contrastantes permitiu a obtenção de 2.334.864 tags que, após análises, resultaram em 54.052 tags únicas, mapeadas no genoma da soja. Além disso, a montagem de 11.783.331 sequências curtas oriundas do sequenciamento de bibliotecas subtrativas de tecidos de folha e raiz submetidos a estresse hídrico resultaram em 5.084 genes (NASCIMENTO et al., 2009).

Com a geração de bibliotecas subtrativas para o estudo de folhas, raízes e outros tecidos submetidos a estresse biótico e abiótico foi possível identificar 5.084 genes diferencialmente regulados sob condições de estresse; tais dados podem ser acessados através de buscas por palavras-chave, comparações estatísticas, visualização de anotação automática e a classificação da ontologia gênica (NASCIMENTO et al., 2009).

Para avaliar esta grande quantidade de dados, tanto em soja quanto em feijão-caupi, ferramentas capazes de gerar alinhamento locais, como BLAST (ALTSCHUL et al., 1990) são essenciais. Também, estes dados podem ser analisados em uma perspectiva de correlação usando perfis de microarranjo *in silico*, que buscam traçar o padrão de expressão de genes ou vias específicas, ou mesmo de todos os transcritos selecionados. Adicionalmente, vários programas e estratégias *in silico* podem ser utilizados no sentido de identificar, caracterizar e comparar o repertório de genes que participam das principais respostas aos estresses abióticos e bióticos em leguminosas. Juntas estas informações devem auxiliar os programas de melhoramento genético da cultura, seja através de métodos convencionais ou da bioengenharia.

De fato, considerando que os métodos de cultivo adotados, na maioria das vezes utilizando pouca tecnologia, reduzem a produtividade e a qualidade de grãos como os de feijão-caupi e soja, o uso de cultivares com características genéticas competitivas e eficientes deverão contribuir para alcançar um padrão de agricultura mais sustentável e com maior produtividade.

7.2. Eucalipto e o projeto FORESTs

Introduzido no Brasil em 1904, o eucalipto apresentou ótima adaptação às condições edafoclimáticas brasileira, de forma que seu cultivo ocupa cerca de 70% da área reflorestada (3,5 milhões de hectares) do Brasil (BACHA e BARROS, 2004). Em função de sua alta diversidade, adaptabilidade, alta produtividade e características físico-mecânicas, as espécies de eucalipto apresentam excelente potencial para a produção de madeira e celulose. Além da produção de papel e celulose, o eucalipto pode ser utilizado na marcenaria (serraria e madeira processada) e na produção de lenha e carvão vegetal (GONZÁLES, 2002)

As principais razões para o uso extensivo de espécies de *Eucalyptus* em reflorestamentos comerciais, em várias partes do mundo, se devem ao seu rápido crescimento (média de 45m³/ha/ano em plantios clonais), quando comparado a outras plantas lenhosas, além de suas excelentes qualidades no que tange à produção de fibras e polpa de madeira de alta qualidade. Tais fatores, somados ao baixo custo e ao curto período de corte (5-7 anos), permitem até três rotações sucessivas, com ciclos de até 21 anos (HO et al., 1998).

A produção de papel e celulose é um componente importante do setor industrial brasileiro, com perspectivas de extensão para os próximos anos. O Brasil apresentou forte crescimento desse setor, tornando-se o principal produtor mundial de celulose a partir do eucalipto como matéria-prima. Segundo estimativas da Bracelpa (2010), em 2009 a produção de celulose no Brasil foi de 13,4 milhões de toneladas, 6,0% maior em relação a 2008. Para o segmento de papéis, a produção em 2009 foi de 9,34 milhões de toneladas, com queda de 0,7% em relação a 2008 (BRACELPA, 2010).

O interesse no melhoramento genético do eucalipto aumentou com a importância das plantações comerciais, o que resultou no desenvolvimento de programas próprios de melhoramento genético por parte das empresas do setor, visando atender às necessidades de cada região. Graças ao intenso trabalho realizado por melhoristas nas últimas décadas, o Brasil possui hoje um dos mais importantes bancos de germoplasma do gênero *Eucalyptus* com alta produtividade e adaptabilidade ao nosso ambiente (ANDRADE, 2001).

Neste contexto foram criados no Brasil dois projetos de sequenciamento do eucalipto (FORESTs e GENOLYPTUS), envolvendo universidades, institutos de pesquisa e empresas do setor. Estes projetos buscaram identificar genes de eucalipto, principalmente aqueles de interesse agrônomo, como resistência às pragas, doenças e estresses abióticos, entre outros, que poderão auxiliar na seleção

assistida de características para o aumento da produtividade junto aos programas de melhoramento genético, bem como na obtenção de plantas transgênicas (ALVES, 2005).

Em novembro de 2001 foi iniciado o projeto ForESTs (Eucalyptus Genome Sequencing Project Consortium), que a partir do sequenciamento de parte do genoma do eucalipto tinha como objetivo o melhoramento da matéria-prima utilizada na produção de papel e celulose (CARRER, 2005). A proposta foi considerada inovadora, por envolver as principais indústrias da área num sistema de consórcio pré-competitivo em parceria com um grupo de indústrias de papel e celulose, as universidades paulistas e a FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo). Adicionalmente, o objetivo final não era apenas identificar genes do eucalipto, mas usar essa informação na análise da expressão gênica, de modo a desenvolver novas tecnologias para a identificação de mecanismos genéticos associados à qualidade da madeira, resistência a doenças e pragas, estresses ambientais e tolerância às deficiências nutricionais (CARRER, 2005).

O Projeto Genoma Eucalipto objetivou identificar 15 mil genes através do sequenciamento de aproximadamente 100 mil ESTs (*Expressed Sequence Tags*, Etiquetas de Sequências Expressas) preparadas a partir de bibliotecas de diferentes tecidos, incluindo plântulas, folhas, raízes, caule e madeira. As sequências de ESTs foram anotadas automaticamente e clusterizadas para a identificação dos genes. Por meio do uso de tecnologias de bioinformática, o projeto realizou a análise dessas sequências, observando a redundância durante o sequenciamento; a comparação das sequências com as depositadas em bancos de dados internacionais e nacionais; e a clusterização das sequências obtidas por programas especiais de análise de dados, como CAP3 e outros. Adicionalmente, pretendeu-se usar as informações obtidas para realizar a análise da expressão gênica usando os microarranjos. A principal intenção do projeto era que essas informações fossem disponibilizadas rapidamente às empresas para modificação das propriedades da madeira no uso industrial ou em outras aplicações (FAPESP, 2011).

7.3. Cana-de-açúcar e o projeto SUCEST

A cana-de-açúcar é uma planta que pertence ao gênero *Saccharum* L. Há pelo menos seis espécies do gênero, sendo que a cana-de-açúcar cultivada compreende um híbrido multiespecífico, que recebe a designação *Saccharum* spp. As espécies de cana-de-açúcar são provenientes do sudeste asiático, tratando-se da principal matéria prima para a fabricação do açúcar e álcool (etanol).

Segundo Vettore e colaboradores (2003), trata-se de uma das mais importantes culturas agrícolas do mundo, sendo cultivada em regiões tropicais e subtropicais em mais de 80 países, com destaque para o Brasil, onde é estimado que o agronegócio sucroalcooleiro seja responsável por 2,4% do PIB nacional e pela geração de 3,6 milhões de empregos diretos e indiretos. Na safra 2009/2010, o Brasil produziu de 603 milhões de toneladas de cana-de-açúcar, o que permitiu a obtenção de 33 milhões de toneladas de açúcar e 2,57 bilhões de litros de álcool (MAPA, 2010). Em 2010 a área plantada com cana-de-açúcar correspondeu a oito milhões de hectares (COHAB, 2011); no entanto, especula-se que este número aumente para 9,1 milhões de hectares até 2013, devido ao crescente interesse internacional na cultura, que possui vantagens sociais, econômicas e ambientais (ALBINO et al., 2006).

Dada sua importância agrícola e econômica, em 1999 foi criado o consórcio SUCEST (*Sugarcane Expressed Sequence Tag Project*; Projeto de ESTs de Cana-de-açúcar) de sequenciamento do genoma funcional da cana-de-açúcar. O projeto sequenciou 237.954 ESTs, derivadas de 13 diferentes tipos de bibliotecas de cDNA geradas a partir de tecidos, estágios de desenvolvimento e condições fisiológicas distintas, as quais foram montadas em 43.141 clusters (VETTORE et al., 2001; 2003). Das sequências montadas, 61,5% (26.525 clusters) apresentam similaridade com sequências existentes em bancos de dados públicos, enquanto 38,5% (16.616 clusters) representaram potenciais novos genes (VETTORE et al., 2001).

REFERÊNCIAS BIBLIOGRÁFICAS

- ABDELNOOR, R. V. et al. GENOSOJA – A Brazilian Soybean Genome Consortium. In: Plant and Animal Genome Conference, XVII, 2009, San Diego - USA. **Anais...**, abstract P034. San Diego, USA; Scherago International. Disponível em: <http://www.intl-pag.org/17/abstracts/P01_PAGXVII_034.html>.
- AGARWAL, P. K.; JHA, B. Transcription factors in plants and ABA dependent and independent abiotic stress signaling. **Biologia Plantarum**, v. 54, n. 2, p. 201-212, 2010.
- ALBINO, J. C.; CRESTE, S.; FIGUEIRA, A. Mapeamento genético da cana-de-açúcar. **Biotecnologia, Ciência & Desenvolvimento**, v. 36, p. 82-91, 2006.
- ALVES, H. S. **Sequenciamento e análise do genoma cloroplastidial de eucalipto (*Eucalyptus grandis*)**. 2005. Dissertação (Mestrado), Escola Superior de Agricultura Luiz de Queiroz, São Paulo, Piracicaba, 2005.
- ALVIN, F. C. et al. Enhanced accumulation of BiP in transgenic plants confers tolerance to water stress. **Plant Physiology**, v. 126, p. 1042-1054, 2001.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, p. 403-410, 1990.
- AMINI, F. et al. Protein pattern changes in tomato under in vitro salt stress. **Russian Journal of Plant Physiology**, v. 54, p. 464-471, 2007.
- AMORIM, L. L. B. et al. Mapa genético de feijão-caupi com marcadores derivados de marcadores gênicos e não codificantes. In: Congresso Brasileiro de Melhoramento de Plantas, 5., 2009, Guarapari, ES. **Anais...** Guarapari, ES: SBMP, 2009, p. 1-4.
- ANDRADE, A. **Avaliação de parâmetros que influenciam a transformação genética do *Eucalyptus grandis* via agrobacterium**. 2001. 82 f. Dissertação (Mestrado)-Escola Superior de Agricultura Luiz de Queiroz, São Paulo, Piracicaba, 2001.
- ASSAD, E.; PINTO, H. S. Aquecimento global e a nova geografia da produção agrícola no Brasil. EMBRAPA, São Paulo, Ago. 2008. Disponível em: <<http://www.embrapa.br/publicacoes/tecnico/aquecimentoglobal.pdf>>. Acesso em: 25 jul. 2011.
- ATHAR, H. R.; ASHRAF, M. 2009. Strategies for crop Improvement against salinity and water stress: an overview. In: ASHRAF, M.; OZTURK, M.; ATHAR, H. R. (Eds.). **Salinity and water stress: improving crop efficiency**. Netherlands: Springer-Verlag, 2009. p. 1-16.
- AYERS, R. S.; WESTCOT, D. W. **A qualidade da água na agricultura**. Tradução de GHEYI, H. R.; DE MEDEIROS J. F.; DAMASCENO F. A. V. Campina Grande: UFPB, 1999. 218 p.
- AZEVEDO-NETO, A. D. et al. Effects of salt stress on plant growth stomatal response and solute accumulation of different maize genotypes. **Brazil Journal of Plant Physiology**, v. 16, p. 31-38, 2004.

BACHA, C. J. C.; BARROS, A. L. M. Reforestation in Brazil: recent evolution and the future. **Scientia Forestalis**, v. 66, p. 191-203, 2004.

BARBOSA, P. K. A. **Interações bióticas e abióticas em feijão-caupi (*Vigna unguiculata*) pela técnica de SAGE (Serial Analysis of Gene Expression)**. 2010. 151 f. Tese (Doutorado em Ciências Biológicas), Universidade Federal de Pernambuco, Pernambuco, Recife, 2010.

BARTELS, D.; SUNKAR, R. Drought and Salt Tolerance in Plants. **Critical Reviews in Plant Sciences**, v. 24, p. 23–58, 2005.

BARTLING, D. et al. A glutathione S-transferase with glutathione-peroxidase activity from *Arabidopsis thaliana*. Molecular cloning and functional characterization. **European Journal of Biochemistry**, v. 216, p. 579-586, 1993.

BECK, E.H. et al. Specific and unspecific responses of plants to cold and drought stress. **Journal of Biosciences**, v. 32, n. 3, p. 501-510, 2007.

BENKO-ISEPPON, A. M. Estudos moleculares e citogenéticos no caupi e em espécies relacionadas: Avanços e perspectivas. **Embrapa Documentos**, v. 56, p. 327-332, 2001.

BENKO-ISEPPON, A. M. et al. Estresses ambientais: danos e benefícios em plantas. In: NOGUEIRA, R. J. M. C. et al. (Eds.). **Genes associados a estresses bióticos e abióticos em feijão-caupi [*Vigna unguiculata* (L.) Walp.] e outras angiospermas**. Recife: MXM Gráfica e Editora. 2005, p. 350-359.

BENKO-ISEPPON, A. M. et al. Genômica funcional, estrutural e comparativa de feijão-caupi (*Vigna unguiculata*). In: Encontro de Avaliação do Programa Rede Nordeste de Biotecnologia – RENORBIO, 1., 2008, Brasília. **Anais...** Brasília: Ministério da Ciência e Tecnologia, 2008, p. 7-8.

BENKO-ISEPPON, A. M. et al. Cowpea (*Vigna unguiculata*) Transcriptome and Differential Expression: The Project and Applications for Breeding Purposes. In: Plant and Animal Genome Conference, XVII, 2009, San Diego - USA. **Anais...**, abstract W304. San Diego, USA; Scherago International. Disponível em: <http://www.intl-pag.org/17/abstracts/W40_PAGXVII_304.html>.

BENSON, D. A. et al. GenBank. **Nucleic Acids research**, v. 39, p. D32-D37, 2011.

BLANCO, F. F. **Tolerância do pepino à salinidade em ambiente protegido e controle da salinização do solo**. 1999. 104 f. Dissertação (Mestrado em Irrigação e Drenagem) - Escola Superior de Agricultura Luis de Queiroz, São Paulo, Piracicaba, 1999.

BLAZQUEZ, M. A. et al. Isolation and molecular characterization of the *Arabidopsis* *TPS1* gene, encoding trehalose-6-phosphate synthase. **Plant Journal**, v. 13, p. 685–689, 1998.

BOHNERT, H. J. et al. A genomic approach towards salt stress tolerance. **Plant Physiology and Biochemistry**, v. 39, p. 295-311, 2001.

BORÉM, A. **Melhoramento de Plantas**. 2. ed. Viçosa: Editora UFV, 1998. 525 p. ISBN 85-7269-028-X.

BORISOVA, T. A. et al. Heat shock increases the tolerance of Plants to UV-B radiation: 1. growth, development, and water supply to tissues. **Russian Journal of Plant Physiology**, v. 48, n. 4, p. 507-513, 2001.

BRACELPA - Associação Brasileira de Celulose e Papel Brasil. **Estatísticas**. São Paulo, SP, 2011. Disponível em: < <http://www.bracelpa.com.br>>. Acesso em: 25 jul. 2011.

BRAY, E. A. Plant responses to water deficit. **Trends in Plant Science**, v. 2, p. 48-54, 1997.

BUCKERIDGE, M. S.; MORTARI, L. C.; MACHADO, M. R. Respostas fisiológicas de plantas às mudanças climáticas: alterações no balanço de carbono nas plantas podem afetar o ecossistema? In: REGO, G. M.; NEGRELLE, R. R. B.; MORELLATO, L. P. C. (Eds.). **Fenologia - ferramenta para a conservação e manejo de recursos vegetais arbóreos**. Colombo, PR: EMBRAPA Florestas, 2007.

CENTURY, K.; REUBER, T. L.; RATCLIFFE, O. J. Regulating the regulators: the future prospects for transcription-factor based agricultural biotechnology products. **Plant Physiology**, v. 147, p. 20-29, 2008.

CIAIS, P. H. et al. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. **Nature**, v. 437, p. 529-533, 2005.

CHEN, C.; PLANT, A. Salt-induced protein synthesis in tomato roots: the role of ABA. **Journal of Experimental Botany**, v. 50, p. 677-687, 1999.

CHEN, L.; BUSH, D. R. LHT1, A Lysine- and Histidine-Specific Amino Acid Transporter in Arabidopsis. **Plant Physiology**, v. 115, p. 1127-1134, 1997.

CHEN, Q.; VIERLING, E. Analysis of conserved domains identifies a unique structural feature of a chloroplast heat shock protein. **Molecular and General Genetics**, v. 226, p. 425-431, 1991.

CHEN, S.; GOLLOP, N.; HEUER, B. Proteomic analysis of salt- stressed tomato (*Solanum lycopersicum*) seedlings: effect of genotype and exogenous application of glycinebetaine. **Journal of Experimental Botany**, v. 60, p. 2005-2019, 2009.

CLOSE, J. et al. HarvEST. **Methods in Molecular Biology**, v. 406, p. 161-177, 2007.

CONAB - Companhia Nacional de Abastecimento. **Levantamento da safra 2010/2011 da cana-de-açúcar**. São Paulo, SP, 2011. Disponível em: <http://www.conab.gov.br/conabweb/download/safra/3cana_09.pdf>. Acesso em: 16 ago. 2011.

DE BRUXELLES, G. L. et al. Absciscic acids induces the alcohol dehydrogenase gene in Arabidopsis. **Plant Physiology**, v. 111, p. 381-391, 1996.

DIETRICH, P.; SANDERS, D.; HEDRICH, R. The role of ion channels in light-dependent stomatal opening. **Journal of Experimental Botany**, v. 52, p. 1959-1967, 2001.

DOOKI, A. D. et al. Proteomic responses of rice young panicles to salinity. **Proteomics**, v. 6, p. 6498-6507, 2006.

ECK, R. V.; DAYHOFF, M. O. **Atlas of protein sequence and structure**. Silver Spring, Maryland: National Biomedical Research Foundation, 1966. p. 161-169.

EISEN, M. B. et al. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, p. 14863–14868, 1998.

EL-SAYED, O. E.; RIZKALLA, A. A.; SABRI, S. R. S. *In vitro* mutagenesis for genetic improvement of salinity tolerance in wheat. **Research Journal of Agriculture and Biological Sciences**, v. 4, n. 5, p. 377-383, 2007.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. **Publicações**. Brasília, DF, 2011. Disponível em: <<http://www.embrapa.br>>. Acesso em 25 jul. 2011.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. **Tecnologias de Produções de Soja - Região Central do Brasil 2008**. Londrina: Embrapa Soja, Embrapa Cerrados, Embrapa Agropecuária Oeste, 2008. 208 p.

FAO – Food and Agriculture Organization. **Publications**. Rome, 2011. Disponível em: <<http://www.fao.org>>. Acesso em: 25 jul. 2011.)

FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo. **Projetos apoiados pela FAPESP**. São Paulo, 2011. Disponível em: <<http://www.fapesp.br>>. Acesso em: 25 jul. 2011.

FITCH, W. M. Toward defining the course of evolution: minimum change for a specific tree topology. **Systematic Zoology**, v. 20, p. 406-416, 1971.

FLOWERS, T. J.; COLMER, T. D. Salinity tolerance in halophytes. **New Phytologist**, v. 179, p. 945–963, 2008.

FREIRE FILHO, F. R. et al. Avanços e perspectivas para a cultura do feijão-caupi. In: ALBUQUERQUE, A. C. S.; SILVA, A. G. da (Eds.). **Agricultura Tropical: quatro décadas de inovações tecnológicas, institucionais e políticas**. Brasília: Embrapa Informação Tecnológica, 2008. v. 1, p. 285-250.

FREIRE FILHO, F. R. et al. Feijão-caupi: melhoramento genético, resultados e perspectivas. In: Simpósio nordestino de genética e melhoramento de plantas, 1., 2009, Fortaleza. **Anais...** Fortaleza: Embrapa Agroindústria Tropical, 2009. p. 25-59.

FROTA, A. B.; PEREIRA, P. R. Caracterização da produção de feijão-caupi na região Meio-Norte do Brasil. In: CARDOSO, M. J. (Org.). **A cultura do feijão-caupi no Meio-Norte do Brasil**. Teresina: Embrapa Meio-Norte, 2000. v. 28, p. 9-25.

GAXIOLA, R. A. et al. The *Arabidopsis thaliana* proton transporters, AtNhx1 and Avp1, can function in cation detoxification in yeast. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 4. p. 1480-1485, 1999.

GENOSOJA – Soybean Genome Project. **[Dados do projeto]**. Campinas, SP, 2011. Disponível em: <<http://bioinfo03.ibi.unicamp.br/soja/>>. Acesso em: 25 jul. 2011.

GHEYI, H. J. Problemas de salinidade na agricultura irrigada. In: OLIVEIRA, T. S. et al. (Eds.). **Agricultura, sustentabilidade e o semi-árido**. Fortaleza: DCS/UFC, 2000. p. 329 -346.

GONG, Q. et al. Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. **Plant Journal**, v. 44, p. 826–839, 2005.

GONG, P. et al. Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. **Journal of Experimental Botany**, v. 61, p. 3563-3575, 2010.

GUO, P. G. et al. Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes in response to drought stress during the reproductive stage. **Journal of Experimental Botany**, v. 60, p. 3531–3544, 2009.

HALL, A. E. et al. Development of cowpea cultivars and germplasm by the Bean/Cowpea CRSP. **Field Crops Research**, v. 82, p. 103-134, 2003.

HARLAN J. R. **Crops & Man**. Madison, Wisconsin: American Society of Agronomy-Crop Science Society, 1992

HARVEST. **HarvEST: Cowpea**. University of California, Riverside, USA, 2011. Disponível em: <<http://harvest.ucr.edu/?wsize=1280x653>>. Acesso em: 20 jul. 2011.

GENOSOJA – Soybean Genome Project. **[Dados do projeto]**. Campinas, SP, 2011. Disponível em: <<http://bioinfo03.ibi.unicamp.br/soja/>>. Acesso em: 25 jul. 2011.

GONZÁLEZ, E. R. **Transformação genética de *Eucalyptus grandis* e do híbrido *E. grandis* x *E. urophylla* via *Agrobacterium***. 2002. 93 f. Tese (Doutorado em Genética e Melhoramento de Plantas). Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, 2002.

HAYANO-KANASHIRO, C. Analysis of Gene Expression and Physiological Responses in Three Mexican Maize Landraces under Drought Stress and Recovery Irrigation. **PLoS ONE**, v. 4, e7531, p. 1-19, 2009.

HIRAYAMA, T.; SHINOZAKI, K. Research on plant abiotic stress responses in the post-genome era: past, present and future. **The Plant Journal**, v. 61, p. 1041–1052, 2010.

HO, C. K. et al. *Agrobacterium tumefaciens* mediated transformation of *Eucalyptus camaldulensis* and production of transgenic plants. **Plant Cell Reports**, v. 17, p. 675-680, 1998.

INAN, G. et al. Salt cress. A halophyte and cryophyte *Arabidopsis* relative model system and its applicability to molecular genetic analyses of growth and development of extremophiles. **Plant Physiology**, v. 135, p. 1718–1737, 2004.

IPCC - Intergovernmental Panel on Climate Change. **IPCC reports. Geneva, 2011**. Disponível em: <<http://www.ipcc.ch/ipccreports/tar/wg1/pdf/TAR-APPENDICES.PDF>>. Acesso em: 25 jul. 2011.

JUNKES, T. H.; CANTOR, C. R. Evolution of protein molecules In: MUNRO, H. N. (Ed.) **Mammalian protein metabolism**. New York: Academic Press, 1969. p. 121-132.

KASUGA, M. et al. Improving plant drought salt and freezing tolerance by gene transfer of a single stress-inducible transcription factor. **Nature Biotechnology**, v. 17, p. 287-291, 1999.

KIDO, E. A. et al. Brazilian cowpea transcriptome project over five million expressed sequence tags to understand salinity and virus resistance. In Congresso Brasileiro de Fisiologia Vegetal, XII, 2009, Fortaleza. **Resumos...** Fortaleza, CE: SPH Comunicação Visual, 2009.

KIM, E. J. et al. *AtKUP1*: An Arabidopsis Gene Encoding High-Affinity Potassium Transport Activity. **Plant Cell**, v. 10, p. 51-62, 1998.

KIMURA, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequence. **Journal of Molecular Evolution**, v. 16, p. 111-120, 1980.

KUBO, A. et al. Primary structure and properties of glutathione reductase from *Arabidopsis thaliana*. **Plant Cell Physiology**, v. 34, p. 1259-1266, 1993.

KUMAR, S.; TAMURA, K.; JAKOBSEN, I.; NEI, M. **MEGA: Molecular Evolutionary Genetics Analysis. Version 2.0**. Pennsylvania: The Pennsylvania State University. University Park, 1993. PA 16802.

LARCHER, W. **Ecofisiologia vegetal**. Tradução de PRADO, C. H. B. A. São Carlos: RiMa Artes e Textos, 2000. 531 p.

LI, W. et al. The vacuolar Na⁺/H⁺ antiporter gene *SsNHX1* from the halophyte *Salsola soda* confers salt tolerance in transgenic alfalfa (*Medicago sativa* L.). **Plant Molecular Biology Reporter**, v. 29, n. 2, p. 278-290, 2010.

LIU, J.; ZHU, J-K. A Calcium Sensor Homolog Required for Plant Salt Tolerance. **Science**, v. 280, p. 1943-1945, 1998.

LIU, J. et al. The *Arabidopsis thaliana* *SOS2* gene encodes a protein kinase that is required for salt tolerance. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 7, p. 3730-3734, 2000.

LIU, Q. et al. Two Transcription Factors, DREB1 and DREB2, with an EREBP/AP2 DNA Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought- and Low-Temperature-Responsive Gene Expression, Respectively, in Arabidopsis. **Plant Cell**, v. 10, p. 1391-1406, 1998.

MAHAJAN, S.; TUTEJA, N. Cold salinity and drought stresses: an overview. **Archives of Biochemistry and Biophysics**, v. 444, p. 139-158, 2005.

MAIA, F. M. M. et al. Proximate composition, amino acid content and haemagglutinating and trypsin-inhibiting activities of some Brazilian *Vigna unguiculata* (L.) Walp cultivars. **Journal of the Science of Food and Agriculture**, v. 80, n. 4, p. 453-458, 2000.

MAPA - Ministério da Agricultura Pecuária e Abastecimento. Disponível em: <<http://www.agricultura.gov.br/>>. Acesso em: 25 jul. 2011.

MARTINS, L. M. V.; NEVES, M. C. P.; RUMJANEK, N. G. Growth characteristics and symbiotic efficiency of rhizobia isolated from cowpea nodules of the north-east region of Brazil. **Soil Biology and Biochemistry**, v. 29, p. 1005-1010, 1997.

MARUR, C. J.; SODEK, L.; MAGALHÃES, A. C. Free aminoacids in leaves of cotton plants under water deficit. **Revista Brasileira de Fisiologia Vegetal**, v. 6, p. 103-108, 1994.

MATSUI, A. et al. Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. **Plant & Cell Physiology**, v. 49, p. 1135-1149, 2008.

MISSÃO, R. M. Soja: origem, classificação, utilização e uma visão abrangente do mercado. **Revista de Ciências Empresariais**, v. 3, p. 7-15, 2006.

MITTLER, R.; BLUMWALD, E. Genetic Engineering for Modern Agriculture: Challenges and Perspectives. **Annual Review of Plant Biology**, v. 61, p. 443-462, 2010.

MOCHIDA, K. et al. *In silico* analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. **DNA Research**, v. 16, p. 353-369, 2009.

MULLINEAUX, P. M. et al. Identification of cDNAs encoding plastid-targeted glutathione peroxidase. **Plant Journal**, v. 13, p. 375-379, 1998.

NASCIMENTO, L. C. Bioinformatics analysis applied to GENOSOJA project. In: International Conference of the Brazilian Association for Bioinformatic and Computational Biology, V, 2009, Rio de Janeiro. **Anais...** Rio de Janeiro: Brazilian Association for Bioinformatics and Computational Biology, 2009.

NEI, M. Relative efficiencies of different tree making methods for molecular data. In: MIYAMOTO, M. M.; CRACRAFT, J. L. (Eds.). **Recent advances in phylogenetic studies of DNA sequences**. New York: Oxford University Press, 1991. p 90-128.

NEPOMUCENO, A. L. et al. Tolerância à seca em plantas: Mecanismos fisiológicos e moleculares. **Biotecnologia, Ciência e Desenvolvimento**, v. 23, p. 12-18, 2001.

PAGE, R. D. M. TREEVIEW: An application to display phylogenetic trees on personal computers. **Computer Applications in the Biosciences**, v. 12, p. 357-358, 1996.

PAPINI-TERZI, F. S. et al. The SUCEST-FUN project: identifying genes that regulate sucrose content in sugarcane plants. **Proceedings - International Society of Sugar Cane Technologists**, v. 26, p. 868-877, 2007.

PARDO, J. M.; SERRANO, R. Structure of a Plasma Membrane H'-ATPase Gene from the Plant *Arabidopsis thaliana*. **Journal of Biological Chemistry**, v. 264, p. 8557-8562, 1989.

PARIDA, A. K.; DAS, A. B. Salt tolerance and salinity effects on plants: a review. **Ecotoxicology and Environmental Safety**, v. 60, p. 324-349, 2005.

PASTORI, G. M.; FOYER, C. H. Common components, networks, and pathways of cross-tolerance to stress. The central role of “redox” and abscisic acid-mediated controls. **Plant Physiology**, v. 129, p. 460-468, 2002.

PELLEGRINI, P. B.; QUIRINO, B. F.; FRANCO, O. L. Plant Cyclotides: an unusual class of defense compounds. **Peptides**, v. 28, p. 1475-1481, 2007.

PEREIRA, P. A. A. et al. Produto feijão: perspectivas de produção, do consumo e do melhoramento genético. In: Reunião nacional de pesquisa de caupi, 2001, Teresina. **Anais...** Teresina, PI: Embrapa Meio Norte, 2001. v. 1, p. 307-311.

PIH, K. T. et al. Characterization of two new channel protein genes in *Arabidopsis*. **Molecules and Cells**, v. 9, n. 1, p. 84-90, 1999.

PIMENTEL C. **A relação da planta com a água**. – Seropédica, RJ: Edur, 2004, 191 p.

PIO-RIBEIRO, G.; ANDRADE, G. P.; ASSIS FILHO, F. M. Doenças do caupi (*Vigna unguiculata*). In: KIMATI, H. et al. (Eds.) **Manual de fitopatologia: doenças das plantas cultivadas**. 4. ed. São Paulo: Agronômica Ceres, 2005. v. 2, p. 215-222.

PROSDOCINI, F. et al. Bioinformática: manual do usuário. **Biotecnologia, Ciência & Desenvolvimento**, v. 29, p. 12-25, 2002.

QIU, Q. S. et al. Regulation of SOS1, a plasma membrane Na⁺/H⁺ exchanger in *Arabidopsis thaliana*, by SOS2 and SOS3. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, p. 8436-8441, 2002.

QUINTERO, F. J.; BLATT, M. R. A new family of K⁺ transporters from *Arabidopsis* that are conserved across phyla. **FEBS Letters**, v. 415, n. 2, p. 206-211, 1997.

QUINTERO, F. J.; BLATT, M. R.; PARDO, J. M. Functional conservation between yeast and plant endosomal Na⁺/H⁺ antiporters. **FEBS Letters**, v. 471, n. 2-3, p. 224-228, 2000.

RAMPINO, P. et al. Drought stress response in wheat: physiological and molecular analysis of resistant and sensitive genotypes. **Plant, Cell and Environment**, v. 29, p. 2143-2152, 2006.

REN, Z. H. et al. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. **Nature Genetics**, v. 37, p. 1141–1146, 2005.

RENTSCH, D. et al. Salt Stress-Induced Proline Transporters and Salt Stress-Repressed Broad Specificity Amino Acid Permeases Identified by Suppression of a Yeast Amino Acid Permease-Targeting Mutant. **Plant Cell**, v. 8, p. 1437-1446, 1996.

RHOADES, J. D.; KANDIAH, A.; MASHALI, A. M. **Uso de águas salinas para produção agrícola**. Tradução de GHEYI, H. R.; SOUSA, J. R. de.; QUEIROZ, J. E. Campina Grande: UFPB, 2000.

- RIZHSKY, L.; LIANG, H.; MITTLER, R. The combined effect of drought stress and heat shock on gene expression in tobacco. **Plant Physiology**, v. 130, p. 1143-1151, 2002.
- ROUZÉ, P.; PAVY, N.; ROMBAUTS, S. Genome annotation: witch tools do we have for it? **Current Opinion in Plant Biology**, v. 2, p. 90-95, 1999.
- RUIZ-LOZANO, J. M. Arbuscular mycorrhizal symbiosis and alleviation of osmotic stress: new perspectives for molecular studies. **Mycorrhiza**, v. 13, p. 309-317, 2003.
- SABBATINI, R. M. E. A bioinformática chegou. **Revista Informática Médica**, v. 2, p. 4-6, 1999.
- 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.
- SALINET, L. H. **Avaliação fisiológica e agrônômica de soja geneticamente modificada para maior tolerância à seca**. 2009. 75 f. Dissertação (Mestrado)-Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, São Paulo, Piracicaba.
- SANGINGA, N. et al. Sustainable resource management coupled to resilient germplasm to provide new intensive cereal-grain-legume livestock systems in the dry savanna. **Agriculture, Ecosystems & Environment**, v. 100, p. 305-314, 2003.
- SANTOS, F. R.; ORTEGA, J. M. Bioinformática. In: BORÉM, A.; DEL GIÚDICE, M.; SEDYIAMA, T. (Org.). **Melhoramento Genômico**. 1. ed. Viçosa, MG: UFV, 2003, v. 1, p. 165-185.
- SEKI, M. et al. Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. **Plant Journal**, v. 31, p. 279-292, 2002.
- SENTENAC, H. et al. Cloning and expression in yeast of a plant potassium ion transport system. **Science**, v. 256, p. 663-665, 1992.
- SHAGAN, T.; BAR-ZVI, D. Nucleotide Sequence of an *Arabidopsis thaliana* Turgor-Responsive cDNA Clone Encoding TMP-A, a Transmembrane Protein Containing the Major Intrinsic Protein Motif. **Plant Physiology**, v. 101, p. 1397-1398, 1993.
- SHI, H. et al. The *Arabidopsis thaliana* salt tolerance gene *SOS1* encodes a putative Na⁺/H⁺ antiporter. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 12, p. 6896-6901, 2000.
- SINGH, S. K. et al. Cowpea (*Vigna unguiculata* [L.] Walp.) genotypes response to multiple abiotic stresses. **Journal of Photochemistry and Photobiology Biology**, v. 100, p. 135-146, 2010.
- SNEATH, P. H. A.; SOKAL, R. R. **Numerical taxonomy**. San Francisco: Freeman, 1973.
- SOARES-CAVALCANTI, N. M. **Mineração da via SOS (Salt Overly-Sensitive) no genoma expresso da cana-de-açúcar**. 2004. 83 f. Monografia (Graduação bacharelado em Ciências Biológicas)-Universidade Federal de Pernambuco, Pernambuco, Recife.

- SOARES-CAVALCANTI, N. M. **Estudo in silico de genes que codificam fatores de transcrição responsivos à seca, salinidade e congelamento nos genomas do eucalipto, cana e arroz.** 2007. 218 f. Dissertação (Mestrado em Genética)-Universidade Federal de Pernambuco, Pernambuco, Recife.
- SOBHANIAN, H. et al. Proteome analysis of soybean leaves, hypocotyls and roots under salt stress. **Proteome Science**, v. 8, p. 19, 2010.
- SONG, S. Q.; LEI, Y. B.; TIAN, X. R. Proline metabolism and cross-tolerance to salinity and heat stress in germinating wheat seeds. **Russian Journal of Plant Physiology**, v. 52, n.6, p. 897-904, 2005.
- TAIZ, L.; ZEIGER, E. **Plant Physiology**. Massachusetts: Publishers Sunderland, 2006. 705 p.
- TAJI, T. et al. Comparative genomics in salt tolerance between Arabidopsis and Arabidopsis-related halophyte salt cress using Arabidopsis microarray. **Plant Physiology**, v. 135, p. 1697–1709, 2004.
- TAJIMA, F.; NEI, M. Estimation of evolutionary distance nucleotide sequences. **Molecular Biology and Evolution**, v. 1, p. 269-285, 1994.
- TAMURA, K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+ C-content bases. **Molecular Biology and Evolution**, v. 9, p. 678-687, 1992.
- TAMURA, K. et al. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Molecular Biology and Evolution**, 2011. Disponível em: < <http://www.megasoftware.net/>>. Acesso em 01 ago. 2011.
- TARAWALI, S. A., SINGH, B. B.; GUPTA, S. C. Cowpea as a key factor for a new approach to integrated crop/livestock systems research in the dry savannas of West Africa. In: FATAKUN, C. A. et al. (Eds.). **Challenges and Opportunities for Enhancing Sustainable Cowpea Production**. Ibadan, Nigeria: International Institute of Tropical Agriculture, 2002. p. 233-251.
- THOMPSON, J. D. et al. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Research**, v. 25, p. 4876-4882, 1997.
- TRAN, L.-S. P. et al. Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought- responsive cis-element in the early responsive to dehydration stress 1 promoter. **The Plant Cell**, v. 16, p. 2481-2498, 2004.
- UOZUMI, N. et al. The *Arabidopsis* *HKT1* Gene Homolog Mediates Inward Na⁺ Currents in *Xenopus laevis* Oocytes and Na⁺ Uptake in *Saccharomyces cerevisiae*. **Plant Physiology**, v. 122, p. 1249-1260, 2000.
- VARSHNEY, R. K. et al. A comprehensive resource of drought- and salinity-responsive ESTs for gene discovery and marker development in chickpea (*Cicer arietinum* L.). **BMC Genomics**, v. 10, p. 523-540, 2009.
- VERSLUES, P. E. et al. Methods and concepts in quantifying resistance to drought salt and freezing abiotic stresses that affect plant water status. **Plant Journal**, v. 45, p. 523-539, 2006.

- VETTORE, A. L. et al. The libraries that made SUCEST. **Genetics and Molecular Biology**, v. 24, n. ¼, p. 1-7, 2001.
- VETTORE, A. L. et al. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. **Genome Research**, v. 13, p. 2725-2735, 2003.
- VOGEL, G. et al. Trehalose-6-phosphate phosphatases from *Arabidopsis thaliana*: identification by functional complementation of the yeast tps2 mutant. **Plant Journal**, v. 13, p. 673-683, 1998.
- WAHID, A.; GHAZANFAR, A. Possible involvement of some secondary metabolites in salt tolerance of sugarcane. **Journal of Plant Physiology**, v. 163, p. 723-730, 2006.
- WANG, W.; VINOCUR, B.; ALTMAN, A. Plant responses to drought salinity and extreme temperatures: towards genetic engineering for stress tolerance. **Planta** v. 218, p. 1-14, 2003.
- WU, C. H. et al. Characterization of an FSP70 cognate gene family in arabidopsis. **Plant Physiology**, v. 88, p. 731-740, 1988.
- YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stress. **Annual Review of Plant Biology**, v. 57, p. 781-803, 2006.
- YANG, Q. et al. Overexpression of SOS (Salt Overly Sensitive) genes increases salt tolerance in transgenic Arabidopsis. **Molecular Plant**, v. 2, n. 1, p. 22-31, 2009.
- YI, Z.; HONG-BO, S. The responding relationship between plants and environment is the essential principle for agricultural sustainable development on the globe. **Comptes Rendus Biologies**, v. 331, p. 321-328, 2008.
- YOSHIBA, Y. et al. Correlation between the induction of a gene for Δ^1 -pyrroline-5-carboxylate synthetase and the accumulation of proline in *Arabidopsis thaliana* under osmotic stress. **Plant Journal**, v. 7, n. 5, p. 751-760, 1995.
- ZENTGRAF, U.; ZINKERNAGEL, I. A gene encoding a catalase (accession no. X94447) isoform from *Arabidopsis thaliana*. **Plant Physiology**, v. 110, p. 1047-1047, 1996.
- ZHU, J. K.; LIU, J.; XIONG, L. Genetic analysis of salt tolerance in arabidopsis. Evidence for a critical role of potassium nutrition. **The Plant Cell**, v. 10, p. 1181-1191, 1998.

CAPÍTULO I

Artigo Publicado no Periódico *Lecture Notes in Bioinformatics*

Artigo Apresentado no 5th International Meeting, CIBB 2008

Comparative *In Silico* Evaluation of MYB Transcription Factors in Eucalyptus, Sugarcane and Rice Transcriptomes

Nina M. Soares-Cavalcanti, Ana C. Wanderley-Nogueira, Luis C. Belarmino, Petra dos Santos Barros, and Ana M. Benko-Iseppon

Universidade Federal de Pernambuco, Center for Biological Sciences (CCB), Genetics Department, Laboratory for Plant Genetics and Biotechnology, Rua Prof. Moraes Rêgo s/no. Recife, PE, Brazil
Tel.: (81) 2126.8520; Fax: (81) 2126.8569
ana.benko.iseppon@pesquisador.cnpq.br

Abstract. Transcription factors are essential in plant growth and development regulation and during response activation to environmental stimuli. This study examined 237,954 sugarcane and 110,000 eucalyptus ESTs in search of genes related to MYB factors active during abiotic stresses in arabidopsis and rice. Searches revealed relative abundance and diversity of MYB candidate sequences in both transcriptomes. In general, 71 and 42 MYB sequences have been selected in eucalyptus and sugarcane, respectively. The generated phenogram, using selected sequences and sequences from other organisms, revealed a sub-grouping according to domain-repeats number; confirming the ancestral origin of *MYB* genes. Expression pattern analysis revealed higher number of reads in wood, flower and bark libraries in eucalyptus, and in tissues infected by pathogens in sugarcane. Splicing pattern comparative analyses of arabidopsis and rice ORFs with eucalyptus and sugarcane ORFs revealed that neither of them presented similarities regarding all exons and intron junctions.

Keywords: Bioinformatics, Salinity Stress, Drought Stress, Saccharum, Crop Evolution.

1 Introduction

Abiotic stresses, like drought and salinity, generally impose limitations to plant growth and crop productivity, also limiting farming activities in many regions throughout the world. Salinity limits vegetative and reproductive growth of plants by inducing severe physiological dysfunctions and causing widespread direct and indirect harmful effects, even at low salt concentrations. Once water is fundamental to almost all aspects of plant physiology, its reduction imposes strong and recurring selective pressure to plants [1].

Drought and salinity stresses are often interconnected, and cause cellular damage and secondary stresses, such as osmotic and oxidative stresses. Plants must

be able to “sense” the environmental cues before being able to respond appropriately to the abiotic stress. Due to the complex nature of stress, multiple sensors, rather than a single sensor, are more likely to be responsible for the stress perception. After initial stress recognition, a signal transduction cascade is induced. Secondary messengers activate stress-responsive genes, like transcription factors (TF), generating the initial stress response [2]. Many TF families existed before the divergence of plants, animals and fungi; however, the size of TF families differs considerably among eukaryote kingdoms. For example, the MADS-box family has two members in *Drosophila*, two in *C. elegans* and four in *S. cerevisiae*. However, it has more than 100 members in *Arabidopsis*, suggesting substantial amplification of this gene in the land plant lineage [3]. A typical plant TF contains a DNA-binding region, an oligomerization site, a transcription-regulation domain, and a nuclear localization signal [4].

The MYB superfamily is the largest TF gene family in plants, also present in all eukaryotes analyzed. The encoded proteins are crucial to control proliferation and differentiation in a number of cell types, and share the conserved MYB DNA-binding domain [5]. This domain generally comprises up to three imperfect repeats, each forming a helix-turn-helix (HTH) structure of about 35 amino acids, with three regularly spaced tryptophan residues, that form a tryptophan cluster in the three dimensional HTH structure [5]. Regarding MYB repeats, the proteins can be classified in three subfamilies R1/2, R2R3 and R1R2R3 [5], while in plants *MYB* genes bearing two repetitions are the most common [3]. The development of tolerant cultivars to abiotic stresses, as well as the development of technologies to over express TFs in transgenic plants may help in the tolerance and surveillance during drought and salinity stress, collaborating to an increase in the world crop production.

Little information is available about *MYB* diversity and abundance in plants with large genomes, as sugarcane or in woody angiosperms, as eucalyptus. The present work aimed to perform a data mining-based identification of *MYB* TF family in FOREST and SUCEST databases, as compared with data available for rice and arabidopsis sequences, comparing them with sequences deposited in public databases and literature data.

2 Methods

Complete full length cDNA sequences of *MYB* transcription factor genes from *A. thaliana* were obtained in fasta format from NCBI, based on Seki *et al.* (2002) [6] results. Each sequence was compared against the SUCEST, FORESTs and TGI (The Gene Index Project) databases using tBLASTn tool. Only sequences with e-value e^{-5} or less were considered. DNA sequences of eucalyptus, sugarcane and rice were translated using ORFfinder program (www.ncbi.nlm.nih.gov/gorf), and screened for similarity to entries in GenBank at NCBI (National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) using mainly BLASTx and BLASTp algorithms [7]. Moreover, the Blink option was used to obtain similar sequences, containing the searched domains, in other organisms. RPS_BLAST' option was used to infer CD presence and integrity.

A multiple alignment was generated using the CLUSTALx program, aiming to evaluate the conservation among MYB sequences. For this alignment only sequences with complete domains were selected. The MEGA software (V. 4.0) allowed the dendrogram construction using Neighbor Joining method [8], gamma distance and Bootstrap test of inferred phylogeny (1,000 replications). The visualization of the trees was made using TreeView [9]. To generate an overall picture of the expression patterns of MYB TFs in sugarcane and eucalyptus, annotated reads were associated to respective library and used for a hierarchical clustering approach with normalized data using CLUSTER program [10]. Dendrograms including both axes (using the weighted pair-group for each gene class and library) were generated. On graphics (Figure 3), light gray means no expression and dark gray/black all degrees of expression.

The analyses of protein migratory behavior, using only the best matches of each seed sequence, were generated with the JvirGel program [11], using analytical modes for serial calculation of MW (molecular weight), pI (Isoelectric Point), pH-dependent charge curves and hydrophobicity profiles, a virtual 2D gel as JavaTM applet.

Aiming to infer about the putative mRNA splicing sites among both dicots (arabidopsis and eucalyptus) and monocots (rice and sugarcane), eucalyptus and sugarcane best hits were translated, having their splicing sites compared with ORFs from arabidopsis and rice, respectively.

3 Results

3.1 Eucalyptus, Sugarcane and Rice Orthologs

MYB candidates were identified in the studied databases, including 89 clusters in FOREST, 81 in SUCEST and 261 in TGI. Only for Ara_MYB_02 seed sequence no match could be uncovered in these databases (Table 1).

Considering the identified CDs in the three mentioned data banks, 149, 13 and 29 sequences presented the SANT and MYB domains combined as follows: 2xSANT, 2xMYB and 1xSANT + 1xMYB (where “2x” means “twice” and “1x” means “once”). Other 78 and 23 sequences presented also SANT and MYB complete CDs, nine with 3xSANT, two and 18 with SANT and MYB incomplete domains among other combinations including incomplete domains and 99 with significant homology but no CD. Moreover, one rice cluster presented two copies of SANT domain and one copy of the maize P_C domain (Table 1). In addition, the results showed that 35.2%, 63.4% and 1.4% of the eucalyptus clusters presented one, two and three MYB repetitions, respectively. In sugarcane and rice clusters with one MYB repetition represented by 35.7% and 32.4%, two MYB repetitions by 59.5% and 64.4% and three repetitions by 4.8% and 3.2%. In a general view 86, 79 and 259 clusters were obtained at FOREST, SUCEST and TGI projects, which participated to salinity+drought category (16, 19 and 59, respectively) and to drought category (70, 60 and 200, respectively). As a whole, results for *MYB* genes showed that the percentage of clusters in each stress category was similar considering all organisms. MYB analysis showed that most

Table 1. Data mining resources and results using arabidopsis seed sequences, matches of eucalyptus, sugarcane and rice with respective domains

<i>Arabidopsis thaliana</i>		Matches of Eucalyptus, Sugarcane and Rice										Domains Found					
Stress Type	Seed Sequence	Higher Match	e-value	Score	Fr.	Other up to E ⁴⁰	Size nt	ORF	Dom 1	Start	End	Dom 2	Start	End	Dom 3	Start	End
Salinity and Drought	AY050961 Ara_MYB_01 type MYB-bind 2684 nt	Euc_MYB_01	4e ⁻³³	135	1	16	1713	409ic	MYB+bind	24	68	-	-	-	-	-	-
		Euc_MYB_02	1e ⁻²⁷	117	3	16	789	195ic	SANT	137	180	-	-	-	-	-	-
		Euc_MYB_03	5e ⁻⁰⁸	52	1	16	1221	263ic	SANT	11	54	SANT	122	167	-	-	-
		Sug_MYB_01	3e ⁻³¹	129	1	19	3291	748cp	SANT	53	96	-	-	-	-	-	-
		Sug_MYB_02	1e ⁻²¹	97	3	19	670	191ic	MYB+bind	96	140	-	-	-	-	-	-
		Sug_MYB_03	2e ⁻⁰⁷	50	-1	19	1483	300cp	SANT	19	66	SANT	127	172	-	-	-
		Ric_MYB_01	2e ⁻²⁹	330	2	59	1007	246ic	MYB+bind	44	88	-	-	-	-	-	-
		Ric_MYB_02	1.1e ⁻²⁸	330	3	59	2805	719cp	SANT	26	69	-	-	-	-	-	-
		Ric_MYB_03	3.8e ⁻⁰⁷	126	1	59	934	182cp	SANT	26	75	SANT	114	159	-	-	-
		no hits	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Salinity, Drought & Cold	AV821864 Ara_MYB_02 no CD, 604 nt	no hits	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Only Drought	AF386932 Ara_MYB_03 type SANT, 1148 nt	Euc_MYB_04	8e ⁻⁷⁸	280	2	70	706	210ic	MYB+bind	14	61	MYB+bind	67	112	-	-	-
		Euc_MYB_05	7e ⁻⁶¹	230	1	70	995	219cp	SANT	16	61	SANT	69	112	-	-	-
		Euc_MYB_06	1e ⁻⁴⁵	177	3	70	796	121cp	SANT	17	61	MYB+bind	67	112	-	-	-
		Euc_MYB_07	3e ⁻²⁴	115	2	70	796	216ic	SANT	67	110	SANT	118	162	SANT	170	213
		Euc_MYB_08	7e ⁻⁴⁸	187	1	70	779	216ic	SANT	24	67	-	-	-	-	-	-
		Euc_MYB_09	1e ⁻²⁵	146	2	70	867	105cp	MYB	14	62	-	-	-	-	-	-
		Sug_MYB_04	2e ⁻⁶¹	232	2	60	687	226ic	SANT	38	83	SANT	91	134	-	-	-
		Sug_MYB_05	3e ⁻⁶¹	231	2	60	613	117ic	SANT	22	66	MYB+bind	72	115	-	-	-
		Sug_MYB_06	1e ⁻²⁵	146	2	60	479	138ic	MYB+bind	24	71	MYB+bind	77	120	-	-	-
		Sug_MYB_07	3e ⁻²⁹	125	1	60	600	109cp	SANT	16	61	-	-	-	-	-	-
		Sug_MYB_08	3e ⁻²²	102	-2	60	1560	274cp	SANT	52	96	SANT	104	148	SANT	156	196
		Sug_MYB_09	4e ⁻²¹	99	1	60	449	145ic	MYB+bind	82	129	-	-	-	-	-	-
		Ric_MYB_04	1.6e ⁻⁶⁰	679	2	200	1331	368cp	SANT	16	61	SANT	69	112	-	-	-
		Ric_MYB_05	2.2e ⁻⁴⁹	518	3	200	1105	255cp	MYB	14	61	SANT	71	112	-	-	-
		Ric_MYB_06	2.4e ⁻⁴²	452	1	200	996	331cp	SANT	16	61	SANT	71	112	P_C	113	331
		Ric_MYB_07	2.8e ⁻⁴¹	352	3	200	536	106ic	SANT	49	92	-	-	-	-	-	-
		Ric_MYB_08	3e ⁻³⁰	422	1	200	1198	102cp	MYB	14	61	MYB	67	112	-	-	-
		Ric_MYB_09	3.5e ⁻²⁶	300	1	200	873	290cp	MYB	94	137	-	-	-	-	-	-
		Ric_MYB_10	6.6e ⁻²⁴	285	1	200	2572	587cp	SANT	70	113	SANT	121	165	SANT	173	216

Abbreviations: nt=nucleotides; fr.=frame; ORF=Open Reading Frame; Dom=Domain; Ara.=Arabidopsis seed sequence; Euc.=Eucalyptus cluster; Sug.=Sugarcane cluster; Ric.=Rice cluster; cp=complete; ic=incomplete; bind=DNA binding element; - =not found.

clusters (ca. 78%) belonged to drought category, against ca. 22% from salinity+drought category.

3.2 Isoelectric Point Analysis

The virtual electrophoresis evaluation of MYB proteins from eucalyptus, sugarcane and rice best matches revealed that all sequences presented isoelectric point from 4.17 to 10.69. Considering the molecular mass, values varied of 5.6 MW to 81.8 MW, with sequences presenting two groupings with few exceptions (Data not shown).

3.3 Phenetic Analysis

Multiple alignment of MYB sequences revealed a considerable degree of conservation among MYB family sequences considering fungi, algae, protozoan, plants and animals. The generated dendrogram using MYB proteins grouped sequences in 24 different branches, with a clear segregation of the organisms in subclades according to their higher taxonomic grouping (Figure 1).

However, in branches XIX and XX, the grouping of distant related species as micro algae (*Ostreococcus tauri*), a gymnosperm (*Sequoia sempervirens*) and a protozoan (*Dictyostelium discoideum*) together with yeast (*Saccharomyces cerevisiae*) was remarkable. Also interesting was the grouping of members exclusively of the animal kingdom into the XVI branch, including insects, amphibian, fishes, birds and mammals (Figure 1). Angiosperms also grouped together, but analyzing internal grouping no distinction was revealed among monocots, dicots and

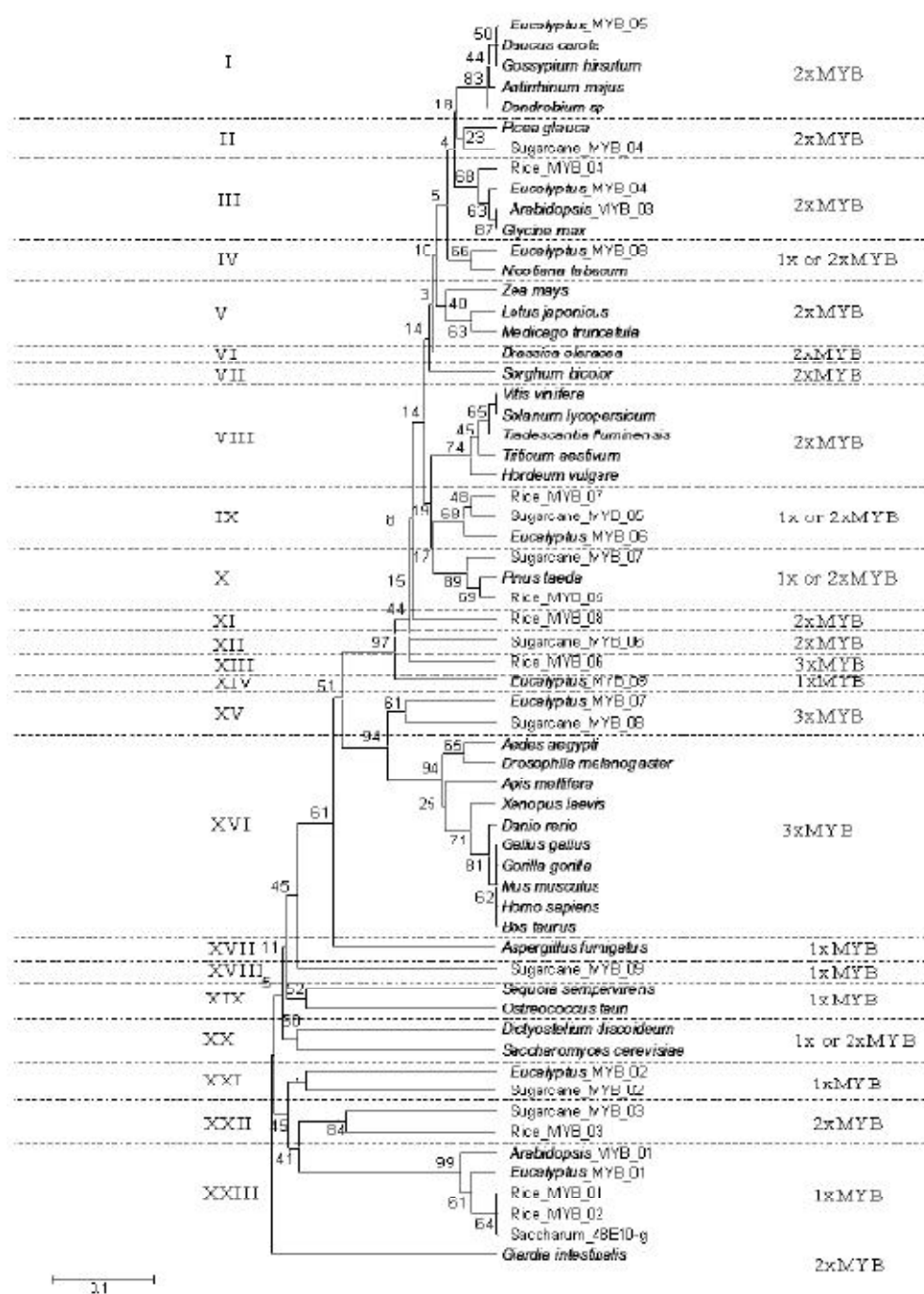


Fig. 1. Dendrogram generated after Neighbor Joining analysis, showing relationships among the MYB seed sequences of Arabidopsis and orthologs of Eucalyptus, Sugarcane, Rice and other organisms. Dotted lines delimit the main taxonomic units and roman numbers on the left of the dendrogram and between the dotted lines refer to the groupings. At the right of the groupings the number of MYB repetitions per group is given. The numbers on the left of the clusters are P values. Bar represents similarity coefficient.

also gymnosperms. Moreover, we could note that inferior organisms like the fungus *Aspergillus fumigatus* and the protozoan *Giardia intestinalis* did not grouped to any other organism, pertaining to individual branches. In most cases, grouping reflected the number of MYB CD repetitions and therefore reflected duplication and divergent evolution events.

3.4 Expression in Different Tissues

It was observed that most tissues presented expression of at least one MYB TF considering FOREST and SUCEST libraries (Figure 2A and B); with exception of a complete absence in the sugarcane NR library (that included all tissues) (Figure 2B). In a general view, MYB family was best represented in eucalyptus by reads from bark/wood tissues from 7-8 year-old trees of *E. grandis* followed by bud, flower and fruit tissues (Figure 2A, bottom). In sugarcane two tissues infected by bacteria presented higher expression (Figure 2B, bottom). Diverse transcripts obtained both in FOREST and SUCEST data bases were used to perform a hierarchical clustering analysis from ESTs permitting an evaluation of expression intensity considering co-expression in different libraries (black upper dendrogram) (Figure 2A and B).

Co-expression was observed regarding the two *MYB* genes in both eucalyptus and sugarcane, with higher expressed WD and AD libraries, respectively (for

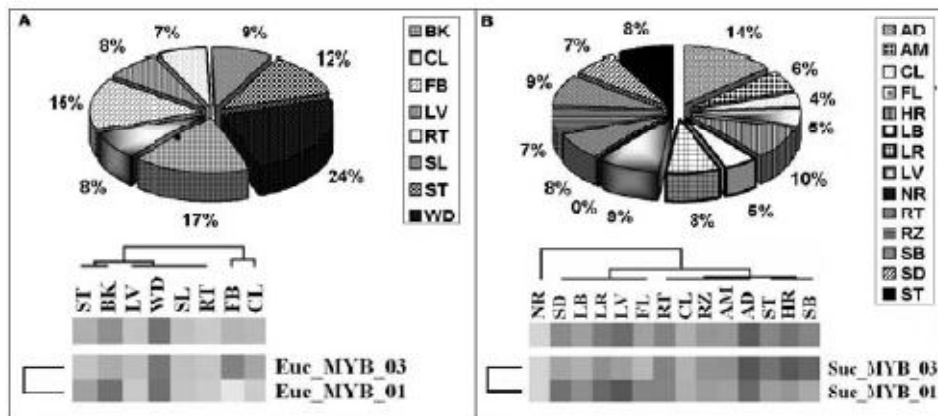


Fig. 2. Prevalence of MYB factors (upper) and *in silico* expression pattern (bottom) in FOREST (A) and SUCEST (B) libraries. Black indicates higher expression, gray lower expression, and light gray absence of expression in the corresponding tissue and cluster. Abbreviations: AD—tissues infected by *Cluconacetobacter diosotrophicans*; AM—Apical meristem; CL=Callus; FL=Flower; HR=tissues infected with *Herbaspirillum rubrisubalbicans*; LB=Lateral bud; LR=Leaf roll; LV=Leaves; RT=Root; RZ=Stem-root transition; SB=Stalk bark; SD=Seeds; ST=Stem; NR=all libraries; BK=bark/duramen/pith and alburnum of *E. grandis*; LV=leaves; RT=roots; FB=flowers; CL=calli; ST=stem; SL=seedlings grown in the dark; WD=wood; Euc_MYB=Eucalyptus MYB candidates; Sug_MYB=sugarcane MYB candidates.

libraries abbreviations see legend of Figure 2). In *Eucalyptus* ST and BK, LV, WD, SL and RT and FB and CL libraries were co-expressed. Regarding sugarcane, co-expression was observed in SD, LB, LR, LV and FL, RT, CL, RZ, AM and AD and ST, HR AND SB libraries.

3.5 Splicing Analysis

Evaluation of *MYB* genes revealed the presence of three exons interrupted by two introns in MYB_03 (Figure 3A). ORF comparison revealed that neither eucalyptus nor sugarcane presented similarities regarding all exons. Moreover, in arabidopsis MYB_01 an alternative splicing was observed, being this sequence MYB_01.1 formed by four introns and four exons and MYB_01.2 formed by seven of each one (Figure 3B). Considering eucalyptus sequences, high similarities were obtained with the three first exons in the first case, and six exons in the second (Figure 3B). This same sequence analysis in rice revealed five exons and five introns with high similarities to sugarcane (Figure 3A and B).

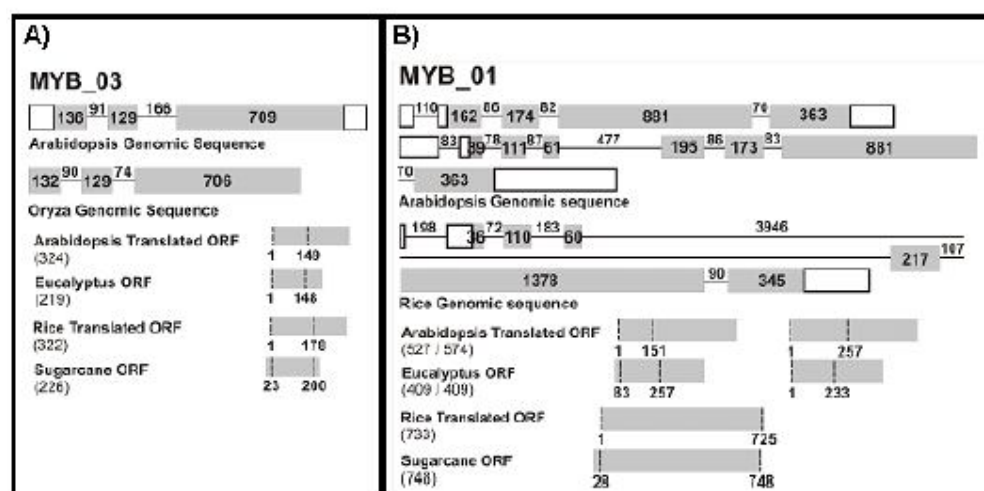


Fig. 3. Comparative analysis of MYB ORFs from arabidopsis and rice genomic sequences with eucalyptus and sugarcane clusters. White and gray boxes indicate UTRs and exons, respectively. Black line indicate introns. Black numbers inside of boxes correspond to the segment size. Black numbers above black line correspond to the segment size. Numbers below the organisms name correspond to ORFs size. Black numbers below the gray boxes (ORFs) correspond to alignment start and end positions. Black bars inside of these boxes delimit the alignment region. Abbreviations: ORF, Open Reading Frame.

4 Discussion

Modulation of transcriptional activity is fundamental for gene expression regulation associated with most biological phenomena and is largely mediated by

proteins that interact directly or indirectly with specific DNA sequences (*cis* elements) in the promoter region of genes. In spite of their importance, such factors present often discrete expression levels enough to activate given genes, and their expression level may not be enough to be measured in microarray experiments. Despite of that, they have been identified in large databases as for *Eucalyptus* (with no previous analysis regarding such factors) and for sugarcane with a single evaluation of TF transcript abundance [12].

MYB family is represented by at least 71, 42 and 189 orthologous sequences in eucalyptus, sugarcane and rice genomes, respectively. Analysis of these sequences showed the presence of two categories of domains: SANT and MYB while comparisons among them showed that they were homologous domains. Thus, *Eucalyptus* and *S. officinarum* transcriptomes presented, at least, 25 and 15 MYBs type R1/R2 and one and two MYBs R1R2R3, respectively, whereas *O. sativa* genome contain 61 members with one repetition, 121 with two repetitions and seven with three repetitions. Of the seven rice candidates to MYB R1R2R3, one presented two SANT domains and one P₁C domain, a similar feature found in maize sequences, a domain combination apparently absent in Magnoliophyta division that may have appeared after separation of Liliopsida and Magnoliopsida classes during divergent evolution.

The obtained results corroborate literature reports, showing high amount of MYB type R2R3, followed by MYB type R1/2 and in few cases R1R2R3. According to Jin and Martin (1999) [5], in plants the predominant family of MYB proteins presents two repeats, but three-repeat MYB together with a growing number of MYB proteins with a single MYB domain were recently identified in plants, as found also in the present work.

Yanhui *et al.* (2006) [13] identified 64 (33%) genes containing one repetition, 126 (65%) containing two repetitions and five (2%) containing three repetitions. A similar proportion to the one found by these authors was revealed in our results, in all transcriptomes studied. Evaluations of transcription profile by microarray using 7,000 genes in *Arabidopsis* revealed the expression of MYB genes are increased 5-fold or more in the first hours after abiotic stress induction [6], with similar results also found in a later experiment in rice [14]. These evaluations revealed that MYB is mainly active during initial salinity and drought stress perception. The identification of such multi function TFs, i.e., active in more than one stress type, are among the most important candidates for improvement of plant resistance to abiotic stress.

Considering our evaluation using virtual bidimensional electrophoresis migration for MYB sequences, two groups could be clearly identified, with only three sequences (two of rice and one of sugarcane) deviating especially with respect to their molecular weight, despite the similar pIs. The first group (with pI from 4.75 to 7.11) included mainly rice sequences, while the second group included most species of the three remaining species. Since the functional domains presented high degree of conservation, probably the divergent pattern of migration for some sequences reflects divergences of the extra-domain regions that are responsible for the acid and basic character of the proteins; these extra-domain variations

are, also, responsible for the diversity of the sequences and for differences in the functionality of the genes.

Regarding generated dendrograms, including MYB proteins from diverse organisms, it was possible to discriminate groupings based on the number of MYB repetitions; the groups XIV, XVII to XIX, XXI and XXIII presented one repetition, the groups I to III, V to VIII, XI to XII, XXII and XXIV presented two repetitions and the groups XII and XV to XVI presented three repetitions. However, some clades containing sequences with one and two MYB repetitions side by side could be observed, as in groups IV, IX, X and XX. The groupings containing two repetition types were formed just by plants, as expected, since in plants the R2R3 type is predominant for MYB family [5], a type absent in *Saccharomyces cerevisiae* whilst only 1-3 copies *R2R3-MYB* genes per haploid genome are known in protists and animals [15]. Besides, the evolutionary pattern of the MYB factor seems to follow a sinarcheomorphic pattern, considering their presence in such different organisms. Regarding the multiple structural forms (number of MYB repetitions), we observed that the presence of three repetitions may be considered a synapomorphy of the animal kingdom, appearing as a monophyletic group. On the other hand, it was not possible to classify remaining types of CD combinations (R1/2 e R2R3), that formed merophyletic polyphyletic groups.

The fact that MYB proteins from animals have grouped separately seems to be related with the number of MYB repetitions. In animals MYB family presents as main characteristic the presence of three imperfect repetitions [16]. In addition, three-repeat MYBs have recently been identified in plants together with a growing number of MYB proteins with a single MYB domain [5]. Our results showed that MYB proteins with three repetitions were not closely related to animal proteins, revealing that these repetitions present specific motifs, in both plants and animals. Despite of that, the neighboring position of both plant sequences with R3 to the animal clade (XIII+XV and XVI, respectively), may indicate that the presence of the third domain is an ancestral feature in plants.

The junction in the same clade of organisms possessing one (R1/2) and two (R2R3) repetitions seem to be related with the functionality of these repetitions. According to Kamiya *et al.* (2002) [17] the DNA-binding specificity is determined primarily by R2 and R3, whereas R1 is not thought to be important for DNA binding specificity. This suggests that selective factors could have collaborated to the diversity of CD combinations in plants. All told, the generated dendrogram may reflect adaptation driven by environmental pressure over the analyzed taxa, rather than stochastic evolutionary patterns of MYB sequences, a gene family that appeared before splitting of the main kingdoms.

Since plants are sessile, transcriptional regulation is also important for adaptation to abiotic stresses, and for protection from biotic stress [18]. The tissues that first “sense” and respond to abiotic stress are roots and root-shoot transition region, for the reduction in water availability causing osmotic imbalance, and the leaves, for the alteration in transpiration and respiration rates [1]. In the FOREST database the *MYB* candidate genes were present in greater amount

in WD (wood), BK (bark, duramen, pith and alburnum) and FB (bud, flower and fruit) libraries, as expected, since these tissues are closely related to plant development. The first one being directly involved in the sustentation, hindering the falling of the plant; the second directly related with plant growth and development, mainly with respect to plant support and cellular water balance; and the last in reproductive processes. In both, eucalyptus and sugarcane, the analysis of expression showed the existence of transcripts in almost all tissues, revealing the crucial importance of transcription factors for plants.

The results from SUCEST database, however, were surprising, with larger amount of reads from libraries generated from tissues infected by nitrogen fixing bacteria (*Herbaspirillum rubrisubalbicans* and by *Gluconacetobacter diazotrophicans*). In these libraries the infection occurred during *in vitro* cultivation. So, the environmental stress can not be discarded as the inducing factor for higher expression, since the *in vitro* environment is largely known as an abiotic stress condition. On the other hand, higher amount of reads from leaves tissues were also found; a plant region closely related with sensing of abiotic stress and responsible for respiration and photosynthesis, processes involved directly with plant growth and development. An association of the here procured TFs with defense also against biotic stress can not be discarded. Tobacco plants presented increased resistance to osmotic stress but also to pathogen attack when a unique *ERF* gene (*Tsi*) was overexpressed [19], showing that responses to different stresses can be regulated by a single 'TF'. Besides activation of *MYB* genes in bacteria infected tissues in sugarcane significant expression was also observed in SB (Stalk Bark) and LV (Leaves) - vascular tissues that need higher amounts of nutrients for plant development and tissues involved in plant respirations respectively - showing the participation of these genes both in defense against pathogen and in acquisition of tolerance to abiotic stresses.

Finally, the fact that MYB TFs were detected in almost all libraries in both FOREST and SUCEST databases, and that a single eucalyptus library was constructed under water deficit, supports the assumption that these genes are expressed not only under stress condition, but also in lower levels or in the absence of stress situations.

Regarding splicing comparative evaluation, TF sequences showed high conservation of splicing patterns. In fact, gene structure analyses of some TFs in different organisms indicated conservation, including the position of putative junction between introns and exons [20,21]. Studying rates and models of intron evolution in green plants Roy and Penny (2007) [22] compared 8,258 pairs of orthologous genes among *A. thaliana* and *O. sativa* and noted that 94.7% of intron positions were conserved comparing the two species and that only 5.3% of introns were species-specific.

The majority of arabidopsis (59%) and rice (53%) *MYB type-R2R3* genes presented conserved splicing patterns, comprising three exons and two introns; however, either one or both of the two introns can be absent. In addition, variable splicing patterns were detected in 22% of arabidopsis and 35% of rice *MYB type-R2R3* genes [23]. Our observations regarding eucalyptus and sugarcane ORFs

revealed a high similarity with three exons present in MYB_03 from *Arabidopsis* and also the corresponding rice paralog, leading to the recognition of at least two introns in MYB sequences of eucalyptus and sugarcane. Additionally, comparisons among ORFs aligned to MYB_01 in *Eucalyptus* and *S. officinarum* exposed at least three exon-exon junction sites revealing that eucalyptus may present two splicing forms for MYB_01 orthologs.

In general, *MYB* TF genes were found in all genomes studied, also considering non induced conditions. However the construction of additional libraries submitted to stress conditions could provide further candidates and give a best coverage of existing TFs in sugarcane and eucalyptus. The recognized sequences and especially their putative splicing sites bring valuable information for future *in vitro* and *in vivo* assays, especially regarding a better understanding of the functional role of their structural diversity.

The identified *MYB* orthologs from sugarcane and eucalyptus are also valuable for comparative studies of evolution among higher plants, including understanding of processes in woody and herbaceous plants. Furthermore, the obtained results are useful for generation of molecular markers in genetic mapping, germplasm screening for breeding purposes and comparative mapping. Since transcription factors are expressed in low levels, the knowledge of the *MYB* gene pool diversity is meaningful for the development of bioinformatic tools and automated assays related to differential expression.

Acknowledgments

The authors acknowledge the FOREST and SUGARCANE consortium for permission to access and mine their databases. To CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) we thank for fellowships and financial support.

References

1. Munns, R.: Comparative Physiology of Salt and Water Stress. *Plant Cell Environ.* 28, 239–250 (2002)
2. Wang, W., Vinocur, B., Altman, A.: Plant Responses to Drought, Salinity and Extreme Temperatures: Towards Genetic Engineering for Stress Tolerance. *Planta* 218, 1–14 (2003)
3. Riechmann, J.L., Herad, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., Yu, G.: *Arabidopsis* Transcription Factors: Genome-Wide Comparative Analysis among Eukaryotes. *Science* 290, 2105–2110 (2000)
4. Liu, L., White, M.J., MacRae, T.H.: Transcription Factors and their Genes in Higher Plants: Functional Domains Evolution and Regulation. *Eur. J. Biochem.* 262, 247–257 (1999)
5. Jin, H., Martin, C.: Multifunctionality and Diversity within the Plant *MYB*-Gene Family. *Plant Mol. Biol.* 41, 577–585 (1999)

6. 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 7,000 *Arabidopsis* Genes under Drought Cold and High-Salinity Stresses Using a Full-Length cDNA Microarray. *Plant J.* 31, 279–292 (2002)
7. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J.: Basic Local Alignment Search Tool. *J. Mol. Biol.* 251, 403–410 (1990)
8. Sneath, P.H.A., Sokal, R.R.: Numerical Taxonomy. Freeman, San Francisco (1973)
9. Page, R.D.: Treeview Program Version 1.6.1. *Comput. Appl. Biosci.* 12, 357–358 (1996)
10. Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D.: Cluster Analysis and Display of Genomic-Wide Expression Pattern. *Proc. Natl. Acad. Sci. USA* 95, 14863–14868 (1998)
11. Hiller, K., Schobert, M., Hundertmark, C., Jahn, D., Munch, R.: JVirGel: Calculation of Virtual Two-Dimensional Protein Gels. *Nucleic Acids Res.* 31, 3862–3865 (2003)
12. Papini-Terzi, F.S., Rocha, F.R., Vêncio, R.Z.N., Oliveira, K.C., Felix, J.M., Vicentini, R., Rocha, C.S., Simões, A.C.Q., Ulian, E.C., Mauro, S.M.Z., Silva, A.M., Pereira, C.A.B., Menossi, M., Souza, G.M.: Transcription Profiling of Signal Transduction-Related Genes in Sugarcane Tissues. *DNA Res.* 12, 27–38 (2005)
13. Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhao Feng, G., Zhiqiang, L.: The MYB Transcription Factor Superfamily of *Arabidopsis*: Expression Analysis and Phylogenetic Comparison with the Rice MYB Family. *Plant Mol. Biol.* 60, 107–124 (2006)
14. Rabbani, M.A., Maruyama, K., Abe, H., Khan, M.A., Katsura, K., Ito, Y., Yoshikawa, K., Seki, M., Shinozaki, K., Yamaguchi-Shinozaki, K.: Monitoring Expression Profiles of Rice Genes under Cold, Drought, and High-Salinity Stresses and Abscisic Acid Application Using cDNA Microarray and RNA Gel-Blot Analyses. *Plant Physiol.* 133, 1755–1767 (2003)
15. Thompson, M.A., Ramsay, R.G.: Myb: an Old Oncoprotein with New Roles. *Bioessays* 17, 341–350 (1995)
16. Lipsick, J.S.: One Billion Years of MYB. *Oncogene* 13, 223–235 (1996)
17. Kamiya, T., Kawabe, A., Miyashita, N.T.: Nucleotide Polymorphism at *Atmyb2* Locus of Wild Plant *Arabidopsis thaliana*. *Genet. Res.* 80, 89–98 (2002)
18. Shikata, M., Matsuda, Y., Ando, K., Nishii, A., Takemura, M., Yekota, A., Kohchi, T.: Characterization of *Arabidopsis* ZIM a Member of Novel Plant-Specific GATA Factor Gene Family. *J. Exp. Bot.* 55, 631–639 (2004)
19. Park, J.M., Park, C.J., Lee, S.B., Ham, B.K., Shin, R., Paek, K.H.: Overexpression of the Tobacco *Tsi1* Gene Encoding an EREBP/AP2-Type Transcription Factor Enhances Resistance Against Pathogen Attack and Osmotic Stress in Tobacco. *Plant Cell* 13, 1035–1046 (2001)
20. Wu, K.-L., Guo, Z.-J., Wang, H.-H., Li, J.: The WRKY Family of Transcription Factors in Rice and Arabidopsis and their Origins. *DNA Res.* 12, 9–26 (2005)
21. Reyes, J.C., Muro-Pastor, M.I., Florêncio, F.L.: The GATA Family of Transcription Factors in Arabidopsis and Rice. *Plant Physiol.* 134, 1718–1732 (2004)
22. Roy, S.W., Penny, D.: Patterns of Introns Loss and Gain in Plants: Intron Loss-Dominated Evolution and Genome Wide Comparison of *O. sativa* and *A. thaliana*. *Mol. Biol. Evol.* 24, 171–181 (2007)
23. Jiang, C., Gu, X., Peterson, T.: Identification of Conserved Gene Structure and Carboxy-Terminal Motifs in the MYB Gene Family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*. *Genome Biol.* 5, 1–11 (2004)



In silico identification of known osmotic stress responsive genes from *Arabidopsis* in soybean and *Medicago*

Nina M. Soares-Cavalcanti¹, Luis C. Belarmino¹, Ederson A. Kido¹, Ana C. Wanderley-Nogueira¹, João P. Bezerra-Neto¹, Rafaela Cavalcanti-Lira¹, Valesca Pandolfi¹, Alexandre L. Nepomuceno², Ricardo V. Abdelnoor², Leandro C. Nascimento³ and Ana M. Benko-Iseppon¹

¹Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Recife, PE, Brazil.

²Embrapa Soja, Londrina, PR, Brazil.

³Laboratório de Genômica e Expressão, Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.

Abstract

Plants experience various environmental stresses, but tolerance to these adverse conditions is a very complex phenomenon. The present research aimed to evaluate a set of genes involved in osmotic response, comparing soybean and medicago with the well-described *Arabidopsis thaliana* model plant. Based on 103 *Arabidopsis* proteins from 27 categories of osmotic stress response, comparative analyses against Genosojá and *Medicago truncatula* databases allowed the identification of 1,088 soybean and 1,210 *Medicago* sequences. The analysis showed a high number of sequences and high diversity, comprising genes from all categories in both organisms. Genes with unknown function were among the most representative, followed by transcription factors, ion transport proteins, water channel, plant defense, protein degradation, cellular structure, organization & biogenesis and senescence. An analysis of sequences with unknown function allowed the annotation of 174 soybean and 217 *Medicago* sequences, most of them concerning transcription factors. However, for about 30% of the sequences no function could be attributed using *in silico* procedures. The establishment of a gene set involved in osmotic stress responses in soybean and barrel medic will help to better understand the survival mechanisms for this type of stress condition in legumes.

Key words: osmotic stress, stress-responsive genes, *Glycine max*, *Medicago truncatula*.

Introduction

In the course of evolution, plants have acquired a myriad of developmental and metabolic strategies to cope with the adverse effects of environmental stresses during vegetative growth and reproduction (Parry *et al.*, 2005), making stress tolerance a complex phenomenon.

Stress perception and the immediate induction of signals that culminate in adaptive responses are key steps leading to plant stress tolerance. Tolerance stress differences between genotypes or different developmental stages of a single genotype may arise from peculiarities in signal perception and transduction mechanisms (Chinnusamy *et al.*, 2004). Under osmotic stress conditions diverse sets of physiological responses are activated, including metabolic and defense systems used to sustain growth and for survival.

The stress-inducible genes are classified into two major groups: one of them protects the plant directly against stresses, whereas the other regulates gene expression and signal transduction (Valliyodan and Nguyen, 2006).

Because plant tolerance against osmotic stress is a complex multigenic trait, a demand exists for genome wide analysis, including 'omics' approaches suitable for uncovering important gene sets involved in this important process (Hirayama and Shinozaki, 2010).

After the 'sequencing era', genetic information was then available for several non-model plants, including some legume species, a group that exhibits unique features, such as the ability to carry the nodulation process. Nitrogen fixation mediated by nodule activities abolishes the need for external nitrogen sources from fertilizers, while providing the so-called 'green manuring' that enriches the soil. Moreover, some legumes, such as soybean, barrel medic and cowpea, are important economic crops that provide humans with food, livestock for feeding purposes, and industry with raw materials (Graham and Vance, 2003).

Send correspondence to Ana M. Benko-Iseppon Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego 1235, 50.670-420 Recife, PE, Brazil. E-mail: ana.iseppon@gmail.com.

Soybean is an example of a non-model plant with plentiful transcriptome information available. Among available databases, the Genosojá platform connects public and restricted data, providing 60,747 unigenes (Nascimento *et al.*, 2012, this issue).

The identification of candidate genes in soybean and barrel medic will provide additional evidence of the response mechanisms for osmotic stresses in Fabaceae, yielding useful information for crop improvement. As osmotic stress cannot be solved solely via remedial land management, tolerant crops – able to maintain cellular turgor and osmotic balance – may contribute significantly to reduce this economic burden. The key to plant engineering for osmotic tolerance lies in the knowledge of the underlying mechanisms of plant adaptive responses (Hariadi *et al.*, 2011).

In the present work the main categories of osmotic stress genes known from *A. thaliana* were identified in the soybean (Genosojá Project) and barrel medic (*M. truncatula*) transcriptomes through an *in silico* approach, in order to contribute to a better understanding of the early molecular adaptation to osmotic (drought and salinity) stress in both leguminous plants.

Materials and Methods

In a previous study based on 7,000 *Arabidopsis* genes, Seki *et al.* (2002) identified 103 coding genes distributed over 27 functional categories (Table 1) whose expression increased more than five times in response to osmotic stress. The protein sequences of these stress-inducible

genes were obtained at the RIKEN Arabidopsis Full-Length Clone Database, and used as query sequences.

After this step, a local bank with the retrieved sequences was generated in order to make searches for similar sequences against the Genosojá platform (Nascimento *et al.*, 2012) and the *M. truncatula* database (Quackenbush *et al.*, 2000) using the tBLASTn algorithm (Altschul *et al.*, 1990) with a cut-off of $1e^{-05}$. The results were annotated in a local databank for further analyses and for comparisons among studied organisms and literature information. In view of the different number of seed sequences per category, the results obtained from each category and organism were normalized. The soybean and *Medicago* genes with unknown function were submitted to the AutoFACT program (Koski *et al.*, 2005), and annotated according to the data available in the largest functional annotation databanks (KEGG, COG, PFAM, SMART, nr). This step was performed in order to categorize these sequences and assign function to them, based on a comparative analysis.

Results and Discussion

The stress-inducible gene products were classified into two main groups: (I) those that are at the front line of defense, protecting the plant against adverse conditions and (II) those that regulate genic expression and signal transduction in the stress response (Seki *et al.*, 2003). The first group included proteins that probably act in the protection of plant cells from dehydration, such as the enzymes required for the biosynthesis of various osmoprotectants, LEA proteins, antifreeze proteins, chaperones and detoxification enzymes. The second group included signaling mol-

Table 1 - Functional categories procured and respective seed-sequence number. Abbreviation: TF = Transcription Factor.

Functional category	# Seed sequence	Functional category	# Seed sequence
bZIP TF	1	WRKY TF	2
Photosynthesis	1	Osmoprotectant	3
Signaling	1	ZincFinger TF	3
Reproductive development	1	Detoxification enzyme	2
Respiration	1	Cellular metabolism	3
DNA nucleus	1	DREB and ERF TF	2
Ferritin	1	Ethylene biosynthesis	2
LEA protein	1	Cytochrome P450	2
MYB TF	1	Fatty Acid metabolism	4
Homeodomain TF	1	Heat Shock protein	2
Membrane protein	2	Kinase protein	2
Senescence-related	1	Carbohydrate metabolism	6
Degradation protein	1	Plant defense	4
Secondary metabolism	1	Transport protein ion channel carrier	4
Water channel protein	1	Cellular struct. organiz. and biogenesis	5
NAC TF	2	Unknown protein	37
Protein phosphatase	2		
Total			103

ecules such as transcription factors and protein kinases, among others (Seki *et al.*, 2003). Twenty-seven categories of these two groups classified according to Seki *et al.* (2002) were analyzed, resulting in 1,088 (soybean) and 1,210 (*Medicago*) sequences (Table S1, supplementary material). In both genomes the 'unknown protein' category was the most representative (Figure 1), with 331 candidates for *Medicago* and 268 for soybean, followed by 'cellular structure organization and biogenesis', 'plant defense' and 'transport protein ion channel carrier' (Figure 1).

The highest number of sequences for genes with 'unknown function' - a very common category in expression essays regarding osmotic stress response in plants - attracting great interest from researchers, since those genes represent a clear source of new candidates for breeding purposes. Previous studies highlighted the importance of analyzing the role of stress-induced genes, not only for a further understanding of the molecular mechanisms of stress tolerance in higher plants, but also for improving crop performance using gene manipulation (Seki *et al.*, 2002).

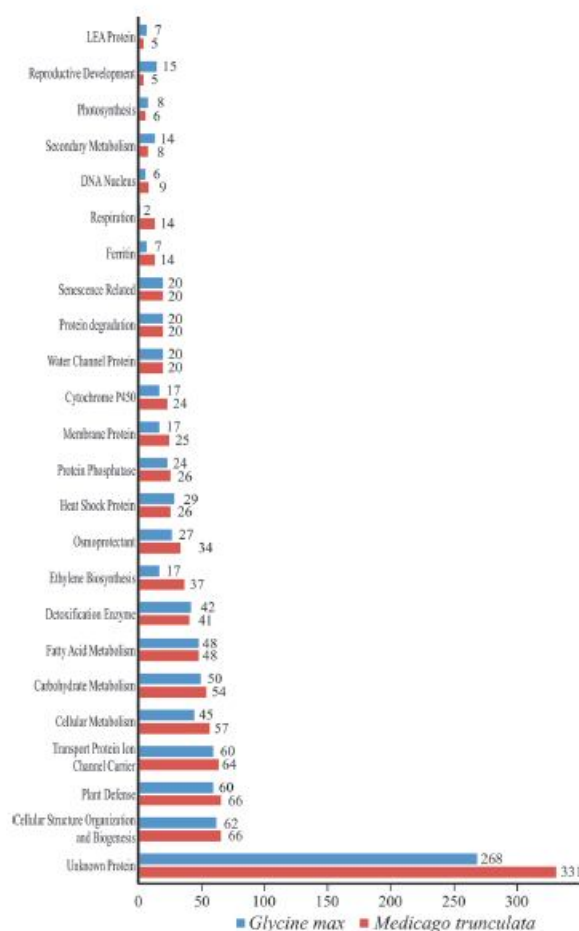


Figure 1 - Main categories of Group I stress-inducible genes (protective molecules), indicating the number of orthologs identified in *Glycine max* and *Medicago truncatula*.

Osmotic stress greatly affects cells both at the micro (*i.e.*, membrane structure), and at the macro level (*i.e.* the physiology of the whole plant), with results that reflect the variety of responses involved in the acquisition of tolerance. At the microcellular level, the activation of genes in the categories 'cellular structure, organization and biogenesis' (soybean: 62; *Medicago*: 66) and 'transport protein ion channel carrier' (soybean: 64; *Medicago*: 60) was observed, showing the importance of the maintenance of cellular structures and of the control of ion exchange with the environment.

Furthermore, we observed the activation of genes in the category 'plant defense' (soybean: 66; *Medicago*: 60), indicating the presence of a cross-talk process between pathways, a common mechanism in plants under stressful conditions. In addition to stress-specific adaptive responses, plants also share responses that protect them from more than one type of stress (Seki *et al.*, 2002; DeFalco *et al.*, 2010; Nuruzzaman *et al.*, 2010), a response also observed in cowpea, another Fabaceae member (Kido *et al.*, 2011).

Amongst the candidates of the second group of responses, composed of genes involved in signal transduction and regulation of expression (203 in soybean and 190 in *Medicago*; Figure 2), the category transcription factor (TF) was the most prevalent, representing up to 80% in soybean and 82% in *Medicago* (Figure 2). The high number of transcription factors suggests that transcriptional regulation is an important mechanism in the signal transduction triggered by osmotic stresses in both legumes.

A surprising result was the absence of a bZIP representative in the soybean database, while in *Medicago* this category was represented by three candidates (Figure 3). This transcription factor has been identified in many plants and is known to participate in various responsive pathways, including abiotic stress response.

Among the transcription factors, the DREB/ERF and Zinc-finger families had the highest number of sequences (Figure 3). This result was expected, since from more than 1,600 transcription factors encoded by *A. thaliana*, 9% are members of the DREB/ERF-like family (Dietz *et al.*, 2010). Due to the versatility of functions that the zinc finger family may have, as well as the variety of their structural proteins, the obtained result was expected. According to

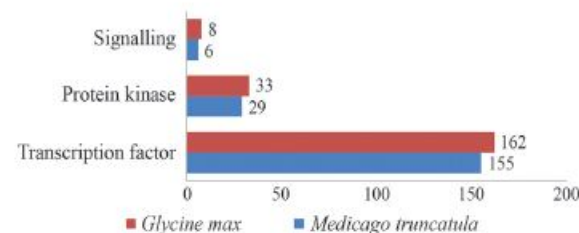


Figure 2 - Percentage of stress-inducible genes (Group II), including cell signaling factors identified in *Glycine max* and *Medicago truncatula*.

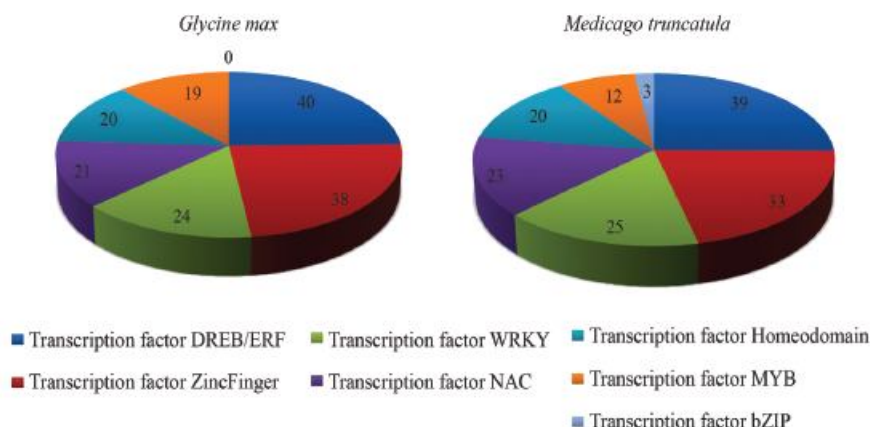


Figure 3 - Graphic representation of transcription factors identified in *Glycine max* and *Medicago truncatula*.

Takatsuji (1998), plants seem to have adopted preexisting prototype zinc-finger motifs, generating new zinc-finger domains to adapt them to various regulatory processes. The zinc finger domain can be present in a number of transcription factors and play critical roles in interactions with other molecules. Mutations in some of the genes coding for zinc-finger proteins have been found to cause profound developmental aberrations or defective responses to environmental cues (Takatsuji, 1998). Zinc finger proteins are required for key cellular processes including transcriptional regulation, development, pathogen defense, and stress responses (Ciftci-Yilmaz and Mittler, 2008). A recent study of rice showed that the C2H2-type zinc finger family alone was represented by 189 members and demonstrated that at least 26 of them respond to different environmental stresses (Agarwal *et al.*, 2007). Moreover, Gong *et al.* (2010), in a study on transcriptional regulation in drought-tolerant tomato genotypes, also identified and characterized the zinc-finger family as the main activated group during the drought response.

It is important to note that the number of seed-sequences used in the search was different for each category; the 'unknown protein' category, for example, was represented by 37 sequences, while the 'bZIP transcription factor' category comprised a single sequence. Thus, it was expected that the more abundant orthologous categories would be those obtained through comparative searches with the categories composed of more query sequences.

As for the remainder, after normalizing the results, proportionally the most representative categories (7% each) were: 'water channel proteins', 'protein degradation' and 'senescence-related' (Figure 4). Without doubt, all categories analyzed may contribute to an improvement in osmotic tolerance, although some functions are more relevant than others. Proteins associated with ion channels and water channels are essential in the acquisition of resistance in the presence of soluble salts and water shortages, the former controlling the entry and exit of ions such as Na^+ , which are

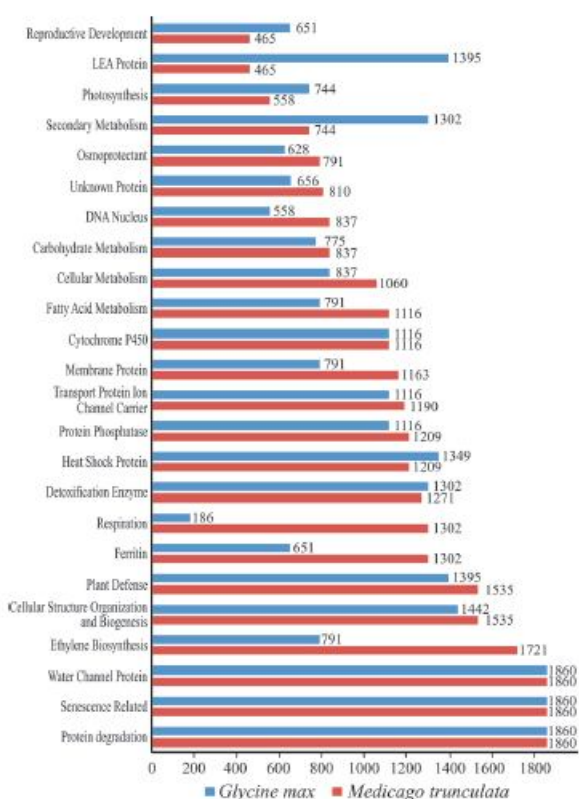


Figure 4 - Number of gene candidates from Group I for *Medicago truncatula* and *Glycine max*.

toxic in high concentrations, and the latter controlling water loss to the environment. Besides these proteins, those falling into the category 'protein degradation' are required for protein turnover and recycling of essential amino acids, while 'senescence-related' genes are key components in the abiotic stress response, with genes controlling subcellular changes that lead to tolerance (Seki *et al.*, 2002).

Table 2 - Sequence description annotated according to the COG (Cluster of Orthologous Groups) functional category in *Glycine max* and *Medicago truncatula*.

COG functional category	Sequence description	Sequence amount	
		<i>G. max</i>	<i>M. truncatula</i>
Amino acid transport and metabolism	Amino acid permease	9	8
Carbohydrate transport and metabolism	Beta-galactosidase	0	2
General function prediction only	Patatin	4	17
Posttranslational modification, protein turnover, chaperones	DnaJ-like protein	14	13
Signal transduction mechanisms	Universal Stress Protein (USP) family protein	15	17
Total		42	57

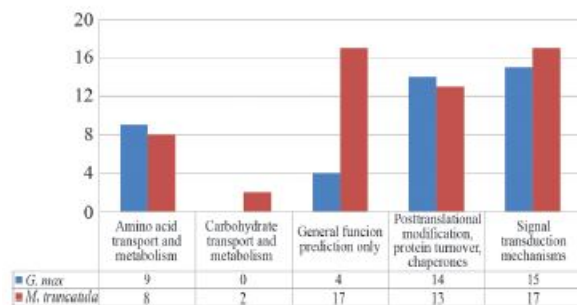
While the normalized results evidenced similar amounts of data in the most representative categories for both organisms, in some categories there were significant variations in the number of sequences between both leguminous species (Figure 4); this difference was even greater than 50% for the categories 'Reproductive development' (soybean: 1,395; *Medicago*: 465), 'Ferritin' (soybean: 651; *Medicago*: 1,392), 'Respiration' (soybean: 186; *Medicago*: 1,302) and 'Ethylene biosynthesis' (soybean: 791; *Medicago*: 1,721). Nevertheless, this variation may be related to the conditions under which the data were generated and deposited, as well as to the number of sequences available in the respective databases. Additionally, species-specific features could be responsible for these variations, to a lesser extent.

Regarding the category 'Unknown Protein', screened candidates from soybean (268) and *Medicago* (331) were subjected to the AutoFACT program in order to assign function to these sequences, allowing the recognition of the function of 174 and 217 sequences, respectively.

As a result, 42 and 57 *G. max* and *M. truncatula* were categorized according to the COG (Cluster of Orthologous Groups) functional database in five categories (Table 2; Figure 5). Within each category, the annotation revealed that they present the same description as the matched sequences deposited in the databank. For example, the 'Amino acid transport and metabolism' functional category was represented just by 'Amino Acid Permease' sequences (Ta-

Table 3 - Description of sequences with unknown function after AutoFACT analysis.

Description	<i>G. max</i>	<i>M. truncatula</i>
Amino acid permease	7	4
ATP binding / kinase / protein serine/threonine kinase	0	3
Auxin-responsive GH3 product [Glycine max]	2	8
BTB/POZ domain-containing protein	0	2
Calcium ion binding	2	4
Calmodulin binding	10	14
CCT_2 domain containing protein	4	5
Copper ion binding / electron transporter	4	1
Cu-binding-like domain containing protein	4	10
Dev_Cell_Death domain containing protein	9	17
DFL1 (DWARF IN LIGHT 1)	1	0
DnaJ-like protein [<i>Phaseolus vulgaris</i>]	3	2
F-box family protein	5	4
Heat shock protein binding	3	4
Herpes BLLF1 domain containing protein	1	0
Hydroxyproline-rich glycoprotein family protein	0	1
IFRD1; interferon-related developmental regulator 1	8	0
Indole-3-acetic acid-amido synthetase GH3.17, putative	3	5
NAC Transcription Factor	4	10
Nucleic acid binding / transcription factor	18	14
Patatin B2 precursor, putative	1	0
PHI-1 (PHOSPHATE-INDUCED 1)	19	20
Plastocyanin-like domain-containing protein	0	1
RCI2A (RARE-COLD-INDUCIBLE 2A)	0	2
SMC_N multi-domain protein	1	3
SPX domain-containing protein	2	0
Stress-inducible protein	0	2
Tify domain containing protein	8	12
Triacylglycerol lipase	5	5
Uncharacterized protein family/Unassigned protein/Protein of unknown function	94	114
Universal stress protein (USP) family protein	1	3
Zinc finger family protein	7	4

**Figure 5** - Categorization of soybean and *Medicago* 'unknown category' candidates based on COG (Cluster of Orthologous Groups) functional database.

ble 2). Two candidates of *Medicago*, which were functionally classified into the 'Carbohydrate transport and metabolism' category, were also annotated on the KEGG database as involved in the beta-galactosidase pathway (Galactose Metabolism Glycan Structure – degradation), (Table 2).

The remaining previously 'unknown' sequences were annotated as shown in Table 3. The analysis through AutoFACT allowed a function assignment to 132 and 160 soybean and *Medicago* sequences, respectively. In general, the highest number of sequences was categorized as transcription factors, essential genes participating in the transcriptional regulation of plants. Although it was possible to record more than 65% of the sequences, 35% of 'unknown' soybean and 34% of 'unknown' *Medicago* sequences remained without their putative function identified. These are relevant data to be worked out in future functional studies, since they may represent new genes not yet described and unique to legumes.

In conclusion, even in the absence of libraries restricted to osmotic stress in the Genosoja databank, this study indicated that most of the genes involved in the osmotic stress pathways were expressed by the non-stressed soybean and *Medicago* libraries at least in a baseline way. The data also revealed that soybean and *Medicago* are a rich source of stress-responsive candidates, which can be also applied to improve soybean and other legumes. It also highlights the existence of significant diversity for most genes, useful for comparative physiological essays. The obtained data are available for gene-targeted functional evaluation using qRT-PCR, as well as other biotechnological approaches. The molecular differences detected between the compared libraries will permit the identification of important candidates by additional approaches including PCR walking, as previously done for other crops (e.g. Coemans *et al.*, 2005).

The identified candidates are also being monitored in further expression assays carried out in the Genosoja project (considering contrasting combinations of tolerant and susceptible plants under drought stress as compared with their negative control in a time frame) providing a more complete picture of genes involved in osmotic stress response and useful for breeding and biotechnological purposes.

Acknowledgments

The authors would like to thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco), and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for their financial support.

References

- Agarwal P, Arora R, Ray S, Singh AK, Singh VP, Takatsuji H, Kapoor S and Tyagi AK (2007) Genome-wide identification of C(2)H(2) zinc-finger gene family in rice and their phylogeny and expression analysis. *Plant Mol Biol* 65:467-485.
- Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410.
- Chinnusamy V, Schumaker K and Zhu J-K (2004) Molecular genetics perspectives on cross-talk and specificity in abiotic stress signaling in plants. *J Exp Bot* 55:225-236.
- Ciftci-Yilmaz S and Mittler R (2008) The zinc finger network of plants. *Cell Mol Life Sci* 65:1150-1160.
- Coemans B, Matsumura H, Terauchi R, Remy S, Swennen R and Sági L (2005) SuperSAGE combined with PCR walking allows global gene expression profiling of banana (*Musa acuminata*), a non-model organism. *Theor Appl Genet* 111:1118-1126.
- DeFalco TA, Bender KW and Snedden WA (2010) Breaking the code: Ca²⁺ 27 sensors in plant signalling. *Biochem J* 425:27-40.
- Dietz K-J, Vogel MO and Viehhauser A (2010) AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma* 245:3-14.
- Gong P, Zhang J, Li H, Yang C, Zhang C, Zhang X, Khurram Z, Zhang Y, Wang T, Fei Z, *et al.* (2010) Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. *J Exp Bot* 61:3563-3575.
- Graham PH and Vance CP (2003) Legumes: Importance and constraints to greater use. *Plant Physiol* 131:872-877.
- Hariadi Y, Marandon K, Tian Y, Jacobsen S-E and Shabala S (2011) Ionic and osmotic relations in quinoa (*Chenopodium quinoa* Willd.) plants grown at various salinity levels. *J Exp Bot* 62:185-193.
- Hirayama T and Shinozaki K (2010) Research on plant abiotic stress responses in the post-genome era: Past, present and future. *Plant J* 61:1041-1052.
- Kido EA, Barbosa PK, Ferreira Neto JCR, Pandolfi V, Houllou-Kido LM, Crovella S and Benko-Iseppon AM (2011) Identification of plant protein kinases in response to abiotic and biotic stresses using SuperSAGE. *Curr Prot Pept Sci* 12:643-656.
- Koski LB, Gray MW, Lang BF and Burger G (2005) AutoFACT: An Automatic Functional Annotation and Classification Tool. *BMC Bioinformatics* 16:151-161.
- Nascimento LC, Costa GGL, Binneck E, Pereira GAG and Carazzolle MF (2012) A web-based bioinformatics interface applied to Genosoja Project: Databases and pipelines. *Genet Mol Biol* 35(suppl 1): 203-211.
- Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H and Kikuchi S (2010) Genome-wide analysis of NAC transcription factor family in rice. *Gene* 465:30-44.
- Parry MAJ, Flexas J and Medrano H (2005) Prospects for crop production under drought: Research priorities and future directions. *Ann Appl Biol* 147:211-226.
- Quackenbush J, Liang F, Holt I, Pertea G and Upton J (2000) The TIGR Gene Indices: Reconstruction and representation of expressed gene sequences. *Nucleic Acids Res* 28:141-145.
- Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, *et al.* (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J* 31:279-292.
- Seki M, Kamei A, Yamaguchi-Shinozaki K and Shinozaki K (2003) Molecular responses to drought, salinity and frost: Common and different paths for plant protection. *Curr Opin Plant Biol* 14:194-199.

- Takatsuji H (1998) Zinc-finger transcription factors in plants. *Cell Mol Life Sci* 54:582-596.
- Valliyodan B and Nguyen HT (2006) Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Curr Opin Plant Biol* 9:189-195.

Internet Resources

- RIKEN Arabidopsis Full-Length Clone Database, <http://www.brc.riken.go.jp/lab/epd/catalog/cdnaclone.html> (May, 2011)
- Genosoja platform, <http://bioinfo03.ibi.unicamp.br/soja/> (May, 2011)
- Medicago truncatula* database, <http://www.medicago.org/> (May, 2011)

Supplementary Material

The following online material is available for this article:

Table S1 - Identified candidates among abiotic stress responsive gene categories in soybean and *Medicago* genomes.

This material is available as part of the online article from <http://www.scielo.br/gmb>.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplementary Material

Table S1. Identified candidates among abiotic stress responsive genes categories in soybean and medicago genomes based on selected arabidopsis seed sequences, as well as number of other hits, e-value and score, against the respective database of *Medicago truncatula* (Mt) and *Glycine max* (Gm).

Arabidopsis Information		Blast Result				
Category	Query Sequence	Best Hit	Organism	Other Hits	E-value	Score
bZIP Transcription Factor	At1g42990	Mt_bZIP_1	<i>M. truncatula</i>	2	7.00 e-12	67.4
	At3g10740	Mt_Carb_Met_1	<i>M. truncatula</i>	7	0.0	959
Carbohydrate Metabolism	At3g06500	Gm_Carb_Met_1	<i>G. max</i>	8	0.0	947
		Gm_Carb_Met_2	<i>G. max</i>	7	6.00 e-121	430
	At3g60130	Mt_Carb_Met_2	<i>M. truncatula</i>	6	0.0	788
		Gm_Carb_Met_3	<i>G. max</i>	2	8.00 e-76	280
	At5g18670	Mt_Carb_Met_3	<i>M. truncatula</i>	5	1.00 e-107	386
		Gm_Carb_Met_4	<i>G. max</i>	3	1.00 e-166	582
	At3g04240	Mt_Carb_Met_4	<i>M. truncatula</i>	3	8.00 e-141	496
		Gm_Carb_Met_5	<i>G. max</i>	19	0.0	919
	At2g43820	Mt_Carb_Met_5	<i>M. truncatula</i>	17	0.0	1166
		Gm_Carb_Met_6	<i>G. max</i>	5	5.00 e-98	353
Cellular Metabolism	At3g53180	Mt_Carb_Met_6	<i>M. truncatula</i>	10	2.00 e-106	382
		Gm_Cell_Met_1	<i>G. max</i>	1	8.00 e-83	304
	At3g45300	Mt_Cell_Met_1	<i>M. truncatula</i>	7	9.00 e-175	610
		Gm_Cell_Met_2	<i>G. max</i>	4	2.00 e-60	229
	At2g39210	Mt_Cell_Met_2	<i>M. truncatula</i>	7	6.00 e-153	536
		Gm_Cell_Met_3	<i>G. max</i>	14	2.00 e-156	352
	At2g42970	Mt_Cell_Met_3	<i>M. truncatula</i>	8	1.00 e-91	333
		Gm_Cell_Met_4	<i>G. max</i>	2	3.00 e-91	331
	At1g68620	Mt_Cell_Met_4	<i>M. truncatula</i>	11	7.00 e-145	510
		Gm_Cell_Met_5	<i>G. max</i>	19	1.00 e-38	155
Cellular Structure Organization and Biogenesis	At1g03220	Mt_Cell_Met_5	<i>M. truncatula</i>	19	7.00 e-40	159
		Gm_Cell_Stru_Org_Biog_1	<i>G. max</i>	16	4.00 e-110	394
	At3g10720	Mt_Cell_Stru_Org_Biog_1	<i>M. truncatula</i>	19	4.00 e-145	510
		Gm_Cell_Stru_Org_Biog_2	<i>G. max</i>	7	1.00 e-123	437
	At5g62350	Mt_Cell_Stru_Org_Biog_2	<i>M. truncatula</i>	7	1.00 e-87	318
		Gm_Cell_Stru_Org_Biog_3	<i>G. max</i>	16	1.00 e-47	184
	At5g20230	Mt_Cell_Stru_Org_Biog_3	<i>M. truncatula</i>	18	9.00 e-47	182
		Gm_Cell_Stru_Org_Biog_4	<i>G. max</i>	19	7.00 e-21	95.9
		Mt_Cell_Stru_Org_Biog_4	<i>M. truncatula</i>	18	1.00 e-22	102

Cytochrome P450	At2g34500	Gm_Cytoch_P450_1	<i>G. max</i>	11	2.00 e-101	365
		Mt_Cytoch_P450_1	<i>M. truncatula</i>	12	0.0	684
	At3g26220	Gm_Cytoch_P450_2	<i>G. max</i>	4	2.00 e-85	311
		Mt_Cytoch_P450_2	<i>M. truncatula</i>	10	8.00 e-97	350
Detoxification Enzyme	At2g31570	Gm_Detox_Enz_1	<i>G. max</i>	14	4.00 e-43	169
		Mt_Detox_Enz_1	<i>M. truncatula</i>	17	9.00 e-40	157
	At2g29450	Gm_Detox_Enz_2	<i>G. max</i>	19	4.00 e-51	196
		Mt_Detox_Enz_2	<i>M. truncatula</i>	19	2.00 e-50	194
	At5g44070	Gm_Detox_Enz_3	<i>G. max</i>	6	4.00 e-107	374
		Mt_Detox_Enz_3	<i>M. truncatula</i>	2	1.00 e-95	346
DNA Nucleus	At2g18050	Gm_DNA_Nuc_1	<i>G. max</i>	5	3.00 e-11	63.2
		Mt_DNA_Nuc_1	<i>M. truncatula</i>	8	1.00 e-11	65.1
DREB/ERF Transcription Factor	At1g22190	Gm_DREB_ERF_TF_1	<i>G. max</i>	19	2.00 e-41	164
		Mt_DREB_ERF_TF_1	<i>M. truncatula</i>	18	2.00 e-37	152
	At4g17500	Gm_DREB_ERF_TF_2	<i>G. max</i>	19	6.00 e-33	136
		Mt_DREB_ERF_TF_2	<i>M. truncatula</i>	19	2.00 e-52	202
Ethylene Biosynthesis	At5g43450	Gm_Ethyl_Bios_1	<i>G. max</i>	7	1.00 e-112	402
		Mt_Ethyl_Bios_1	<i>M. truncatula</i>	19	4.00 e-106	380
	At1g17020	Gm_Ethyl_Bios_2	<i>G. max</i>	8	5.00 e-61	230
		Mt_Ethyl_Bios_2	<i>M. truncatula</i>	16	7.00 e-101	363
Fatty Acid Metabolism	At1g73480	Gm_Fatty_Acid_Met_1	<i>G. max</i>	14	6.00 e-96	347
		Mt_Fatty_Acid_Met_1	<i>M. truncatula</i>	18	5.00 e-26	115
	At4g09760	Gm_Fatty_Acid_Met_2	<i>G. max</i>	8	3.00 e-94	340
		Mt_Fatty_Acid_Met_2	<i>M. truncatula</i>	6	3.00 e-124	441
	At1g73920	Gm_Fatty_Acid_Met_3	<i>G. max</i>	3	2.00 e-58	222
		Mt_Fatty_Acid_Met_3	<i>M. truncatula</i>	1	3.00 e-134	446
	At1g07720	Gm_Fatty_Acid_Met_4	<i>G. max</i>	19	2.00 e-92	335
		Mt_Fatty_Acid_Met_4	<i>M. truncatula</i>	19	8.00 e-168	586
Ferritin	At5g01600	Gm_Ferritin_1	<i>G. max</i>	6	1.00 e-85	311
		Mt_Ferritin_1	<i>M. truncatula</i>	13	1.00 e-87	318
Heat Shock Protein	At3g46230	Gm_HSF_1	<i>G. max</i>	19	2.00 e-51	197
		Mt_HSF_1	<i>M. truncatula</i>	19	1.00 e-51	198
	At1g16030	Gm_HSF_2	<i>G. max</i>	8	0.0	996
		Mt_HSF_2	<i>M. truncatula</i>	5	0.0	994
Homeodomain Transcription Factor	At2g35940	Gm_Homeodom_TF_1	<i>G. max</i>	19	8.00 e-88	320
		Mt_Homeodom_TF_1	<i>M. truncatula</i>	19	4.00 e-116	415
LEA Protein	At4g02380	Gm_LEA_1	<i>G. max</i>	6	6.00 e-09	55.5
		Mt_LEA_1	<i>M. truncatula</i>	4	2.00 e-07	50.4
Membrane Protein	At5g54170	Gm_Memb_Prot_1	<i>G. max</i>	7	1.00 e-116	415
		Mt_Memb_Prot_1	<i>M. truncatula</i>	6	3.00 e-113	332
	At1g30360	Gm_Memb_Prot_2	<i>G. max</i>	8	6.00 e-44	174
		Mt_Memb_Prot_2	<i>M. truncatula</i>	17	0.0	855

MYB Transcription Factor	At1g01060	Gm_MYB_TF_1	<i>G. max</i>	18	8.00 e-57	217
		Mt_MYB_TF_1	<i>M. truncatula</i>	11	4.00 e-25	112
NAC Transcription Factor	At5g63790	Gm_NAC_TF_1	<i>G. max</i>	10	1.00 e-89	325
		Mt_NAC_TF_1	<i>M. truncatula</i>	11	4.00 e-91	330
	At4g27410	Gm_NAC_TF_2	<i>G. max</i>	9	2.00 e-95	344
		Mt_NAC_TF_2	<i>M. truncatula</i>	10	2.00 e-94	341
Osmoprotectant	At2g47180	Gm_Osmoprot_1	<i>G. max</i>	4	2.00 e-142	500
		Mt_Osmoprot_1	<i>M. truncatula</i>	4	3.00 e-157	550
	At1g09350	Gm_Osmoprot_2	<i>G. max</i>	-	2.00 e-11	65.5
		Gm_Osmoprot_3	<i>G. max</i>	-	5.00 e-09	57.8
	At1g60470	Mt_Osmoprot_2	<i>M. truncatula</i>	-	9.00 e-16	80.9
		Gm_Osmoprot_4	<i>G. max</i>	8	9.00 e-155	543
	At3g57520	Mt_Osmoprot_3	<i>M. truncatula</i>	9	0.0	865
		Gm_Osmoprot_5	<i>G. max</i>	10	0.0	1410
Photosynthesis	At5g20830	Mt_Osmoprot_4	<i>M. truncatula</i>	17	0.0	1384
		Gm_Photosynt_1	<i>G. max</i>	7	6.00 e-143	504
	At4g15530	Mt_Photosynt_1	<i>M. truncatula</i>	5	0.0	1102
		Gm_Plant_Defen_1	<i>G. max</i>	19	6.00 e-86	313
Plant Defense	At3g55430	Mt_Plant_Defen_1	<i>M. truncatula</i>	19	2.00 e-133	472
		Gm_Plant_Defen_2	<i>G. max</i>	19	4.00 e-42	166
	At4g13580	Mt_Plant_Defen_2	<i>M. truncatula</i>	19	1.00 e-67	252
		Gm_Plant_Defen_3	<i>G. max</i>	-	1.00 e-53	206
	At2g40000	Mt_Plant_Defen_3	<i>M. truncatula</i>	5	9.00 e-130	459
		Gm_Plant_Defen_4	<i>G. max</i>	18	6.00 e-109	389
Protein degradation	At5g06860	Mt_Plant_Defen_4	<i>M. truncatula</i>	19	2.00 e-100	362
		Gm_Prot_Degrad_1	<i>G. max</i>	19	0.0	634
	At1g47128	Mt_Prot_Degrad_1	<i>M. truncatula</i>	19	7.00 e-161	330
		Gm_Prot_Kinase_1	<i>G. max</i>	15	7.00 e-166	309
Protein Kinase	At2g31880	Mt_Prot_Kinase_1	<i>M. truncatula</i>	14	9.00 e-114	407
		Gm_Prot_Kinase_2	<i>G. max</i>	16	5.00 e-146	242
	At5g25110	Mt_Prot_Kinase_2	<i>M. truncatula</i>	13	3.00 e-143	504
		Gm_Prot_Phosphat_1	<i>G. max</i>	13	3.00 e-108	387
Protein Phosphatase	At4g26080	Mt_Prot_Phosphat_1	<i>M. truncatula</i>	13	7.00 e-98	353
		Gm_Prot_Phosphat_1	<i>G. max</i>	9	1.00 e-50	196
	At3g11410	Mt_Prot_Phosphat_1	<i>M. truncatula</i>	11	2.00 e-91	332
		Gm_Reprod_Develop_1	<i>G. max</i>	14	3.00 e-162	566
Reproductive Development	At5g56750	Mt_Reprod_Develop_1	<i>M. truncatula</i>	4	5.00 e-74	274
		Gm_Reprod_Develop_2	<i>G. max</i>	1	3.00 e-142	500
	At3g22370	Mt_Reprod_Develop_2	<i>M. truncatula</i>	13	9.00 e-143	502
		Gm_Second_Metabol_1	<i>G. max</i>	13	3.00 e-67	251
Secondary Metabolism	At2g38240	Mt_Second_Metabol_1	<i>M. truncatula</i>	7	1.00 e-125	445
Senescence-	At5g13170	Gm_Senesc_Relat_1	<i>G. max</i>	19	4.00 e-69	257

Related		Mt_Senesc_Relat_1	<i>M. truncatula</i>	19	1.00 e-74	275
Signalling	At5g33380	Gm_Siganlling_1	<i>G. max</i>	7	8.00 e-57	215
		Mt_Siganlling_1	<i>M. truncatula</i>	5	2.00 e-43	171
		Gm_Transp_Prot_Ion_1	<i>G. max</i>	19	4.00 e-116	414
Transport Protein Ion Channel Carrier	At1g58360	Mt_Transp_Prot_Ion_1	<i>M. truncatula</i>	19	1.00 e-180	629
	At1g08930	Gm_Transp_Prot_Ion_2	<i>G. max</i>	19	3.00 e-81	298
	At5g20380	Gm_Transp_Prot_Ion_3	<i>G. max</i>	5	5.00 e-93	337
		Mt_Transp_Prot_Ion_2	<i>M. truncatula</i>	12	1.00 e-86	317
	At2g22500	Gm_Transp_Prot_Ion_4	<i>G. max</i>	13	2.00 e-117	417
		Mt_Transp_Prot_Ion_3	<i>M. truncatula</i>	10	1.00 e-112	402
	At5g22290	Gm_Unknown_1	<i>G. max</i>	7	6.00 e-42	167
		Mt_Unknown_1	<i>M. truncatula</i>	11	4.00 e-62	234
	At1g11210	Mt_Unknown_2	<i>M. truncatula</i>	-	8.00 e-11	64.3
	At1g15430	Gm_Unknown_2	<i>G. max</i>	3	5.00 e-28	120
		Mt_Unknown_3	<i>M. truncatula</i>	7	1.00 e-31	132
	At1g55280	Gm_Unknown_3	<i>G. max</i>	-	9.00 e-34	140
		Mt_Unknown_4	<i>M. truncatula</i>	-	6.00 e-39	157
	At1g63720	Gm_Unknown_4	<i>G. max</i>	5	2.00 e-20	95.5
		Mt_Unknown_5	<i>M. truncatula</i>	1	2.00 e-28	122
	At1g69890	Gm_Unknown_5	<i>G. max</i>	1	2.00 e-76	281
		Mt_Unknown_6	<i>M. truncatula</i>	7	2.00 e-66	248
	At1g76600	Gm_Unknown_6	<i>G. max</i>	1	4.00 e-16	80.5
		Mt_Unknown_7	<i>M. truncatula</i>	2	1.00 e-14	75.9
Unknown Protein	At2g26560	Gm_Unknown_7	<i>G. max</i>	4	3.00 e-124	441
		Mt_Unknown_8	<i>M. truncatula</i>	19	3.00 e-147	518
	At2g32240	Gm_Unknown_8	<i>G. max</i>	1	2.00 e-39	160
		Mt_Unknown_9	<i>M. truncatula</i>	2	2.00 e-67	254
	At2g38820	Gm_Unknown_9	<i>G. max</i>	14	2.00 e-56	214
		Mt_Unknown_10	<i>M. truncatula</i>	19	3.00 e-81	298
	At2g41190	Gm_Unknown_10	<i>G. max</i>	15	1.00 e-61	233
		Mt_Unknown_11	<i>M. truncatula</i>	11	6.00 e-50	154
	At3g17800	Gm_Unknown_11	<i>G. max</i>	12	6.00 e-79	290
		Mt_Unknown_12	<i>M. truncatula</i>	9	1.00 e-87	320
	At4g21570	Gm_Unknown_12	<i>G. max</i>	13	3.00 e-106	380
		Mt_Unknown_13	<i>M. truncatula</i>	4	4.00 e-58	202
	At4g25670	Gm_Unknown_13	<i>G. max</i>	3	5.00 e-29	123
		Mt_Unknown_14	<i>M. truncatula</i>	1	1.00 e-25	112
	At4g27520	Gm_Unknown_14	<i>G. max</i>	7	7.00 e-20	94.0
		Mt_Unknown_15	<i>M. truncatula</i>	12	3.00 e-30	129
	At4g30650	Gm_Unknown_15	<i>G. max</i>	6	2.00 e-13	70.1
		Mt_Unknown_16	<i>M. truncatula</i>	6	4.00 e-18	86.3
	At4g38060	Gm_Unknown_16	<i>G. max</i>	-	9.00 e-18	84.7
		Mt_Unknown_17	<i>M. truncatula</i>	3	2.00 e-15	77.8

	At5g02020	Gm_Unknown_17	<i>G. max</i>	1	8.00 e-10	58.2
		Mt_Unknown_18	<i>M. truncatula</i>	3	2.00 e-07	50.8
	At5g42050	Gm_Unknown_18	<i>G. max</i>	8	8.00 e-66	246
		Mt_Unknown_19	<i>M. truncatula</i>	17	2.00 e-70	262
	At5g50100	Gm_Unknown_19	<i>G. max</i>	1	3.00 e-60	197
		Mt_Unknown_20	<i>M. truncatula</i>	1	2.00 e-46	181
	At3g61060	Gm_Unknown_20	<i>G. max</i>	19	5.00 e-102	366
		Mt_Unknown_21	<i>M. truncatula</i>	19	2.00 e-101	365
	At4g37390	Gm_Unknown_21	<i>G. max</i>	5	0.0	510
		Mt_Unknown_22	<i>M. truncatula</i>	12	0.0	754
	At5g630160	Gm_Unknown_22	<i>G. max</i>	3	3.00 e-20	95.1
		Mt_Unknown_23	<i>M. truncatula</i>	4	5.00 e-32	134
	At5g43260	Mt_Unknown_24	<i>M. truncatula</i>	1	3.00 e-36	146
	At1g76650	Gm_Unknown_23	<i>G. max</i>	1	4.00 e-11	63.5
		Mt_Unknown_25	<i>M. truncatula</i>	5	6.00 e-25	109
	At1g29395	Mt_Unknown_26	<i>M. truncatula</i>	-	2.00 e-40	161
	At2g40140	Gm_Unknown_24	<i>G. max</i>	19	1.00 e-161	565
		Mt_Unknown_27	<i>M. truncatula</i>	19	5.00 e-153	537
	At4g36040	Gm_Unknown_25	<i>G. max</i>	20	1.00 e-28	121
		Mt_Unknown_28	<i>M. truncatula</i>	19	2.00 e-26	114
	At4g33050	Gm_Unknown_26	<i>G. max</i>	9	1.00 e-59	225
		Mt_Unknown_29	<i>M. truncatula</i>	14	7.00 e-74	273
	At5g09440	Gm_Unknown_27	<i>G. max</i>	18	3.00 e-87	317
		Mt_Unknown_30	<i>M. truncatula</i>	19	2.00 e-79	291
	At1g19180	Gm_Unknown_28	<i>G. max</i>	9	9.00 e-38	152
		Mt_Unknown_31	<i>M. truncatula</i>	14	3.00 e-23	105
	At1g17380	Gm_Unknown_29	<i>G. max</i>	1	4.00 e-11	64.3
		Mt_Unknown_32	<i>M. truncatula</i>	1	7.00 e-14	74.3
	At1g02660	Gm_Unknown_30	<i>G. max</i>	4	4.00 e-59	224
		Mt_Unknown_33	<i>M. truncatula</i>	5	7.00 e-168	587
	At2g21620	Gm_Unknown_31	<i>G. max</i>	2	6.00 e-66	245
		Mt_Unknown_34	<i>M. truncatula</i>	1	2.00 e-65	244
	At1g27760	Gm_Unknown_32	<i>G. max</i>	7	2.00 e-126	448
		Mt_Unknown_35	<i>M. truncatula</i>	1	6.00 e-107	384
	At1g63010	Gm_Unknown_33	<i>G. max</i>	4	3.00 e-77	285
		Mt_Unknown_36	<i>M. truncatula</i>	3	8.00 e-89	324
	At2g41640	Gm_Unknown_34	<i>G. max</i>	1	1.00 e-56	216
		Mt_Unknown_37	<i>M. truncatula</i>	2	9.00 e-160	559
	At1g11360	Gm_Unknown_35	<i>G. max</i>	14	5.00 e-67	249
		Mt_Unknown_38	<i>M. truncatula</i>	19	1.00 e-59	225
Water Channel		Gm_Water_Chanel_1	<i>G. max</i>	19	2.00 e-116	414
Protein	At2g37180	Mt_Water_Chanel_1	<i>M. truncatula</i>	19	3.00 e-115	410
WRKY	At2g30250	Gm_WRKY_TF_1	<i>G. max</i>	10	4.00 e-65	244

Transcription Factor	At5g13080	Mt_WRKY_TF_1	<i>M. truncatula</i>	15	3.00 e-62	235
		Gm_WRKY_TF_2	<i>G. max</i>	12	9.00 e-43	167
Mt_WRKY_TF_2		<i>M. truncatula</i>	8	1.00 e-40	161	
Zinc Finger Transcription Factor	At2g19580	Gm_ZF_TF_1	<i>G. max</i>	7	9.00 e-31	129
		Mt_ZF_TF_1	<i>M. truncatula</i>	15	4.00 e-32	134
	At5g59820	Gm_ZF_TF_2	<i>G. max</i>	11	1.00 e-24	107
	At2g31380	Gm_ZF_TF_3	<i>G. max</i>	17	7.00 e-68	252
		Mt_ZF_TF_2	<i>M. truncatula</i>	16	4.00 e-67	250



Overall picture of expressed Heat Shock Factors in *Glycine max*, *Lotus japonicus* and *Medicago truncatula*

Nina M. Soares-Cavalcanti¹, Luís C. Belarmino¹, Ederson A. Kido¹, Valesca Pandolfi¹, Francismar C. Marcelino-Guimaraes², Fabiana A. Rodrigues², Gonçalo A.G. Pereira³ and Ana M. Benko-Iseppon¹

¹Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Recife, PE, Brazil.

²Embrapa Soja, Londrina, PR, Brazil.

³Laboratório de Genômica e Expressão, Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.

Abstract

Heat shock (HS) leads to the activation of molecular mechanisms, known as HS-response, that prevent damage and enhance survival under stress. Plants have a flexible and specialized network of Heat Shock Factors (HSFs), which are transcription factors that induce the expression of heat shock proteins. The present work aimed to identify and characterize the *Glycine max* HSF repertory in the Soybean Genome Project (GENOSOJA platform), comparing them with other legumes (*Medicago truncatula* and *Lotus japonicus*) in view of current knowledge of *Arabidopsis thaliana*. The HSF characterization in leguminous plants led to the identification of 25, 19 and 21 candidate ESTs in soybean, *Lotus* and *Medicago*, respectively. A search in the SuperSAGE libraries revealed 68 tags distributed in seven HSF gene types. From the total number of obtained tags, more than 70% were related to root tissues (water deficit stress libraries vs. controls), indicating their role in abiotic stress responses, since the root is the first tissue to sense and respond to abiotic stress. Moreover, as heat stress is related to the pressure of dryness, a higher HSF expression was expected at the water deficit libraries. On the other hand, expressive HSF candidates were obtained from the library inoculated with Asian Soybean Rust, inferring crosstalk among genes associated with abiotic and biotic stresses. Evolutionary relationships among sequences were consistent with different HSF classes and subclasses. Expression profiling indicated that regulation of specific genes is associated with the stage of plant development and also with stimuli from other abiotic stresses pointing to the maintenance of HSF expression at a basal level in soybean, favoring its activation under heat-stress conditions.

Key words: HSF, Fabaceae, bioinformatics, abiotic stress, transcription factor.

Introduction

Heat stress is one of the major factors limiting the productivity and adaptation of crops, especially when temperature extremes coincide with critical stages of plant development. The major developmental performance of plants occurs at a temperature regime between 10° and 40 °C. Temperatures below or above this range generally cause temperature-induced stresses (Treshow, 1970; Hsu *et al.*, 2010). In the case of heat stress, both the rate of temperature change and the duration and degree of high temperatures contribute to the intensity of heat stress. The degree of inherent adaptedness to heat stress of a plant is an important

determinant of its ability to survive a stress period (Efeolu, 2009). However, the expression of HSF and HSP genes has been also observed under other abiotic and biotic stresses, as cited by Pirkkala *et al.* (2001). In response to various inducers such as elevated temperatures, salinity, drought, oxidants, heavy metals, bacterial and viral infections, most HSFs acquire DNA binding activity to the heat shock element (HSE), thereby mediating transcription of the heat shock factor genes, which results in accumulation of heat shock proteins (HSPs). Among important transcription factors, heat shock factors (HSFs) are essential for the transcription of many HSP coding genes that are active in response to sublethal heat stress leading to increased tolerance against a subsequent, otherwise lethal, heat shock (Treshow, 1970; Hsu *et al.*, 2010).

Send correspondence to Ana M. Benko-Iseppon. Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego 1235, 50.670-420 Recife, PE, Brazil. E-mail: ana.benko.iseppon@pq.cnpq.br.

After stress perception, intracellular changes lead to a molecular cascade of events, initiated by HSF activation and subsequent expression of HSPs limiting stress damage (Hsu *et al.*, 2010). In general, HSF proteins have a common core structure comprising a N-terminal DNA binding domain (DBD) characterized by a central helix-turn-helix (HTH) motif, an adjacent domain with a heptad hydrophobic repeat (HR-A/B) which is involved in oligomerization, a short peptide motif essential for nuclear import [nuclear localization signal (NLS)] and export [nuclear export signal (NES)], and a C-terminal AHA type activation domain (Mittal *et al.*, 2009; Hsu *et al.*, 2010).

Through the DNA binding domain, activated HSFs bind to conserved cis-acting elements called heat shock elements (HSEs). HSEs are located in the promoters of HSP genes and are defined as adjacent and inverse repeats of the motif 5-nGAAn-3, for instance 5-nGAAnnTTCnnGAAn-3 (Schöffl *et al.*, 1998).

Some HSFs have been cloned and characterized from various plant species (Nover *et al.*, 2001; Baniwal *et al.*, 2007) revealing that the network of HSF genes is highly flexible and specialized in this group. Details regarding the overall HS response network were initially not clear. However, studies in *Arabidopsis* revealed that 21 HSFs form a complex network, in which AtHSFA1a and AtHSFA1b play important roles in the induction of HSP genes in the early stage of HSR (Nover *et al.*, 2001).

An insight into the response of HSPs and HSFs to different abiotic stresses was provided through a number of genome-wide microarray datasets. *Arabidopsis* HSFs and HSPs were strongly induced by heat, cold, salinity and osmotic stresses. Furthermore, overlapping responses of HSPs and HSFs to heat and other abiotic stresses are reported, indicating that these genes are important elements in the crosstalk among different response pathways (Hu *et al.*, 2009). In rice, over-expression of OsHsp17.7 enhanced rice tolerance to heat UV-B as well as to drought (Sato and Yokoya, 2008).

Hu *et al.* (2009) identified rice *HSF* and *HSP* genes and analyzed their expression profiles under different abiotic stresses. A whole-genome microarray analysis was carried out to investigate expression changes of rice *HSFs* and *HSPs* genes in response to heat stress. By comparing their experimental data with other expression data under salt, cold, and drought conditions, Hu *et al.* (2009) found that the rice HSF and HSP families responded to different stresses in an overlapping relationship. The analysis also indicated that some *HSF* and *HSP* genes exhibited specific expression patterns in response to distinct stress types.

In *Arabidopsis*, for example, the major role of the representatives of the HsfA4/A5 group, which is generally not involved in the conventional heat stress response, may reside in cell type-specific functions connected with the control of cell death triggered by pathogen infection and/or reactive oxygen species (Baniwal *et al.*, 2007).

Although the flexible network of HSF genes has been well studied in plants, there is little information available regarding the structure and function of HSF genes in legumes. Additionally, no comparison of HSF orthologs has been carried out up until now among legumes. In this study, we used well-described *Arabidopsis* HSF proteins as seed sequences in order to identify and characterize the pool of HSF genes present in the *Glycine max* genome and perform a comparative analysis against *Lotus japonicus* and *Medicago truncatula* genomes, so as to trace the panorama of the HSF genes in these leguminous plants.

Material and Methods

Based on 21 well-described *Arabidopsis* HSF genes in the AfTDB database, BLASTp searches (Altschul *et al.*, 1990) were carried out for similar sequences against the GENOSOJA database. GENOSOJA connects public and project soybean data (Nascimento *et al.*, 2012). In total, the initiative provides information on 60,747 unigenes from the NBCI, Phytozome and Soybean full-length cDNA databases (Nascimento *et al.*, 2012). Comparative searches were made in the *Medicago truncatula* and *Lotus japonicus* (databases). After searching the GENOSOJA databank, only orthologs presenting the fully characteristic HSF DNA-Binding Domain (DBD) were considered for subsequent analysis. In view of the obtained soybean, *Medicago* and *Lotus* HSF candidates together with the *Arabidopsis* seed-sequences, a comparative analysis with 69 aligned proteins was performed, enabling the generation of a dendrogram, using the Neighbor-Joining (NJ) method with 2,000 bootstrap replications using the program MEGA version 5.0 (Tamura *et al.*, 2011), to infer about HSF groups and classes within the analyzed legumes. For this purpose the sequence-coding genes from *Arabidopsis* that did not present similarity (orthology) with the studied legumes were excluded from the phenetic analysis. To prevent the influence of different sequence sizes, the alignments were trimmed aiming to exclude unequal 5' and 3' extremities.

To evaluate the HSF-related tags represented in the SuperSAGE libraries generated by the GENOSOJA project, a comparative analysis using the same seed sequences and the MegaBLAST algorithm was carried out according to Altschul *et al.* (1990). For this purpose the parameters were adjusted to an e-value equal to or less than 0.1 and word size equal to 7. The low complexity filter was deactivated. Results considered only tags with identity equal to or larger than 23 bp.

The GENOSOJA databank is comprised of six SuperSAGE libraries and allowed the generation of three comparisons, including two from root tissues subjected to water deficit stress and one inoculated with Asian Soybean Rust fungus (*Phakopsora pachyrhizi*). For the water deficit libraries, seeds of a drought tolerant cultivar (Embrapa 48) and a drought susceptible cultivar (BR 16) were germinated on filter paper for four days in a growth chamber at

25 ± 1 °C and 100% relative humidity (RH). Seedlings were placed in 36 L boxes containing 50% Hoagland's solution (Hoagland and Arnon, 1950) continuously aerated and replaced on a weekly basis. These boxes were then transferred to a greenhouse under natural photoperiod of approximately 12/12 h light/dark cycle, temperature of 30 ± 5 °C and 60 ± 10% RH. The plants were allowed to grow until the V4 stage (Fehr *et al.*, 1971). The experimental plan was a randomized complete block 2x7 factorial design with three repetitions. The treatments included two cultivars (BR 16 and Embrapa 48) and seven water deficit periods (0, 25, 50, 75, 100, 125 and 150 min). Water stress was applied by removing the plants from the hydroponic solution and leaving them in boxes without nutrient solution for up to 150 min under ambient-air exposure. For each stress exposure time, roots from 10 plants were collected, pooled and frozen in liquid nitrogen before storage at -80 °C. The abovementioned exposure times were bulked together generating a library from drought tolerant genotype Embrapa 48 after stress as compared with the negative control (T0); the same procedure was also applied to the drought sensitive genotype (BR16 cultivar). The comparison regarding Asian Soybean Rust infection was generated from leaves of the resistant cultivar PI561356 collected at different times (12, 24 and 48 h) after spraying with a *P. pachyrhizi* spore suspension (6 x 10⁵ uredospores.mL⁻¹). The urediniospores were collected from *Phakopsora pachyrhizi* infected soybean fields in the state of Mato Grosso, Brazil, and maintained for over 10 generations on the susceptible cv. BRSMS-Bacuri. The suspension of spores was sprayed onto three plants per pot at the V2 to V3 growth stage (Fehr and Caviness, 1977). The same solution without the spores was used for the false inoculations (Mock). The different times were bulked together to form a single resistant library, which was compared with the false inoculated negative control collected at the same time points.

Considering the identified *G. max* EST transcripts, standard statistical methods (see Eisen *et al.*, 1998) were used to arrange the HSF genes according to their gene expression pattern, generating a graphic with colors (green, red and black) indicating their quantitative and qualitative expression (down-, up- and unregulated genes, respectively), while gray stood for absence of information. The gene expression data analyzed were collected from soybean during a variety of challenging and control conditions available at the GENOSOJA database. So as to obtain a picture of how HSFs contribute to sensing the environmental up-shifts in temperature, we applied Self-Organizing Maps followed by pairwise average-linkage cluster analysis to normalized gene expression data (Eisen *et al.*, 1998). Relationships among genes and libraries were represented by dendrograms in which branch lengths reflect the degree of gene co-expression.

An available genome browser for soybean (Phytozome database) was used to anchor identified EST candidate sequences on *G. max* virtual chromosomes, aiming to identify their distribution, relative position, and abundance. For this purpose the MegaBLAST tool was used to identify the exact location of the HSF genes in the genome, using at least 80% identity as a parameter. For the construction of a virtual karyotype representation, a CorelDRAW12 graphic application was used. The soybean chromosome information for the schematic representation was obtained from the SOYBASE site. For the design of chromosomes, considering the need for high-resolution bands (data anchored in the genome), a proportion of 1:1 (cm:Mb) was adopted for all chromosomes; thus, for the sequence positioning, each millimeter corresponded to 100,000 bp. On the representation each transversal black line corresponded to an HSF gene.

Results and Discussion

Heat and cold can have damaging consequences on both vegetative and reproductive tissues. Temperature changes can also regulate plant movements, resetting internal clocks and diurnal synchronization, flowering and germination in some species (Ruelland and Zachowski, 2010). Moreover, temperature changes can induce metabolic changes so that plants adapt and tolerate moderate cold, freezing and heat stresses (Ruelland and Zachowski, 2010). HSFs are important components of the heat shock regulatory network, with a single gene identified for yeast and drosophila, while vertebrates accounted with only four genes of this category (Swindell *et al.*, 2007). Nevertheless, unlike other organisms, plant genomes encode extraordinarily complex HSF families, both in terms of the total number of genes (usually more than 20), as well as in terms of their structural and functional diversification (Nover *et al.*, 2001). This abundance and diversity can be also seen in legumes. An extensive BLAST search of *Arabidopsis* HSF orthologs in soybean, *Lotus* and *Medicago* EST databases led to the identification of a total of 25, 19 and 21 expressed sequences, respectively (Table 1).

HSF expressed sequence tags

The characteristic HSF domains were complete in 24, 13 and 17 orthologous candidates identified among the three species, respectively (Table 1). From the 21 types of *Arabidopsis* HSF genes only 13 types were identified in soybean and *Lotus* and 14 in *Medicago* (Table 1). In our evaluation, HSFA1B, HSFA6A, HSFA7B and HSFA9 were absent in all species analyzed (Table 1). HSFA1A and HSFA1B interact as regulators responsible for immediate-early transcription of a subset of HS genes in *Arabidopsis*, and are independently important for the initial phase of HS-responsive gene expression, while their interaction enhances the expression of their target genes (Li *et al.*, 2010). The absence of HSFA1B may render soybean more sensitive to heat stress but another class An HSF may

Table 1 - *Arabidopsis thaliana* heat shock factors used as seed sequences and respective matches from *Glycine max*, *Lotus japonica* and *Medicago truncatula*, with corresponding domain information. Abbreviations: HSF = Heat Shock Factor; ID, identification; Gm, *Glycine max*; Lj, *Lotus japonicus*; Mt, *Medicago truncatula*.

<i>Arabidopsis</i> seed-sequence (Loci)	Orthologous information			BLAST results	
	Fabaceae species	Sequence ID	E-value	Score	HSF domain
ATHSFA1A (AT4G17750.1)	<i>Glycine max</i>	Gm_HSFA1A.1	1.00 e ⁻⁷⁷	166	Complete
	<i>Lotus japonicus</i>	Lj_HSFA1A.1	3.00 e ⁻³⁰	88	Incomplete
	<i>Medicago truncatula</i>	Mt_HSFA1A.1	1.00 e ⁻¹⁰⁶	381	Complete
ATHSFA1B (AT5G16820.1)	-	-	-	-	-
ATHSFA1D (AT1G32330.1)	<i>Lotus japonicus</i>	Lj_HSFA1D.1	4.00 e ⁻⁶⁷	406	Incomplete
ATHSFA1E (AT3G02990.1)	<i>Glycine max</i>	Gm_HSFA1E.1	2.00 e ⁻³²	159	Complete
		Gm_HSFA1E.2	5.00 e ⁻²³	124	Complete
	<i>Lotus japonicus</i>	Lj_HSFA1E.1	4.00 e ⁻⁹¹	487	Complete
	<i>Medicago truncatula</i>	Mt_HSFA1E.1	6.00 e ⁻⁹⁸	354	Complete
		Mt_HSFA1E.2	6.00 e ⁻⁹⁸	354	Complete
ATHSFA2 (AT2G26150.1)	<i>Glycine max</i>	Gm_HSFA2.1	6.00 e ⁻³⁰	68.9	Complete
	<i>Lotus japonicus</i>	Lj_HSFA2.1	4.00 e ⁻⁹¹	387	Complete
		Lj_HSFA2.2	1.00 e ⁻⁶⁴	373	Incomplete
	<i>Medicago truncatula</i>	Mt_HSFA2.1	4.00 e ⁻⁹²	334	Complete
ATHSFA3 (AT5G03720.1)	<i>Lotus japonicus</i>	Lj_HSFA3.1	1.00 e ⁻⁷⁴	324	Complete
ATHSFA4A (AT4G18880.1)	<i>Glycine max</i>	Gm_HSFA4A.1	1.00 e ⁻⁸²	143	Complete
		Gm_HSFA4A.2	7.00 e ⁻¹²	121	Complete
	<i>Lotus japonicus</i>	Lj_HSFA4A.1	3.00 e ⁻⁹⁰	401	Complete
		Lj_HSFA4A.2	6.00 e ⁻⁸⁹	393	Complete
	<i>Medicago truncatula</i>	Mt_HSFA4A.1	1.00 e ⁻⁸⁴	309	Complete
ATHSFA4C (AT5G45710.1)	<i>Medicago truncatula</i>	Mt_HSFA4C.1	9.00 e ⁻⁴⁷	183	Complete
ATHSFA5 (AT4G13980.1)	<i>Glycine max</i>	Gm_HSFA5.1	2.00 e ⁻⁵⁸	117	Complete
		Gm_HSFA5.2	1.00 e ⁻¹⁰⁹	392	Complete
	<i>Medicago truncatula</i>	Mt_HSFA5.1	1.00 e ⁻¹⁰⁴	374	Complete
		Mt_HSFA5.2	2.00 e ⁻⁷²	269	Complete
		Mt_HSFA5.3	2.00 e ⁻⁶¹	232	Complete
		Mt_HSFA5.4	1.00 e ⁻²¹	100	Incomplete
		Mt_HSFA5.5	2.00 e ⁻¹⁵	79.7	Incomplete
ATHSFA6A (AT5G43840.1)	-	-	-	-	-
ATHSFA6B (AT3G22830.1)	<i>Glycine max</i>	Gm_HSFA6B.1	1.00 e ⁻⁴⁹	153	Complete
	<i>Lotus japonicus</i>	Lj_HSFA6B.1	9.00 e ⁻⁸²	291	Complete
		Lj_HSFA6B.2	7.00 e ⁻⁸¹	311	Complete
		Lj_HSFA6B.3	7.00 e ⁻⁹⁷	347	Complete
	<i>Medicago truncatula</i>	Mt_HSFA6B.1	1.00 e ⁻⁶⁶	249	Complete
ATHSFA7A (AT3G51910.1)	<i>Glycine max</i>	Gm_HSFA7B.1	5.00 e ⁻²¹	117	Incomplete
ATHSFA7B (AT3G63350.1)	-	-	-	-	-
ATHSFA8 (AT1G67970.1)	<i>Glycine max</i>	Gm_HSFA8.1	3.00 e ⁻⁶³	131	Complete
		Gm_HSFA8.2	0.0	678	Complete
	<i>Lotus japonicus</i>	Lj_HSFA8.1	9.00 e ⁻⁶⁵	318	Complete
	<i>Medicago truncatula</i>	Mt_HSFA8.1	1.00 e ⁻⁶¹	233	Complete
ATHSFA9 (AT5G54070.1)	-	-	-	-	-

Table 1 (cont.)

<i>Arabidopsis</i> seed-sequence (Loci)	Orthologous information			BLAST results	
	Fabaceae species	Sequence ID	E-value	Score	HSF domain
ATHSFB1 (AT4G36990.1)	<i>Glycine max</i>	Gm_HSFB1.1	e^{-144}	508	Complete
		Gm_HSFB1.2	e^{-127}	450	Complete
	<i>Lotus japonicus</i>	Lj_HSFB1.1	$3.00 e^{-18}$	187	Incomplete
		Lj_HSFB1.2	$1.00 e^{-26}$	68	Incomplete
	<i>Medicago truncatula</i>	Mt_HSFB1.1	$4.00 e^{-51}$	197	Complete
		Mt_HSFB1.2	$3.00 e^{-36}$	148	Complete
		Mt_HSFAB1.3	$7.00 e^{-19}$	90.9	Incomplete
ATHSFB2A (AT5G62020.1)	<i>Glycine max</i>	Gm_HSFB2A.1	$3.00 e^{-48}$	80.5	Complete
		Gm_HSFB2A.2	$5.00 e^{-73}$	270	Complete
	<i>Lotus japonicus</i>	Lj_HSFB2A.1	$9.00 e^{-30}$	197	Complete
	<i>Medicago truncatula</i>	Mt_HSFB2B.1	$3.00 e^{-06}$	48.9	Incomplete
ATHSFB2B (AT4G11660.1)	<i>Glycine max</i>	Gm_HSFB2B.1	$2.00 e^{-44}$	222	Complete
		Gm_HSFB2B.2	$5.00 e^{-81}$	296	Complete
	<i>Lotus japonicus</i>	Lj_HSFB2B.1	$8.00 e^{-68}$	280	Complete
	<i>Medicago truncatula</i>	Mt_HSFB2B.1	$6.00 e^{-67}$	251	Complete
ATHSFB3 (AT2G41690.1)	<i>Glycine max</i>	Gm_HSFB3.1	$1.00 e^{-47}$	160	Complete
		Gm_HSFB3.2	$6.00 e^{-32}$	97.4	Complete
	<i>Lotus japonicus</i>	Lj_HSFB3.1	$5.00 e^{-56}$	194	Complete
	<i>Medicago truncatula</i>	Mt_HSFB3.1	$1.00 e^{-51}$	199	Complete
ATHSFB4 (AT1G46264.1)	<i>Glycine max</i>	Gm_HSFB4.1	$3.00 e^{-61}$	96.3	Complete
		Gm_HSFB4.2	$1.00 e^{-44}$	86.3	Complete
		Gm_HSFB4.3	$1.00 e^{-126}$	448	Complete
		Gm_HSFB4.4	$1.00 e^{-120}$	427	Complete
	<i>Lotus japonicus</i>	Lj_HSFB4.1	$1.00 e^{-66}$	202	Incomplete
		Lj_HSFB4.2	$1.00 e^{-19}$	52	Complete
	<i>Medicago truncatula</i>	Mt_HSFB4.1	$2.00 e^{-89}$	325	complete
		Mt_HSFB4.2	$8.00 e^{-65}$	243	Complete
ATHSFC1 (AT3G24520.1)	<i>Glycine max</i>	Gm_HSFC1.1	$8.00 e^{-60}$	73.6	Complete

alternately play this role (Sung *et al.*, 2003; Kotak *et al.*, 2004; Li *et al.*, 2004). Whether another gene substitutes the role of HSFA1B in soybean could be tested by heterologous expression of HSFA1B; in the case of the existence of different pathways, the over-expression of HSFA1B might change the performance of soybean plants, especially under heat stress.

Other members of class A, such as HSFA9, are less active or may be active only under certain conditions. The reason seems to be the presence of interesting regulators (HSFs or other transcription factors) with specialized functions. In fact, HSFA9 was found to be specific to seed development in sunflower and was exclusively detected in yellow siliques of *Arabidopsis* mRNA (Kotak *et al.*, 2004). Hence the lack of identification of some HSF classes may correlate with specialized functions other than those represented among the conditions analyzed herein.

A similar result was reported by Nover *et al.* (2001) after carrying out an analysis of HSFs in *A. thaliana*. Among the 21 described genes, HSFs A3, A6A, A6B, A7B, B2A and B3 could not be detected in any of the tissues analyzed (etiolated seedlings, roots, leaves from vegetative plants stems, flowers, siliques, and developing seeds) or conditions (heat stressed leaves and cell cultures *vs.* control). According to the authors it was not surprising that no matching EST was found in libraries created exclusively from RNA isolated from control tissues; a serious limitation of the data from EST libraries for these studies is the lack of samples from heat stressed tissues.

Comparing the obtained results with the data available in the Legume Transcription Factor database (Legume TFDB, Mochida *et al.*, 2009a) an increased number of *Lotus* and *Medicago* HSF representatives were observed, since the LegumeTFDB includes 18 and 16 genes, respec-

tively, and our searches identified 19 and 21, respectively, revealing that both organisms presented a similar number of HSFs as *Arabidopsis*. However, the results considering the ESTs deposited at the GENOSOJA platform revealed a surprisingly low number of HSFs (25 sequences) as compared to the LegumeTFDB information (65 sequences). This may be due to the type of databases (LegumeTFDB is sourced from large-scale shotgun sequencing whereas GENOSOJA is sourced from transcriptomic approaches), besides the fact that the LegumeTFDB bank considers both HSF and HSF-like sequences with data annotation based on different databanks (NCBI nr, *A. thaliana*, TIGR rice, *L. japonicus*, *M. truncatula*, *Populus trichocarpa* and UniProt). On the other hand, Kotak *et al.* (2004) listed 34 soybean sequences, a higher number of HSF representatives than those in GENOSOJA, but these authors did not indicate the methods and procedures used in the acquisition of these HSFs. Finally, the soybean candidates identified herein represent the active (expressed) HSFs bearing the complete DBD-domain. This set size was similar to that described for *Arabidopsis* and also for the *Lotus* and *Medicago* orthologs identified in this study; both being evolutionarily closely related species when compared to soybean (Fabaceae family, Papilionoideae subfamily).

Notwithstanding, it is important to highlight that evolutionary studies and haploid genome analysis suggested that the soybean genome experienced a tetraploidization event approximately 10-15 million years ago. Since then, the soybean genome has gone through gene rearrangements and deletions, reverting to diploid state. Therefore, soybean multigene families, including the heat shock factor family, may contain highly related but diversified genes (Mochida *et al.*, 2009b).

HSF matching to SuperSAGE tags

Regarding SuperSAGE, 68 different tags could be identified, including 26 tags unique to water deficit experiments with the tolerant comparison (water deficit stressed Embrapa 48 vs. control), 28 tags unique to water deficit experiments with the susceptible comparison (water deficit stressed BR16 vs. control) and 14 regarding Asian Soybean

Rust (PI561356 inoculated vs. control) (Table 2; Figure 1). No common tags were identified. It is important to note that among 25 HSF EST clusters, 18 had no representative in the tags database, while five clusters were represented in all libraries. The sequence Gmax_HSF1_SJ09-E1-R06-064-B09-UC.F was not identified in 'Embrapa 48' and 'PI561356' libraries, and Gmax_HSF3_Contig20961 was present in the water deficit stressed libraries only. When looked at from a different point of view, from the 14 HSF types compared, only six HSF types (HSFB1, HSFA1E, HSFB2A, HSFB3, HSFA8 and HSFA4A) were identified (Table 2; Figure 1), indicating their induction during the stress response.

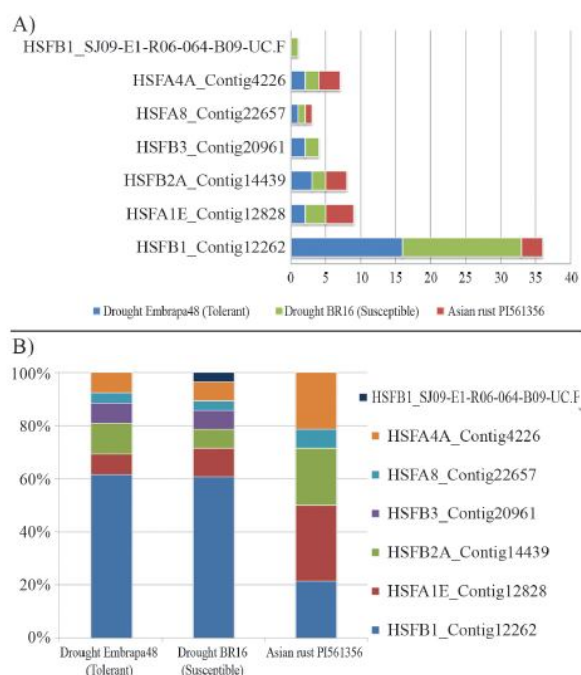


Figure 1 - Graphic representation of SuperSAGE tag distribution per cluster and compared libraries. A) Quantitative analysis (number of units per category). B) Qualitative analysis of tag prevalence (in %).

Table 2 - Distribution of unique SuperSAGE tags in the three comparisons considered, as compared with the identified soybean EST contigs.

Soybean Contig	Drought Embrapa 48 (tolerant) vs. Control	Drought BR16 (susceptible) vs. Control	Asian Rust PI561356 vs. Control	Total
Gmax_HSF1_Contig12262	16	17	3	36
Gmax_HSFA1E_Contig12828	2	3	4	9
Gmax_HSF2A_Contig14439	3	2	3	8
Gmax_HSF3_Contig20961	2	2	0	4
Gmax_HSFA8_Contig22657	1	1	1	3
Gmax_HSFA4A_Contig4226	2	2	3	7
Gmax_HSF1_SJ09-E1-R06-064-B09-UC.F	0	1	0	1
Total	26	28	14	68

Despite the small number of identified sequences in the Asian rust 'PI561356' stress analysis, when compared to water deficit experiments, the presence of HSF representatives indicates the involvement of HS-response also during biotic stresses. The stress condition by itself can activate non-specific stress-responsive-pathways, due to the debility caused to plants by biotic stressful conditions, which can activate a crosstalk among different stress related pathways, as observed in other plants (Glombitza *et al.*, 2004; Kido *et al.*, 2011). Moreover, it is important to consider the tissue from which the library was generated, since leaves are among the first organs to present stress symptoms (especially to abiotic ones). These are necessary for the maintenance of photosynthesis and evapotranspiration processes to ensure plant survival. Moreover, the Gmax_HSFB3_Contig20961 gene seems to be expressed specifically under abiotic stress, such as water deficit.

The analysis of SuperSAGE transcript abundance revealed a higher number of orthologous tags for the Gmax_HSFB1_Contig12262 cluster (more than 50% of the identified SuperSAGE tags), followed by Gmax_HSFA1E_Contig12828 (Table 2; Figure 1A). There is evidence suggesting that HsfB1 plays a special role in gene activation as a cooperative partner of HsfA1 and that coexpression of low levels of HSFB1 with HSFA1 can result in strong synergistic effects in reporter gene activation. Experiments in tomato showed that HSFB1 acts as a novel type of coactivator and may be able to cooperate with HSFA1a or other activators to control expression of certain housekeeping genes (Bharti *et al.*, 2004).

Evaluating the results for the comparisons among water deficit libraries (susceptible X tolerant), a similar proportion of HSF genes was observed, with the exception of the Gmax_HSFB1_SJ09-E1-R06-064-B09-UC.F transcript, which was recorded exclusively in the susceptible genotype. In both libraries, Gmax_HSFB1_Contig12262 (Figure 1B) was more represented, indicating that HSF genes are expressed under water stress conditions in a similar way in both susceptible and tolerant cultivars.

As expected, most SuperSAGE tags were identified from water deficit libraries. However, it is worth noting that more than 60% of the HSF gene types obtained from soybean ESTs were not identified in the SuperSAGE comparisons, suggesting that the seed EST sequences used were not complete, lacking the necessary 3' extremity for anchoring of SuperSAGE tags. This opens the possibility of identifying additional candidates upon using other annotation approaches. A role of these factors in water deficit response may exist, since their expression was reported also in association with other abiotic stresses (Kotak *et al.*, 2007). Moreover, the 68 identified tags could be potentially useful for 3' RACE (3'-rapid amplification of cDNA ends) experiments to identify the complete transcript, besides expression validation using RT-qPCR with the same mRNA samples.

Structure and evolution of HSF candidates in soybean, *Medicago*, *Lotus* and *Arabidopsis*

The functional properties of HSFs are attributed to conserved structural domains, with the highest degree of conservation being observed for the DNA-binding domain (DBD) composed of helix-turn-helix (HTH) structures, and an adjacent domain with a heptad hydrophobic repeat (HR-A/B) which is involved in oligomerization. In addition, there are two further characteristic components: (i) the short peptide motif essential for nuclear import (NLS: nuclear localization signal) and export (NES: nuclear export signal), and (ii) a C-terminal AHA type activation domain (Li *et al.*, 2010). Primarily based on the structural features of the oligomerization domain, plant HSFs are classified into three evolutionarily-conserved classes, viz. A, B and C, bearing 14 sub-classes (Nover *et al.*, 2001). The high degree of conservation within the HSF family is corroborated by our *in silico* analysis, as in the generated dendrogram it was possible to observe the differentiation of sequences according to their classes, and within each class there was a grouping of sequences according to their sub-classes (Figure 2). A clear differentiation among the HSF classes A and B classes from a basal ancestral sequence has been established, as expected, since class B- and non-plant-HSFs differ from class A- and C-HSFs by an additional 21 or 7 amino acids, respectively, which separate the two sub-domains HR-A and HR-B located in the hydrophobic regions (Nover *et al.*, 2001). Furthermore, the AHA type acidic activation domain is exclusively represented by class A members (Mittal *et al.*, 2009).

With respect to class A, two main groups emerged in the present evaluation: one (I) with HSFA4 and HSFA5 representatives and the other (II) with the remaining HSFA and HSFC members (Figure 2). This is a predictable result, since HSFs A4 and A5 form a group distinct from the remaining HSFs by structural features of their oligomerization domains and by a number of conserved signatures. This is also consistent with their role, since A4 HSFs are potent activators of heat stress-related gene expression, whereas A5 HSFs act as a specific repressor of HSFA4 activity, while other members of class A are not affected due to the high specificity of their oligomerization domains (Baniwal *et al.*, 2007).

The second group included three branches, with a basal one including HSFA8 and HSFC1 (Figure 2). Although class C is more similar to class A than to B, it was expected that this class would behave as a separate group. Nevertheless, the high diversity in the response of different HSF genes to different stresses suggests that there is a high degree of specialization regarding the response of specific HSFs to a particular stress condition. This is consistent with the fact that both HSFA8 and HSFC presented increased expression under cold stress (Miller and Mittler, 2006), indicating that this adaptive response to tolerate cold conditions may be responsible for characteristics shared by

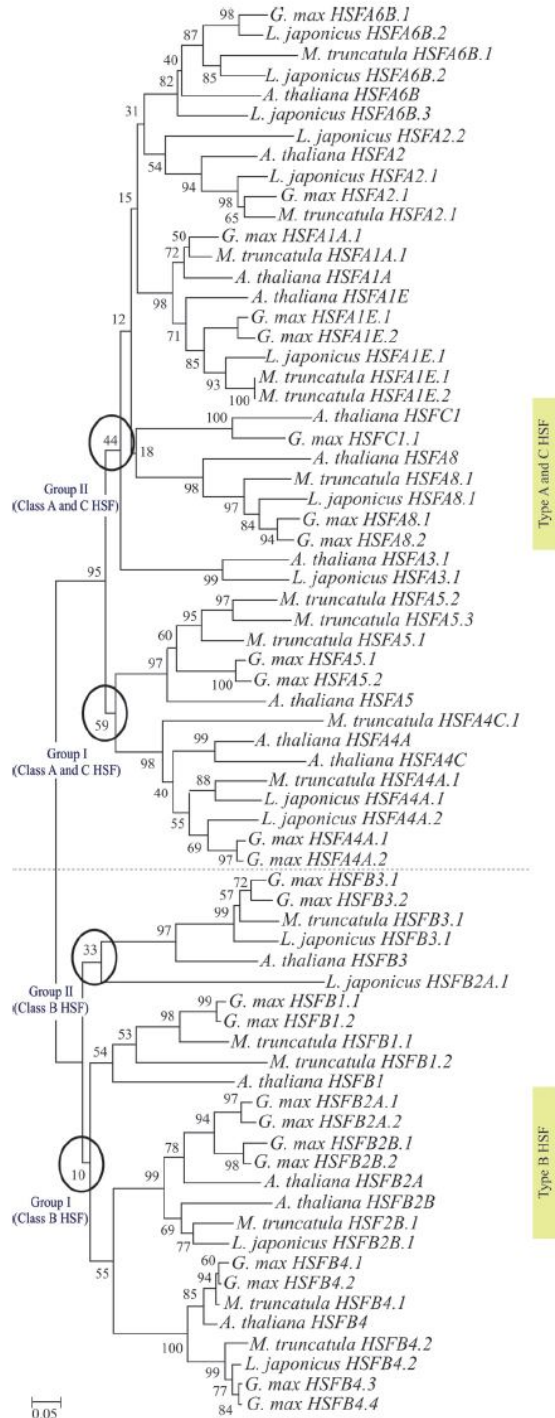


Figure 2 - Dendrogram showing relationships among HSF proteins of *A. thaliana*, *G. max*, *L. japonicus* and *M. truncatula*. Numbers on the base of the nodes represent bootstrap values. Dotted line divides Type A and C from type B. Gray circles indicate each node shared by HSF groups: 'Group I (Class A and C HSFs)' = HSFA4 + HSFA5, 'Group II (Class A and C HSFs)' = HSFA1 + HSFA3 + HSFA6 + HSFA8 + HSFC, 'Group I (Class B HSF)' = HSFB1 + HSFB2 + HSFB4 and 'Group II (Class B HSF)' = HSFB3 + *L. japonicus* HSFB2A.1. Bar represents similarity coefficient.

these two genes. In fact, in the multiple alignment analysis, two regions comprising 15 residues each (amino acid positions 125 to 139 and 154 to 168) were shared by both HSFA8 and HSFC protein sequences, though absent in other class A HSF members. Furthermore, peculiarities shared by HSFCs, such as deletions of six amino acids at position 106-111 and probable mutations in two segments (intervals: 161-168 and 195-220) may justify the differentiation of class C proteins from class A ones, as evidenced in the dendrogram.

Regarding the specific function of class C, remarkable little information is currently available. According to Nover *et al.* (2001), HSFCs were well represented in expressed sequence tags (ESTs) from libraries of tomato, soybean, potato, barley and *Arabidopsis*. The HSFC type is clearly separated from all others by sequence details of the DBDs and by the characteristics of the HR-A/B region. However the significance of these extended oligomerization domains in class A and C HSFs for the coiled-coil structure and oligomerization behavior is not yet clear (Nover *et al.*, 2001).

We denoted a conservation in the position and function of AHA motifs and NES in the C-terminal regions of class A. These regions, in addition to the flanking amino acid residues, were sufficient to identify the HSFs without prior knowledge about the respective DBDs or HR-A/B regions (Kotak *et al.*, 2004). Furthermore, the results were positive for ESTs encoding representatives of HSF groups A1, A2 and A6 (Kotak *et al.*, 2004). Thus, it can be inferred that the observed grouping formed by HSFA1, HSFA2 and HSFA6B in the dendrogram (Figure 2) was based on the similarity of AHA motifs and NES in the C-terminal regions.

It is noteworthy that the C-terminal domains (CTDs) of class B HSFs are completely different (Nover *et al.*, 2001), justifying their isolation in a separate branch, composed of two main groups. The first one includes the B3 sub-class members together with a single member of the B2 sub-class from *L. japonicus*. This unexpected grouping of the *Lotus* B2 sub-class member seems to result from a deletion in a region rich in alanine, valine, isoleucine and methionine. Apparently, this deletion was responsible for the exclusion of this sequence from the branch including the remaining class B members. The second group includes B1, B2 and B4 sub-classes, these being separated in different branches according to their sub-classes (Figure 2). This grouping may be explained by differences observed in a cluster containing arginine and lysine residues close to the C-terminus of HSFB1, probably responsible for permanent nuclear localization (Heerklotz *et al.*, 2001) and also by the fact that similar motifs were found in other representatives of this group and also in groups B2 and B4 (with the exception of the HSFB3 sub-class) which is the smallest of all HSFs identified so far.

Although our knowledge is still limited, functional diversification seems to be the main reason for the coexis-

tence of more than 20 HSF types in plants (Baniwal *et al.*, 2007). A systems analysis of tomato HSFs revealed two interesting peculiarities: (i) there are at least four different HSF groups (Scharf *et al.*, 1990, 1993; Treuter *et al.*, 1993; Bharti *et al.*, 2000) belonging to two classes (*i.e.*, class A with HSFs A1, A2, and A3 and class B with HSFB1), and (ii) two of the four HSFs (HSFA2 and B1) are heat stress-inducible proteins (Nover *et al.*, 2001; Kotak *et al.*, 2004). In most cases, all identified gene classes and sub-classes were expressed and identified in the four evaluated legumes, suggesting that the family members diverged before the species differentiated. Alternatively, such gene classes and sub-classes may have already functioned as independent genes in the common ancestor, thus favoring divergent evolution.

HSF expression in soybean

Plant cells constitutively express a pool of HSF proteins that are maintained in an inactive state. Certain results suggest that heat-induced protein denaturation participates in the activation of these HSFs (Yamada *et al.*, 2007). This molecular device is normally based on changes in protein conformation and can respond very quickly, playing therefore a central role in transcriptomic remodeling induced upon heat exposure. Accordingly, all HSFs expressed in soybean identified in this study were derived from experiments in the absence of heat stress.

Moreover, it is well known that heat often occurs in combination with drought or other stresses that cause extensive agricultural losses worldwide. HSFs serve as the terminal components of signal transduction, mediating the expression of HSPs and other HS-induced transcripts, but their diverse temporal and spatial expression has also been demonstrated under the influence of other abiotic stresses (Kotak *et al.*, 2007).

HSFs are involved in stress sensing and signaling but can also be part in the regulation of other cellular processes, including development, where a role is strongly suggested by expression profiles in libraries of tissues from young stages. The only exceptions seen herein were mature adult and drought-stressed leaves where the expression of HSFB1 and HSFB2A1 was diametrically and remarkably down- and up-regulated, respectively (Figure 3).

Plant HSFs may also function as H₂O₂ sensors, as is also the case in humans and *Drosophila*, where HSFs directly sense H₂O₂ and assemble into homotrimers in a redox-regulated manner. HSFA2 controls expression under prolonged HS and recovery conditions. Interestingly, its expression is induced by high luminosity and exposition to H₂O₂, emphasizing its importance under various stress conditions (Miller and Mittler, 2006). HSFA4A and HSFA8 are likely to act as sensors of reactive oxygen species (ROS), with HSFA5 acting as a repressor of HSFA4. Indeed, in soybean the profiles of HSFA4A and HSFA8 were quite similar, considering the number of libraries where

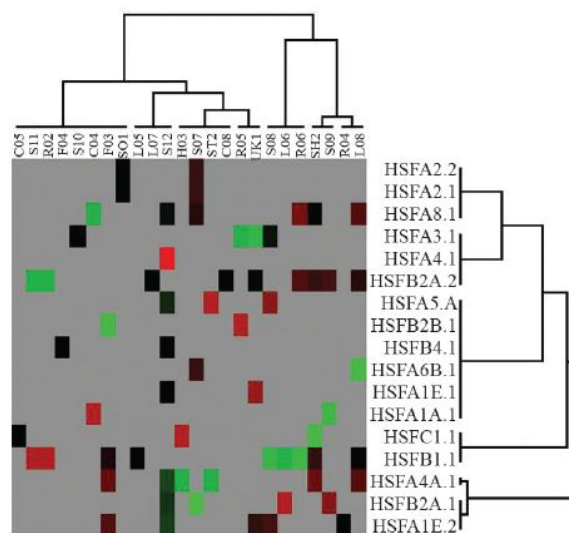


Figure 3 - Hierarchical clustering (Cluster3.0) of up-regulated (red), down-regulated (green) and non-regulated (black) soybean EST clusters ($p < 0.05$) related to HS response; gray stands for absence of information. Dendrograms above and to the left of the graph show the relationships among libraries and expressed genes, respectively. Library codes: C04, immature cotyledons of greenhouse grown plants; C05, 8-day-old cotyledons; C08, 3- and 7-day-old cotyledons; F03, mature flowers of field grown plants; F04, floral meristem; H03, hypocotyl and plumule, germinating seeds; L05, unexpanded leaves and shoot tips of 2-week-old seedlings; L06, drought stressed leaf tissue; L07, leaf, 3-week-old, greenhouse grown; L08, leaf; R02, roots of 7-day-old plants; R04, roots of bulked times; R05, roots of 8-day-old seedlings; R06, root; S07, seed coats of greenhouse grown plants; S08, 11-day-old seedlings; S09, whole seedlings of greenhouse grown plants; S10, seedlings; S11, seedlings, minus the cotyledons; S12, Seeds containing globular-stage embryos; SH2, germinating shoots; SO1, *in vitro* cultivated somatic embryos; ST2, stem tissue of greenhouse grown plants.

they were detected. On the other hand, and considering the same libraries, HSFA5 was absent, except in immature seeds containing globular embryo stages where none of the three genes were detectably expressed (Figure 3). It is also interesting to note that HSFB1.1 was up-regulated in seven-day-old root libraries (R02) and in seedlings (without cotyledons) (S11), situations in which HSFB2A.2 was down-regulated, indicating that these genes may act as antagonists during the initial phases of plant development. This assumption is corroborated by the fact that HSFB1.1 was down-regulated, while HSFB2A.2 was up-regulated in the mature root library (L08).

The similarity in expression patterns of HSF genes in specific libraries (in specific developmental stages or conditions) indicates that the activation of these genes might be evoked by the same *cis*-regulatory elements in their promoters. Such co-expression was observed for HSFA2.1, HSFA2.2, HSFA6B.1 and HSFA4A.1 in the library S07 from 'seed coats of greenhouse grown plants'. Co-expression could indicate that these genes play the same role or are co-participants in the same pathway.

The induction of transcriptomic remodeling through the HSF network is very important but complex, as it in-

volves several HSFs. This network is only a part of the orchestration that contributes to survival under high temperature stress. The panel exposed by our work suggests that HSFs also mediate cross-talk between signaling cascades in soybean for HS and other abiotic stresses, with possible roles in soybean development. Nevertheless, the questions raised here may have to be addressed in subsequent experiments in which the tissues and conditions should be pooled for different and sequential time points.

Distribution of HSF genes in the soybean genome

The comparative analysis of *G. max* EST sequences (25 in total) and genomic sequences enabled the identification of 62 loci bearing HSF genes (Table 3; Figure 4) from 65 HSFs previously described for soybean (Mochida *et al.*, 2009a), a crop with a supposed polyploid recent past (McClellan *et al.*, 2010). From the 25 obtained candidates, two did not align significantly with the characterized heat

Table 3 - Correspondence among identified GENOSOJA expressed sequence tags (ESTs) and characterized genes of *Glycine max*. Abbreviations: ID = identification; HSF, Heat Shock Factor; Gm, *Glycine max*.

Genosojia ID	Gene ID	Genosojia ID	Gene ID
Gm_HSFA1A.1	Glyma01g01990	Gm_HSFA1A.4	Glyma10g07620
Gm_HSFB1.1	Glyma01g22910	Gm_HSFA2.3	Glyma10g09460
Gm_HSFB3.1	Glyma01g34490	Gm_HSFA5.2	Glyma10g36910
Gm_HSFB4.1	Glyma01g34490	Gm_HSFB2B.1	Glyma10g38240
Gm_HSFB1.2	Glyma01g39260	Gm_HSFA1E.2	Glyma10g38930
Gm_HSFB2A.1	Glyma01g42640	Gm_HSFA1E.3	Glyma11g01190
Gm_HSFA1E.1	Glyma01g44330	Gm_HSFB2A.2	Glyma11g02800
Gm_HSFB4.2	Glyma02g44670	Gm_HSFB1.5	Glyma11g06010
Gm_HSFB3.2	Glyma03g29190	Gm_HSFA2.4	Glyma11g33630
Gm_HSFA1A.2	Glyma03g31380	Gm_HSFA1A.5	Glyma13g16510
Gm_HSFA6B.1	Glyma03g31380	Gm_HSFA1A.6	Glyma13g21490
Gm_HSFA7A.1	Glyma03g31380	Gm_HSFB1.6	Glyma13g24860
Gm_HSFA1A.3	Glyma03g34900	Gm_HSFA4A.3	Glyma13g29760
Gm_HSFB4.3	Glyma04g04200	Gm_HSFB4.7	Glyma14g04070
Gm_HSFB4.4	Glyma04g04200	Gm_HSFB4.8	Glyma14g09190
Gm_HSFA2.1	Glyma04g05500	Gm_HSFA1A.7	Glyma14g11030
Gm_HSFA2.2	Glyma04g05500	Gm_HSFA4A.4	Glyma15g09280
Gm_HSFB1.3	Glyma05g20460	Gm_HSFA1A.8	Glyma16g13400
Gm_HSFB1.4	Glyma05g20460	Gm_HSFA1E.4	Glyma16g19500
Gm_HSFA5.1	Glyma05g28460	Gm_HSFB3.3	Glyma16g29750
Gm_HSFA4A.1	Glyma05g28460	Gm_HSFB2A.3	Glyma16g32070
Gm_HSFA4A.2	Glyma05g29470	Gm_HSFA1A.9	Glyma17g06160
Gm_HSFA8.1	Glyma05g34450	Gm_HSFB1.7	Glyma17g20070
Gm_HSFB4.5	Glyma06g04390	Gm_HSFA1A.10	Glyma17g34540
Gm_HSFC1.1	Glyma07g09510	Gm_HSFB4.9	Glyma17g35980
Gm_HSFC1.2	Glyma07g09520	Gm_HSFB2A.4	Glyma18g14700
Gm_HSFB4.6	Glyma07g36370	Gm_HSFA1E.5	Glyma19g26460
Gm_HSFA8.2	Glyma08g05220	Gm_HSFA5.4	Glyma19g26750
Gm_HSFA5.3	Glyma08g11460	Gm_HSFB3.4	Glyma19g31940
Gm_HSFA4A.5	Glyma08g12630	Gm_HSFB3.5	Glyma19g31940
Gm_HSFB2A.5	Glyma09g26510	Gm_HSFA6B.3	Glyma19g34210
Gm_HSFB2A.6	Glyma09g26510	Gm_HSFA8.3	Glyma19g37580
Gm_HSFC1.3	Glyma09g32300	Gm_HSFB4.10	Glyma20g08250
Gm_HSFA1A.11	Glyma09g33920	Gm_HSFA1A.13	Glyma20g28870
Gm_HSFA1A.12	Glyma10g00560	Gm_HSFB2B.2	Glyma20g29610
Gm_HSFA6B.2	Glyma10g03530	-	-

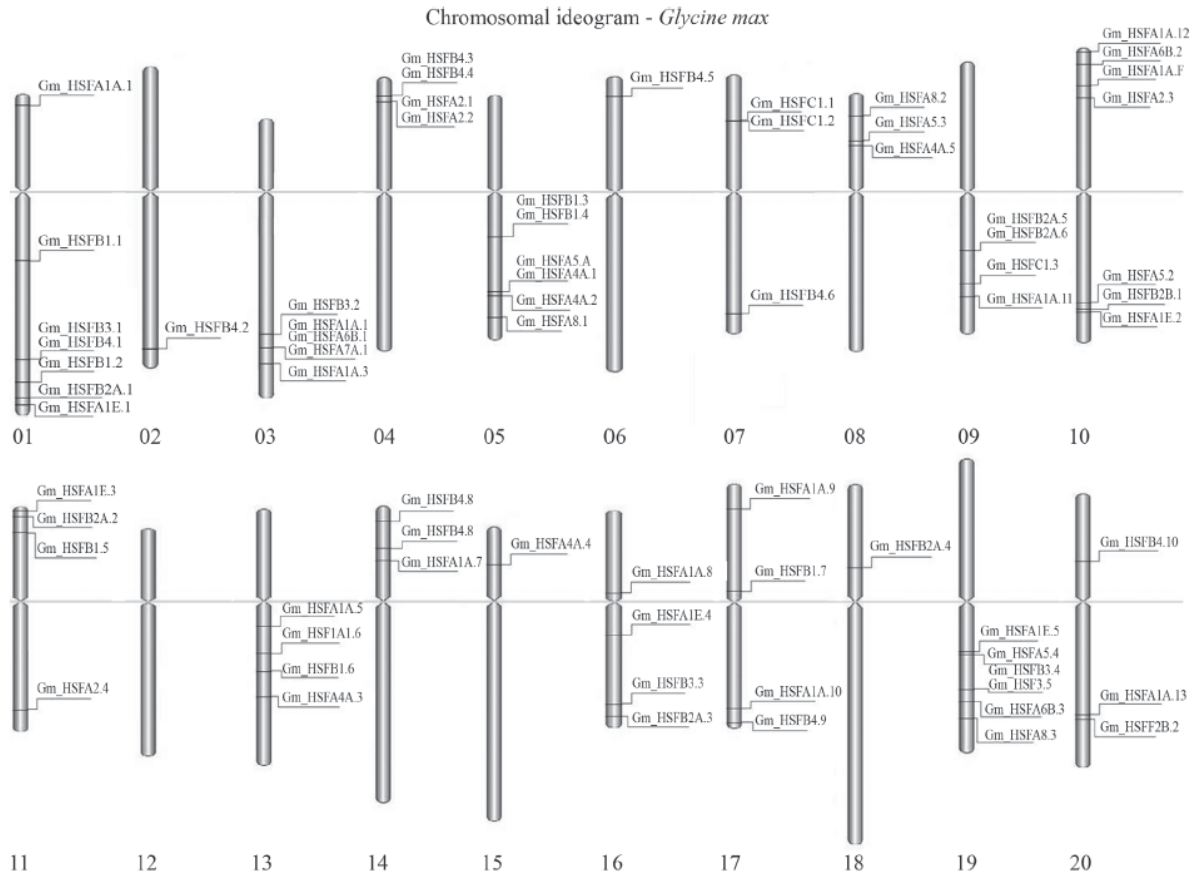


Figure 4 - *In silico* hybridization of HSF sequences against the SOYBASE database. Schematic representation of clusters that were anchored in soybean based on BLAST similarity results (see Table 3 for correspondence between EST cluster identification and HSF described genes).

shock factor genes, which can be justified by differences in the cultivars used in genomic and expression sequencing projects. In addition, three described genes for soybean were not identified among the EST sequences, indicating a lack of expression of these genes in the libraries of the GENOSOJA database. Differences among the analyzed cultivars may also explain this lack of similarity.

With respect to the genomic distribution of the HSF family, nine gene clusters could be identified in chromosomes 01, 03, 04, 05, 10, 11 and 19 (Figure 4). According to Mochida *et al.* (2009a) these clusters may consist of paralogous genes. In soybean, the relative physical distribution of transcription factor genes is of interest, and two types of clusters can be distinguished based on their evolutionary history. The first type consists of a series of genes that arose through repeated tandem duplications (originated from a founding locus). The second type, which is not considered as consisting of paralogous genes, probably arose independently and then relocated to form these duplications and clusters (Mochida *et al.*, 2009b). Pairs of duplicated genes on different chromosomes are common and gene clusters of three or more highly related genes are also widely found

(Mochida *et al.*, 2009a). Considering the distance of their occurrence, a few of the duplicated genes could be classified arbitrarily as either genes that were not duplicated in tandem on the same chromosome, or genes that were so (Mochida *et al.*, 2009a).

Moreover, none of the EST clusters aligned on chromosome 12. This was expected, since in this chromosome there is no description of HSF family members (Mochida *et al.*, 2009b), while the other chromosomes (02, 06, 15 and 18) presented a single representative of the group.

Concluding Remarks

Results from the present investigation indicate that gene duplication and diversification occurred during plant evolution, whilst differences in their expression patterns caused species-specific variability in the composition of the HSF family members, which can be divided into three different classes and several sub-classes according to their particular motifs and residue-specific rich regions. Although not all of the previously described genes could be for the three species studied when using a transcriptomic approach, we expect that experiments directed at heat-

stress conditions may provide additional sequences related to the HS response, including other HSF genes. Furthermore, the absence of soybean ESTs for some HSF members did not impair the evaluation of the distribution of the HSF family in the soybean genome. The family is present in 19 of the 20 chromosomes, including clustered distribution in some.

To understand the complexity of a plant's HSF family and stress response systems in general, it is important to consider that when plants became adapted to terrestrial habitats they evidently had to face and become specialized to rapidly changing and extreme environmental conditions. The present approach represents the first evaluation considering only expressed HSF genes, revealing 25 expressed ESTs and 68 SuperSAGE tags, with emphasis on root tissue (water deficit) libraries. Some HSF candidates present in *Arabidopsis*, that are apparently missing in the transcriptome of the evaluated legumes (for example HSFA1B), may be important candidates for biotechnological approaches in soybean and other legumes directed towards increasing their performance under temperature stress conditions. Moreover, some genes found to be induced under water deficit may constitute interesting target genes for inferences regarding the association of heat and cold stresses, especially considering current climate change scenarios.

References

- Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410.
- Baniwal SK, Chan KY, Scharf K-D and Nover L (2007) Role of heat stress transcription factor HsfA5 as specific repressor of HsfA4*. *J Biol Chem* 282:3605-3613.
- Bharti K, Schmidt E, Lyck R, Bublak D and Scharf K-D (2000) Isolation and characterization of HsfA3, a new heat stress transcription factor of *Lycopersicon peruvianum*. *Plant J* 22:355-365.
- Bharti K, von Koskull-Döring P, Bharti S, Kumar P, Tintschl-Körbitzer A, Treuter E and Nover L (2004) Tomato heat stress transcription factor HsfB1 represents a novel type of general transcription coactivator with a histone-like motif interacting with HAC1/CBP. *Plant Cell* 16:1521-1535.
- Efeoglu B (2009) Heat shock proteins and heat shock response in plants. *G U J Sci* 22:67-75.
- Eisen MB, Spellman PT, Brown PO and Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863-14868.
- Fehr WR, Caviness CE, Burmood DT and Pennington IS (1971) Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. *Crop Sci* 11:929-931.
- Fehr WR and Caviness CE (1977) Stage of Soybean Development. Special Report n. 80. Ames, Iowa State University of Science and Technology, Iowa, 12 pp.
- Glombitza S, Dubuis P-H, Thulke O, Welzl G, Bovet L, Götz M, Affenzeller M, Geist B, Hehn A, Asnaghi C, *et al.* (2004) Crosstalk and differential response to abiotic and biotic stressors reflected at the transcriptional level of effector genes from secondary metabolism. *Plant Mol Biol* 54:817-835.
- Heerklotz D, Döring P, Bonzelius F, Winkelhaus S and Nover L (2001) The balance of nuclear import and export determines the intracellular distribution and function of tomato heat stress transcription factor HsfA2. *Mol Cell Biol* 21:1759-1768.
- Hoagland D and Arnon DI (1950) The water culture method for growing plants without soil. *Calif Agric Exp Stn Circ* 347:1-32.
- Hsu S-F, Lai H-C and Jinn T-L (2010) Cytosol-localized heat shock factor-binding protein, AtHSBP, functions as a negative regulator of heat shock response by translocation to the nucleus and is required for seed development in *Arabidopsis*. *Plant Physiol* 153:773-784.
- Hu W, Hu G and Han B (2009) Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. *Plant Sci* 176:583-590.
- Kido EA, Barbosa PK, Ferreira Neto JCR, Pandolfi V, Houllou-Kido LM, Crovella S and Benko-Iseppon AM (2011) Identification of plant protein kinases in response to abiotic and biotic stresses using SuperSAGE. *Curr Prot Pept Sci* 12:643-656.
- Kotak S, Port M, Ganguli A, Bicker F and von Koskull-Döring P (2004) Characterization of C-terminal domains of *Arabidopsis* heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization. *Plant J* 39:98-112.
- Kotak S, Larkindale J, Lee U, von Koskull-Döring P, Vierling E and Scharf KD (2007) Complexity of the heat stress response in plants. *Curr Opin Plant Biol* 10:310-316.
- Li H-Y, Chang C-S, Lu L-S, Liu C-A, Chan M-T and Charng Y-Y (2004) Over-expression of *Arabidopsis thaliana* heat shock factor gene (*AtHsfA1b*) enhances chilling tolerance in transgenic tomato. *Bot Bull Acad Sin* 44:129-140.
- Li M, Berendzen KW and Schoffl F (2010) Promoter specificity and interactions between early and late *Arabidopsis* heat shock factors. *Plant Mol Biol* 73:559-567.
- McClellan PE, Mamidi S, McConnell M, Chikara S and Lee R (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics* 11:e184.
- Miller G and Mittler R (2006) Could heat shock transcription factors function as hydrogen peroxide sensors in plant? *Ann Bot* 98:279-288.
- Mittal D, Chakrabarti S, Sarkar A, Singh A and Grover A (2009) Heat shock factor gene family in rice: Genomic organization and transcript expression profiling in response to high temperature, low temperature and oxidative stresses. *Plant Physiol Biochem* 47:785-95.
- Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K and Tran L-SP (2009a) *In silico* analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. *DNA Res* 16:353-369.
- Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K and Tran L-SP (2009b) LegumeTFDB: An integrative database of *Glycine max*, *Lotus japonicus* and *Medicago truncatula* transcription factors. *Bioinformatics* 26:290-291.

- Nascimento LC, Costa GGL, Binneck E, Pereira GAG and Caraz-
zolle MF (2012) A web-based bioinformatics interface ap-
plied to Genosojá Project: Databases and pipelines. *Genet*
Mol Biol 35(suppl 1): 203-211.
- Nover L, Bharti K, Doring P, Mishra SK, Ganguli A and Scharf
K-D (2001) *Arabidopsis* and the heat stress transcription
factor world: How many heat stress transcription factors do
we need? *Cell Stress Chap* 6:177-189.
- Pirkkala L, Nykanen I and Sistonen L (2001) Roles of the heat
shock transcription factors in regulation of the heat shock re-
sponse and beyond. *FASEB J* 15:1118-1131.
- Ruelland E and Zachowski A (2010) How plants sense tempera-
ture. *Environ Exp Bot* 69:225-232.
- Sato Y and Yokoya S (2008) Enhanced tolerance to drought stress
in transgenic rice plants overexpressing a small heat-shock
protein, sHSP17.7. *Plant Cell Rep* 27:329-334.
- Scharf K-D, Rose S, Thierfelder J and Nover L (1993) Two
cDNAs for tomato heat stress transcription factors. *Plant*
Physiol 102:1355-1356.
- Scharf K-D, Rose S, Zott W, Schoffl F and Nover L (1990) Three
tomato genes code for heat stress transcription factors with a
region of remarkable homology to the DNA-binding domain
of the yeast HSF. *EMBO J* 9:4495-4501.
- Schöff F, Prändl R and Reindl A (1998) Regulation of the heat-
shock response. *Plant Physiol* 117:1135-1141.
- Sung D-Y, Kaplan F, Lee K-J and Guy CL (2003) Acquired toler-
ance to temperature extremes. *Trends Plant Sci* 8:179-187.
- Swindell WR, Huebner M and Weber AP (2007) Transcriptional
profiling of *Arabidopsis* heat shock proteins and transcrip-
tion factors reveals extensive overlap between heat and
non-heat stress response pathways. *BMC Genomics* 8:e125.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar
S (2011) MEGA5: Molecular Evolutionary Genetics Analy-
sis using maximum likelihood, evolutionary distance, and
maximum parsimony methods. *Mol Biol Evol* 28:2731-
2739.
- Treshow M (1970) *Environment and Plant Response*. McGraw-
Hill Company, New York, 421 pp.
- Treuter E, Nover L, Ohme K and Scharf K-D (1993) Promoter
specificity and deletion analysis of three tomato heat stress
transcription factors. *Mol Gen Genet* 240:113-125.
- Yamada K, Fukao Y, Hayashi M, Fukazawa M, Suzuki I and
Nishimura M (2007) Cytosolic HSP90 regulated the heat
shock response that is responsible for heat acclimation in
Arabidopsis thaliana. *J Biol Chem* 282:37794-37804.

Internet Resources

- AtTFDB - Arabidopsis transcription factor database at Agris,
<http://arabidopsis.med.ohio-state.edu/AtTFDB/> (Jun/2011).
- GENOSOJA database, Brazilian Soybean Genome Consortium,
<http://bioinfo03.ibi.unicamp.br/soja/> (Jun/2011).
- Lotus japonicas* database, <http://www.kazusa.or.jp/lotus/>
(Jun/2011).
- Medicago truncatula* database, <http://www.medicago.org/>
(Jun/2011).
- Phytosome Soybean genome browser,
<http://www.phytozome.net/cgi-bin/gbrowse/soybean/>
(Jun/2011).
- SOYBASE site,
[http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01/#se
arch](http://soybase.org/gbrowse/cgi-bin/gbrowse/gmax1.01/#search) (Jun/2011).
- Legume Transcription Factor Database, LegumeTFDB,
<http://legumetfdb.psc.riken.jp/> (Jun/2011).

License information: This is an open-access article distributed under the terms of the
Creative Commons Attribution License, which permits unrestricted use, distribution, and
reproduction in any medium, provided the original work is properly cited.

Differential gene expression in the prospection of tolerance genes against drought and salinity in native and cultivated plants

Nina da Mota Soares-Cavalcanti¹, Luis Carlos Belarmino¹, João Pacífico Bezerra-Neto¹, Ana Carolina Wanderley-Nogueira¹, Rafaela Cavalcanti Lira¹, Hayana M. Arruda Azevedo¹, Ederson A. Kido¹, Valesca Pandolfi², Ana Maria Benko-Iseppon^{1*}

¹Genetics Department, Universidade Federal de Pernambuco, Recife, PE, Brazil.

*Corresponding author: Ana Maria Benko Iseppon. Laboratory for Plant Genetics and Biotechnology. Av. Prof. Moraes Rego 1235, CEP 50.670-420, Recife, PE, Brazil. Phone +55 (81) 21268520. E-mail: ana.benko.iseppon@pesquisador.cnpq.br

ABSTRACT. The present work discusses the impact of gene discovery for the breeding and biotechnological purposes regarding plants adapted to drought and salinity on agricultural production in both the semi-arid and humid regions of the world. Important gene candidates may be recognized in both native and cultivated plants. The present work brings a general review of important gene families for this purpose, bringing an overview regarding their prevalence and diversity in both model (arabidopsis and rice) and non model plant species, with emphasis on legumes, cowpea and soybean. We also bring a mini-review presenting mainly biotechnological strategies to minimize the impact of drought and salinity using candidates from the presented gene families. In cowpea 1,010 candidates from 27 categories associated to abiotic stresses have been found, including transcription factors (TF), transport protein ion channel, plant defense, among other. With respect to TF, DREB/ERF and Zinc Finger were the most representative families. Moreover expression profile from the TF results indicated high expression in root tissues. Evaluations carried out with SOS pathway members showed the conservation of this pathway both in cowpea and soybean. Regarding specific efforts for underprivileged regions as the Brazilian Northeastern region, there are still scarce efforts, especially considering that extensive areas are deteriorating, representing a serious threat to food security.

Key words: Target genes; Abiotic stresses; Leguminous; Transcriptome

Running title: Abiotic stress induced genes in plants

Introduction

Drought and salinity as agricultural scenarios

Annual losses estimated at billions of dollars have been increased due to climate changes and global warming pointing to prolonged drought and heat episodes; these episodes represent a serious challenge for agricultural production worldwide that affects plant growth and yield. Reports about the European heatwave of 2003 estimated losses in the crop production by around 30% (Ciais et al., 2005). Abiotic stress is a major limiting factor for plant growth and food production in many regions of the world and its effects increase as desertification claims more of the world's arable land (Hayano-Kanashiro et al., 2009).

Many reports have proved that water is one of the key factors influencing plant production, being essential for photosynthetic reactions and to finish the accumulation of photosynthetic products, which are impacted by physiological pathways and environmental factors (Yi and Hong-Bo, 2008). Loss of water in the soil leads to great reduction in crop production, which has been reflected by the total grain yield of many countries in the world. Moreover, the continued depletion of ground water level and the increased demands for irrigation have led to the salinization of arable lands. The geographical distribution of most highly salinity-sensitive plants including legumes is delimited by high NaCl concentrations in soil, which cause salinity stress. For example, salinity is no less an important constraint for chickpea yield reduction (Varshney et al., 2009).

The relationship between plants and environment involves many processes, which are the essential bases for sustainable agricultural development. As cited by Yi and Hong-Bo (2008) the adequate combination of plant species and the corresponding environment (in particular the soil) is the key problem to modern sustainable agriculture development under global climate changes. In this context, drought and salinity are the main abiotic stresses and their directly effects on plants are being extensively studied. Also, an urgent need exists to develop drought and salinity-tolerant varieties either by conventional breeding or by genetic engineering in order to cope with the abiotic stress injuries.

CONSEQUENCES OF DROUGHT AND SALINITY FOR PLANT GROWTH

During their long period of evolution and artificial domestication plants have evolved and acquired a myriad of developmental and metabolic strategies to optimize water uptake and ionic control for efficiently balance this processes during vegetative growth and reproduction, making drought and salinity tolerance a complex multigenic trait (Hayano-Kanashiro et al., 2009). The earliest responses to drought that protect the plant from extensive water loss and toxic ion accumulation include reduction in vegetative growth, stomatal closure and a decrease in the photosynthetic rate. Physiological studies have shown that sugars, sugar alcohols, amino acids and amines function as osmolytes, protecting cellular functions from the effects of dehydration and are known to accumulate under drought and salinity stresses conditions in different plant species (Hayano-Kanashiro et al., 2009).

Moreover, water deficit and excess of salt in the soil influence solute transport (ion and nutrient uptake) to large extent, that are reflected on photosynthetic reactions in plant chloroplasts (Yi and Hong-Bo, 2008). Regulatory network systems based on the response to crop developmental cues regulate and control spatially and temporally the responsive reactions and processes of the plants to environmental changes. Such network systems have been formed by the plant-biome interaction for a long time in the

course of evolution, and it will continue to evolve with environmental succession (Yi and Hong-Bo, 2008).

Responsive molecular components include environmental signal recognition (input), signal transduction (cascade of biochemical reactions), signal output, signal responses and phenotype realization, which is a multi-dimension network system with many levels of gene expression and regulation (Yi and Hong-Bo, 2008). In this way, aiming to increase their resistance to stress, plants can sense, process, respond to environmental stress and activate related-gene expression. Environmental stress-inducible genes can be divided into two types in terms of their protein products: (i) the first type include coding products that directly confer the function of crop cells to resist to environmental stress, such as LEA protein, anti-freezing protein, osmotic regulatory protein, enzymes for synthesizing betaine, proline and other osmoregulators; the second type (ii) comprehend products that play an important role in regulating gene expression and signal transduction, such as MAP protein kinases and transcription factors, as bZIP and MYB, among others.

Gene expression profile under abiotic stresses: drought and salinity

The ability of cells to respond to external stimuli in a timely and specific manner is a recurring theme in molecular biology. Seki et al. (2002) prepared a full-length cDNA microarray containing 7,000 genes to evaluate their expression profiles under drought, cold and high-salinity conditions. The transcripts of 53, 277 and 194 genes increased after cold, drought and high-salinity treatments, respectively, more than fivefold when compared with the control. Moreover, a strong relationship between the expression of these stress-responsive genes was observed; 22 stress-inducible genes responded to all three stresses. As arabidopsis (Seki et al., 2002) and rice (Rabbani et al., 2003) had their whole-genome sequence released earlier, a number of studies focus on gene expression patterns in response to salt stress were monitored in these species. Based on those studies, it is known that a number of genes show dynamic expression changes at a genomic-level scale while response to salt stress (Li et al., 2009). Taji et al. (2004) noted that a lower number of genes was induced by 250 mM NaCl stress in *Thellungiella halophila* (salt cress, a salt-tolerant relative of arabidopsis), in contrast to arabidopsis, indicating that the stress tolerance of salt cress may be due to the constitutive overexpression of many genes that function in stress tolerance and are only stress inducible in arabidopsis. A larger spectrum of gene expression changes was noted between *Atnhx1* knockout mutant (Apse et al. 2003) and wild type plants (Sottosanto et al., 2004), showing that salt-sensitive and salt-tolerant arabidopsis phenotypes differ markedly in expression of their genetic machinery.

To study the molecular mechanisms of drought responses in tomato, gene expression profiles of two drought-tolerant lines and one drought-sensitive cultivar were investigated under drought stress using microarray experiments. Around 400 genes identified were responsive to drought stress only in the drought-tolerant lines; among these genes a number of transcription factors and signaling proteins was included, which could be global regulators of the response to drought stress (Gong et al., 2010). Moreover, monitoring the changes in gene expression at the transcriptional level in leaves of two drought-tolerant and one drought-sensitive barley genotypes during the reproductive stage under drought, Guo et al. (2009) observed that 17 genes were expressed exclusively in the drought-tolerant genotypes under drought, including genes controlling stomatal closure via carbon metabolism, or synthesis of the osmoprotectant glycine-betaine, generating protectants against reactive-oxygen species scavenging, and stabilizing membranes and proteins.

Integrated data obtained with various 'omics' approaches and research on stress responses in various plant species have increased the knowledge regarding the mechanisms of plant stress tolerance in nature

and provided a more comprehensive picture of abiotic stress responses. Based on this progress, improvements in crop stress tolerance have been attempted by means of gene transfer and marker-assisted breeding (Hirayama and Shinozaki, 2010).

Databanks and bioinformatics

Identification of stress-responsive genes can be undertaken by expression essays, a type of evaluation carried out for many species in the last decade, with emphasis on data like expressed sequence tags (ESTs) from stress-challenged tissues. EST sequencing projects have been contributing to gene discovery and marker development as simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs), as well as providing insights into the complexities of gene expression patterns in several crop species (Varshney et al., 2009).

The largest coverage and a wide dynamic range of gene expression information are provided by transcript profiling and can often be performed genome wide. Microarrays are popular for transcript profiling and can be readily afforded by many laboratories. Various commercial and academic microarray platforms exist that vary in genome coverage, availability, specificity, and sensitivity. Serial analysis of gene expression (SAGE) and massively parallel signature sequencing (MPSS) are well-established alternatives to microarrays and can also be superior because they do not depend on prior probe selection. More recently, direct sequencing of transcripts by high-throughput sequencing technologies (RNA-Seq) has become an additional alternative to microarrays and is superseding SAGE and MPSS (Busch and Lohmann, 2007). RNA-Seq does not depend on genome annotation for prior probe selection and avoids biases introduced by the mentioned methods. On the other hand, RNA-Seq poses novel algorithmic and logistic challenges, and current wet-lab RNA-Seq strategies require lengthy library preparation procedures.

A major challenge for genome-scale transcription analysis regards the large number of predictors (genes) compared to a small number of measurements (microarrays). Without appropriate statistical measures to correct for multiple testing and including false discovery rates, almost any approach will yield significant genes, including many false positives. The creation of large databases in recent years has brought an additional layer of complexity and precautions to take. For example, large databases such as Genevestigator (HRUZ et al., 2008) not only profile a large number of genes, but also allow contextual meta-analysis of several hundred of conditions, each of which is covered by only a small number of replicates (usually 3–5). While some genes respond to a small number of conditions and therefore their expression is easier to contextualize and interpret, other genes respond to dozens or hundreds of specific conditions (Baginsky et al., 2010). On the other hand, cDNA sequencing projects examine specifically the genes expressed under different conditions. Using cellular mRNA as template the cDNA is synthesized allowing sampling of the pool of active genes in a specific cellular condition. The fragments of sequenced cDNA are called ESTs (Expressed Sequence Tags). The ESTs are often assembled to obtain the current set of full sequence of cDNAs, to identify expression patterns from different studied conditions and to improve the length and the quality of the EST (Carazzolle et al., 2007).

Databases have to be constructed to store and analyze the high volume of transcripts, providing data organization and also the tools for their analysis. Although there are many databases with restricted access, the recent tendency is that they become publicly available, as it is the case of the GenBank, further increasing their availability for research. GenBank, a public database built by the National Center for Biotechnology Information (NCBI) is a comprehensive databank, containing, for example, nucleotide sequences for more than 380,000 organisms, obtained primarily through submissions from

individual laboratories and batch submissions from large-scale sequencing projects, including ESTs (Expressed Sequence Tags) and whole genome shotgun (WGS) (Benson et al., 2011).

Regarding to Brazilian databases, recent efforts have resulted in the development of genomic and transcriptomic data banks, including important plant transcriptome projects: the Sugarcane Expressed Sequence Tag Project (SUCEST), Soybean Genome Project (GENOSOJA) and Cowpea Transcriptome Project (NordEST). The first project regarded the SUCEST (Sugarcane Expressed Sequence Tags; <http://watson.fapesp.br/sucest.htm>) database, established in 1999 and launched by the ONSA network (Organization for Nucleotide Sequencing and Analysis). This project generated a total of 291,689 cDNA sequences from 26 libraries, organized in 43,141 clusters, from which at least 33% regard full-length cDNA sequences (Vettore et al., 2001). Currently, a large amount of data generated by SUCEST is already available at GenBank.

The NordEST network was established in 2005, including 12 Laboratories, to generate genomic data useful for cowpea breeding and biotechnology programs. Efforts include generation of 30.000 ESTs from 12 EST libraries for biotic (severe mosaic virus and potyvirus) and abiotic (salinity) stresses. Also eight SuperSAGE libraries for severe mosaic virus and drought have been generated, allowing the analysis of more than five million of transcripts (SuperTags) under different conditions. Physical chromosome maps are being developed, including microsatellites, rDNA and BAC probes, allowing the rapid transference of data among both crops for breeding purposes (Benko-Iseppon et al., 2009).

The Brazilian Soybean Genome Consortium (GENOSOJA) was established in 2009 with the goal of integrating several institutions currently working with soybean genomics in Brazil. This project aims to investigate several aspects of soybean genomics, including studies with structural and functional genomics, transcriptomics and proteomics of soybean. Most essays include the evaluation of sequences under biotic and abiotic stresses, such Asian rust, nematodes and drought (Abdelnoor et al., 2009). To evaluate such large amount of data, tools able to generate local alignments, such as BLAST algorithm (Atlschul et al., 1990), have been essential. Additionally, the data can be analyzed in a correlative view using microarray profiles of signaling cascade mutants that can be compared from a wide range of organisms, e.g. by using the Genevestigator tool (<https://www.genevestigator.com/gv/index.jsp>). Mutants with overlapping transcriptome profiles are likely to act in common signaling cascades (Pitzschke and Hirt, 2009).

Once selected genes with variable expression data are interesting for clustering analysis, such as the integrated pair of tools CLUSTER (Eisen et al., 1998) and TREEVIEW (Page, 1996), that allow the identification of genes with similar expression pattern in the condition evaluated, allowing the visualization of data in graphics and dendrograms. Spots in red represent genes induced, permitting the identification of genes linked to the same physiological process or metabolic pathway, allowing inferences regarding their roles. For each gene, the expression profile reflects its dynamic behavior during a time-course experiment or its behavior under distinct conditions. Genes with similar expression profiles are considered “co-expressed”. To identify sets of co-expressed genes from high-throughput expression data, clustering techniques are required. In addition to standard cluster algorithms, such as hierarchical clustering, K-means, and self-organization maps, more advanced algorithms are also being developed, which are specially fine-tuned for biological applications (Rombauts et al., 2003). SAGE experiments also permit differential gene expression evaluation, normally in association with programs as CLUSTER and Discovery Space (Wang et al., 2005) that are able to calculate this variation in expression, through the application of Audic and Claverie (1997) statistic test.

Together these approaches contribute to the construction of a panel of genes that are differentially expressed under abiotic stress conditions, including drought and salinity. The importance of these information are perceived when prospecting genetic breeding of agricultural crops as well as for

cultivate plants in soil with water deficit and/or salt accumulation, as those found in Brazilian northeastern.

Biotechnological approaches to overrule abiotic stresses

Plant transformation has become heavily adopted as a method for understanding the plants functioning and the improvement of specific characteristics. To obtain highly desirable attributes, modern tools of biotechnology have shown an almost inexhaustible potential. Although the aims of the development of new varieties through biotechnology are basically the same as conventional plant breeding (i.e., productivity, vigor, disease resistance, nutritional quality, etc.), it has been observed that the non-conventional breeding techniques offer additional possibilities for generating variability previously not considered for by conventional breeding (Borém, 1998).

Current studies indicate the importance of functional genomic studies and of the application of genetic engineering to introduce desirable characteristics for different individuals. In addition, there are several experiments in the literature regarding the successful genetic transformation in many plant species. For example, studies in sugarcane based on cDNA microarrays containing sugarcane ESTs, identified two members of the aquaporin family of transporters as highly associated with sugar content (Papini-Terzi et al., 2007), indicating that these genes could be used to distinguish and characterize superior varieties or develop new varieties with increased sucrose content, minimizing the osmotic challenges that demands efficient control of solute compartmentalization.

Also related to abiotic stress tolerance, the transcription factor-based technologies are likely to be a prominent part of the next generation of successful biotechnology-derived crop (Century et al., 2008). Agarwal and Jha (2010) discussed the role of ABA in stress signalling and its involvement in different regulatory systems during abiotic stress in ABA-dependent and independent manner. Since some transcription factors (TFs) follow an ABA-dependent signal transduction pathway, while others appear ABA-independent. However, they claim that the ABA-dependent and ABA-independent pathways act in parallel and also interact, thereby providing added coordination between stress signals and ABA in the regulation of stress-responsive genes.

Taking advantage of the recent completion of the soybean genomic sequence, Tran e Mochida (2010) carried out a computational analysis of the soybean genomic sequence to identify the full complement of TF-encoding genes. This study could help identify stress responsive TF genes in a systematic manner. The authors believe that their results will rapidly accelerate the progress in “transcription factoromics” of soybean, comparative genomics of regulatory TF proteins in legumes, facilitating genetic engineering programs to improve the productivity of soybean grown in adverse conditions.

Another topic related to abiotic stress is the future climate change scenario. Thinking about genetic solutions to the problem, Singh et al. (2010) analyzed the effects of multiple abiotic stressors on the growth, physiology and reproduction of cowpea genotypes to identify genotypic tolerance to these factors. They concluded that regardless of CO₂ enrichment, a combined effect of UVB radiation and temperature possibly will pose a serious problem for cowpea and most likely for many summer-grown crops in future climatic conditions. However, the identified tolerant cowpea genotypes and groups of plant attributes could be used for the selection and development of tolerant cultivars to multiple abiotic stresses by trait-based plant breeding or genetic engineering programs.

Recently, many vacuolar Na⁺/H⁺ antiporter genes (*NHX1*) have been characterized and identified from different plant species. Li et al. (2010) cloned a novel vacuolar Na⁺/H⁺ antiporter gene from the halophyte *Salsola soda* using the rapid amplification of cDNA ends (RACE) method and introduced it into alfalfa (*Medicago sativa*), which showed a strong salt-tolerant ability and could grow in conditions

with high levels of NaCl. The transgenic alfalfa lines survived in serious saline soils and could be helpful in exploiting and utilizing degraded areas. According to Yang et al. (2009), despite the large value of NHX1 antiporter for osmotic control in plants, the first gene isolated from the family nhx, *AtNHX1*, to its full activity in transgenic plants the action of some genes of SOS pathway is essential. They hypothesized that transgenic plants co-expressing *SOS1*, *SOS2*, and *SOS3* together may improve salt tolerance more than transgenic plants overexpressing any of the single genes and provided evidence to show that overexpression of *AtNHX1* in *A. thaliana* did not improve plant salt tolerance significantly above that of control plants. However, salt tolerance in transgenic plants overexpressing *SOS1* + *SOS2* + *SOS3* was improved only marginally compared to transgenic plants overexpressing only single genes. Considering that the intervention of the plant transformation may facilitate the selection and isolation of tolerant lines (El-Sayed et al., 2007), current strategies have been clearly successful in incorporating tolerance levels in different species (Novak and Brunner, 1992).

Legumes, as cowpea and soybean, are the main protein sources in tropical countries, having also important economic value. Therefore, the identification of new gene candidates regarding their performance under drought and salinity is urgent, having the potential to allow the acquisition of new features of interest not only for these species, but also for the transference to other important crops in drought prone areas.

Material and methods

Interested in evaluating the genes involved in responsive pathways to abiotic stresses our group analyzed databases of expressed sequences (ESTs and SuperSAGE Libraries) in an attempt to identify the orthologs/paralogs genes expressed in cowpea and soybean. Searches were carried out against the EST datasets using tBLASTn algorithm (Altschul et al., 1999) with a cut-off e-value of e^{-5} or less using seed-sequences of genes that had their expression increased more than fivefold in response to abiotic stresses in *A. thaliana*, according to Seki et al. (2002) and genes integrating the SOS (Salt Overly Sensitive) pathway. The expression pattern of transcription factor proteins based on EST results were sampled generating a graphic showing the direct correlation between the contigs/singlets and their respective libraries. For the SOS pathway orthologs the expressed genes were annotated by direct counting of reads that composed each cluster. The CLUSTER (Eisen et al., 1998) and TreeView (Page, 1996) programs (Eisen et al., 1998) were used to generate a Hierarchical Clustering showing the normalized differential expressions of genes and their relations, as well as the relation between libraries.

In the soybean databank a similar analysis was conducted using genes of the SOS pathway as seed sequences against the GENOSOJA ESTs collection. Again, the CLUSTER and TreeView programs were used to evaluate the expression pattern of orthologs in different conditions and tissues.

Results and discussion

Considerations regarding the overall response against abiotic stresses

The products of stress-inducible genes can be classified into two main groups: (i) those that directly protect plants against environmental stresses, and (ii) those that regulate the genic expression and signal transduction in the abiotic stress response (Seki et al., 2002). The first group included proteins with protecting function against dehydration, as LEA proteins and detoxifying enzymes. The second one included transcription factors and kinase proteins (Seki et al., 2003). Our analysis at cowpea EST database resulted in 1,010 highly similar sequences distributed in 27 categories. The most represented

categories were “Unknown Proteins”, “Transcription Factors”, “Cellular Structure, Organization and Biogenesis”, “Transport Protein Ion Channel”, “Plant Defense”, and “Carbohydrate Metabolism”, with 234, 127, 69, 65, 61 and 61 sequences respectively (Figure 1).

The greater representativeness of “Unknown Proteins” category indicated a rich resource of new genes involved in the abiotic stress responses; from these data it is important to analyze its function and structure, not only to better understand the molecular mechanisms of response and stress tolerance in higher plants, but also to improve stress tolerance to crops by genetic engineering (Seki et al., 2002). Considering the members of the second group, whose are involved in a higher regulation of signal transduction and genic expression, the high abundance of transcription factors (12.6%) suggests that there are many mechanisms of transcriptional regulation in the signal transduction pathways triggered by drought and saline stresses. High concentration of salt in the soils and hydric deficit are factors that greatly affect plants both at micro and macro level, as promoting changes at cell membranes and delaying floral development, respectively, it is considered that the here obtained results reflected the variety of responses involved in the acquisition of tolerance.

At microcellular level the activation of genes from categories: “Cellular Structure, Organization and Biogenesis” and “Transport Protein Ion Channel” was observed, showing the importance of preserving the cellular structures and organization, and controlling ion exchange with the environment. Considering the plant as a whole, the activation of genes in the “Carbohydrate Metabolism” and “Plant Defense” was observed; the first is directly related to plant growth and development, and the second seems to indicate the importance of cross-talk of genes between pathways, a common mechanism of plants submitted to stress. Plants in addition to having stress-specific adaptative responses, also share strategies that protect them from more than one type of stress.

The here presented data are preliminary results from a larger study that aim at the identification and characterization of the responsive gene repertoire under abiotic stress in legumes. Thus, these preliminary data will support the generation of an expression profile, allowing inferences about the differential regulation of target genes expression, including specific changes according to the tissue, stage of life cycle and treatment, involved in the response to abiotic stresses, as drought and salinity.

Transcription factors & transcriptomics essays

External stress factors as abiotic stresses exert strong evolutionary pressure on all organisms. To survive, plants must develop tolerance or resistance mechanisms, which involve sophisticated changes in metabolic and physiological responses. Among these changes there is the activation of many gene groups as transcription factors. Among the ESTs results regarding transcriptions factors, prevalent genes were, as expected, DREB/ERF (26%) and Zinc-finger (20,5%) (Figure 2). This result was expected, since from more than 1,600 transcription factors (TFs) encoded by *A. thaliana*, 9% regarded members of DREB/ERF-like family (Dietz et al., 2010). Moreover, also Gong et al. (2010), in a research to understand the transcriptional regulation in drought-tolerant tomato genotypes, identified and characterized the zinc-finger family as the main activated group during the drought response. A similar result was found by our group examining the soybean transcriptome (Soares-Cavalcanti NM et al., in press).

The EST clusters identification led to the prospection of cowpea transcription factor expression pattern. After direct counting of the reads that composed each cluster it was possible to generate a bar graphic showing the distribution of reads of each TF family per library (tissue/condition) (Figure 3). Considering the transcription factors as whole, more expression was noted at ST08 library (Root of genotype tolerant to salinity after 8 hours of stress), followed by SS02 (Root of genotype sensitive to salinity after 2 hours

of stress) and SS08 (Root of genotype sensitive to salinity after 8 hours of stress), with 39, 36 and 32 reads, respectively (Figure 3). As expected, additional reads were found at libraries from salinity stress experiments. Furthermore, the results indicated that both tolerant and susceptible varieties presented differential and/or specific transcription regulation processes, having their expression altered after 8h of exposure to stress. Moreover, it is important to note that transcription factor expression could be observed in all libraries.

In a recent research our group analyzed a collection of 330,000 cowpea SuperSAGE tags (26 nt), from five libraries in a specific search of 10 transcription factors families putatively highly active in the signal transduction related to drought and salinity in arabidopsis using BLASTn algorithm, considering an identity of ≥ 18 nt and cutoff-value 10, suitable for small sequences, consisting of only 26 nucleotides. Furthermore, comparative analysis between the libraries was performed to identify differential expression. The results indicated 81 tags, of which 5, 19, 2, 10, 4, 24, 9, 6 and 2 showed high identities to *DREB+ERF*, Zincfinger, *WRKY*, *MYB*, *bHLH*, *NAC*, Homeodomain, *bZIP* and *HSF* genes, respectively. In relation to the best matches, in all families, was found that almost all tags were up-regulated in BRC1 library (absolute control) and BRC2 (injury control), but not in the mosaic infected libraries. So, the results revealed relative abundance and diversity of transcription factors genes in the cowpea genome, although lower expression was identified of transcription factors in infected plants; it was speculated that the obtained results were related to the fact that these genes originally have basal expression under normal conditions and are up-regulated during abiotic stresses (Soares-Cavalcanti et al., 2009).

Salt overly-sensitive pathway

In plants tolerant or susceptible to salinity or drought, the stressful condition initiate a complex of responses, starting with the perception of stress, which triggers a molecular events cascade that completed at various levels of physiology, metabolic and growth. A simple example of this chain reaction is the SOS (Salt Overly-Sensitive) pathway, one of the best responses to the effects of high salt concentrations in the soil. The SOS genes are involved in maintaining the cellular homeostasis through the exclusion of sodium excess. The excess of sodium in the soil triggers the increase of intracellular calcium that stimulates SOS3/SOS2 complex, which activates the *SOS1* gene, which encodes a Na^+/H^+ antiporter protein, allowing exclusion of the excess of sodium ion and minimizing the toxic effects of salinity on the plant (Zhu, 2002). The components of the SOS pathway are relatively well known in *A. thaliana*, being also found in the SUCEST (sugarcane, *Saccharum officinarum*) genome database (Benko-Iseppon et al., 2005), with exception of the coding-gene for the HKT1 protein.

With respect to the here studied leguminous genomes, the majority of the SOS pathway genes (Table 1) was identified. The most representative genes in cowpea and soybean were SOS2, ATPase, and ABI2, with 20 orthologs in each of both organisms. This great abundance may be related to the fact that these proteins are widely employed in different cellular processes; (i) the first one is a kinase protein used in the regulation of several other proteins and enzymes (Seki et al., 2002), (ii) the second is involved in energy transformation of virtually all cellular activities, not exclusive of SOS pathway (Silva and Gerós, 2009), and (iii) the third is a protein induced by ABA, whose expression is regulated not only by salt and hydric stress (Parida and Das, 2005). Moreover, in the SOS pathway there is a direct interaction between SOS2 and ATPase, where SOS2 promotes the transport activity of H^+ -ATPase and also enhances abiotic stress tolerance (Mahajan et al., 2008).

The large amount of SOS2 and ABI2 genes and the less abundance of the SOS1 representative may be explained by the interactions of these genes in the SOS pathway. Under osmotic stress conditions the

increased level of cytosolic Ca^{2+} caused by the stress is sensed by a calcium sensor such as SOS3, which interacts with the SOS2 protein kinase. Then the SOS3-SOS2 complex phosphorylates the plasma membrane Na^+/H^+ antiporter SOS1 and the vacuolar exchanger NHX1. In the cell, phosphorylation events are often coupled with dephosphorylation events for the maintenance of homeostasis; so, SOS2 also interacts with ABI2. It may be speculated that ABI2 acts dephosphorylating the proteins that are phosphorylated by SOS2 in order to restore homeostasis. CAX1 ($\text{H}^+/\text{Ca}^{2+}$ antiporter) has been identified as an additional target of SOS2 activity reinstating cytosolic Ca^{2+} homeostasis under stress (Mahajan et al., 2008). In this view, it is reasonable to assume that under the conditions in which the ESTs experiments were conducted the interaction between SOS2 and ABI2 down-regulated the activation of SOS1.

In a previous evaluation carried out by our group (Benko-Iseppon, et al., 2005) no *HKT1* transcripts could be identified in sugarcane, similarly to the occurred to cowpea and soybean in the present approach. The absence of potassium transporter proteins in the transcriptome projects can be related to their function as low affinity Na^+ transporter which can lead to *HKT1* gene inactivation during SOS pathway response (Chinnusamy et al., 2004), although the availability of intracellular potassium represents a vital role in various processes such as metabolism, growth, and stress adaptation (Mahajan et al., 2008).

From 123 soybean sequences matching to members of the SOS pathway a total of 4,690 reads could be counted. The quantitative results indicated a higher expression in S12 library (Seeds with the embryo in globular stage), followed by L08 (leaf) and R06 (root) libraries. However all 50 libraries, that include many vegetable tissues, were represented by at least one read, such as L04 (immature leaves of greenhouse grown plants) and C02 (cotyledons of 18 day-old). The results demonstrated the importance of osmotic adjustment in the initial stages of development in the tissues that first are affected by the injuries, since the roots represent the first site to sense the environmental changes and activate the tolerance mechanisms, while the leaves are directly involved to the photosynthesis and transpiration processes, with changes in these vital processes slowing or retarding plant development.

The present results point to the conservation of the SOS pathway in the soybean and cowpea genomes, as presumed, since the ionic exchanges, mainly the sodium exclusion is one of the first and most important mechanisms activated in the abiotic stresses responses.

Challenges and perspectives

The complex environmental conditions with its heterogeneous conditions, including different abiotic stress combinations are some of the challenges to be faced by the modern agriculture. To significantly improve the abiotic stress tolerance of crops in the field a combination of approaches like mechanistic understanding and subsequent utilization of stress response and stress acclimation networks will likely be needed (Mittler and Blumwald, 2010). Currently, research efforts should be directed to extensive testing in the laboratory, in greenhouses, and in the field, including the use of innovative approaches that take into consideration the genetic background and the physiology of different crops, the use of enzymes and proteins from other organisms, and the integration of QTL mapping and other genetic and breeding tools (Mittler and Blumwald, 2010).

Major progresses to understand abiotic stress responses have come from the application of molecular biology and are now thought to be one of the most important topics in plant science. After this methodology was employed, many abiotic stress-inducible genes were isolated and their functions were precisely characterized in transgenic plants. The availability of these data broadened and deepened the view of abiotic stress responses and tolerance in plants (Hirayama and Shinozaki, 2010).

During the analysis of large gene expression datasets, several questions arise for the researcher and despite it is not possible to answer certainly if correlated genes are necessarily co-regulated or if they are involved in the same pathway or biological process or even, how do we interpret a mathematical relationship between genes and environmental conditions, today, several approaches like expression analysis are available to help us to propose new answers. Although this approach appear to be rather descriptive, is an excellent starting point to study novel cellular processes and to formulate novel hypotheses (Baginsky et al., 2010).

References

- Abdelnoor RV, Nepomuceno AL, Barros EG, Sá MFG, et al. (2009). GENOSOJA – a brazilian soybean genome consortium. In: Plant and Animal Genome Conference, San Diego – USA, abstract P034. Available at [http://www.intl-pag.org/17/abstracts/P01_PAGXVII_034.html].
- Altschul SF, Gish W, Miller W, Myers EW, et al. (1990). Basic local alignment search tool. *J.Mol. Biol.* 215: 403-410.
- Apse MP, Sottosanto JB and Blumwald E (2003). Vacuolar cation/H⁺ exchange, ion homeostasis, and leaf development are altered in a T-DNA insertional mutant of AtNHX1, the *Arabidopsis* vacuolar Na⁺/H⁺ antiporter. *Plant J.* 36: 229–239.
- Agarwal PK and Jha, B (2010). Transcription factors in plants and ABA dependent and independent abiotic stress signaling. *Biol. Plant.* 54(2): 201-212.
- Audic S and Claverie JM (1997). The significance of digital gene expression profiles. *Genome Res.* 7(10): 986-995.
- Baginsky S, Henning L, Zimmermann P and Gruissen W (2010). Gene expression analysis, proteomics, and network discovery. *Plant Physiol.* 152: 402–410.
- Benko-Iseppon AM, Soares-Cavalcanti NM, Wanderley-Nogueira AC, Belarmino LC, et al. (2005). Genes associados a estresses bióticos e abióticos em feijão-caupi [*Vigna unguiculata* (L.) Walp.] e outras angiospermas. In: Estresses ambientais: danos e benefícios em plantas (Nogueira RMC, Araújo EL, Willadino LG and Cavalcante UMT, Org.). : MXM Gráfica e Editora, Recife/PE, Brazil, 350-359.
- Benko-Iseppon AM, Kido EA, Monte SJH, Pandolfi V, et al. (2009). Cowpea (*Vigna unguiculata*) transcriptome and differential expression: the project and applications for breeding purposes. In: Plant and Animal Genome Conference, San Diego – USA, abstract P034. Available at [http://www.intl-pag.org/17/abstracts/W40_PAGXVII_304.html].
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. (2011). GenBank. *Nucleic Acids Res.* 39: D32-D37.
- Borém, A (1998). Melhoramento de Plantas. 2. ed. Editora UFV, Viçosa. 525 p.
- Busch W and Lohmann JU (2007). Profiling a plant: expression analysis in *Arabidopsis*. *Curr. Opin. Plant Biol.* 10: 136–141.
- Carazzolle MF, Formighieri EF, Digiampietri LA, Araujo MRR, et al. (2007). Gene projects: a genome web tool for ongoing mining and annotation applied to CitEST. *Genet. Mol. Biol.* 30(3): 1030-1036.
- Century K, Reuber TL and Ratcliffe, OJ (2008). Regulating the regulators: the future prospects for transcription-factor based agricultural biotechnology products. *Plant Physiol.* 147: 20-29.
- Chinnusamy V, Schumaker K and Zhu J-K (2004). Molecular genetics perspectives on cross-talk and specificity in abiotic stress signaling in plants. *J. Exp. Bot.* 55: 225-236.
- Ciais PH, Reichstein M, Viovy N, Granier A, et al. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* 437: 529–533.

- Dietz KJ, Vogel MO and Viehhauser A (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma* 245: 3–14.
- Eisen MB, Spellman PT, Brown PO and Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U.S.A.* 95: 14863–14868.
- El-Sayed OE, Rizkalla AA and Sabri SRS (2007). *In vitro* mutagenesis for genetic improvement of salinity tolerance in wheat. *Res. J. Agric. Biol. Sci.* 4(5): 377-383.
- Gong P, Zhang J, Li H, Yang C, et al. (2010). Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. *J. Exp. Bot.* 61: 1-13.
- Guo P, Baum M, Grando S, Ceccarelli S, et al (2009). Differentially expressed genes between drought-tolerant and drought-sensitive barley genotypes in response to drought stress during the reproductive stage. *J. Exp. Bot.* 60: 3531–3544.
- Hayano-Kanashiro C, Calderón_Vázquez C, Ibarra-Laclette E, Herrera-Estrella L, et al. (2009). Analysis of gene expression and physiological responses in three mexican maize landraces under drought stress and recovery irrigation. *PLoS ONE* 4: e7531.
- Hirayama T and Shinozaki K (2010). Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant J.* 61: 1041–1052.
- Hruz T, Laule O, Szabo G, Wessendorp F, et al. (2008). Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinformatics* 2008:420747.
- Li D, Su Z, Dong J and Wang T (2009). An expression database for roots of the model legume *Medicago truncatula* under salt stress. *BMC Genomics* 10: 517.
- Li W, Wang D, Jin T, Chang Q, et al. (2010). The vacuolar Na⁺/H⁺ antiporter gene SsNHX1 from the halophyte *Salsola soda* confers salt tolerance in transgenic alfalfa (*Medicago sativa* L.). *Plant Mol. Biol. Rep.* 29(2): 278-290.
- Mahajan S, Pandey GK and Tuteja N (2008). Calcium- and salt-stress signaling in plants: shedding light on SOS pathway. *Arch. Bioch. Biophys.* 471: 146–158.
- Mittler R and Blumwald E (2010). Genetic Engineering for Modern Agriculture: Challenges and Perspectives. *Annu. Rev. Plant Biol.* 61: 443–462.
- Novak FJ and Brunner, H (1992). Plant breeding: induced mutation technology for crop improvement. *IAEA Bulletin* 4: 25-33.
- Page RDM (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12: 357-358.
- Papini-Terzi FS, Felix JM, Rocha FR, Waclawovsky EC, et al. (2007). The SUCEST-FUN project: identifying genes that regulate sucrose content in sugarcane plants. *Proceedings - International Society of Sugar Cane Technologists* 26: 868-877.
- Parida AK and Das AB (2005). Salt tolerance and salinity effects on plants: a review. *Ecotoxicol. Environ. Saf.* 60: 324-349.
- Pitzschke A and Hirt H (2009). Bioinformatic and systems biology tools to generate testable models of signaling pathways and their targets. *Plant Physiol.* 152: 460-469.
- Rabbani MA, Maruyama K, Abe H, Khan AM, et al. (2003). Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiol.* 133: 1755–1767.
- Rombauts S, Florquin K, Lescot M, Marchal K, et al. (2003). Computational Approaches to Identify Promoters and cis- Regulatory Elements in Plant Genomes. *Plant Physiol.* 132: 1162–1176.

- Seki M, Narusaka M, Ishida J, Nanjo T, et al. (2002). Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31: 279-292.
- Seki M, Kamei A, Yamaguchi-Shinozaki K, Shinozaki K (2003). Molecular responses to drought, salinity and frost: common and different paths for plant protection. *Curr. Opin. Plant Biol.* 14: 194–199.
- Silva P and Gerós H (2009). Regulation by salt of vacuolar H⁺-ATPase and H⁺-pyrophosphatase activities and Na⁺/H⁺ exchange. *Plant Signal. Behav.* 4(8): 718-726.
- Singh SK, Kakani VG, Surabhi GK and Reddy KR (2010). Cowpea (*Vigna unguiculata* [L.] Walp.) genotypes response to multiple abiotic stresses. *Journal of Photochemistry and Photobiology Biology* 100: 135-146.
- Soares-Cavalcanti NM, Wanderley-Nogueira AC, Kido EA, Pandolfi V, et al. (2009). Quantitative analysis of cowpea SAGE tags related to abiotic and biotic responses. In: II Simpósio Brasileiro de Genética Molecular de Plantas, Buzios-RJ, Brazil.
- Sottosanto JB, Gelli A and Blumwald E (2004). DNA array analyses of *Arabidopsis thaliana* lacking a vacuolar Na/H antiporter: impact of AtNHX1 on gene expression. *Plant J.* 40: 752–771.
- Taji T, Seki M, Satou M, Sakurai T, et al (2004). Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using *Arabidopsis* microarray. *Plant Physiol.* 135: 1697–1709.
- Tran LSP and Mochida K (2010). Identification and prediction of abiotic stress responsive transcription factors involved in abiotic stress signaling in soybean. *Plant Signal. Behav.* 5(3): 255-257.
- Varshney RK, Hiremath PJ, Lekha P, Kashiwagi J, et al. (2009) A comprehensive resource of drought- and salinity-responsive ESTs for gene discovery and marker development in chickpea (*Cicer arietinum* L.). *BMC Genomics* 10: 523-540.
- Vettore AL, Silva FR, Kemper EL and Arruda P (2001) The libraries that made SUCEST. *Genet. Mol. Biol.* 24(¼): 1-7.
- Wang Y-H, Li YD, Luo GZ, Tian AG, et al. (2005). Cloning and characterization of an HDZip I gene GmHZ1 from soybean. *Planta* 221: 831–843.
- Yang Q, Chen Z-Z, Zhou X-F, Yin H-B, et al. (2009) Overexpression of SOS (Salt Overly Sensitive) genes increases salt tolerance in transgenic *Arabidopsis*. *Mol. Plant* 2(1): 22-31.
- Zhou Y and Shao HB (2008) The responding relationship between plants and environment is the essential principle for agricultural sustainable development on the globe. *C. R. Biol.* 331: 321–328.
- Zhu, J-K. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53: 247-273.

Table 1. SOS pathway orthologous genes in cowpea and soybean genomes

Organism	Number of Ortologous Genes										
	SOS1	SOS2	SOS3	ATPase	HKT1	AKT1	NHX1	CAX1	V_ATPase	V_PPase	ABI2
<i>V. unguiculata</i>	2	20	8	20	0	12	4	6	3	3	20
<i>G. max</i>	1	20	8	20	0	19	6	11	13	5	20

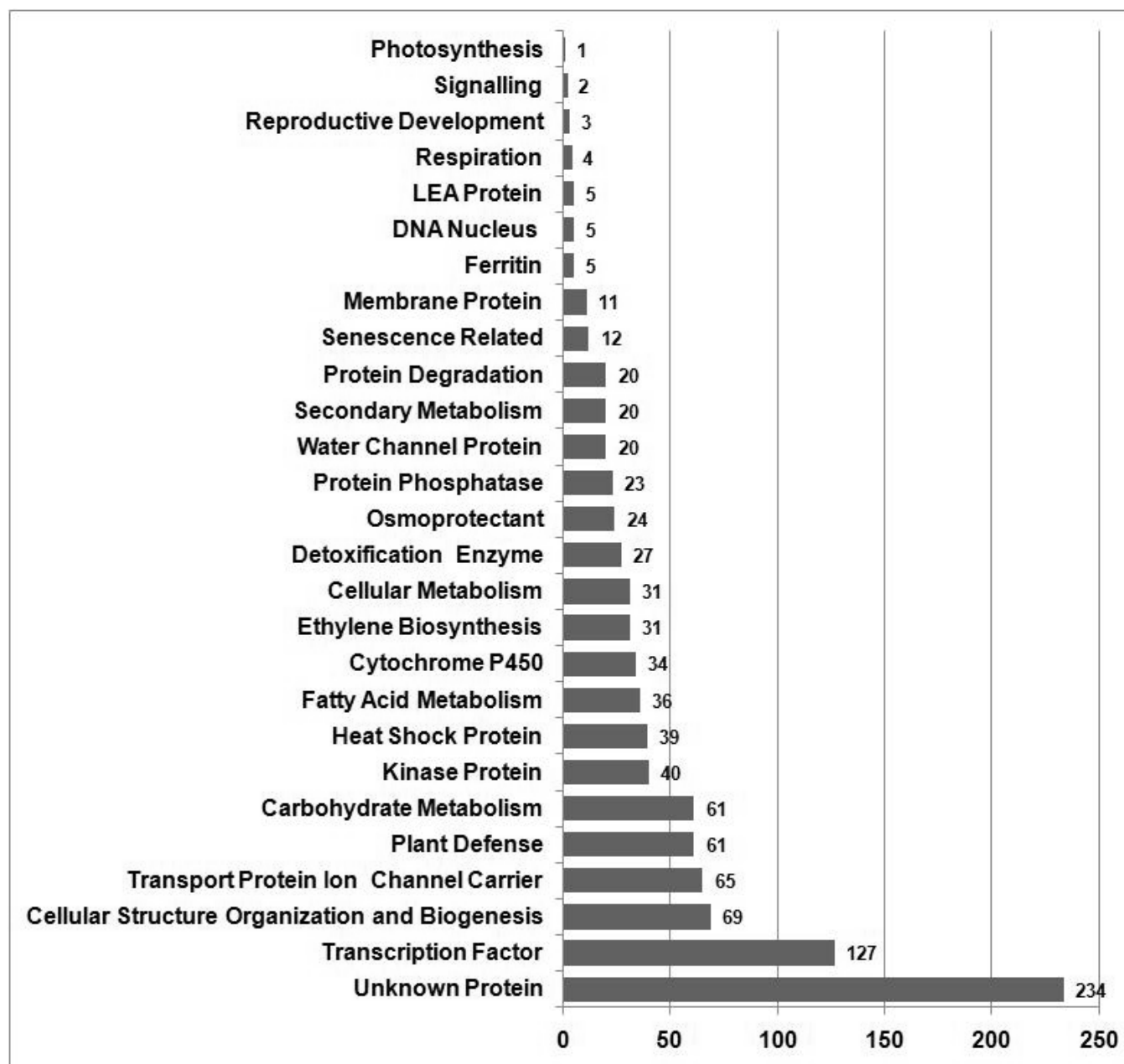


Figure 1. Diversity and abundance stress-inducible genes, considering 1,010 candidates from different categories identified in cowpea.

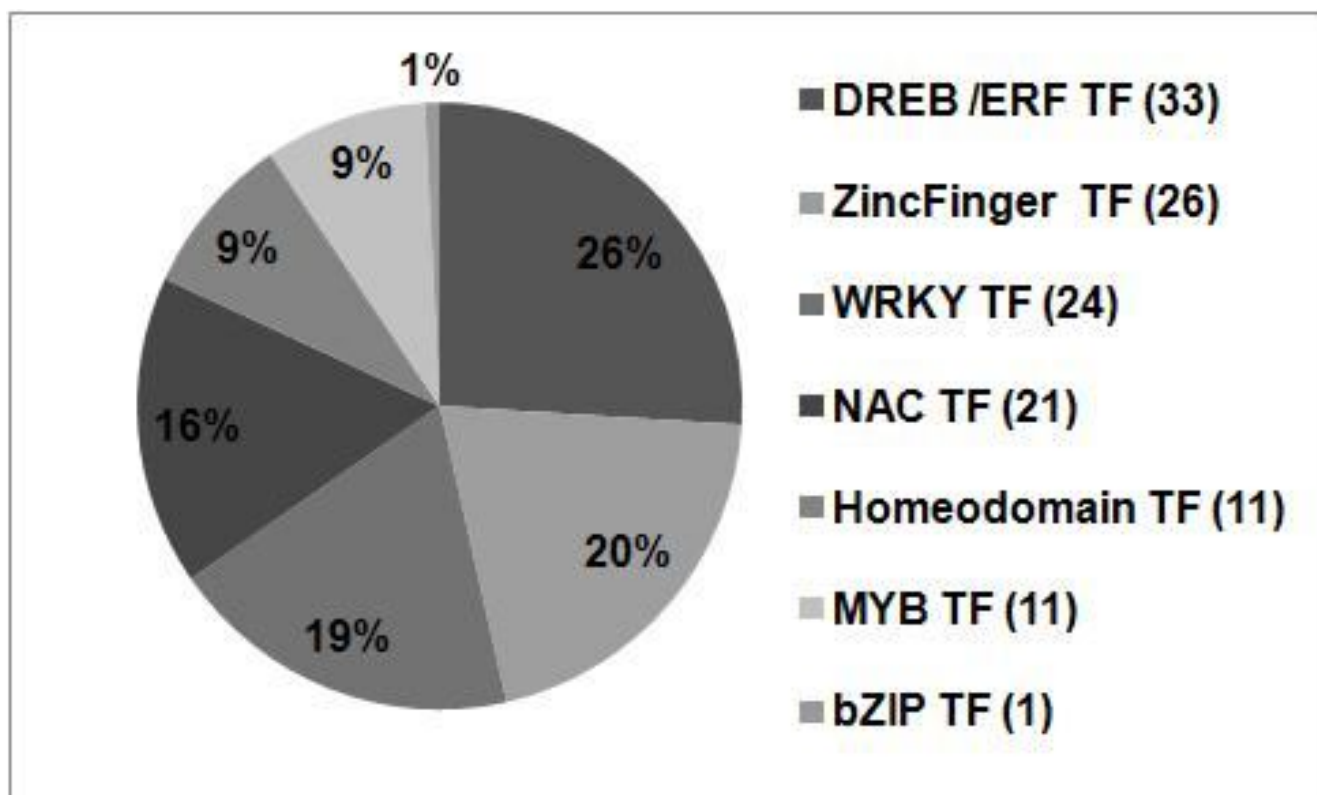


Figure 2. *Vigna unguiculata* transcription factor repertory; percentage of transcripts for each genic family. Numbers in parenthesis indicates the amount of orthologous sequences per transcriptions Factor family.

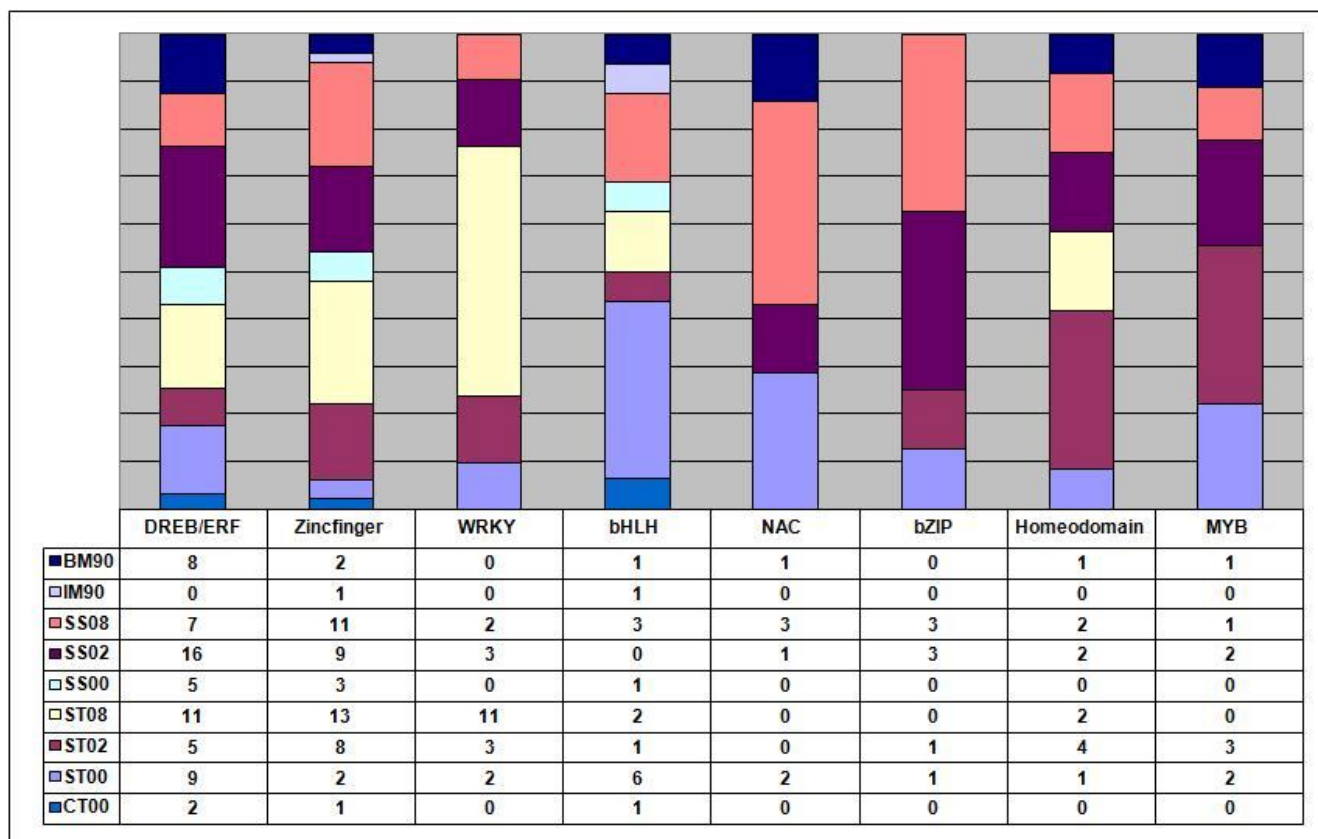


Figure 3. Graphic representation of transcription factor families' expression in cowpea generated in the NordEST Transcriptome Project. Abbreviations: BM90, leaves of BR14-Mulato genotype; IM90, leaves of IT85F genotype collected 90 minutes after mosaic virus infection; SS00, Root of genotype sensitive to salinity without salt stress; SS02, Root of genotype sensitive to salinity after 2 hours of stress; SS08, Root of genotype sensitive to salinity after 8 hours of stress; ST00, Root of genotype tolerant to salinity without salt stress; ST02, Root of genotype tolerant to salinity after 2 hours of stress; ST08, Root of genotype tolerant to salinity after 8 hours of stress; CT00, absolute control.

Plant-Specific Transcription Factors in *Vigna unguiculata* Transcriptome: Sequence Conservation and DNA-Binding Domain Integrity

Running head: Cowpea Transcriptions Factors

Nina da Mota Soares-Cavalcanti¹, Ederson Akio Kido¹, Luis Carlos Belarmino¹, Ana Carolina Wanderley-Nogueira¹, Valesca Pandolfi¹, Adriano Barbosa-Silva²; Semiramis Jamil Hadad do Monte³, Ana Maria Benko-Iseppon^{1§}

¹Laboratory for Plant Genetics and Biotechnology, Genetics Department, Center for Biological Sciences (CCB), Universidade Federal de Pernambuco, Rua Prof. Moraes Rêgo s/nº. Recife, PE, Brazil.

²PostGraduate Program in Bioinformatic, Universidade Federal do Paraná, Rua Dr. Alcides Vieira Arcoverde nº 1225, Jardim das Américas. Curitiba, PR, Brazil.

³Laboratory of Immunogenetics and Molecular Biology, Universidade Federal do Piauí, Campus Petrônio Portela, bloco 16. Teresina, PI, Brazil.

[§]Corresponding author

Email addresses:

NMSC:	ninamotasc@gmail.com
EAK:	ederson.kido@gmail.com
LCB:	lucabellarmino@gmail.com
ACWN:	carolwn@gmail.com
VP:	valescapandolfi@gmail.com
ABS:	adriano.bioinfo@gmail.com
SJHM:	libpi@ufpi.br
AMBI:	ana.benko.iseppon@pq.cnpq.br

Abstract

Plants devote a large portion of their genome capacity to transcription, with the arabidopsis genome coding about 1,500 transcription factors. Transcriptional control of gene expression is a crucial part of the plant response to a range of abiotic and biotic stresses. These transcription factors often belong to large gene families, which in some cases are unique to plants, like DREB (Drehydration Responsive Element Binding Factor), ERF (Ethylene Responsive Factor), NAC and WRKY. In this report, we searched for sequence-specific DNA-binding TFs exclusive to plants, using a bioinformatic approach and aiming to identify the TF repertoire within the cowpea genome. The sequences were classified as a TF if they presented a significant match to a model annotated at the DNA-Binding Domain (DBD) database. The searches were carried in the NordEST database (comprising ~30.000 ESTs generated by our project and ~150.000 from public databases) and showed a great abundance and diversity of transcription factors in cowpea, with the identification of 106 plant-exclusive TFs. In relation to the differential regulation regarding SuperSAGE, 53 tags (56.4% of the transcripts) seemed to be repressed when submitted to salinity stress, while 41 (43.6%) were activated. The identified transcription factors of cowpea regard interesting candidates for improvement of legume resistance to abiotic stress. Evaluations including their identification, diversity, mapping and transient expression may help in their manipulation considering both, conventional and biotechnological breeding programs.

Keywords: Transcription Factors, Cowpea, Abiotic Stress, Bioinformatic

Background

Although plants share fundamental biological mechanisms with other eukaryotic kingdoms, they acquired different systems and organs after their arrival on land 0.4-0.5 billion years ago. Relating to their immobility, plants possess distinct systems of adaptation to the environment, defense against pathogens, growth of specific organs such as flowers, among other [1].

Within cells, some genes are expressed constitutively, whereas other regard a given life step, organ or stimulus. These patterns depend on the interaction of transcription factors with cis-acting elements and/or with other transcription factors required for gene expression, and they are important in the regulation of cell activities [2]. Therefore, changes in the expression of transcription factor genes normally result in dramatic modifications to an organism and structural changes to these genes may represent a significant evolutionary force [3].

Many transcription factors (TF) families existed before the divergence of plants, animals and fungi; however, the size of TF families differs considerably among eukaryote kingdoms. For example, the MADS-box family has two, four and two members in *Drosophila*, *S. cerevisiae* and *C. elegans*, respectively. There are more than 100 members in arabidopsis, suggesting substantial amplification of this gene in the land plant lineage [2]. Nevertheless, interestingly many of the TF showed no obvious sequence similarity to those of other eukaryotes or bacteria. In ~1,500 probable specific transcription factors identified in the *Arabidopsis thaliana* genome, ~45% were classified into those plant-specific TF [4].

A typical plant transcription factor, such as DREB/ERF, NAC and WRKY, contains a DNA-binding region, an oligomerization site, a transcription-regulation domain and a nuclear localization signal. Most TFs exhibit only one type of DNA-binding and oligomerization domain, occasionally in multiple copies, but some contain two distinct types [4].

The AP2 (Apetala2)/EREBP (Ethylene-Responsive-Element-Binding Protein) superfamily, that includes DREB (Dehydration Responsive Element Binding) and ERF (Ethylene Responsive Factor) subfamilies, is composed by important transcription factors that induce a set of abiotic stress-related genes and impart stress endurance to plants [5, 6].

The DREB/ERF proteins share the AP2 DNA-binding domain, formed by 58-59 aminoacids residues, that binds to two cis-elements, the GCC box and the CRT (C-Repeat)/DRE (Dehydration Responsive Element) motif. In this quite conserved domain the 14th valine and 19th glutamic acid play a

crucial role in the determination of DNA-binding specificity. Moreover, a conserved ser/thr-rich region adjacent to the AP2 domain is considered to be responsible for phosphorylation of DREB sub-family [6].

NAC domain-containing proteins are transcription factors that can be expressed in several tissues and developmental stages. These TFs are organized into a general structure that consists of a highly conserved N-terminal domain involved in DNA-binding (called NAC domain) and a C-terminal region highly divergent in sequence and length that function as an activation domain [7]. The NAC domain comprises nearly 160 aminoacids residues, divided into five sub-domains (A-E) of conserved blocks intercalated with heterogeneous blocks or gaps [8]. In a general view, they comprise a large family of genes represented by 120 members in the *A. thaliana* genome [9]; of which five seem to be differentially expressed under abiotic stress conditions (salinity, drought and freezing), according Seki *et al.* (2002) [10].

WRKY TFs are well-known components of biotic stress responses [11]. In 2006 abiotic stress responsive WRKY genes have been reported [12], with four representatives in arabidopsis [10]. These TFs present a WRKY domain, formed by a region with 60 aminoacids, highly conserved amongst family members, with a binding preference for the W-box, the DNA motif (T)(T)TGAC(C/T). The WRKY domain is defined by the conserved WRKYGQK aminoacid sequence at its N-terminal end, together with a novel zinc-finger-like motif [13]. Members of WRKY family can be classified into three groups based in the number of WRKY domains (one or two) and in the features of their zinc-finger-like motif. Representatives including these two domains belong to group I, whereas most proteins with one domain belong to group II. Generally, the WRKY domains of groups I and II have the same type of zinc finger motif, whose pattern of potential zinc ligands (C-X₄₋₅-C-X₂₂₋₂₃-H-X₁-H) is unique among all described zinc-finger-like motifs [13]. The single zinc finger motif of a small subset (group III) of WRKY proteins is distinct from the members of group I and II. Instead of a C₂-H₂ pattern, their WRKY domains contain a C₂-HC motif (C-X₇-C-X₂₃-H-X₁-C) [13].

Until recently, little information about the genetic background of *Vigna unguiculata* – a crop of high economic importance in semi-arid regions – was available. A transcriptome project of this plant was carried out in the course of the NordEST project (<http://bioinfo03.ibi.unicamp.br/vigna/>), including EST and SuperSAGE sequences (26-bp tag fragments from expressed transcripts), allowing the identification of an important part of the expressed gene pool under different conditions (salinity, pathogen infection and control). In the present work we used the HT-SuperSAGE “High-Throughput

(HT-) SuperSAGE” (SuperSAGE technology combined with next-generation sequencing), a highly sensitive, reproducible and accurate method for digital gene expression data [14].

The present work aimed to perform a data mining-based identification of plant-specific transcription factors families in the NordEST database, using well known arabidopsis sequences involved in the abiotic stresses response as template. Moreover, a detailed description of cowpea WRKY factors, including domain-based classification, evolutionary relationships and proteins migratory behavior, was carried out.

Results

Cowpea Orthologs

ESTs Database Analysis

In the searches for candidates to the 11 DREB/ERF genes, it was possible to identify similarity (e-value $2e-44$ to $1e-15$) with 58 different cowpea clusters, from which 51 (87.9%) presented the procured AP2 domain (Table 1). Moreover, the results of reverse and reciprocal alignments showed great conservation between sequences of *V. unguiculata* and other Fabaceae plants (82.7% and 48.3%, respectively). Among these, *Glycine max* showed the greatest similarity in both analyses.

The results for the five members of the NAC family pointed the presence of 42 possible candidates in cowpea (Table 1). Complementary results, through BLASTn and BLASTp tools revealed 32 (76.2%) and 22 (52.4%) clusters, respectively, with high similarity with organisms from the genera *Glycine*, *Phaseolus* and *Medicago*. Structurally, the rpsBLAST allowed the identification of the NAM conserved domain in 36 sequences.

Regarding the WRKY family, a search using three seed sequences revealed 27 transcripts at the NORDEST database (Table 1). However, the structure analysis of the sequences showed that 19 transcripts (70.4%) presented the characteristic conserved domain of the family. From these clusters, 17 and two (respectively) presented one and two repetition of the WRKY domain, as expected. Furthermore, in five of the clusters only one WRKY repetition was found in association with a zinc binding domain. In respect to the reciprocal and reverse alignments, a high identity (56% to 97%) with *Glycine max* sequences was observed.

SuperSAGE Tags Database Analysis

The obtained results in the local database of contrasting tags (salinity versus control) for the DREB/ERF orthologs clusters allowed the identification of 759 different sequences, including 273 tags exclusive to control experiments, 312 tags exclusive to salinity treatments and 174 were common to both libraries. It is important to note that only 12 clusters presented no representative in the local tags database (Figures 1 and 2A).

In the analysis of orthologs tags to the cowpea NAC candidates were identified 531 tags, which 77 and 175 were exclusive from the control and stressed libraries, respectively (Figure 1 and Figure 2B). In respect to the comparative analysis of WRKY, clusters against the SuperSAGE tags databank, the results pointed to a high similarity with 358 tags, including 107 exclusive to control experiments, 180 exclusive to salinity experiments and 71 found in both experiments (Figures 1 and 2C).

Quantitatively, the analyses of the transcript abundance in the control and salinity stressed libraries were divided into categories according with their frequency (Figure 2). For the three families of transcription factors, the data revealed that most tags regarded singlets, represented by just one copy (~65.45%), being therefore excluded of the present evaluation. Also, considering the DREB/ERF, NAC and WRKY families together, it was possible to note that the transcript category represented for more than 100 copies was the one that presented lower number of tags, 1.83% (Figure 2). It is worth mentioning that when these analyses were compared separately for each family this distribution pattern remained similar.

Differential Regulation

Considering all studied families, it was observed that from 1,648 analyzed tags, 94 presented significant difference in their expression ($p < 0.05$). From these differentially expressed tags, 56.4% of the potential transcription factors had their expression repressed after salt stress perception, while 43.6% of the transcripts had their expression increased after plant exposition to high concentrations of soluble salt (100 mM NaCl). With regard to the individual families (DREB/ERF, NAC and WRKY - Figure 3), it was viewed, respectively, that 17, 25 and 11 transcripts presented their expression repressed behind the perception of salt concentration changes, while 11, 19 and 11 transcripts, respectively, had their expression increased when in contact to high salinity.

Furthermore, it was possible to note that 457 tags, identified only in the control library, presented their expression completely repressed in the stressed condition, while 667 tags, exclusive to the salinity library, seemed to be activated just after the salt stress perception.

According to the transcript pattern, comparing the absolute frequency of tags in the control library in relation to the stressed library (salinity), it was noted that of 1,648 tags, 31.79% presented constitutive expression, not presenting significant differences in their frequencies, according to statistic value $p < 0.05$ measured by the DiscoverySpace program.

Phylogeny

The dendrogram generated with aminoacids sequences and including only the c-terminal WRKY repetition from different organisms allowed the differentiation, as expected, of two main groups (A and B) (Figure 4), the first one including representative of WRKY28 and WRKY75 proteins, both presenting in their structure one WRKY repetition, and the second was formed by WRKY33 members, which present two WRKY repetition in their structure. Moreover, in the first group (A) it was possible to perceive the segregation of the paralogous proteins in different clades (Figure 4). Although the phenetic character of the analysis, within the different branches there was a separation of the organisms following the classic plant classification, with the members of monocots class grouping in a separated clade from those of dicots class. Moreover, the segregation of the organisms followed their taxonomic classification (subclass, order and family). The exception occurred with *V. unguiculata* in the WRKY28 branch, which despite grouped with organisms from the same subclass (Rosids), their sequence presented lower degree of conservation when compared with sequences from other Fabids. In fact, while *Ricinus communis* and *Populus trichocarpa* shared 46.52% of the amino-acids, cowpea shared with these organisms just 27.02% of the amino-acids.

With regard to the multiple alignments, considering all orthologs and paralogous, a larger frequency of lysine, glycine and serine amino-acids was observed, which represented, respectively, 7.00, 9.76 and 11.43% of all residues. The comparative analysis of the sequences showed that 15.32% of all studied residues (359) were quite conserved and that 62.67% of the sites were parsimony-informative. In relation to each genic grouping individually, it was possible to observe the presence of 23.12, 41.23 and 27.85% of conserved sites between the WRKY28, WRKY33 and WRKY75 sequences, respectively.

Moreover, the multiple alignment with the complete proteins from all orthologous and paralogous studied allowed the identification of the conservation pattern (Figure 4), involving both domain and extra domain regions, what aided in the characterization of the three studied sequences. Despite this conservation pattern, it was possible to classify the different amino-acid sequences in the

well know Classes (I, II and III) of WRKY proteins family [13], according to the amount of domain repetition and to the type of zinc binding domain.

Virtual Electrophoresis

Corroborating the analysis previously performed, the virtual 2D electrophoresis results (Figure 5) showed the segregation of the proteins in three different groups, characterizing the studied WRKY gene products (28, 33 and 75). In a general way, the proteins presented variation of pI from 6.13 to 9.61, however, in relation to the molecular weight (MW), it were identified proteins with low MW (16.08) as well as proteins with high MW (61.08). As expected, the proteins with higher molecular mass were those belonging to WRKY33 grouping, which are characterized by the presence of two WRKY repetitions. Furthermore, based on the electrophoretic pattern it was not possible, within each one of the three genic groups, draw a parallel between the organisms classification and their pI and MW.

Discussion

Modulation of transcriptional activity is fundamental for gene expression regulation associated with most biological phenomena and it is largely mediated by proteins that interact directly or indirectly with specific DNA sequences (*cis*-elements) in the promoter region of genes. Thus, one critical step controlling stress responses involves transcriptional regulation [15]. Plant genomes, for example, dedicate a large number of their coding sequences to TFs, reaching about 6.91% (>1,500 TF genes) in the fully sequenced arabidopsis genome [4] and 6,56% (4,342 TF genes) in soybean [16].

In spite of their importance, such factors present often discrete expression levels enough to activate given genes, and their expression level may not be enough to be measured in microarray experiments. However, transcriptomic studies have been used to identify TFs regulated by abiotic stresses potentially linked to specific genetic programs activated to cope with stress [10, 17]. In model legumes, extensive sequencing highlighted around 2,000 TFs per genome, of which less than 1% have being genetically characterized [18]. In fact, very few TFs have been described in association to abiotic stress tolerance in legumes [19].

According to Libault et al. (2009) [20], studying legume genes, no legume-specific TF gene family was found, which may reflect the paucity of knowledge about legume TFs but more likely indicates conservation of TF families among plants. Consistent with this notion, the distribution of legume genes among the various TF families is similar to that of other plant species. Overall, based on

this census of genes in the various TF families, there does not appear to be any enrichment of known TF families in legumes. The expression of legume TF genes in arabidopsis or tobacco plants also supports a conservation of functions for TF homologs across plant species, e.g. in plant resistance to abiotic stresses (GmWRKY [21] and GmDREB3 [22]).

The *V. unguiculata* EST data mining, as expected, revealed high abundance of cowpea transcription factors with higher similarity with the available soybean sequences, although the seed sequences used were from arabidopsis; this higher similarity appears to be related to the evolutionary proximity between the Fabaceae species.

Considering the DREB/ERF family, it was possible to identify high similarity between arabidopsis seed sequences and *V. unguiculata* putative DREB/ERF sequences related to the abiotic stress, with the identification of 51 clusters containing the AP2 domain; it is important to note that the functionality of these proteins is directly related to the integrity of this domain. The ERF family proteins contain a single AP2/ERF domain, and are sometimes further divided into two major subfamilies, the CBF/DREB subfamily and the ERF subfamily [23], which can play a crucial role in the response of plants to abiotic stresses. According to Zhang et al (2008) [24], comparative analysis of aminoacid residues of the AP2/ERF domains in the soybean ERF family proteins with those of arabidopsis and rice suggested the AP2/ERF domains were well conserved among the plant species. Moreover, the authors identified 98 unigenes that contained a complete AP2/ERF domain, probably involved in the biotic and abiotic stress responses signaling pathway.

Taking into consideration that proteins sharing conserved domains and motifs are likely to have similar functions, it may be speculated that cowpea presented similar mechanisms of abiotic stress responses to the well described model plant *A. thaliana*. Furthermore, the large amount of cowpea DREB/ERF sequences emphasizes the importance of these transcription factors also in cowpea stress response cascade; however, as in all in silico study, these function need to be validated.

With respect to NAC proteins, in rice, the expression levels of this family were also checked under various abiotic stresses including drought, salinity and low temperature. A preliminary evaluation based on microarray data suggested that more than 40 transcripts were responsive to drought and/or salt stresses [25]. Although a fraction of the NAC genes has been characterized in plants, such as arabidopsis and rice, the characterization of these factors in legumes is still rudimentary. Here, 36 *NAC-like* genes presumably related to abiotic stresses pathways were identified. The comparative analysis in cowpea, that showed high degree of aminoacids similarity, suggested that protein functions have been both

conserved and diverged during evolution of transcription factors families. In view of the fact that a very limited number of genes of the NAC family have been characterized in legume plants, our data provide a very useful reference for functional analysis of this family in Fabaceae organisms.

In 2007, Meng et al. [26] reported for the first time the molecular cloning of soybean NAC-like genes, but the functional relevance of these genes remained obscure. More recently, Pinheiro et al [9] performed a bioinformatic analysis in the soybean genome and identified 101 putative *NAC* genes, although these genes had not been yet functionally categorized.

In respect to the *WRKY* family, characterized by the presence of one or two *WRKY* domain together with a zinc binding motif [13], several studies have indicated their involvement in a range of biological processes, revealing that a large number of these genes are transcriptionally regulated under conditions of abiotic stresses. In arabidopsis, microarray studies have revealed that the expression of some *WRKY* genes is regulated by abiotic stresses, such as salinity, drought and cold [10, 27].

In rice, real-time PCR analyses showed that 41 *WRKY* genes were significantly regulated by abiotic stresses including cold, drought and salinity. Among them, 22 *WRKY* genes were regulated by only one of these stresses (nine by cold or salinity and four by drought). Another set of 17 genes were regulated by two of these three stresses. The remaining two genes were regulated by all three stresses [28]. In relation to legumes *WRKY* data, Zhou et al. (2008) [21] reported the identification of 64 *WRKY*-type transcription factor genes in soybean and showed that 25 of them presented differential expression in response to at least one experiment (abiotic stress).

In cowpea, the results showed the presence of 19 *WRKY* clusters, of which 17 were composed by one *WRKY* domain repetition and in the structure of two clusters two repetitions were viewed; a interesting result once the majority of the *WRKY* proteins contained two *WRKY* domains, being classified in the group I, while a fewer number of genes presented only one *WRKY* repetition, being classified in group II and III [13]. Moreover, the results showed in five sequences the presence of a zinc binding domain similar to those from non-legumes plants. It has been speculated that the Fabaceae could present a singular zinc binding domain, once nitrogen fixing plants - as leguminous - presented particularities in their genome.

In 2005, the three-dimensional structure of the carboxy-terminal *WRKY* domain of arabidopsis *WRKY4* was determined, and it was strongly suggested that the conserved *WRKYGQK* residues are directly involved in DNA binding [29]. Once cowpea clusters presented this essential domain it is

speculated that they can binding to the W-box in the promoter region of downstream genes [13], regulating the expression of genes involved in the abiotic stress responses.

Moreover, considering that the cowpea transcriptome sequencing project (NordEST) is still being finalized, the preliminary results obtained in this studied, 19 putative WRKY, were similar to those achieved in the soybean project (25) by Zhou et al. (2008) [21].

Overexpression of legume TF genes in non-legume plants often produces phenotypes consistent with strong conservation of TF function among higher plant species. Despite the redundancy of TF gene families across the plant kingdom, it is interesting that legumes nodulate while most other plant species do not. In fact, although specific TF genes have been implicated in nodule development (*SNF*), these belong to families common to non-legumes. This suggests that neofunctionalization of TF genes was important in the evolution of legume TF repertory, rather than the invention of novel TF genes/families [20]; a hypothesis supported by the obtained results and corroborated by the authors.

The absence of significant differences in TF gene distribution across TF families between legume and non-legume plants, as cited by Libault et al. [20], suggests that legume-specific traits are likely dependent on TF gene expression patterns and TF protein function. Since the set of TF families is almost unchanged among plants, it is speculated that the alterations in the activity and regulatory specificity of these TFs are the major source of diversity and evolutionary adaptation, resulting in the specific characteristics of each species.

In the analysis of SuperSAGE tags database it was found more amount of transcripts in the salinity library, when compared to the negative control. In both libraries, only around 0.78% to 1.41% tags presented frequency upper to 100, while more than 90% of the tags presented frequency up to 20-fold; similar values were found in the Carneiro (2007) [30] results, using the SAGE methodology in eucalyptus. In relation to the differential expression it was not possible to perceive a significant difference between the quantity of up and down-regulated genes. Apparently there was a balance between genes that was repressed (56.4%) and activated (43.6%); when plants are subjected to abiotic stress conditions there is a expression alteration (activation and/or repression) of these TFs in order to further regulation of other genes involved in the response signaling pathways.

The 524 tags expressed in both libraries, i.e. with constitutive expression, are related to transcription factors involved in the basal metabolic and physiologic processes in legume plants with a probable participation in vital pathways for plant survival. In a general view, the amount of tags involved in the stressed condition was higher, showing the importance of these TF in the responsive

pathways. In the other hand, the salinity-specific tags may be potentially useful for the design of specific primers in order to validate the genic expression through RT-qPCR using the same mRNA samples, which would enable the utilizations of these genes to obtain tolerant cowpea cultivars.

According to Berri et al. 2009 [31], the WRKY transcription factor gene family seems to have a very ancient origin and have undergone extensive duplications in the plant kingdom. In fact, the WRKY genes have experienced an incredible evolutionary success in the plant kingdom where successive duplication events have resulted in large gene families, as viewed in the arabidopsis and rice genome, with up to 74 and over one hundred members, respectively. This large variety observed for the WRKY family justifies their diversity of actions, functioning in many responsive pathways, as to abiotic and biotic stresses, as well as in developmental, metabolic and physiological processes; this article explore the WRKY proteins involved in the abiotic stress response.

The generated dendrogram allowed the identification of WRKY amino-acid sequences in three separated clades (WRKY 28, 33 and 75), in accordance to their distinct conserved domains and motifs. Despite the grouping of the sequences in accordance with the classic taxonomic classification, in the WRKY 28 clade a higher divergence between cowpea sequence and those from other Rosids was evident, a fact that could be attributed to the nitrogen-fixing nature of the leguminous plants.

In the Rosid clade, arabidopsis and legumes fall in the Malvidae (Eurosidae II) and the Fabidae (Eurosidae I) subclades, respectively. These two subclades probably diverged approximately 115 to 93 million years ago (Mya). According to Wang et al. (2009) [32], a diversification of Rosids occurred suddenly (in less than 15 million years) shortly after the divergence between Fabidae and Malvidae. Such diversification might explain the vast heterogeneity in Rosid habitats, morphology, and development. This evolutionary distance between crop plants and arabidopsis is also highlighted by the strong changes in the size and organization of their genomes, such as the ancient genome duplication in legumes occurring around 45 to 55 Mya. Despite the evolutionary distances existing between legumes and arabidopsis, microsynteny is still found in some regions [20].

As described, all known WRKY are sub-divided into three groups on the basis of the number of WRKY domains and the features of their zinc-finger-like motif: group 1 contains two WRKY domains including a C₂-H₂ motif (C-X4-5-C-X22-23-H-X1-H); group 2 contains one WRKY domain including a C₂-H₂ motif; and group 3 contains one WRKY domain including a different zinc-finger-like motif C₂-H-C (C-X7-C-X23-H-X1-C) [13]. Although their DNA-binding domain is highly conserved, the overall structures of WRKY proteins are highly divergent. In fact, in our analysis it was possible to see that,

excluding the well characterized WRKY domains, only little conserved motifs were identified; however, this reduced number of conserved regions were enough to classify the sequences in three different genic groupings (WRKY 28, 33 and 75). In a similar way, Eulgem et al. (2000) [13] described 11 different models of gene structure, considering the WRKY repetitions and other motifs, using 58 arabidopsis WRKY sequences. Moreover, no cowpea sequence belonged to the class III, which presented the C₂-HC distinct zinc binding domain.

Furthermore, once the WRKY genes presented characteristic motifs, it was expected that the proteins migration behavior, viewed in the virtual 2D electrophoresis gel, allowed the segregation of different paralogs (WRKY 28, 33 and 75) in separated groups.

Finally, the identification, characterization and classification of TFs at the genome-wide level will provide an important resource for researchers who are interested in studying the regulation of gene expression. Using bioinformatic approaches, as in the present work, computational studies have documented valuable TF repertoires by searching for genes containing DBDs within individual organisms ranging from prokaryotes to eukaryotes or by searching across all completely sequenced genomes [4, 33, 34, 35, 36, 37].

It is desired that this study will contribute in the elucidation of the complex regulatory mechanisms associated with abiotic stress responses in legumes. However, further functional studies are required to understand the role of specific TF networks activated. The greater understanding of how TFs operate may be translated into their potential applications to enhance plant productivity, seed quality, nitrogen fixation and the sensing/response and adaptation to the environment.

Conclusions

The cowpea transcriptome seems to present the main plant-exclusive transcriptions factors that are involved in the perception and, subsequently, repression/activation of genes that participated in the abiotic response signaling pathways. Despite of the similarity of cowpea clusters with arabidopsis sequences, the results showed that legume plants presented some particularities in their sequences. Considering that no legume-specific gene was identified, it seem that cowpea have succeeded in developing specific traits by diverting some TF genes for a more specialized function.

Grouping TFs by their structural domains has been extremely useful in gaining insights into how they recognize and bind specific DNA sequence. In this sense, it is plausible to assume that the genes identified in this work are functional, participating, among others, in the main pathways responsive to

abiotic stress. Moreover, the DBD may provide clues to their biological function. This strategy has also been proven successfully for characterizing the cowpea WRKY evolutionary history as well. With respect to expression data, it is noteworthy that after the salinity stress perception many genes are repressed, while some are activated. Moreover, the majority genes were expressed in the stressed library, with a high amount of TFs being constitutively expressed.

The phenetic (NJ) analysis was enough to separate the sequences according to their grouping and also allowed the differentiation, within each clade, of the organisms based on their taxonomic classification.

Since cowpea is one of the most important legume crops, functional characterization of transcription factors genes has important significance for the genetic improvement of cowpea for abiotic stress tolerance. Thus, this study is significant in the elucidation of the role of DREB/ERF, NAC and WRKY-type transcription factors in the responses of legume plants to the salinity stress. The development of tolerant cultivars to abiotic stresses, as well as the development of technologies to over express TFs in transgenic plants may help in the tolerance and surveillance during drought and salinity stress, collaborating to an increase in the world crop production.

Methods

Complete protein sequences of transcription factor genes from *A. thaliana* (Table 2) were obtained in fasta format from TAIR database (<http://www.arabidopsis.org>), based on Seki *et al.* (2002) [10], Yamasaki *et al.* (2008) [1] and Jiang and Deyholos (2009) [12] data. Each sequence was compared against the partial data from *V. unguiculata* sequencing project (NORDEST) using tBLASTn tool. Only sequences with significant similarity (e-value e^{-15} or less) were considered.

Cowpea DNA sequences were translated with the aid of ORFfinder program (www.ncbi.nlm.nih.gov/gorf/gorf.html) and screened for similarity to entries in GenBank (NCBI; <http://www.ncbi.nlm.nih.gov/>), using mainly BLASTx and BLASTp algorithms function of basic and advanced BLAST search [38]. Moreover, the RPS_BLAST option was used to inquire the presence and integrity of the CDs.

Based on the obtained clusters from the searches at the local database of *V. unguiculata* ESTs, a screening using BLASTn algorithm was carried out for sequences with identity equal to or greater than 24 bp at the SuperSAGE tags database, which was composed by 631,135 tags, being 293,847 exclusive to control experiments and 239,211 exclusive to salinity experiments. These results allowed quantitative

and qualitative analyses of the tags per library and family of transcription factor. Afterward, the tags, and the information about their frequency, were retrieved from the cited local database and submitted to the DiscoverySpace 4.0 program [39], a tool that has particular strength in the analysis of SAGE data based on the Audic-Claverie [40] significance test, with the aim to identify differential expression patterns of the tags, and consequently the genes target of this studied.

A multiple alignment at CLUSTALW program [41] was generated using EST sequences converted to aminoacids, aiming to evaluate the conservation among WRKY sequences of *V. unguiculata* and other organisms; for these purpose, only the c-terminal WRKY domain region of the sequences was used. The MEGA software (Version 4.0) allowed the generation of dendrograms using the Neighbor-Joining method [42] and Bootstrap test of inferred phylogeny with 2,000 replications. Moreover, multiple alignments using the complete aminoacids sequences were performed at CLUSTALW program [41] to evaluate the degree of similarity within and between the groups of orthologs and paralogous sequences.

The analysis of WRKY proteins migratory behavior, using cowpea best hits, arabidopsis seed sequences and sequences from other organisms, was made in the JvirGel program [43], a tool of the Swiss-Prot database, which offers different analytical modes, like a serial calculation of MW (Molecular Weight) and pI (Isoelectric Point). Afterwards, graphic representation was constructed correlating the isoelectric points and the molecular weight.

Acknowledgements

The authors thank to CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FACEPE (Fundação de Amparo à Ciência e tecnologia do Estado de Pernambuco), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), BNB (Banco do Nordeste do Brasil) and FINEP (Financiadora de Estudos e Pesquisas) for the financial support.

References

1. Yamasaki K, Kigawa T, Inoue M, Watanabe S, Tateno M, Seki M, Shinozaki K, Yokoyama S (2008) Structure and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiol Biochem* 46: 394-401.
2. Liu L, White MJ, MacRae TH (1999) Transcription factors and their genes in higher plants: functional domains, evolution and regulation. *Eur J Biochem* 262: 247-257.

3. Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell* 10: 1075-1082.
4. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C-Z, et al. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Sci* 290: 2105-2110.
5. Singh KB, Foley RC, Oñate-Sánchez L (2002) Transcription factors in plant defense and stress response. *Curr Opin Plant Biol* 5: 430-436.
6. Agarwal PK, Agarwal P, Reddy MK, Sopory SK (2006) Role of DREB transcription factors in abiotic stress tolerance in plants. *Plant Cell Rep* 25: 1263-1274.
7. Tran LS, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, et al. (2004) Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell* 16: 2481-2498.
8. Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, et al. (2003) Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Res* 10: 239-247.
9. Pinheiro GL, Marques CS, Costa MDBL, Reis PAB, Alvez MS, et al. (2009) Complete inventory of soybean NAC transcription factors: sequence conservation and expression analysis uncover their distinct roles in stress response. *Gene* 444: 10-23.
10. Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, et al. (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought cold and high-salinity stresses using a full-length cDNA microarray. *Plant J* 31: 279-292.
11. Chinnusamy V, Schumaker K, Zhu J-K (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signaling in plants. *J Exp Bot* 55:225-236.
12. Jiang Y-Q, Deyholos MK (2006) Comprehensive transcriptional profiling of NaCl-stressed Arabidopsis roots reveals novel classes of responsive genes. *BMC Plant Biol* 6: 25.
13. Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci* 5:199-206.
14. Matsumura H, Yoshida K, Luo S, Kimura E, Fujibe T, et al. (2010) High-Throughput SuperSAGE for Digital Gene Expression Analysis of Multiple Samples Using Next Generation Sequencing. *PLoS ONE* 5(8): e12010. doi:10.1371/journal.pone.0012010
15. Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218: 1-14.
16. Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran L-S P (2009) In silico analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. *DNA Res* 16: 353-369
17. Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS et al. (2002) Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell* 14: 559-574
18. Udvardi MK, Kakar K, Wandrey M, Montanari O, Murray J, et al. (2007) Legume transcription factors: global regulators of plant development and response to the environment. *Plant Physiol* 144: 538-549.
19. Gruber V, Blanchet S, Diet A, Zahaf O, Boualem A, et al. (2009) Identification of transcription factors involved in root apex responses to salt stress in *Medicago truncatula*. *Mol Genet Genomics* 281: 55-66.
20. Libault M, Joshi T, Benedito VA, Xu D, Udvardi MK, et al. (2009) Legume transcription factor genes: what makes legumes so special? *Plant Physiol* 151: 991-1001.

21. Zhou QY, Tian AG, Zou HF, Xie ZM, Lei G, et al. (2008) Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. *Plant Biotechnol J* 6: 486–503.
22. Chen M, Xu Z, Xia L, Li L, Cheng X, et al. (2009) Cold-induced modulation and functional analyses of the DRE-binding transcription factor gene, GmDREB3, in soybean (*Glycine max* L.). *J Exp Bot* 60: 121–135.
23. Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, et al. (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem Biophys Res Commun* 290: 998-1009.
24. Zhang G, Chen M, Chen X, Xu Z, Guan S, et al. (2008) Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (*Glycine max* L.). *J Exp Bot* 59: 4095-4107.
25. Fang Y, You J, Xie K, Xie W, Xiong L (2008) Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Mol Genet Genomics* 280: 547-563.
26. Meng Q, Zhang C, Gai J, Yu D (2007) Molecular cloning, sequence characterization and tissue-specific expression of six NAC-like genes in soybean (*Glycine max* (L.) Merrill). *J Plant Physiol* 164: 1002-1012.
27. Karam BS, Rhonda CF, Luis O-S (2002) Transcription factors in plant defense and stress response. *Curr Opin Plant Biol* 5: 430-436.
28. Ramamoorthy R, Jiang S-Y, Kumar N, Venkatesh PN, Ramachandran S (2008) A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormony treatment. *Plant Cell Physiol* 49: 865-879.
29. Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T et al. (2005) Solution structure of an Arabidopsis WRKY DNA binding domain. *Plant Cell* 17: 944-956.
30. Carneiro RT (2007) Análise serial da expressão gênica do caule de plantas de *Eucalyptus grandis* com 6 meses de idade. Thesis. Universidade de São Paulo, São Paulo, SP, Brazil.
31. Berri S, Abbruscato P, Faivre-Rampant O, Brasileiro ACM, Fumasoni I et al. (2009) Characterization of WRKY co-regulatory networks in rice and Arabidopsis. *BMC Plant Biol* 9: 120.
32. Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF et al. (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc Natl Acad Sci U S A* 106: 3853-3858.
33. Iida K, Seki M, Sakurai T, Satou M, Akiyama K et al. (2005) RARTE: database and tools for complete sets of arabidopsis transcription factors. *DNA Res* 12: 247-256.
34. Perez-Rueda E, Collado-Vides J (2000) The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acid Res* 28: 1838-1847.
35. Riaño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B (2007) PlnTFDB: an integrative plant transcriptional factor database. *BMC Bioinformatics* 8: 42.
36. Park J, Park J, Jang S, Kim S, Kong S et al. (2008) FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics* 24: 1024-1025.
37. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA (2008) DBD-taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acid Res* 36: D88-92.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 251: 403-410.
39. Robertson N, Oveisi-Fordorei M, Scott D Zuyderduyn, Richard J et al. (2007) DiscoverySpace: an interactive data analysis application. *Genome Biol* 8:R6

40. Audic S, Calverie JM (1997) The significance of digital gene expression profiles. *Genome Res* 7: 986-995.
41. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.
42. Sneath PHA, Sokal RR: Numerical taxonomy. San Francisco: Freeman 1973
43. Hiller K, Schobert M, Hundertmark C, Jahn D, Münch R (2003) JVirGel: calculation of virtual two-dimensional protein gels. *Nucleic Acids Res* 31: 3862-3865.

Figures

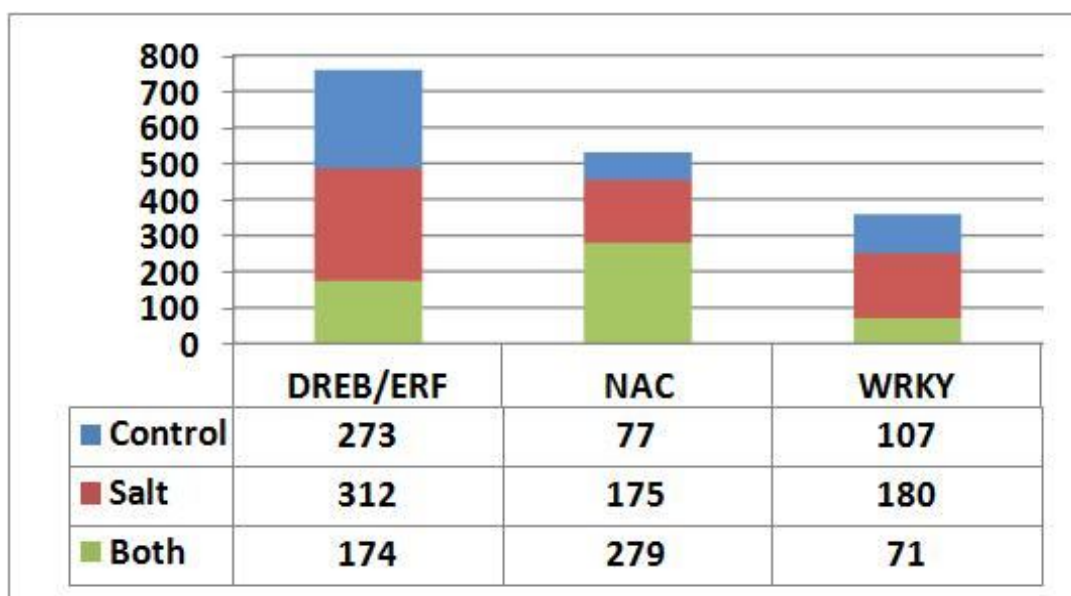


Figure 1. Graphic representation of the diversity and abundance of SuperSAGE tags for each transcription factor family. Numbers refer to absolute values.

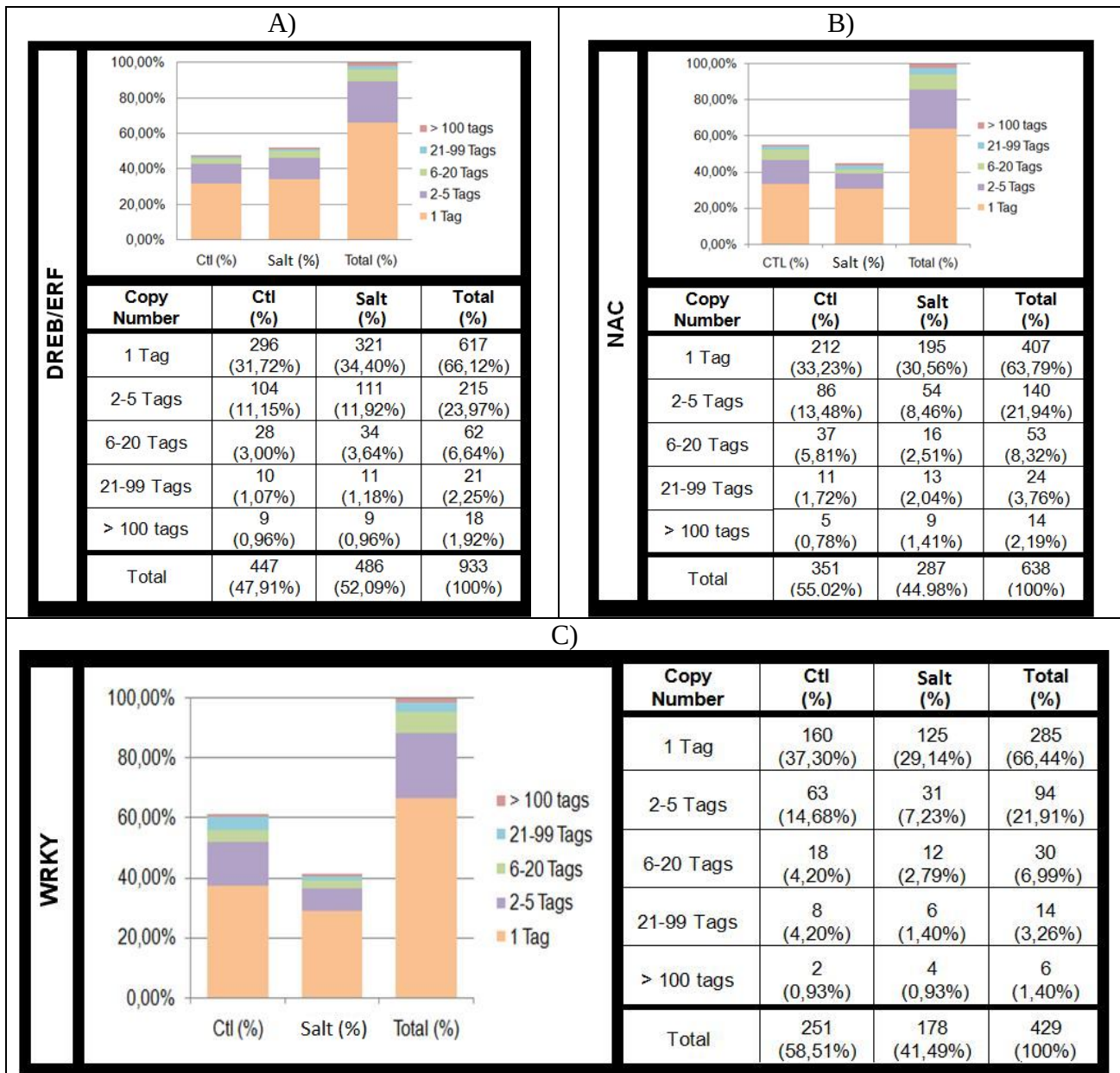


Figure 2. Quantitative distribution of SuperSAGE tags. Graphic visualization of tags quantitative distribution (number of copies and libraries, values in percentage) and table showing the tags frequencies (absolute and percentage values) in relation to the counting number in libraries; both analysis are evidenced for DREB/ERF (A), NAC (B) and WRKY (C) families.

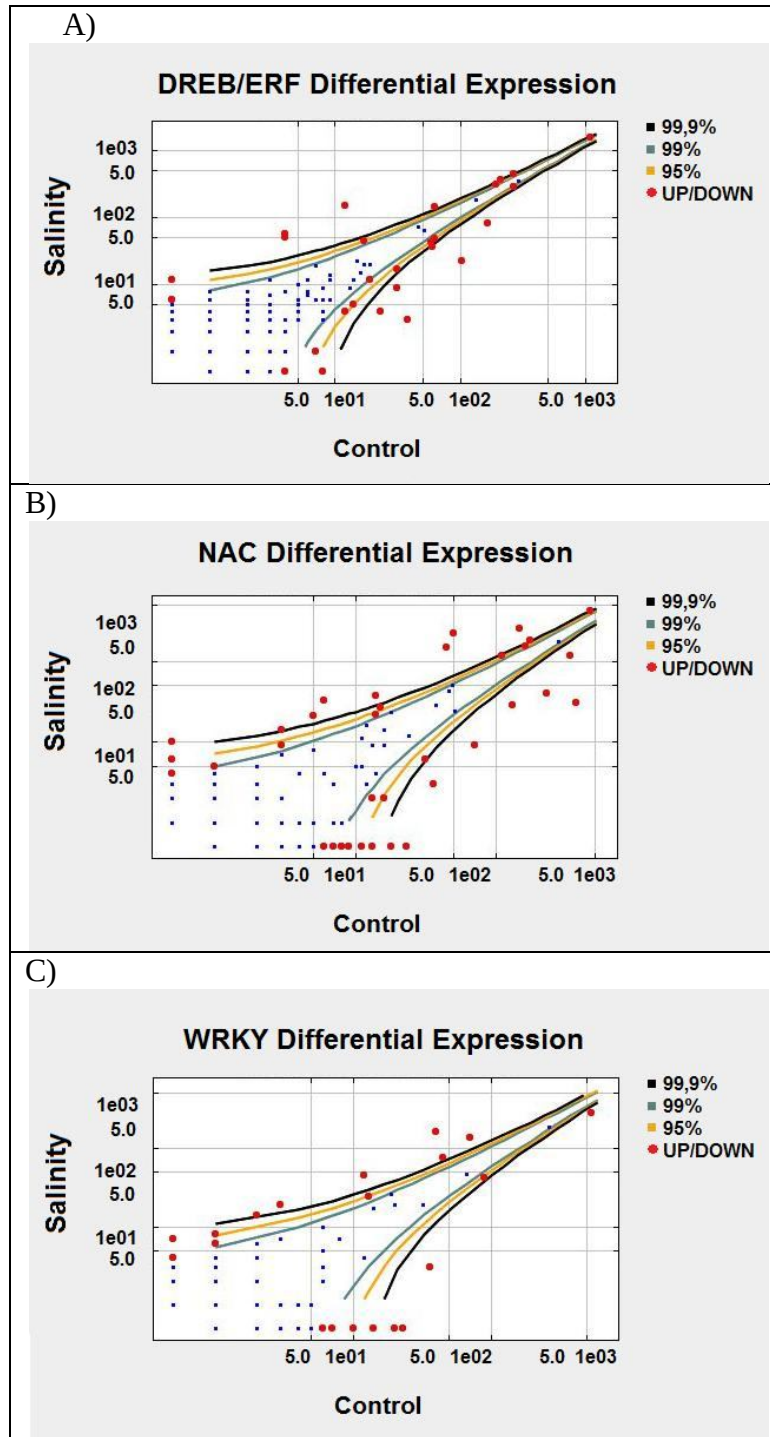


Figure 3. Graphic representation showing the distribution of the differentially expressed transcripts (red circles), considering the libraries comparatively for DREB/ERF (A), NAC (B) and WRKY (C) families.

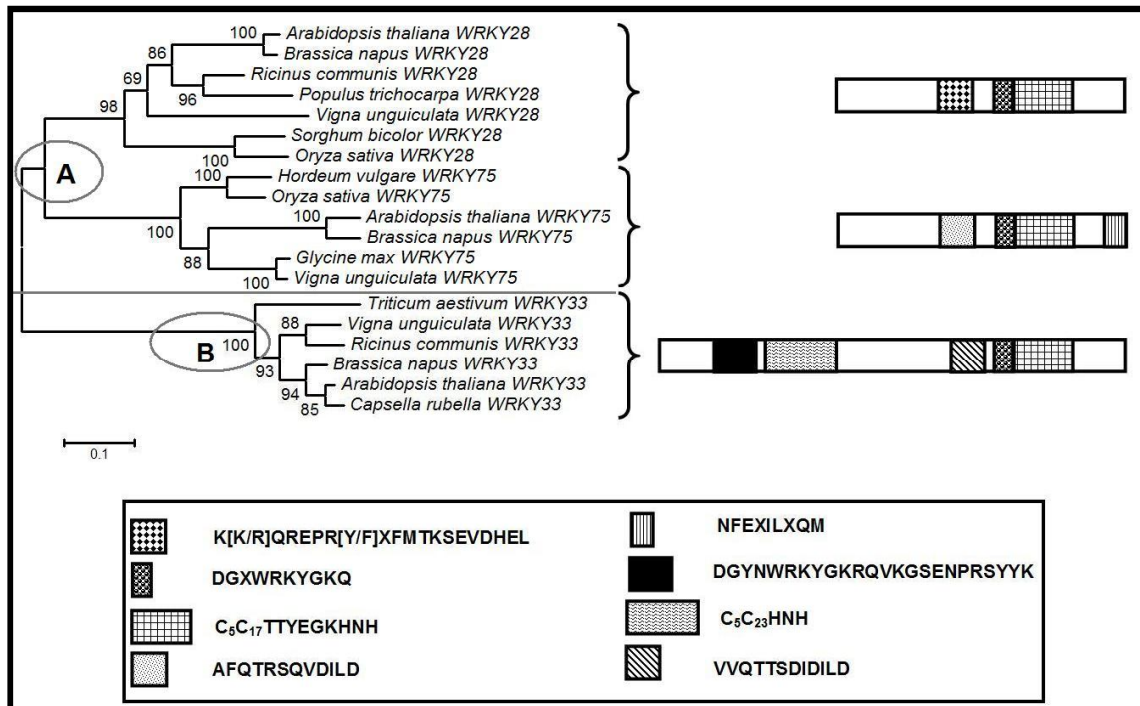


Figure 4. Dendrogram generated after Neighbor-Joining analysis, showing relationships among amino-acid sequences of WRKY candidates. The numbers on the left of the branch nodes are p values. Letters A and B in the nodes indicate the grouping formed by sequences from Class II and I, respectively. Bar represents similarity coefficient. The Graphic representation on the right side of the dendrogram represents genes structure, including their patterns of conservation, viewed at the multiple alignments.

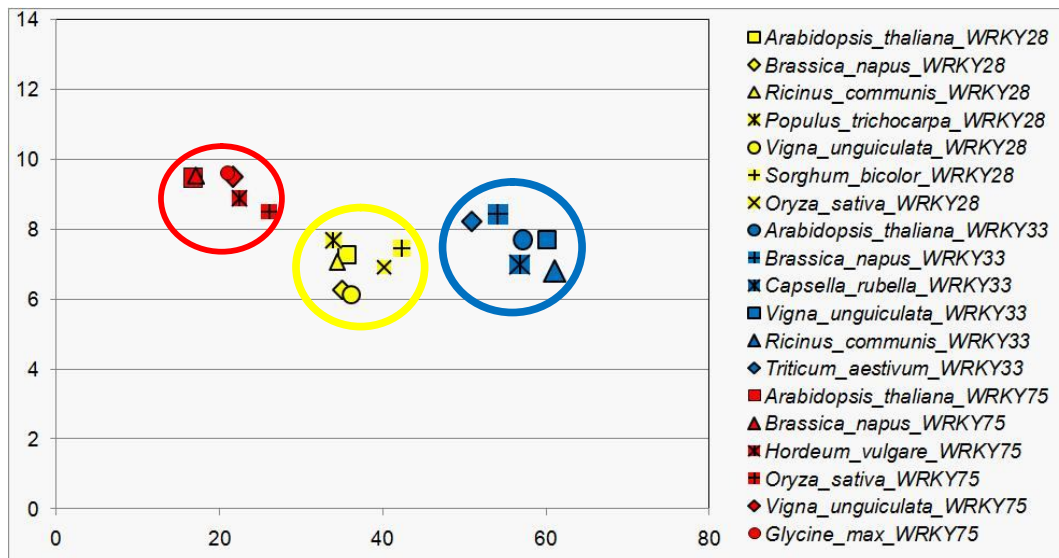


Figure 5. Graphic representation of the virtual 2D electrophoresis of WRKY transcription factor. Proteins delimited by a yellow circle include the WRKY28 members. Proteins delimited by a blue circle include the WRKY33 members. Proteins delimited by a red circle include the WRKY75. Abscissa (x-axis) represents isoelectric point, Ordinate (y-axis) represents molecular weight.

Tables

Table 1. Data mining resources and results using *Arabidopsis thaliana* seed sequences, matches of cowpea with respective informations about the alignments against NORDEST and GenBank databases, ORF and domains. Abbreviations: DREB, Dehydration Responsive Element Binding; ERF, Ethylene Responsive Factor; G. max, *Glycine max*; P. trichocarpa, *Populus trichocarpa*; L. japonicus, *Lotus japonicus*; M. truncatula, *Medicago truncatula*; V. vinifera, *Vitis vinifera*; V. thumbergii, *Vitis thumbergii*; cp, complete; inc, incomplete.

Transcription Factor Information			tBLASTn - NORDEST database				BLASTn - GenBank database				ORF		Domain	BLASTp - GenBank Database		
Family	Gene	Subject id	E-value	Score	Other	NCBI Access/ Organism	Score	E-value	Frame / Length / Integrity					NCBI Access/ Organism	Score	E-value
DREB/ERF	DREB1A	CL8528Contig1	8E-41	162	7	gb AF37C735.1 / G. max	776	0.0	3 / 219 / cp	AP2				gb AAQ02703.1 / G. max	345	2,00E-93
	DREB1B	CL8528Contig1	1E-39	159	7	gb AF37C735.1 / G. max	776	0.0	3 / 219 / cp	AP2				gb AAQ02703.1 / G. max	345	2,00E-93
	DREB1C	CL8528Contig1	2E-44	175	9	gb AF37C735.1 / G. max	776	0.0	3 / 219 / cp	AP2				gb AAQ02703.1 / G. max	345	2,00E-93
	DREB2A	CL9195Contig1	6E-28	121	2	gb BT095431.1 / G. max	508	3E-138	2 / 185 / cp	AP2				gb AAT12423.1 / G. max	266	5E-70
	DREB2B	CL9195Contig1	1E-29	126	2	gb BT095431.1 / G. max	508	3E-138	2 / 185 / cp	AP2				gb AAT12423.1 / G. max	266	5E-70
	DREB_61403.m00C80	CL8066Contig1	1,00E-37	152	7	gb BT093306.1 / G. max	518	4E-143	1 / 292 / cp	AP2				ref XP_002307098.1 / P. trichocarpa	279	2E-73
	DREB_51596.m00120	VUPISS08022E07.b00	3,00E-27	117	8	gb DQ208969.1 / G. max	244	5,00E-61	2 / 196 / cp	AP2				gb ABB366.46.1 / G. max	167	3E-40
	DREB_49260.m00C84	CL382Contig3	2,00E-28	122	37	db AK2385635.1 / G. max	566	4,00E-127	3 / 361 / cp	AP2				gb AAZ03388.1 / G. max	371	6E-101
	DREB_67169.m00C02	CL382Contig3	1,00E-24	110	38	db AK2385635.1 / G. max	566	4,00E-127	3 / 361 / cp	AP2				ref XP_002307098.1 / P. trichocarpa	279	2E-73
	ERF_50980.m00070	gil190506633	2,00E-23	105	11	db JAB378637.1 / L. japonicus	356	9,00E-95	2 / 154 / inc	-				gb AAZ03388.1 / G. max	371	6E-101
WRKY	ERF_68169.m01270	CL2369Contig1	1,00E-38	155	16	db JAK243009.1 / G. max	660	0.0	3 / 248 / cp	AP2				emb CAN66025.1 / V. vinifera	84	5E-15
	WRKY28	CL5559Contig1	2,00E-37	152	21	gb DC322698.1 / G. max	1254	0.0	1 / 320 / cp	WRKY				gb ABD040237.1 / M. truncatula	331	3E-89
	WRKY33	CL5023Contig1	2,00E-93	339	23	db JAK244489.1 / G. max	2239	0.0	1 / 548 / cp	2x WRKY				gb IACB26920.1 / G. max	509	1E-142
	WRKY75	CL8611Contig1	2,00E-41	163	11	gb BT098641.1 / G. max	643	0.0	1 / 195 / cp	WRKY				emb CAP08303.1 / V. trunbergii	709	0.0
NAC	NAC_51786.m00193	CL1656Contig1	1,00E-115	412	44	AK244623.1 / G. max	974	0.0	1 / 295 / cp	NAM superfamily				ACD39369.1 / G. max	517	6e-145
	NAC_51903.m00352	gil182651054	1,00E-47	184	32	XM_002315300.1 / P. trichocarpa	316	8E-83	3 / 228 / inc	NAM superfamily				NP_566374.1 / A. thaliana	317	5e-85
	NAC_67165.m00132	CL4019Contig1	1,00E-89	326	40	DQ202772.1 / G. max	857	0.0	1 / 219 / inc	NAM superfamily				AA05961.1 / G. max	398	2e-109
	NAC_67745.m00110	gil182647558	2,00E-86	315	41	EU661925.1 / G. max	924	0.0	2 / 228 / inc	NAM superfamily				ACD39383.1 / G. max	453	7e-126
	NAC_67913.m00149	CL1Contig251	7,00E-92	333	41	EU661926.1 / G. max	1009	0.0	1 / 295 / cp	NAM superfamily				ACU39384.1 / G. max	529	1e-148

Table 2. Information about *Arabidopsis thaliana* seed sequences, including gene name, and locus (sequences available at <http://arabidopsis.org>).

Transcription Factor Family	Gene	Locus
DREB/ERF	DREB1A	At4g25480
	DREB1B	At4g25490
	DREB1C	At4g25470
	DREB2A	At5g05410
	DREB2B	At3g11020
	DREB_61403.m00080	At1g22190
	DREB_51596.m00120	At1g74930
	DREB_49260.m00084	At2g20880
	DREB_67169.m00002	At4g28140
	ERF_50980.m00070	At1g43160
	ERF_68169.m01270	At4g17500
NAC	NAC_51786.m00193	At1g01720
	NAC_51903.m00352	At1g32870
	NAC_67165.m00132	At4g27410
	NAC_67745.m00110	At5g39610
	NAC_67913.m00149	At5g63790
WRKY	WRKY28	At4g18170
	WRKY33	At2g38470
	WRKY75	At5g13080

Deciphering the Plant Transcription Factor World

Running Title: Plant Transcription Factor World

Nina M. Soares-Cavalcanti¹; Luis C. Belarmino¹; Ana C. Wanderley-Nogueira¹; Ederson A. Kido¹; Sergio Crovella¹; Ana M. Benko-Iseppon^{1*}.

¹Universidade Federal de Pernambuco, Center of Biological Sciences, Department of Genetics, Laboratory of Plant Genetics and Biotechnology, R. Prof. Moraes Rêgo, 1235, CEP 50670-420, Recife, PE, Brazil.

*Corresponding Author: ana.benko.iseppon@pq.cnpq.br

Abstract

Transcriptional networks mediated by Transcription Factors (TFs) control the responsive processes to a variety of stresses in multicellular organisms, by specifying where and when genes will be expressed. The activation/repression of key genes is fundamental to expression control of most pathways. Actually, it is observed that transcriptional regulation occurs first binding the TFs to *cis*-regulatory modules within upstream noncoding sequences (called ‘promoters’) through the DNA Binding Domains (DBDs) that targets specific motifs in the sequence of interest. However, there are many reports in plants and animals in which regulatory elements, within a given transcribed region, are necessary for the correct establishment of the mRNA pattern. Moreover, many post-transcriptional regulatory mechanisms have been described. Bioinformatics is a powerful tool to disclose these networks mediated by transcription factors; the use of different softwares allow the elucidation of the function of given genes, including regulation and synthesis modes, post-transcriptional modifications, interaction partners, and cellular localization, among other. However, since more and more high quality data are available, the scientists are faced with the challenge of integrating different approaches in a single intelligent software. An overall evaluation of such factors in plants will be presented, regarding their diversity and performance and showing how these mechanisms interact to modulate diverse molecular pathways and how they can be analyzed using bioinformatics and systems biology tools.

Keywords: bioinformatics, *cis*-acting binding elements, signaling, transcription regulation.

Introduction

Mutant research in plants has shown that informative phenotypes are currently correlated to mutations in genes coding transcription factors, and an increasing number of TFs have been described as key regulators of various plant functions. The importance of the transcription factors role in plants could be verified by their amount in *Arabidopsis thaliana*, where more than 1,600 TF genes were described [1]. Furthermore, plants present exclusive factors which are not shared with other organisms. Thus, as expected, this high amount of transcription factors is kept up with an equally high amount of diversified specific binding sites.

Elucidating the molecular mechanisms of abiotic stress tolerance in plants is a critical point for strategy creation aimed to increase crop production and quality.

This paper reviews the main characteristics of plant TFs, describing their main families, activation mechanisms and those involved in the target gene regulation, also describing a pipeline for their identification and classification.

1. Transcription factor structure and function

The regulation of gene expression is an essential process that is highly conserved from humans to bacteria. In the course of the evolution complex mechanisms were developed in order to control the transcription. Gene expression at the level of mRNA is regulated through different processes, however the most widely studied is the regulation by transcription factors (TFs). The nucleus of eukaryotic cells contains similar mass quantities of DNA and proteins; many of these, such as the histones, participate in organizing the DNA into higher-order structures, while others, like the transcription factors, are involved in regulating the expression of a given gene at a particular time and place [2].

All biological processes are directly regulated or influenced by TFs. These factors regulate the first step of gene expression, being usually defined as proteins containing a DNA-binding domain (DBD) that recognizes a specific DNA sequence. In addition, proteins without a DBD, which interact with a DNA-binding protein to form a transcriptional complex, are often categorized as TFs [3].

One way to achieve the expression control is to regulate the initiation of transcription. This step requires the assembly of two main components: (I) the basal/general machinery, that is common to all expressed genes and is composed of the RNA polymerase (RNA pol) and general transcription factors (GTFs), e.g. TFIIA, and (II) a specific/regulatory machinery, which differs among genes and is the responsible for proper gene expression in response to environmental or developmental signals. The specific apparatus consists mainly of transcription factors (trans-acting factors, TFs), proteins that regulate the initiation of transcription, and thus its rate, in a spatio-temporal manner [4].

Transcription factors are DNA-binding proteins that command the initiation of transcription, one of the most important control points in the gene expression regulation. With few exceptions, a typical plant transcription factor contains a DNA-binding domain, a transcription regulation domain (activation domain-AD or repression domain-RD), an oligomerization site and a nuclear localization domain. Most TFs exhibit only one DNA-binding type and oligomerization domain, sometimes in multiple copies, but some contain two distinct types. The regions of DNA-binding are normally adjacent to or overlap with the oligomerization sites, and their combined tertiary structure determines critical aspects of transcription activity [5]. Normally, a TF has a DBD and a transcriptional AD or RD domains. TFs having only a DBD are likely to be passive repressors [6].

It is important to note that these factors were identified and classified based on their DBDs and that the TF families were denominated according to the names of respective DBDs (e.g. WKY, NAC, and HSF, among other) [7].

A critical component of transcription control depends on sequence-specific binding of multiple TF to short DNA sequences (13 bp on average) in the relative vicinity of target gene [8]. Mutations in the transcription factor binding sites (TFBS) are known to underlie a substantial component of phenotypic variability within and across species [9]. TFs exert gene-specific and/or tissue-specific functions by binding to specific cis-acting regulatory elements (CAREs) in the promoter of target genes, enhancing or repressing their transcriptional rates. They can bind not only near or far away, but also up- or downstream of the TSS (transcription start site) of the gene they control. Through interaction with the basal apparatus these factors are responsible to regulate the levels of transcription in response to different stimuli [4; 10].

These CAREs are short conserved motifs of five to 20 nucleotides found in the neighborhood of the 5' end of genes, the so called promoter region. The promoter sequence is usually located upstream from the TSS, but regulatory elements can also be located downstream (into first intron of the gene itself, for example) [11].

The promoter region can be divided into two parts: a proximal part - referred to as the core - and a distal part. The first one is believed to be responsible for correctly assembling the RNA polymerase II complex at the right position and for directing a basal level of transcription; it is mediated by elements, such as TATA and Initiator boxes through the binding of the TATA box-binding protein, and other general TF proteins specific for the RNA polymerase II [12]. On the other hand, the distal part is believed to contain those elements that regulate the spatio-temporal expression [13]; how far upstream (or downstream) such a distal part reaches is still not defined [14].

TFs can act both as transcriptional activators and as repressors. In some eukaryotes TFs containing domains rich in the acidic amino acids glutamine or proline (such as TOC1, DREBs, ARFs and GBF1) are transcriptional activators [15; 16]. Also the AHA motif, characterized by an aromatic pattern and large hydrophobic amino acid residues embedded in an acidic context, seems to act as an activation domain (AD) of heat shock factors of plants [17].

Transcriptional repressors are categorized into passive or active. Passive repressors have neither an activation domain (AD) nor a repression domain (RD). A few act repressing the transcription by binding to the promoter of the target gene, thereby competing with an activator that interacts with cis-elements [6].

Active repressors have distinct RDs which confer repressive activity to the TF. For example, the EAR motif is a component of repression domain specific to plant. The minimum unit of the EAR-motif RD has just six amino acids, which includes an amphiphilic feature composed of acidic amino acids and leucine [18]. Because the fusion of EAR motif, repression domain can be converted from a transcriptional activator into a repressor [19], the TFs containing this motif are considered as transcriptional repressors. Database analysis indicated the presence of EAR motif RD in 404 loci among 2,620 putative TFs [6]. Most RDs are conserved in many plant species, including dicots and monocots; however these domains in other organisms are less represented (N. Mitsuda *et al* unpublished results). Together these data indicate that the EAR motif RD and its mechanism of action are specific to plants [6].

Bifunctional TFs have been reported acting either as transcriptional activators or as repressors, as the plant WRKY53 in which the activation or repression function depends on the sequence surrounding the W-box [6].

To characterize the transcription regulation processes of a given pathway, two fundamental points should be mentioned: (I) not necessarily the high homology between DBDs is related to the similarity of target motif; for example, bZIP factors from tobacco (RSG2) and tomato (VSF-1) present a highly conserved bZIP domain, however they bind to completely different motifs [20; 21], and (II) as the stress sensors are not known and most signaling intermediates are not identified, there is little information about cross-talk between different stress signal transduction pathways in plants; for example, some WRKY TFs participate in different pathways in response to abiotic and biotic stress.

In plants, a large number of transcription factors (TFs) mediate the expression control of tens or hundreds of target genes in various signal transduction cascades [22; 23]. For example, the arabidopsis genome presented at least 1,554 TFs, corresponding to more than 5.9% of the coding genes, a higher amount than that of *Drosophila melanogaster* (4.7 %) and of *Caenorhabditis elegans* (3.6 %) [24]. By the fact that arabidopsis, besides presenting a larger number of TF genes, presents a high variety of factors with a greater diversity of DNA binding specificities, when compared with *D. melanogaster* or *C. elegans*, it is believed that transcriptional regulation plays a more important role in plants than in animals [6].

In rice, Xiong *et al.* (2005) [25] identified 1,611 TF coding genes and perceived that compared to 1,510 TF genes from arabidopsis nearly half of the genes formed clear orthologous pairs or clusters, which seemed to be derived from 383 genes in the common ancestor of angiosperms.

To evaluate the abundance of TF genes, Mitsuda *et al* (2009)[6] analyzed four arabidopsis TF databases, RARTF [26], AGRIS [27], DATF [28] and PlnTFDB [29], and observed, respectively, a total of 51, 51, 64 and 67 TFs families (72 families in total) and 1,965, 1,837, 1,914 and 1,949 loci (2,620 loci in total).

Since the majority of the “plant-specific” DBDs originated from non-plant species and expanded their population along with the evolution of higher plants, to elucidate the functions and evolutionary relationships of the transcription factors the structures of DBDs should be revealed [7]. In 2005 Shiu *et al.* [30], through comparative analysis, observed that expansion rates of plant TFs were much higher than those of animals; this finding may be primarily justified by the relatively high frequency of polyploidization in plants. It was also observed in plants genomes that expansion rates of TFs are higher than other genes, and that this feature was probably a requirement that resulted in determining the population of transcription factors. Seki *et al.* (2002) [31] showed that transcription factors belonging to different families can be involved in the same category of reaction, such as drought response. In addition, transcription factors sharing the same DBD may have a cross-reaction within each other through binding the same sequence of the target gene promoter [7].

While in prokaryotes or other non-plants eukaryotes the majority of DBDs recognize DNA through helices, such as the helix-turn-helix, many “plant-specific” DBDs bind to DNA through b-sheet regions. The extensive lineage specific expansion of transcription factors in plants seems to justify this difference [7].

Finally, once various ‘omes’ (proteome, metabolome and phenome) could be affected by transcriptional regulation, the structural and functional analysis of TFs is an important and necessary process for omics studies and for the elucidation of whole functional networks in plants [6].

1.1. DREB/ERF families

DREB (Dehydration Responsive Element-Binding) and ERF (Ethylene-Responsive Factor) families integrate the AP2/EREBP (Ethylene-Responsive-Element-Binding Protein) superfamily of transcription factors exclusive of plants [32].

The proteins from these two families present the DNA-binding AP2 domain, that is composed by 60 amino acids and possess two distinct regions: (I) the YRG element, with 20 amino acids rich in basic residues in the N-termini portion,

and (II) the RAYD element, that forms an amphipathic alpha helix in the C-termini portion [33]. The DNA-binding is YRG-dependent, while RAYD may contribute to this interaction or mediate the protein-protein interactions [34]. In addition, two residues in the AP2 domain, valine-14 and glutamate-19, play crucial role in the specificity determination of DNA binding [16].

The AP2 DBD can bind to DNA through cis-acting elements: the GCC Box (5'-AGCCGCC-3'), found in pathogen resistance genes, and CRT/DRE (C-Repeat/Dehydration Responsive Element) Box (5'-TACCGACAT-3'), involved in cold and dehydration responses [35].

1.2. MYB family

The MYB TF gene family is present in all eukaryotic kingdoms. Although presenting few members per haploid genome in non-plant organisms, the amount of representatives in plants is considerably higher. At least 85 R2R3 MYB genes were observed in arabidopsis [36].

Even though such genes present discrete expression, a search in EST databases of non model plants as sugarcane and eucalyptus revealed 42 and 71 representatives, respectively [37].

In animals MYB proteins present three MYB repetitions (R1, R2, and R3). In plants, almost all proteins possess two imperfect repetitions; however, both proteins with one or three repetitions have been already reported. By the fact that R2R3 MYB genes are unique to plants, it is speculated that the same are involved in the regulation of plant-specific processes and functions. Moreover, comparisons among MYB proteins from animals and plants revealed that there is more conservation between repetitions from different organisms than between the repetitions from the same protein [38].

These transcription factors are characterized by a modular structure, comprising functional domains that enable the hard regulation of transcription activation. The DNA-binding domain can be formed by one, two or three variables helix-turn-helix motifs, which is composed by 51-52 amino acids. Such motifs present three tryptophan residues spaced regularly that play a role in the folding of the domain hydrophobic core [39]. The target sequence of MYB domain in the promoter region is a motif called MYB-binding site (MBS), which may be of three types: MBS1, MBS2 or MBS2G [40].

1.3. WRKY family

WRKY transcription factors are one of the largest families of transcriptional regulators in plants and form integral parts of signaling pathways that modulate many plant processes such as senescence, morphogenesis, abiotic stresses, among other [41]. According Wu *et al.* (2005) [42] and Robatzek and Somssich (2001) [43] the *Oryza sativa* and *A. thaliana* genomes include 102 and 74 putative WRKY genes, respectively.

WRKY proteins often act as repressors as well as activators, and members of the family play roles in both the repression and de-repression of important plant processes. Furthermore, it is becoming clear that a single WRKY transcription factor might be involved in regulating several seemingly disparate processes [44]. The family name is derived from its main characteristic, the WRKY domain, a region of 60 highly conserved amino acids. This domain possesses the WRKYGQK conserved sequence followed by a zinc finger motif C2H2 or C2HC [45], which seems to participate of the DNA binding process [46].

Based on the coding proteins ability to bind specifically to the (T)(T)TGAC(C/T) motif, also called W Box, the first WRKY cDNAs, from *Ipomoea batatas* (SPF1) and *Arabidopsis thaliana* (ZAP1), were cloned [45]. The proteins can present one or two WRKY CDs, being classified as group I, when one domain is found, or group II, when presenting two domains [45].

1.4. bZIP family

All analyzed eukaryotic organisms present the bZIP (Basic Region/Leucine Zipper) transcription factors. Characteristically their sequences contain a basic region mediating sequence-specific DNA binding site and a leucine zipper dimerization motif [47]. The arabidopsis genome exhibits approximately four fold more bZIP genes than yeast, worms (*C. elegans*) and humans [24]. In *A. thaliana* molecular and genetic studies of these genes showed that they function in the regulation of many biological processes such as defense against pathogen attack, seed maturation, development of flowers and abiotic stress tolerance [47].

The bZIP domain consists of two characteristic structures located in adjacent alpha helices: (I) a basic region, composed by approximately 16 amino acids containing a nuclear localization signal followed for the invariably motif N-X7-R / K interacting with DNA; and (II) periodic repetition of leucine residues, or other hydrophobic amino acids, every seven positions by a distance that contains eight helical turns, creating an amphipathic helix; these repeats are positioned exactly nine amino acids from the C-terminal portion of the basic region [47].

Plant bZIP proteins preferentially bind to DNA through the central ACGT sequence, in A BOX (TACGTA), C BOX (GACGTC) and G BOX (CACGTG). Furthermore, the specificity of binding is regulated by flanking nucleotides [48].

2. Transcription factors activation

Changes in the TF activity dynamically alter the transcriptome, leading to metabolic and phenotypic modifications. Thus, analysis of TFs structure and function is one of the most important approaches of the post-genome era [6].

Several processes regulate the activity of transcription factors, including transcriptional regulation, post-translational modification and nuclear transport. In some cases, TFs are synthesized and remain as dormant forms in the cytoplasm; being activated through noncovalent interactions with other factors and after post-translational modifications (e.g. phosphorylation), when stimulated [49].

The expression of most TFs is controlled at the transcriptional level by other TFs, while several TFs are regulated post-transcriptionally, such as EIN3 that is controlled by the Ub/26S proteasome pathway [50]. Small RNAs that target TF genes are also important regulators of gene expression; the miR164 family from loci AT2G47585 (miR164A), AT5G01747 (miR164B) and AT5G27807 (miR164C), for example, regulate the expression of some NAC TFs (CUC1, CUC2, ATNAC4, ATNAC5, NAC1 and ORE1) [6]. In addition, TFs regulated at the post-transcriptional level may be regulators acting at early stages of transcriptional cascades [6].

An important molecular feature of the regulatory networks that govern diverse cellular functions is the ability to activate dormant transcription factors. In several families, TF proteins need a physical interaction with an identical molecule or with another molecule within the same family to form a functional dimer and bind to the DNA. The creation of a functional protein complex that is composed of two subunits, is defined as dimerization process. Usually in signal transduction pathways dimeric interactions are not very stable. Rather, they are dynamic and act as reversible switches in the process of information flow [51]. The families of TFs can both form heterotypic dimers (dimer formed between two proteins derived from different gene families) or homotypic dimerization (dimerization among members of the same TF family). Furthermore, two monomers (identical or not) might first dimerize in solution and then bind to the DNA (dimer pathway) or, alternatively, they might first bind to the DNA as monomers and then dimerize (monomer pathway) [52].

Members dimerizing TF families mediate their DNA-binding and dimerization activities through highly conserved domains. The TF groups that form homotypic dimers include, among others, bHLH, bZIP, MADS-box and HD-ZIP families. These TFs form a large number of dimers with distinct biological properties and create an elaborate control circuit that seems to be central to the evolution and generation of organismal complexity [53].

TF dimerization has as most important implications the differential regulation by heterodimerization, facilitated proximity, orientation and enhancement of DNA-binding specificity. Moreover, various DNA elements may be recognized by different combinations of heterodimers, whereas the monomer-to-dimer transition can become an additional point of regulation. Other implications of dimerization are sharp temporal responses to changes in protein concentration, allostery and synergy [54].

Each dimer triggers a sequence of regulatory events leading to a particular cellular fate depending on the choice of partner and the cellular context as, for example, the proliferation or the differentiation. Recent evaluation of genomic and functional data revealed that partner choice is not random; instead, there are some dimerization specificities strongly linked to the protein family evolution [53].

One TF monomer can have multiple binding partners and form dimers that have distinct properties and act in specific functions, interfering directly in the differential gene regulation. In such cases, which dimer will be formed and, consequently, which signaling process going to prevail will be determined by the concentration of each monomer in the cell, its posttranslational modifications (e.g. phosphorylation) and its binding affinity to other monomers [53]. As heterodimers combine different types of DNA-binding domains, each protein complex has different DNA-binding specificities. The various heterodimeric complexes promote the fine-tuning of gene expression through the combination of a limited number of TFs that recognize different promoter elements with various degrees of affinity [53].

So, in a genome the actual number of sequence-specific TF genes could be even higher because most TF families can combine in homotypic/heterotypic and dimeric protein complexes creating a large number of TF dimers with distinct properties [53].

Beyond that, the controlled proteolytic activation of membrane-bound transcription factors (MTFs) is an adaptive strategy which ensures that transcriptional responses to abrupt environmental changes occur rapidly. In this case, the activity of transcription factors is regulated by sequestering these factors within cells; by their physical association with the intracellular membranes the factors are rendered dormant and released by controlled proteolytic cleavage. Intramembrane proteases or ubiquitin-proteasome pathway are the main ways that MTFs are activated [49].

When the TFs are not needed and to provide a direct and rapid mean of regulating their activities, the TF membrane association is a usual strategy that serves to sequester them away from the nucleus. The perception and interpretation of external signals directly by cellular membranes are crucial both as structural barriers, that ensure cellular maintenance, and for their ability to sense the environment [49].

Many plant transcription factors seem to be associated with membranes. Some NAC (NAM, ATAF, and CUC) members are associated with intracellular membranes, such as the endoplasmic reticulum and plasma membranes. Certain

analyses, involving transgenic plants and knockout mutants, have indicated that the NAC MTFs, called NTLs (NTM1-Like), are closely related with the responsive pathways to abiotic stress [55]. Moreover, according to genome-scale analysis, over 10% of all transcription factors described are membrane bound, which indicates that activation of MTFs occur at the genomic level [49].

No less interesting, a frequent mechanism employed by cells to regulate transcription activity is protein phosphorylation. After transcription, phosphorylation events not only modulate the entrance of transcription factor into the nucleus, but also change the activities or DNA-binding ability of TFs [56].

The transcription factor function can be regulated by phosphorylation and dephosphorylation of proteins through at least five different ways: (i) control the length of time transcription factors (or protein kinases and protein phosphatases that regulate their activity) spend in the nucleus; (ii) transcription factors or their co-regulators targeting for proteolytic degradation; (iii) protein-protein interactions modulations between transcription factors, co-regulators and the basal transcription complex; (iv) transcription factor DNA binding regulation; and (v) chromatin structure modification [57].

With respect to regulation of cellular location, the constitutively nuclear transcription factors are phosphorylated and dephosphorylated by protein kinases and protein phosphatases within the nucleus. Otherwise, in many cases, transcription factors transiting between the cytoplasm and the nucleus are regulated by protein phosphorylation/dephosphorylation. This shuttling is an active process that is dependent on the nuclear localization signals (NLSs) and nuclear export signals (NESs) recognition within the primary sequence of transcription factor by proteins of the nuclear import and export machinery [58]. The Accessibility of NLSs and NESs regulation to nuclear import and export proteins via phosphorylation and dephosphorylation processes occur in at least two ways: (i) masking or unmasking the transcription factor signal sequences, or (ii) transcription factor binding modulation to other regulatory proteins which mask or unmask the signal sequences [57].

Rather than protecting against degradation, in proteolysis regulation the mechanisms of phosphorylation of some transcription factors seem to be necessary to facilitate degradation [57]. Protein phosphorylation could regulate the interactions of TFs with many different proteins, including components of the basal transcription complex, co-regulators and other transcription factors. Many transcription factors contain phosphorylation-dependent transcriptional activation domains and, in a few cases, transcriptional repression domains. Despite the information on how phosphorylation potentiates or represses the activity of transcription still poorly known, probably they modulate the transcription factor affinity for co-activators, co-repressors or the basal transcription complex; a process that could occur: (i) blocking directly the interaction by a phosphorylated residue, (ii) conformational change induced by phosphorylation that unmasks or masks a binding surface or (iii) dissociation caused by phosphorylation of an inhibitor molecule to unmask a surface of binding [57].

3. Regulation of Target Genes

A substantial portion of morphological and functional cell attributes is determined at the transcriptional level and involves several steps, including chromatin remodelling in order to allow access to the promoter, RNA pol II recruiting to the gene promoter, transcription initiation, RNA pol II clearing of the promoter, elongation of the nascent RNA molecule and termination of transcription [59].

In eukaryotic cells the transcriptional control involves several levels of regulation. The cellular differentiation and the responses to hormones and signals from neighboring cells regulate the concentrations and activities of activators and repressors that control transcription of many protein-coding genes. The changes in chromatin structure and histone acetylation and deacetylation are, in turn, regulated by these activators and repressors, influencing the ability of general transcriptions factors to bind to promoters. In addition, activators and repressors directly regulate assembly of transcription-initiation complexes and the rate at which they initiate transcription [60].

The explanation of how histone modifications and chromatin structure are intimately linked to the regulatory activity of many TFs is one of the first steps to understand target gene regulation; it has been perceived that TFs classically associated with the plant metabolism control are also associated with factors involved in chromatin function [61].

The chromatin structure control regulates the ability of transcription factors to interact with long stretches of DNA sequences. The basic structural unit of chromatin is the nucleosome, composed of DNA and histone proteins. In condensed chromatin, the amino acid residues at the N-terminus of each histone extend from the surface of the nucleosome; because histone N-termini are rich in lysine residues reversibly modifications can happen by acetylation, phosphorylation and methylation, as well as by the addition of a single ubiquitin molecule [60].

Initially, TFs mediate, through protein-protein interactions, the assembly of the basal transcription machinery resulting in the activation of RNA polymerase II and mRNA synthesis [61]. Transcription factors of eukaryotic organism contain diverse structural motifs that interact with specific DNA sequences. In the DNA-binding domain α -helices are oriented to lie in the major groove of DNA where protein atoms make specific hydrogen bonds and van der Waals interactions with atoms in the DNA. Interactions with sugar-phosphate backbone atoms and, in some cases, with atoms in the DNA minor groove also contribute to binding [60].

The promoters of target genes can be classified as core promoters, defined as the region within 100 bp around the TSS, and proximal promoters, defined as regions further away from the TSS. Core promoters contain binding sites for ubiquitous GTFs, which are critical for polymerase recruiting to the TSS. The proximal promoters present binding sites for activator proteins that interact with the GTFs and can drive the tissue-specific expression. In this last case, the so called CAREs are localized, appearing as single elements, or as modules, where TFs can bind cooperatively. It is important to note that gene promoters with similar patterns of CAREs will have identical or highly similar expression patterns, and will likely be regulated by common TFs [62].

An enhancer, or CARE, usually contains binding sites of multiple activators spatially limited to allow for a stable DNA–protein complex. Two mechanisms are proposed for enhancer activity. The looping theory affirms that once activators bind to the enhancer, the DNA between the enhancer and the core promoter loops out, leaving the activators close to the promoter. Specific protein–protein interactions between activators binding the enhancer and the promoter ensure that the correct target gene is activated. Scanning the DNA is the alternative mechanism which after binding enhancers the activators move continuously along the DNA until they find their target promoter [62].

The access of the transcriptional apparatus to the DNA template and initiation of transcription are, respectively, limited and inhibited by the packaging of DNA into chromatin. So, chromatin conformation is directly related to the repression of transcription. De-repression of transcription by partial unfolding of the chromatin structure probably constitutes an important part of gene regulation, and several TFs and transcriptional co-activators have been shown to disrupt or remodel the chromatin structure [63]. In fact, the three-dimensional configuration adopted after the TF-DNA binding; for example, enable the interaction with other parts to activate the basal transcriptional machinery [64].

Even more, the three-way connection between methylation, gene activity and chromatin structure is a well known process. DNA methylation represses transcription initiation by interfering directly with the binding of transcriptional activators or indirectly by binding proteins with affinity for methylated DNA; methylation seems to actively participate in the regulation of gene expression [65].

During transcription elongation, RNA polymerase and DNA template must rotate relative to each other inducing rotary constraints. Scaffold or matrix attachment regions are involved at this level of gene expression by stabilizing the formation of heterochromatin. These repetitive regions enable the formation of Z-DNA dividing the DNA into topological domains, which are delineated by torsionally locked boundaries [66].

4. Characterizing the transcription factors: an *in silico* routine

Biochemical identification and validation of individual genes involved in such complex regulatory events is a hard work to perform. On the other hand, the bioinformatic is emerging as a useful approach to acquire information by integrating diverse data sets and tools. Large scale data integration from multiple experimental and bioinformatics resources can provide a robust platform to understand the major molecular players behind a biological problem [67]. Likewise, the rapid accumulation of genome-wide data describing both genome sequences and functional properties of genes has facilitated the development of systems biology approaches [23].

In fact, some bioinformatic programs can aid in many ways to elucidate the function of a gene of interest, including modes of regulation and synthesis, its post-translational modifications and potential interaction partners, and last but not least processes that are regulated by its gene products [68].

The identification and functional characterization of TFs has become essential for the reconstruction of transcriptional regulatory networks, which govern major cellular pathways in response to biotic (e.g. response against pathogen attack) and abiotic (e.g. salinity) stimuli and intrinsic developmental processes (e.g. growth of organs) [29].

The identification of promoters and their regulatory elements is one of the major challenges in bioinformatics and integrates comparative, structural and functional genomics. Due to the fact that promoters are defined functionally but not structurally, their modeling is even harder and complex. Furthermore, detection of CAREs in the promoter is not self-evident, because such short motifs are statistically expected to occur at random every few hundred base pairs. *In silico* promoter prediction is still in the beginning; so, it is important to attempt that existing programs could predict too many false positives. The MEME (Multiple Em for Motif Elucidation; http://meme.nbcr.net/meme4_5_0/cgi-bin/meme.cgi) [69] is a tool for discovering motifs, considering only the sequence patterns that occur in promoter regions. MEME represents motifs as position-dependent letter-probability matrices which describe the probability of each possible letter at each position in the pattern, using statistical modeling techniques to automatically choose the best width, number of occurrences, and description for each motif [69].

Additionally, analysis of the putative TFs subcellular localization is fundamental because TFs cannot function outside the nucleus. Some NAC domain TFs possess a transmembrane motif at the C-terminus and are liberated by proteolytic cleavage to move into the nucleus [70; 71]. Sub-cellular localization of proteins can be efficiently predicted by computer programs such as WoLF PSORT [72].

Several TFs bind their target sequences as homo- or heterodimers, leading to co-occurring binding sites. Cooperative binding (or synergy) is another reason for co-occurrence of motifs. Per definition, classical cooperative binding occurs when protein-protein interactions lead to a more efficient control of the promoter [61]. Co-Bind [73] is an example of method that can be used for discovering co-occurring motifs.

In addition, a number of TFs are known to interact with kinases, resulting in TF phosphorylation [74]. Protein phosphorylation at serine, threonine or tyrosine residues affects the function of TFs. PhosphoSitePlus (PSP) is an open, dynamic, continuously curated, and highly interactive systems biology resource for studying Post-Translational Modifications (PTMs) in the regulation of biological processes. PSP, while still providing comprehensive coverage of protein phosphorylation, includes coverage of other commonly studied PTMs including acetylation, methylation, ubiquitination, and O-glycosylation [75].

Transcription factors can be identified and grouped into different families according to their domain architecture, mainly taking their DNA-binding domains into account. Therefore, conserved domain databases, including DBDs motifs, as the DBD Transcription Factor Prediction Database [76] are fundamental in the preliminary categorization of TFs. This database, which is provided through <http://transcriptionfactor.org>, is composed of predicted sequence-specific DNA-binding transcription factors (TFs) for all publicly available proteomes; all predicted TFs must contain a significant match to a hidden Markov model representing a sequence specific DNA-binding domain family [76].

Homology searches, performed by BLAST algorithm [77], are at the same time one of the most simple and, probably, the most important tool for the bioinformatic study of TFs. Proteins sharing high homology not only in their DBDs but also in other regions are likely to be functionally redundant. BLAST searches against several plants species, performed through the NCBI website (National Center for Biotechnology Information; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>), can enable comparative analysis in order to create a consensus DBD, useful in posterior analyses. Because evolutionary analysis suggests that regulatory modules are conserved between species belonging to different plant families [70], comparative analysis seems to be an effective approach to infer about TFs.

In addition, combining these analyses with searches at the plant specific TF database PlnTFDB (The Plant Transcription Factor Database; <http://plntfdb.bio.uni-potsdam.de/v3.0>) [78] is a useful strategy to identify putative TFs, once it is an integrative database that provides putatively complete sets of transcription factors (TFs) and other transcriptional regulators (TRs) in plant species (*sensu latu*) whose genomes have been completely sequenced and annotated [78].

Finally, for the functional characterization of TFs is advisable to use an integrative tool able to assemble more than one analysis. For this purpose, Koski *et al.* (2005) [79] created the AutoFACT (An Automatic Functional Annotation and Classification Tool) research tool, a fully automated and customizable annotation tool that assigns biologically informative functions to a sequence. Key features of this tool are that it (1) analyzes nucleotide and protein sequence data; (2) determines the most informative functional description by combining multiple BLAST [77] reports from several user-selected databases; (3) assigns putative metabolic pathways, functional classes, enzyme classes, GeneOntology [80] terms and locus names; and (4) generates output in HTML, text and GFF formats for the user's convenience.

With the great scientific interest focused on transcriptional regulation via transcription factors, an increasingly amount of TF databases, including sequences, conserved domains/DBD, promoter regions, regulatory motifs and CAREs was created, comprising tools that help to identify these factors and bind-elements in the target sequence. In this context, the purpose of our group was to deliver an integrative script, the TFIT (Transcription Factor Integrative Tool), able to assemble open source databases and research tools enabling the generation of several analyses automatically from a single input data (putative TF), in order to identify and characterize TFs structurally and functionally. Based on the programs listed in this section and considering three types of input sequence, genomic DNA, cDNA and protein, a pipeline is proposed (Figure 1). For more information on how to obtain TFIT send inquiries to ana.benko.iseppon@pq.cnpq.br.

5. TF mining and crop production: a perspective view

Increase crop productivity in view of the escalating population maintaining cultivable land and natural resources has become vital. Likewise, environmental stresses, as drought, salinity, and high or low temperatures, along with biotic agents like pests and diseases, reduce agricultural yields significantly. In this context the development of crops that tolerate environmental stresses, while maintaining productivity, will become a critical requirement for enhancing agriculture [67].

Advances in plant genomics and systems biology, including the availability of the complete genome/transcriptome sequences of plant model organisms have offered an unprecedented opportunity to identify regulatory genes and networks that control important traits related to abiotic and biotic responses [82]. Transcription factors regulating gene expression is an important facet of abiotic and biotic stresses responsive signal transduction cascades; not to mention its role in the developmental responses in plants. Transcriptional regulation of these responsive genes is an emerging area of plant research [67]. In this scenario not only the quality of the induced genes is vital. Especially the velocity with which a signal cascade

may be activated is determinant of life or death under a sudden stress condition, a situation where fast responding TFs are essential.

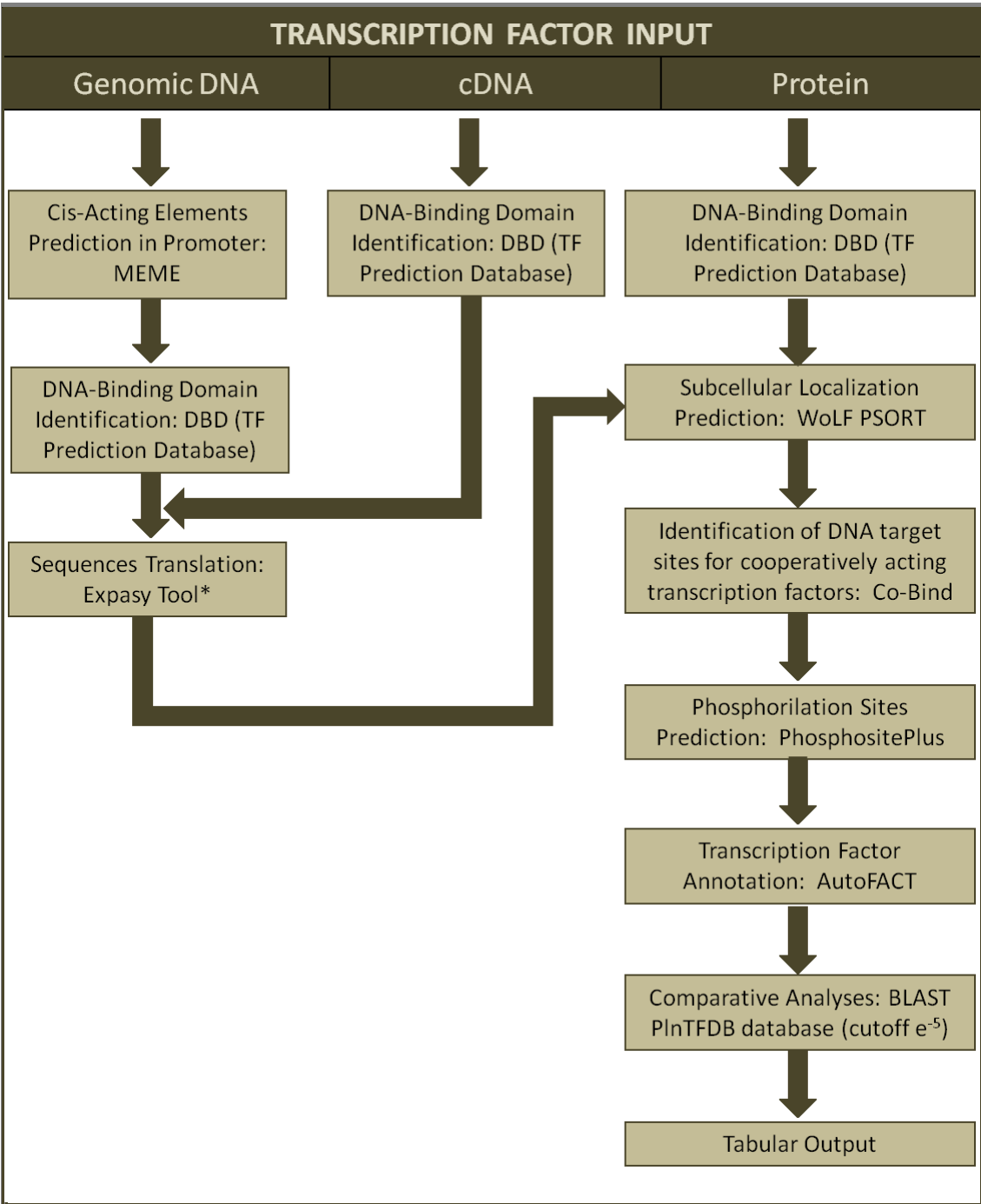


Figure 2. Schamatic representation of the integrative script for transcription factor identification, characterization and annotation. *([81]; available at: <http://www.expasy.ch/tools/dna.html>).

Because manipulation of a TF often induces drastic phenotype changes, altering the proteome, the metabolome and the phenome, in addition to the transcriptome, it may be possible to predict events in the ‘omics’ induced by manipulation of a gene of interest in order to elucidate the whole TF functional network in plants [67]. Thus, they are expected to be excellent

candidates for modifying complex traits in crop plants, and TF-based technologies are likely to be a prominent part of the next generation of successful biotechnology crops [82].

Even before the onset of the genomics era, by the early 1990s there was extensive evidence that TFs act as master regulators that coordinate the expression of entire response networks of target genes, including numerous examples where a change in activity of a single TF was observed to have a profound effect on important aspects of plant biology [82]. All eukaryotic organisms have a diversity of transcription factor gene families that are strongly conserved across species, especially in the case of plants. Deciphering the relationships among TF and CARE players in the control of developmental programs and responses to environmental conditions, including functions of key importance to agronomic performance, is one of the goals of functional genomics and of systems biology [83].

Well-known examples include the NAC multi-functional proteins in the plant life cycle, such as maintenance of the shoot apical meristem [55] and lateral root development [84]. In addition, some specific examples using TFs to engineer abiotic stress tolerance include many different TFs from large TF families like AP2/EREBP, bZIP, NAC, MYB, MYC, and WRKY [85].

Probably the most studied group of TFs involved in drought and cold tolerance are the CBF/DREB (C-repeat binding factor/Dehydration-Responsive Element Binding factor) genes. According to Zhang *et al.* (2004) [86], ectopic expression of these genes in arabidopsis, as well as in heterologous systems such as wheat, tomato, tobacco, rice, and potato [87; 88] has produced enhanced tolerance to one or more types of abiotic stress.

In many researches of individual TFs it is possible to observe that genes from the same family often regulate similar physiological functions even among very different plant species. Under the control of the promoter from the stress-inducible *rd29a* gene, significant stress tolerance without strong growth retardation by expressing DREB1A/CBF3 was reported in arabidopsis by Kasuga *et al.* (1999) [89]. Also Oh *et al.* (2005) [90] showed enhanced drought tolerance in rice plants that constitutively overexpressed either CBF3 or ABF3 (a bZIP TF from arabidopsis), with no obvious negative side effects.

Apart from abiotic stress, another greatest limitation for agricultural productivity worldwide is plant disease, both by damaging host plant tissues and by diverting resources to the growth of the pathogen. As TFs have been found to play a role in conserved pathogen response pathways in various plants, they are therefore useful components for engineering disease resistance; the tomato *Pti4* and *Pti5* ERF genes have been shown to enhance disease resistance when overexpressed [91].

In particular, the members of the ERF subfamily (from AP2/EREBP family) have been implicated in the plant pathogen response [33]. In several cases, the overexpressed ERF provided enhanced resistance to multiple unrelated pathogens, which would be essential for a viable commercial product.

The few cases presented here are just a small sample to highlight the exciting potential of TF for engineering both biotic and abiotic stress tolerance in plants. Moreover, as mentioned by Century *et al.* (2008) [82], it seems reasonable to expect that TFs will be a significant component of the next round of agricultural biotechnology products, conferring enhanced intrinsic yield and yield stability, which will hit the market during the next decade.

6. Conclusion

Transcription factor genes encoding regulatory proteins are present in all eukaryotic organisms. Through interactions with specific regions, the cis-acting regulatory elements in the promoter of target sequences, these factors are able to activate or repress the expression in a spatio-temporal manner, allowing cells the generation of responses to changes in their environment (e.g. biotic and abiotic stresses) and the regulation of the cell cycle. As well described by Carroll (2001) [92], the expansion of regulatory protein numbers and interactions, as well as changes to their spatial and temporal expression, is part of the evolutionary process leading to increasingly complex organisms. Therefore, determining the repertoire of TF genes in genomes, the regulation of their expression, and their biochemical properties (i.e. DNA and protein-binding affinities) is important to understand TF regulatory networks and organism evolution. The here presented script although very useful for the characterization and identification of putative TFs just intends to fill the existing gaps in respect to integrated-tools analysis of these factors. In future, it is expected that joint efforts will create tools capable to identify the target genes which are activated/repressed by TFs through prediction of the interaction between the regulatory motifs with the DNA binding domains.

7. Acknowledgements

The authors thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil) and FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco, Brazil) for financial support and the concession of fellowships.

8. References

- [1] Dietz, K-J.; Vogel, M.O.; Viehhauser, A. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma*, 2010, 245, 3-14.
- [2] Grotewold, E.; Auer, H. The DNA-proteome: recent advances towards establishing the protein-DNA integration space. *Int.l J. Comput. Biosci.*, 2010, 1, 3.
- [3] Wang, Z.; Libaut, M.; Joshi, T.; Valliyodan, B.; Nguyen, HT.; Xu, D.; Stacey, G.; Cheng, J. SoyDB: a knowledge database of soybean transcription factors. *BMC Plant Biol.*, 2010, 10, 14-25.
- [4] Riaño-Pachón, D.M. Identification of transcription factor genes in plants. PhD Thesis, Institute of Biochemistry and Biology, University of Potsdam: Germany, 2008.
- [5] Liu, L.; White, M.J.; MacRae, T.H. Transcription factors and their genes in higher plants: functional domains evolution and regulation. *Eur. J. Biochem.*, 1999, 262, 247-257.
- [6] Mitsuda, N.; Ohme-Takagi, M. Functional analysis of transcription factors in arabidopsis. *Plant Cell Physiol.*, 2009, 50, 1232-1248.
- [7] Yamasaki, K.; Kigawa, T.; Inoue, M.; Watanabe, S.; Tatenno, M.; Seki, M.; Shinozaki, K.; Yokoyama, S. Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiol. Biochem.*, 2008, 46, 394-401.
- [8] Kadonaga, J.T. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, 2004, 116, 247-257.
- [9] Wray, G.A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.*, 2007, 8, 206-216.
- [10] Romeuf, I.; Tessier, D.; Dardevet, M.; Branlard, G.; Charmet, G.; Rave, L.C. wDBTF: a integrated database resource for studying wheat transcription factor families. *BMC Genomics*, 2010, 11, 185-200.
- [11] Rombauts, S.; Florquin, K.; Lescot, M.; Marchal, K.; Rouzé, P.; Van de Peer, Y. Computational approaches to identify promoters and cis-regulatory elements in plant genomes. *Plant Physiol.*, 2003, 132, 1162-1176.
- [12] Featherstone, M. Coactivators in transcription initiation: here are your orders. *Curr. Opin. Genet. Dev.*, 2002, 12, 149-155.
- [13] Fessele, S.; Maier, H.; Zischek, C.; Nelson, P.J.; Werner, T. Regulatory context is a crucial part of gene function. *Trends Genet.*, 2002, 18, 60-63.
- [14] Wasserman, W.W.; Palumbo, M.; Thompson, W.; Fickett, J.W.; Lawrence, C.E. Human-mouse genome comparisons to locate regulatory sites. *Nat. Genet.*, 2000, 26, 225-228.
- [15] Strayer, C.; Oyama, T.; Schultz, T.F.; Raman, R.; Somers, D.E.; Más, P.; Panda, S.; Kreps, J.A.; Kay, S.A. Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Sci.*, 2000, 289, 768 - 771.
- [16] Sakuma, Y.; Liu, Q.; Dubouzet, J.G.; Abe, H.; Shinozaki, K.; Yamaguchi-Shinozaki, K. DNA binding specificity of the ERF/AP2 domain of Arabidopsis DREB's transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.*, 2002, 290, 998-1009.
- [17] Döring, P.; Treuter, E.; Kistner, C.; Lyck, R.; Chen, A.; Nover, L. Role of AHA motifs for the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. *Plant Cell*, 2000, 12, 265-278.
- [18] Hiratsu, K.; Mitsuda, N.; Matsui, K.; Ohme-Takagi, M. Identification of the minimal repression domain of SUPERMAN shows that the DLELRL hexapeptide is both necessary and sufficient for repression of transcription in Arabidopsis. *Biochem. Biophys. Res. Commun.*, 2004, 321, 172-178.
- [19] Hiratsu, K.; Matsui, K.; Koyama, T.; Ohme-Takagi, M. Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. *Plant J.*, 2003, 34, 733-739.
- [20] Ringli, C.; Keller, B. Specific interaction of the tomato bZIP transcription factor VSF-1 with a non-palindromic DNA sequence that controls vascular gene expression. *Plant Mol. Biol.*, 1998, 37, 977-988.

- [21] Fukazawa, J.; Sakai, T.; Ishida, S.; Yamaguchi, I.; Kamiya, Y.; Takahashi, Y. Repression of shoot growth, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. *Plant Cell*, 2000, 12, 901–915.
- [22] Wellmer, F.; Riechmann, J.L. Gene network analysis in plant development by genomic technologies. *Int. J. Dev. Biol.*, 2005, 49, 745–759.
- [23] Vandepoele, K.; Quimbaya, M.; Casneuf, T.; De Veylder, L.; Van de Peer, Y. Unraveling transcriptional control in arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol.*, 2009, 150, 535–546.
- [24] Riechmann, J.L.; Herad, J.; Martin, G.; Reuber, L.; Jiang, C-Z.; Keddie, J.; Adam, L.; Pineda, O.; Ratcliffe, O.J.; Samaha, R.R.; *et al.* Arabidopsis transcription factors Genome-wide comparative analysis among eukaryotes. *Sci.*, 2000, 290, 2105–2110.
- [25] Xiong, Y.; Liu, T.; Tian, C.; Sun, S.; Li, J.; Chen, M. transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots. *Plant Mol. Biol.*; 2005, 59, 191–203.
- [26] Iida, K.; Seki, M.; Sakurai, T.; Satou, M.; Akiyama, K.; Toyoda, T.; Konagaya, A.; Shinozaki, K. RARTF: Database and Tools for Complete Sets of Arabidopsis Transcription Factors. *DNA Res.*, 2005, 12(4), 247–256.
- [27] Davuluri, R.V.; Sun, H.; Palaniswamy, S.K.; Matthews, N.; Molina, C.; Kurtz, M.; Grotewold, E. AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. *BMC Bioinformatics*, 2003, 4, 25.
- [28] Guo, A.; He, K.; Liu, D.; Bai, S.; Gu, X.; Wei, L.; Luo, J. DATF: a database of Arabidopsis transcription factors. *Bioinformatics*, 2005, 21(10), 2568–2569.
- [29] Riaño-Pachón, D.M.; Ruzicic, S.; Dreyer, I.; Mueller-Roeber, B. PInTFDB: an integrative plant transcription factor database. *BMC Bioinformatics*, 2007, 8, 42–52.
- [30] Shiu, S.H.; Shih, M.C.; Li, W.H. Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol.*, 2005, 139, 18–26.
- [31] Seki, M.; Narusaka, M.; Ishida J.; Nanjo, T.; Fujita, M.; Oono, Y.; Kamiya, A.; Nakajima, M.; Enju, A.; Sakurai, T.; *et al.* Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.*, 2002, 31(3), 279–292.
- [32] Singh, K.B.; Foley, R.C.; Oñate-Sánchez, L. Transcription factors in plant defense and stress response. *Curr. Opin. Plant Biol.*, 2002, 5, 430–436.
- [33] Gutterson, N.; Reuber, T.L. Regulation of disease resistance pathways by AP2/ERF transcription factors. *Curr. Opin. Plant Biol.*, 2004, 7, 465–471.
- [34] Kizis, D.; Lumberras, V.; Pagés, M. Role of Ap2/EREBP transcription factors in gene regulation during abiotic stress. *FEBS Lett.*, 2001, 498, 187–189.
- [35] Agarwal, P.K.; Agarwal, P.; Reddy, M.K.; Sopory, S.K. Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Rep.* 2006, 25, 1263–1274.
- [36] Romero, I.; Fuertes, A.; Benito, M.J.; Malpica, J.M.; Leyva, A.; Paz-Ares, J. More than 80 R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana*. *Plant J.*, 1998, 14, 273–284.
- [37] Soares-Cavalcanti, N.M.; Wanderley-Nogueira, A.C.; Belarmino, L.C.; Barros, P.S.; Benko-Iseppon, A.M. Comparative *in silico* evaluation of MYB transcription factors in eucalyptus, sugarcane and rice transcriptomes. *Lect. Notes Comput. Sci.*, 2009, 5488, 44–55.
- [38] Martin, C.; Paz-Ares, J. MYB transcription factors in plants. *Trends Genet.*, 1997, 13, 67–73.
- [39] Tanikawa, J.; Yasukawa, T.; Eanri, M.; Ogata, K.; Nishimura, Y.; Ishii, S.; Sarai, A. Recognition of specific DNA sequences by c-myc proto-oncogene product: role of three repeat units in the DNA-binding domain. *Proc. Natl. Acad. Sci. U.S.A.*, 1993, 90, 9320–9324.
- [40] Kamiya, T.; Kawabe, A.; Miyashita, N.T. Nucleotide polymorphism at Atmyb2 locus of wild plant *Arabidopsis thaliana*. *Genet. Res. Camb.*, 2002, 80, 89–98.

- [41] Yamasaki, k.; Kigawa, T.; Inoue, M.; Tateno, M.; Yamasaki, T.; Yabuki, T.; Aoki, M.; Seki, E.; Matsuda, T.; Tomo, Y.; *et al.* Solution structure of an Arabidopsis WRKY DNA binding domain. *Plant Cell*, 2005, 17, 944-956.
- [42] Wu, K-L.; Guo, Z-J.; Wang, H-H.; Li, J. The WRKY family of transcription factors in rice and Arabidopsis and their origins. *DNA Res.*, 2005, 12, 9-26.
- [43] Robatzek, S.; Somssich, I.E. A new member of the Arabidopsis WRKY transcription factor family AtWRKY6 is associated with both senescence- and defense-related processes. *Plant J.*, 2001, 28, 123-133.
- [44] Rushton, P.J.; Somssich, I.E.; Ringler, P.; Shen, Q.J. WRKY transcription factor. *Trend Plant Sci.*, 2010, 15, 247-258.
- [45] Euglen, T.; Rushton, P.J.; Robatzek, S.; Somssich, I. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.*, 2000, 5, 199-206.
- [46] Maeo, k.; Hayashi, S.; Kojima-Suzuki, H.; Morikami, A.; Nakamura, K. Role of conserved residues of the WRKY domain in the DNA-binding to tobacco WRKY family proteins. *Biosci. Biotechnol. Biochem.*, 2001, 65, 2428-2436.
- [47] Jakoby, M.; Weisshaar, B.; Droge-Laser, W.; Vicente-Carbajosa, J.; Tiedemann, J.; Kroj, T.; Parcy, F. bZIP transcription factors in Arabidopsis. *Trends Plant Sci.*, 2002, 7, 106-111.
- [48] Choi, H.; Hong, J.; Ha, J.; Kang, J.; Kim, S.Y. ABFs, a family of ABA-responsive element binding factors. *J. Biol. Chem.*, 2000, 275, 1723-1730.
- [49] Seo, P.; Kim, S-G.; Park C-M. Membrane-bound transcription factors in plants. *Trends Plant Sci.*, 13(10), 550-556.
- [50] Yanagisawa, S.; Yoo, S-D.; Sheen, J. Differential regulation of EIN3 stability and ethylene signaling in plants. *Nature*, 2003, 425, 521-525.
- [51] Nooren, I.M.; Thornton, J.M. Diversity of protein–protein interactions. *EMBO J.*, 2003, 22, 3486-3492.
- [52] Kohler, J.J.; Schepartz, A. Kinetic studies of Fos. Jun. DNA complex formation: DNA binding prior to dimerization. *Biochem.*, 2001, 40, 130-142.
- [53] Amoutzias, G.D.; Robertson, D.L.; Van De Peer, Y.; Oliver, S.G. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.*, 2008, 33(5), 220-229.
- [54] Desvergne, B.; Michalik L.; Wahli W. Transcriptional regulation of metabolism. *Physiol. Rev.*, 2006, 86, 465-514.
- [55] Kim, S-Y.; Kim, S-G.; Kim, Y-S.; Seo, P.J.; Bae, M.; Yoon, H-K.; Park, C-M. Exploring membrane-associated NAC transcription factors in Arabidopsis: implications for membrane biology in genome regulation. *Nucleic Acids Res.*, 2007, 35, 203-213.
- [56] Liu, Q.; Zhang, G.; Chen, S. Structure and regulatory function of plant transcription factors. *Chin. Sci. Bull.*, 2001, 46, 1-8.
- [57] Whitmarsh, J.; Davis, R.J. Regulation of transcription factor function by phosphorylation. *Cell. Mol. Life Sci.*, 2000, 57, 1172-1183.
- [58] Hood, J.K.; Silver, P.A. In or out? Regulating nuclear transport. *Curr. Opin. Cell. Biol.*, 1999, 11, 241-247.
- [59] Hannenhalli, S. Eukaryotic transcription factor binding sites – modeling and integrative search methods. *Bioinformatics*, 2008, 24, 1325-1331.
- [60] Lodish, H.; Berk, A.; Zipursky, L.S.; Matsudaira, P.; Baltimore, D.; Darnell, J. Molecular Cell Biology, 4th ed.; W. H. Freeman: New York, 2000.
- [61] Grotewold, E. Transcription factors for predictive plant metabolic engineering: are we there yet? *Curr. Opin. Biotechnol.*, 2008, 19, 138-144.
- [62] Narlikar, L.; Ovcharenko, I. Identifying regulatory elements in eukaryotic genomes. *Brief. Funct. Genomic Proteomic*, 2009, 8, 215-230.
- [63] Brower-Toland, B.D.; Smith, C.L.; Yeh, R.C.; Lis, J.T.; Peterson, C.L.; Wang, M.D. Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 2002, 99, 1960-1965.
- [64] Struhl, K. Gene regulation: a paradigm for precision. *Sci.*, 2001, 293, 1054-1055.

- [65] Meyer, P. Transcriptional transgene silencing and chromatin components. *Plant Mol. Biol.*, 2000, 43, 221–234.
- [66] Bentin, T.; Nielsen, P.E. *In vitro* transcription of a torsionally constrained template. *Nucleic Acids Res.*, 2002, 30, 803–809.
- [67] Shameer, K.; Ambika, S.; Verghese, S.M.; Karaba, N.; Udayakumar, M.; Sowdhamini, R. STIFDB – Arabidopsis stress responsive Transcription Factor Database. *Inter. J. Plant Genomic*, 2009, 1-8.
- [68] Pitzschke, A.; Hirt, H. Bioinformatic and systems biology tools to generate testable models of signaling pathways and their targets. *Plant Physiol.*, 2010, 152(2), 460-460.
- [69] Bailey, T.L.; Williams, N.; Misleh, C.; Li, W.W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.*, 2006, 34, W369–W373.
- [70] Kim, Y.S.; Kim, S.G.; Park, J.E.; Park, H.Y.; Lim, M.H.; Chua, N.H.; Park, C-M. A membrane-bound NAC transcription factor regulates cell division in Arabidopsis. *Plant Cell*, 2006, 18, 3132 – 3144.
- [71] Kim, S.G.; Lee, A.K.; Yoon, H.K.; Park, C.M. A membrane-bound NAC transcription factor NTL8 regulates gibberellic acid-mediated salt signaling in Arabidopsis seed germination. *Plant J.*, 2008, 55, 77 –88.
- [72] Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.*, 2007, 35, W585-W587.
- [73] GuhaThakurta, D.; Stormo, G.D. Identifying target sites for cooperatively binding factors. *Bioinformatics*, 2001, 17, 608-621.
- [74] Robertson, M.; Helliwell, C.A.; Dennis, E.S. Post-translational modifications of the endogenous and transgenic FLC protein in Arabidopsis thaliana. *Plant Cell Physiol.*, 2008, 49, 1859-1866.
- [75] Hornbeck, P.V.; Chabra, I.; Kornhauser, J.M.; Skrzypek, E.; Zhang, B. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics*, 2004, 4(6), 1551–1561.
- [76] Wilson, D.; Charoensawan, V.; Kummerfeld, S.K.; Teichmann, S.A. DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.*, 2008, 36, D88–D92.
- [77] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.*, 1990, 215, 403-410.
- [78] Pérez-Rodríguez, P.; Riaño-Pachón, D.M.; Corrêa, L.G.G.; Rensing, S.A.; Kersten, B.; Mueller-Roeber, B. PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.*, 2010, 38, D822-D827.
- [79] Koski, L.B.; Gray, M.W.; Lang, B.F.; Burger, G. AutoFACT: an automatic functional annotation and classification tool. *BMC Bioinformatics*, 2005, 6, 151-161.
- [80] Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; *et al.* Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, 2000, 25, 25–29.
- [81] Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, 2003, 31, 133784-133788.
- [82] Century, K.; Reuber, T.L.; Ratcliffe, O.J. Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. *Plant Physiol.*, 2008, 147, 20-29.
- [83] Libault, M.; Joshi, T.; Benedito, V.A.; Xu, D.; Udvardi, M.K. Stacey G. Legume transcription factor genes: what makes legumes so special? *Plant Physiol.*, 2209, 151, 991-1001.
- [84] He, X.J.; Mu, R.L.; Cao, W.H.; Zhang, Z.G.; Zhang, J.S.; Chen, S.Y. AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *Plant J.*, 2005, 44, 903–916.
- [85] Bhatnagar-Mathur, P.; Vadez, V.; Sharma, K.K. Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. *Plant Cell Rep.*, 2008, 27, 411–424.
- [86] Zhang, J.Z.; Creelman, R.A.; Zhu, J.K. From laboratory to field. Using information from Arabidopsis to engineer salt, cold, and drought tolerance in crops. *Plant Physiol.*, 2004, 135, 615–621.

- [87] Behnam, B.; Kikuchi, A.; Celebi-Toprak, F.; Kasuga, M.; Yamaguchi-Shinozaki, K.; Watanabe, K.N. Arabidopsis rd29A: DREB1A enhances freezing tolerance in transgenic potato. *Plant Cell Rep.*, 2007, 26, 1275–1282.
- [88] Pino, M.T.; Skinner, J.S.; Park, E.J.; Jeknic, Z.; Hayes, P.M.; Thomashow, M.F.; Chen, T.H. Use of a stress inducible promoter to drive ectopic AtCBF expression improves potato freezing tolerance while minimizing negative effects on tuber yield. *Plant Biotechnol. J.*, 2007, 5, 591–604.
- [89] Kasuga, M.; Liu, Q.; Miura, S.; Yamaguchi-Shinozaki, K.; Shinozaki, K. Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat Biotechnol* 1999, 17, 287–291.
- [90] Oh, S.J.; Song, S.I.; Kim, Y.S.; Jang, H.J.; Kim, S.Y.; Kim, M.; Kim, Y.K.; Nahm, B.H.; Kim, J.K. Arabidopsis CBF3/DREB1A and ABF3 in transgenic rice increased tolerance to abiotic stress without stunting growth. *Plant Physiol.*, 2005, 138, 341–351.
- [91] Gu, Y.Q.; Wildermuth, M.C.; Chakravarthy, S.; Loh, Y.T.; Yang, C.; He, X.; Han, Y.; Martin, G.B. Tomato transcription factors *pti4*, *pti5*, and *pti6* activate defense responses when expressed in Arabidopsis. *Plant Cell*, 2002, 14, 817–831.
- [92] Carroll, S.B. Chance and necessity: the evolution of morphological complexity and diversity. *Nature*, 2001, 409, 1102–1109.

- As plantas estudadas apresentam os principais mecanismos de tolerância à seca e salinidade, indicando a existência destes mecanismos no ancestral comum das plantas;
- As características genéticas observadas nos organismos estudados parecem ser resultantes do acúmulo de mutações, principalmente *indels*, ocorridos durante o curso da evolução e especiação;
- Os genomas da soja e do feijão-caupi apresentam sequências similares aos dos principais genes envolvidos nas vias responsivas aos estresses causados pela seca e salinidade nos organismos modelo analisados, sejam os que atuam diretamente nas respostas, ou aqueles que funcional em etapas intermediárias;
- As diferenças com relação aos genes identificados quando comparados aos da planta modelo *Arabidopsis thaliana* podem estar relacionadas com particularidades das leguminosas;
- A modulação da ativação e/ou repressão de genes através dos fatores de transcrição é uma característica presente em todos os organismos estudados, desta feita herdada do ancestral comum das plantas;
- No caso específico dos genes MYB, observa-se que os mesmos são conservados não apenas em plantas, mas também no reino animal;
- Com relação aos perfis de expressão traçados, percebe-se que muitos dos genes identificados também participam das respostas aos estresses bióticos, num processo de resposta cruzada (*crosstalk*);
- Os tecidos que apresentam maior expressão são aqueles que primeiro percebem ou sentem os estresses ambientais, como raízes e folhas;
- De forma geral, a ocorrência dos genes estudados em organismos diferentes indica que os mesmos são essenciais na aquisição de tolerância.

Anexo I. Instrução para autores: Revista Genetics and Molecular Research

Genetics and Molecular Research (GMR) publishes Book Review, Brief Note, Case Report, Comment, Correction, Errata, Homage, In Memoriam, Letter to the Editor, Methodology, Mini-Review, Obituary, Opinion, Point of View, Research Note, Retraction, Review, Review Article, Short Communication, and Thesis Abstract, with regard to genetics, evolution, molecular biology, and bioinformatics. Review articles are normally received by invitation only. If you would like for us to consider a review article, please consult the editor first; send a proposed title, a brief outline and a list of papers relevant to the review published by the author(s). GMR is an exclusively online journal.

The journal is maintained by the not-for-profit scientific foundation Ribeirão Preto Foundation for Scientific Research (FUNPEC-RP) and the articles are open access. The fee per accepted submission is R\$ 1.380.00 for Brazilian authors and US\$820.00 for authors from other countries. The US dollar amount reflects the approximate current foreign exchange rate and is subject to change. This fee covers part of the expenses for final language and technical revision, for page setup, and for publishing online.

Payment of the publishing fee should be made by the authors only after receiving a letter of acceptance. After payment is received by our office, the manuscript will be processed further for publication.

Payment, both from within or outside Brazil, should be made by bank transfer (Banco do Brasil or City Bank) or by credit card (Visa). Instructions for payment of the publishing fee will be sent together with the letter of acceptance for publication.

Please contact the editorial office [gmr@geneticsmr.com] if you have any questions.

All GMR articles must meet the highest standards of scientific quality, both in terms of originality and significance, and the research findings reported should make substantial advances. As it is a journal serving a wide and varied scientific community, article abstracts, introductions and conclusions should be comprehensible to the non-specialist, stressing any wider implications of the study. However, the papers should not compromise on the scientific rigor and detail demanded by an international research journal. The broad readership that GMR attracts gives authors an opportunity to convey to a large audience, as well as to specialists, the importance of their research. The journal is currently indexed in over 64 services; see [<http://www.geneticsmr.com>].

Contributions should be sent either by e-mail as attachments to [gmr@geneticsmr.com] or by regular mail (in files on a CD or DVD) to: Prof. Dr. Francisco A. Moura Duarte, Editor. Rua: Floriano Peixoto, 2444 - Alto da Boa Vista, 14025-220 Ribeirão Preto, SP, Brasil.

It is a fundamental condition that submitted manuscripts have not been previously published and will not be simultaneously published elsewhere. With the acceptance of a manuscript for publication, the publishers acquire full and exclusive copyright for all languages and countries. The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

All papers should be prepared in U.S. English. An initial evaluation of the language will be made upon receipt of each manuscript. Those that are considered inadequate for initial review will be returned or sent out for correction, at the discretion of the author. The manuscript will be considered officially received when the corrected version is ready to be sent to the referees.

Before final acceptance, a submission letter with the title of the article and names and signatures of all the authors should be posted to the above address or faxed to the journal at 55 (16) 3621-1991. Galley proofs will be sent in "pdf" form via e-mail for final revision. All authors are co-responsible for their submissions and they should make every effort to check the paper before this final step to avoid costly reformatting and possible introduction of new errors.

GMR articles have no rigid length restrictions. They should contain sufficient technical detail for an expert reader to understand and assess the methods and results. There is no page limit for GMR articles, but authors should still be concise, for two main reasons. First, our electronic refereeing system relies on e-mail, and very large files occasionally cause problems. Second, lengthy manuscripts can be cumbersome to read and study. Referees tend to dislike them, and they take longer to process. In addition, readers of electronic journals often print articles to read them. Remember that a 10,000-word article takes up around 11 pages.

How many pages would you be willing to read on-screen or print out?

Editorial policies: GMR is a refereed journal. Only original manuscripts will be considered for publication. Manuscripts will be reviewed by at least two independent reviewers before a decision is made on publication. The whole process is conducted electronically to speed progress and final publication. Papers will be published (placed online), once they are fully processed. Papers accepted in their final form from January 1 to March 31 constitute the first issue of each volume, and so on. There are four issues per year.

Manuscripts (in U.S. English), together with a cover letter from the author responsible for all correspondence, should be submitted to the Editor at [gmr@geneticsmr.com] in electronic format as .doc files saved in Microsoft Word 97 for Windows, or later version. Do not use formatting such as Word's "Heading" or "Style Sheets". Spelling, punctuation, sentence structure, spacing, length, and consistency of usage in form and descriptions should be checked before submission. Please also check references for accuracy. Ensure that all figures and tables are mentioned in the text, and that all references are cited in the text. Figure and table files (see below) should be separate.

Submission information

Authors are required to provide the following information with their electronic submissions: Author submitting the article; article title; authors (full list); article type and session; status of article (e.g., new, revised, etc.); postal address; e-mail address; phone number; fax number; names and types of the files sent.

Brazilian authors should not translate their institutional addresses. These should remain in the original (Portuguese) language.

Revised versions: Authors submitting a revised version of an article, must remember to include a list of changes, and replies to the referees (or technical editor). All the files, not just those revised, for the final draft of paper should be sent.

Acknowledgment of electronic submissions: Successful receipt and processing of the author's submission will be acknowledged by e-mail when the submitted manuscript has been checked. If no reply has been received within one week, the author should contact the editor at [gmr@geneticsmr.com].

Review: Articles are reviewed anonymously by independent referees. Authors are encouraged to suggest names of expert reviewers, but selection remains the prerogative of the editors. To facilitate the review process, the authors can send supplementary material, such as cited accepted but not yet published papers, which may be important for assessment of the manuscript.

A review article should contain: an abstract of 250 words or less, no more than six key words, a running title and no more than 60 references. It should be divided into sections with appropriate titles and subtitles.

Preparation of the manuscript

Order the sections comprising the manuscript as follows: title, running title, author, address, abstract, key words, introduction, material and methods, results, discussion, acknowledgments, and references.

Title Page: The title page should include the title of the article, authors' names (names and initials (only) thinking in indexing services), and authors' affiliation. The affiliation should comprise the department, institution (usually university or company), city, and state (or nation). The title page should include the name and complete mailing address, telephone number, fax

number, and e-mail address of the author designated to review proofs. A running title of no more than 60 characters (including spaces) should be provided.

Abstract: An abstract of up to 250 words, single-spaced, is required of research articles and reports and should be arranged in one paragraph. The following information (without headings) should be included: purpose, methods, results (please report numerical data (means $\pm$ SE) for significant results), and conclusions. Review articles also require an abstract, which need not include all of these items.

Key words: A list of key words or indexing terms (up to six) should be included.

Text

Format: Headings should be bold, and first letters capitalized and left-aligned. All text should be set in Times New Roman font, 12 point, left-aligned, single-spaced. Do not justify the right margin. Leave only one (1) space after periods. Paragraphs should not be indented; there should not be any blank lines between them. Use line returns only at the end of paragraphs. Do not use tabs or spaces to create indents. Use the Symbol font for symbols and special characters. Do not use equation editors or footnoting utilities. Save equations as images. Equations should be numbered consecutively with Arabic numerals in parentheses on the right hand side of the page.

Footnotes: Footnotes should be avoided. When their use is absolutely necessary, footnotes should be numbered consecutively using Arabic numerals and should be placed at the bottom of the page to which they refer. Place a line above the footnote, so that it is set off from the text.

Tables/Charts: Special care should be taken to ensure that all tables are properly formatted. Scientific symbols used should be in Symbol or Times New Roman. Tables should be on a separate page, numbered consecutively (with Arabic numerals) referred to by number in the text and designed to fit the column or page size of the journal. Use tables with cells to separate columns. Do not use spaces, tabs or vertical lines. Left justify the title above the table. Indicate each table's location within the manuscript.

Illustrations: Illustrations/figures (photographs, drawings, diagrams, and charts) should each be in a single file, numbered in a consecutive series of Arabic numerals in the order in which they are cited in the text. Illustrations must be submitted as separate files. All illustrations are to be supplied in JPEG (jpg) format in either color or black and white. Images must be saved as separate, stand-alone files. The image resolution should be 300 dpi. Do not embed images within the text file. The placing of graphics in the paper should be indicated in the text and should include the captions for the figures. The authors should also send, by mail, a printed version of the figures. These should be at least 10 x 15 cm, up to US letter size, so that figures can be scanned (in case the figure files are not adequate) to guarantee good quality for publishing online.

Abbreviations: Try to use abbreviations in the text sparingly. Write out abbreviations in full before the first time they are used in the text. Use the metric system for all measurements without periods (cm, mL, s). Define all symbols used in equations and formulas. Do not abbreviate the word "Figure" or "Table" in titles or text.

Acknowledgments: All acknowledgments (including those for grant and financial support) should be typed in one paragraph directly preceding the reference section. Authors of manuscripts submitted to GMR are requested to state the source of all funding that enabled the described research to be undertaken.

References: References in the text should include the name of the author and the year in parentheses, e.g. (Searle, 1961) or (King and Wilson, 1975). When a reference with more than two authors is cited, only the first author is named, e.g. (Comstock et al., 1958). The references must be cited in the text in chronological order, e.g. (Ideber, 2001; Uetz, 2002; Ottavai, 2004). References to "unpublished results" and "submitted papers" should appear in the text in parentheses following the name(s) of the individual(s). Example: (Pereira KS, Martins PK and Silva TM, unpublished results). No more than 40 references should be cited in a Full-length paper, 20 references in a Short Communication and 60 references in a Review article.

References, under the heading "References", should include only works referred to in the text. This section should be arranged in alphabetical order under the first author's last name. References should be cited as follows: journal papers -

names and initials of the first four authors (after that using et al.), year, journal title abbreviated according to PubMed or Web of Science, volume number, first and last page numbers; books - names of authors, year, full title, edition, publishers, address (city); articles published in symposia - names of authors, year, full title of book, name(s) of editor(s) in parentheses, publisher, address (city), first and last page numbers.

The references should consist mainly of articles from indexed journals. References for techniques that are essential for understanding or repeating the methods should always be in easily accessible (indexed) journals.

Reference style: The list of references at the end of the paper should follow the format requested by GMR. The link below can be accessed to see how the references should appear.

Examples of bibliographic style

JOURNAL (Standard article). Ex.:

Chao EC and Lipkin SM (2006). Molecular models for the tissue specificity of DNA mismatch repair-deficient carcinogenesis. *Nucleic Acids Res.* 34: 840-852.

Ranke MB, Lindberg A, Ferrandez Longas A, Darendeliler F, et al. (2007). Major determinants of height development in Turner syndrome (TS) patients treated with GH: analysis of 987 patients from KIGS. *Pediatr. Res.* 61: 105-110.

BOOK, WHOLE: authors, year, book title, edition or volume number, publisher, city. Ex.:

Bartholomei-Santos ML and Sokal RR (1973). *Numerical taxonomy*. W.H. Freeman and Company, San Francisco.

Cicotti and Tildesley DJ (1987). *Computer simulation of liquids*. 1st edn. Oxford University Press, New York.

BOOK CHAPTER: authors, year, book title, chapter title, editors, edition, publisher, city, pages of citation. Ex.:

Drosopoulou E (1981). The Drosophilidae: A Taxonomic Overview. In: *The Genetics and Biology of Drosophila* (Ashburner M, Carson HL and Thompson JN Jr, eds.). Academic Press, New York, 1-97.

REPORT. Ex.:

CIAT (International Center for Tropical Agriculture) (1980). *Annual Report*. CIAT, Cali.

FAO (Food and Agricultural Organization) (2004). *Production Yearbook*. FAO, Rome.

THESIS. Ex.:

Hellmann K and Gömori G (1997). Estudo citoquímico em glândulas salivares de triatomíneos do gênero *Rhodnius*. Master's thesis, UNESP, São José do Rio Preto.

Pepin L (2003). Aplicação de marcadores microsatélites de *Sus scrofa* doméstica na caracterização genética de populações de *Sus scrofa* sp (porco-monteiro) e *Tayassu pecari* (queixada). Doctoral thesis, Faculdade de Medicina de Ribeirão Preto, USP, Ribeirão Preto.

CONFERENCE, SYMPOSIUM PROCEEDINGS: cite papers only from published proceedings. Ex.:

Lehninger AL, Silva MB, Lombardi F, Zamaro PJA, et al. (2002). Análise do comportamento eletroforético das frações de hemoglobinas de *Rhinoclemys punctularia* (Chelonia) em diferentes pH. In: *Anais do 48º Congresso Nacional de Genética*, Ribeirão Preto.

Olsen GJ, Ludwig T, Meier H and Wolf MJ (2002). AxML: a fast program for sequential and parallel phylogenetic tree calculations based on the maximum likelihood method. Proceedings of 1st IEEE Computer Society Bioinformatics Conference (CSB2002), Palo Alto, 21-28.

Pacheco EB and Miyazawa CS (2002). Estudos citogenéticos em Cheirodontinae (Characiformes, Characidae) do pantanal do Mato Grosso. In: IX Simpósio de Citogenética e Genética de Peixes, Maringá, 38.

Schaeffer LR, Swalve HH and Dekkers JCM (1994). Random regressions in animal models for test-day production in dairy cattle. Proceedings of the 5th World Congress of Genetic and Applied Livestock Production, Guelph, 433-446.

ELECTRONIC CITATIONS (Online Journals): ensure that URLs are active and available. Ex.:

Rozen S and Skaletsky HJ (2000). Primer3: Bioinformatics Methods and Protocols. In: Methods in Molecular Biology (Krawetz S and Misener S, eds.). Humana Press, New Jersey, 365-386. Available at [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi].

INTERNET COMMUNICATION: Ensure that URLs are active and available. Ex.:

HTML. HyperText Markup Language. Available at [<http://www.w3.org/TR/html401/>]. Accessed August 23, 2007.

NCICB. National Cancer Institute Web Services. Available at [http://ncicb.nci.nih.gov/infrastructure/cacore_overview]. Accessed August 23, 2007.

Wolf MJ (2005). Global review of commercialized transgenic crops: 2005. The International Service for the Acquisition of Agri-biotech Applications (ISAAA). Available at [<http://www.isaaa.org/kc/bin/briefs34/pk/index.htm>]. Accessed August 21, 2006.

Anexo II. Instrução para autores: Revista PLoS Computational Biology

Contents:

1. About PLoS Computational Biology
2. Open Access
3. Publication Charges
4. Criteria for Publication
5. Overview of the Editorial Process
6. Presubmission Inquiries
7. Preparation of Research Articles
 - 7.1. Organization of the Manuscript
 - 7.2. Nomenclature
 - 7.3. Accession Numbers
 - 7.4. Abbreviations
8. Software Articles
9. Materials Required for Manuscript Submission
 - 9.1. Cover Letter
 - 9.2. Author Status
 - 9.3. Financial Disclosure
 - 9.4. Competing Interests
 - 9.5. Electronic Formats
 - 9.5.1. Manuscript and Tables File
 - 9.5.2. LaTeX files
 - 9.5.3. Figure Files
 - 9.5.4. Multimedia Files and Supporting Information
 - 9.6. Ready to Submit Your Manuscript?
10. Other Types of Articles
11. Overview of the Production Process

12. Blogs, Wikis, Embargoes, and the Media

1. About PLoS Computational Biology

PLoS Computational Biology features works of exceptional significance that further our understanding of living systems at all scales—from molecules and cells, to patient populations and ecosystems—through the application of computational methods. Readers include life and computational scientists, who can take the important findings presented here to the next level of discovery.

Research articles that primarily describe new methods and offer only limited biological insight will be considered only in those exceptional instances in which the method is expected to significantly impact the field of computational biology, typically making substantial breakthroughs in areas of demonstrated importance. Research articles modeling aspects of biological systems should demonstrate both scientific novelty and profound new biological insights. Research articles describing improved or routine methods, models, software, and databases will not be considered by PLoS Computational Biology, and may be more appropriate for PLoS ONE.

Generally, reliability and significance of biological discovery are validated and enriched by experimental studies. Experimental validation is not required for publication, however, nor does experimental validation of a modest biological discovery render a manuscript suitable for PLoS Computational Biology.

For all submissions, authors must clearly provide detail, data, and software to ensure readers' ability to reproduce the models, methods, and results.

2. Open Access

The Public Library of Science (PLoS) applies the Creative Commons Attribution License (CCAL) to all works we publish. Under the CCAL, authors retain ownership of the copyright for their article, but authors allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in PLoS journals, so long as the original authors and source are cited. No permission is required from the authors or the publishers.

3. Publication Charges

To provide open access, PLoS journals use a business model in which our expenses—including those of peer review, journal production, and online hosting and archiving—are recovered in part by charging a publication fee to the authors or research sponsors for each article they publish. For PLoS Computational Biology the publication fee is US\$2250. Authors who are affiliated with one of our Institutional Members are eligible for a discount on this fee.

We offer a complete or partial fee waiver for authors who do not have funds to cover publication fees. Editors and reviewers have no access to payment information, and hence inability to pay will not influence the decision to publish a paper.

For further information, see our Publication Fee FAQ.

4. Criteria for Publication

To be considered for publication in PLoS Computational Biology, any given manuscript must satisfy the following criteria:

Originality

High importance to researchers in the field

Significant biological insight and general interest to life scientists

Rigorous methodology

Substantial evidence for its conclusions

5. Overview of the Editorial Process

Our aim is to provide all authors with an efficient, courteous, and constructive editorial process. To achieve its required level of quality, PLoS Computational Biology is highly selective in the manuscripts that it publishes; rejection rates are high. To ensure the fairest and most objective decision-making, the editorial process is run as a partnership between the PLoS Computational Biology Editor-in-Chief (EIC), a team of Deputy Editors, and a group of academic experts who serve as Associate Editors (AEs). These individuals are leaders in their fields and represent the full breadth of expertise in computational biology.

Submitted manuscripts are first reviewed by the EIC or one of the Deputy Editors, who may decide to reject the paper or send it on to an AE for further review. The AE is most often a member of the PLoS Computational Biology Editorial Board, but occasionally a guest of the Board is invited to serve in this capacity. The AE evaluates the paper and decides whether it describes a sufficient body of work to support a major advance in a particular field. If so, the paper is sent out for external peer review, at which stage the technical and scientific merits of the work are carefully considered. Once the reviews have been received and considered by the editors, a decision letter to the corresponding author is drafted and sent.

The decision will be within one of the following categories:

Reject

Major revision

Minor revision

Accept

Appeals of Decisions

PLoS Computational Biology encourages input from all community members regarding editorial and publishing policy. However, appeals of manuscript decisions should be a) limited to the specific manuscript in question, b) made only by the corresponding author, and c) sent by e-mail to [ploscompbiol \[at\] plos.org](mailto:ploscompbiol@plos.org). Telephone or other informal appeals will not be considered. Editors involved in considering an appeal will include the Associate Editor, Editor-in-Chief and Deputy Editor; additional editorial board members may also be consulted. The editors will try to handle an appeal expeditiously; however, each appeal is unique and the journal cannot guarantee the turnaround time or the outcome. Appeals of decisions made before review will only be considered in exceptional circumstances.

6. Presubmission Inquiries

When authors are unsure whether their work satisfies the basic requirements for publication in PLoS Computational Biology, we are happy to consider presubmission inquiries. If you would like to submit an informal presubmission inquiry to see if a manuscript is appropriate in principle, please login or register for a new account within our online submission system, and then submit your query by choosing 'Submit Presubmission Inquiry' from the list of article types. Required for all Presubmission Inquiries are contact information for the corresponding author, a cover letter (as a separate file), and an abstract.

Responses to these inquiries are normally provided within one week. Responses may take longer if consultation between members of the editorial board is required. If you are invited to submit your manuscript, we will do our best to provide an expeditious initial assessment of the complete manuscript for suitability and then, if warranted, external peer review.

7. Preparation of Research Articles

PLoS Computational Biology publishes original research that clearly demonstrates novelty, importance to a particular field, biological significance, and conclusions that are justified by the study.

Our aim is to make the editorial process rigorous and consistent, and to offer the best possible support to our authors throughout this process. Authors are encouraged to decide how best to present their ideas, results, and conclusions. The writing style should be concise and accessible. Editors may make suggestions for how to achieve this, as well as suggestions for cuts or additions that could be made to the article to strengthen the argument.

Although we encourage submissions from around the globe, we require that manuscripts be submitted in English. As a step towards overcoming language barriers, we encourage authors fluent in other languages to provide copies of their full articles or abstracts in other languages. Translations should be submitted as supporting information and listed, together with other supporting information files, at the end of the article text.

Organization of the Manuscript

Articles published in PLoS Computational Biology are organized into the following sections: Title, Authors, Affiliations, Abstract, Author Summary, Introduction, Results, Discussion, Materials and Methods, Acknowledgments, References, Figure Legends, and Tables. Uniformity in format facilitates the experience of readers and users of the journal. To provide flexibility, however, authors are also able to include the Materials and Methods section before the Results section or before the Discussion section. Please also note that the Results and Discussion can be combined into one Results/Discussion section. Although we have no firm length restrictions for the entire manuscript, we urge authors to present and discuss their findings concisely.

Title (150 characters or less)

The title should be specific to the project, yet concise. It should be comprehensible to readers outside your field. Avoid specialist abbreviations, if possible. Titles should be presented in title case, meaning that all words except for prepositions, articles, and conjunctions should be capitalized. Please also provide a brief Short Title (or "running head") of no more than 50 characters.

Example title:

Detection of Specific Sequences among DNA Fragments Separated by Gel Electrophoresis.

Authors and Affiliations

Provide the first names or initials (if used), middle names or initials (if used), surnames, and affiliations—department, university or organization, city, state/province (if applicable), and country—for all authors. One of the authors should be designated as the corresponding author. It is the corresponding author's responsibility to ensure that the author list, and the summary of the author contributions to the study, is accurate and complete. If the article has been submitted on behalf of a consortium, all consortium members and affiliations should be listed after the Acknowledgments.

Abstract

The abstract of the paper should be succinct; it must not exceed 300 words. Authors should mention the techniques used without going into methodological detail and should summarize the most important results. While the abstract is conceptually divided into three sections (Background, Methodology/Principal Findings, and Conclusions/Significance), please do not apply these distinct headings to the abstract within the article file. Please do not include any citations and avoid specialist abbreviations.

Author Summary

We ask that all authors of research articles include a 150–200 word non-technical summary of the work as part of the manuscript to immediately follow the abstract. This text is subject to editorial change, should be written in the first-person voice, and should be distinct from the scientific abstract. Aim to highlight where your work fits within a broader context; present the significance or possible implications of your work simply and objectively; and avoid the use of acronyms and complex terminology wherever possible. The goal is to make your findings accessible to a wide audience that includes both

scientists and non-scientists. Authors may benefit from consulting with a science writer or press officer to ensure they effectively communicate their findings to a general audience. Examples are available at:

Systems Analysis of Chaperone Networks in the Malarial Parasite *Plasmodium falciparum*

Protein–Protein Interaction Hotspots Carved into Sequences

Elucidating the Altered Transcriptional Programs in Breast Cancer using Independent Component Analysis

Introduction

The introduction should put the focus of the manuscript into a broader context. As you compose the introduction, think of readers who are not experts in this field. Include a brief review of the key literature. If there are relevant controversies or disagreements in the field, they should be mentioned so that a non-expert reader can delve into these issues further. The introduction should conclude with a brief statement of the overall aim of the experiments and a comment about whether that aim was achieved.

Results

The results section should provide details of all of the experiments that are required to support the conclusions of the paper. There is no specific word limit for this section, but details of experiments that are peripheral to the main thrust of the article and that detract from the focus of the article should not be included. The section may be divided into subsections, each with a concise subheading. Large datasets, including raw data, should be submitted as supporting files; these are published online alongside the accepted article. The results section should be written in the past tense.

Discussion

The discussion should spell out the major conclusions of the work along with some explanation or speculation on the significance of these conclusions. How do the conclusions affect the existing assumptions and models in the field? How can future research build on these observations? What are the key experiments that must be done? The discussion should be concise and tightly argued. The results and discussion may be combined into one section, if desired.

Materials and Methods (also called "Methods" or "Models")

This section should provide enough detail for reproduction of the findings. Protocols for new methods should be included, but well-established protocols may simply be referenced. While we do encourage authors to submit all appendices, detailed protocols, or details of the algorithms for newer or less well-established methods, please do so as Supporting Information files. These are not included in the typeset manuscript, but are downloadable and fully searchable from the HTML version of the article.

Acknowledgments

People who contributed to the work but do not fit the criteria for authors should be listed in the Acknowledgments, along with their contributions. You must also ensure that anyone named in the Acknowledgments agrees to being so named.

Details of the funding sources that have supported the work should be confined to the funding statement provided in the online submission system. Do not include them in the Acknowledgments.

References

Only published or accepted manuscripts should be included in the reference list. Meetings abstracts, conference talks, or papers that have been submitted but not yet accepted should not be cited. Limited citation of unpublished work should be included in the body of the text only. All personal communications should be supported by a letter from the relevant authors.

Download EndNote output style file for EndNote 8 and above or EndNote 4-7 (right-click the link to download the file to your computer)

Download BibTeX style file (right-click the link to download the file to your computer)

PLoS uses the numbered citation (citation-sequence) method. References are listed and numbered in the order that they appear in the text. In the text, citations should be indicated by the reference number in brackets. Multiple citations within a single set of brackets should be separated by commas. Where there are three or more sequential citations, they should be given as a range. Example: "...has been shown previously [1,4-6,22]." Make sure the parts of the manuscript are in the correct order for the relevant journal before ordering the citations. Figure captions and tables should be at the end of the manuscript.

Because all references will be linked electronically as much as possible to the papers they cite, proper formatting of the references is crucial. Please use the following style for the reference list:

Published Papers

1. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-5467.

Please list the first five authors and then add "et al." if there are additional authors. Use of a DOI number for the full-text article is acceptable as an alternative to or in addition to traditional volume and page numbers.

Accepted Papers

Same as above, but "In press" appears instead of the page numbers. Example: *Adv Clin Path*. In press.

Electronic Journal Articles

1. Loker WM (1996) "Campesinos" and the crisis of modernization in Latin America. *Jour Pol Ecol* 3. Available: http://www.library.arizona.edu/ej/jpe/volume_3/ascii-lokeriso.txt. Accessed 11 August 2006.

Books

1. Bates B (1992) *Bargaining for life: A social history of tuberculosis*. Philadelphia: University of Pennsylvania Press. 435 p.

Book Chapters

1. Hansen B (1991) New York City epidemics and history for the public. In: Harden VA, Risse GB, editors. *AIDS and the historian*. Bethesda: National Institutes of Health. pp. 21-28.

Figure Legends

The aim of the figure legend should be to describe the key messages of the figure, but the figure should also be discussed in the text. An enlarged version of the figure and its full legend will often be viewed in a separate window online, and it should be possible for a reader to understand the figure without switching back and forth between this window and the relevant parts of the text. Each legend should have a concise title of no more than 15 words that can stand alone, without the use of figure part labels. The overall legend itself should be succinct, while still explaining all figure parts, symbols and abbreviations. Avoid lengthy descriptions of methods.

Tables

Tables should be included at the end of the manuscript file and cited sequentially in the text. All tables should have a concise title. Footnotes can be used to explain abbreviations. Citations should be indicated using the same style as outlined above.

Tables should not occupy more than one printed page; larger tables can be published as online supporting information. Tables must be cell-based; do not use picture elements, text boxes, tabs, or returns in tables. Please ensure that all tables conform to our Guidelines for Figure and Table Preparation when preparing them.

Nomenclature

The use of standardized nomenclature in all fields of science and medicine is an essential step toward the integration and linking of scientific information reported in published literature. We will enforce the use of correct and established nomenclature wherever possible:

We strongly encourage the use of SI units. If you do not use these exclusively, please also provide the SI value in parentheses after each value.

Species names should be italicized (e.g., *Homo sapiens*) and the full genus and species must be written out in full, both in the title of the manuscript and at the first mention of an organism in a paper; after that, the first letter of the genus name, followed by the full species name may be used.

Genes, mutations, genotypes, and alleles should be indicated in italics. Use the recommended name by consulting the appropriate genetic nomenclature database, e.g., HUGO for human genes. It is sometimes advisable to indicate the synonyms for the gene the first time it appears in the text. Gene prefixes such as those used for oncogenes or cellular localization should be shown in roman: v-fes, c-MYC, etc.

The Recommended International Non-Proprietary Name (rINN) of drugs should be provided.

Accession Numbers

All appropriate datasets, images, and information should be deposited in public resources. Please provide the relevant accession numbers (and version numbers, if appropriate). Accession numbers should be provided in parentheses after the entity on first use. Suggested databases include, but are not limited to:

ArrayExpress
BioModels Database
Database of Interacting Proteins
DNA Data Bank of Japan [DDBJ]
DRYAD
EMBL Nucleotide Sequence Database
GenBank
Gene Expression Omnibus [GEO]
Protein Data Bank
UniProtKB/Swiss-Prot
ClinicalTrials.gov

In addition, as much as possible, please provide accession numbers or identifiers for all entities such as genes, proteins, mutants, diseases, etc., for which there is an entry in a public database, for example:

Ensembl
Entrez Gene
FlyBase
InterPro
Mouse Genome Database (MGD)
Online Mendelian Inheritance in Man (OMIM)
PubChem

Providing accession numbers allows linking to and from established databases and integrates your article with a broader collection of scientific information.

Abbreviations

Please keep abbreviations to a minimum and define them upon first use in the text. Non-standard abbreviations should not be used unless they appear at least three times in the text.

8. Software Articles

PLoS Computational Biology publishes articles describing outstanding open source software of exceptional importance that has been shown to provide new biological insights, either as a part of the software article, or published elsewhere. The software must already be widely adopted, or have the promise of wide adoption by a broad community of users. Enhancements to existing published open source software will only be considered if those enhancements bring exceptional new capabilities. The software must be downloadable anonymously in source code form and licensed under an Open Source Initiative (OSI) compliant license. The source code must be accompanied with documentation on building and installing the software from source, as well as for using the software, including instructions on how a user can test the software on supplied test data. Software articles require a presubmission inquiry that includes explanations on how the above criteria are met.

Instructions for Authors:

Format: Articles should be concise (less than 3500 words, not including supplementary material) and be organized as follows:

Title which includes the name of the software.

Authors and affiliations.

Abstract – Fundamental task(s) which the software accomplishes, examples of biological insights from the use of the software, details of availability, including where to download the most recent source code, the license, any operating system dependencies, and support mailing lists.

Introduction – A description of the problem addressed by the software and of its novelty and exceptional nature in addressing that problem.

Design and Implementation – Details of the algorithms used by the software, how those algorithms have been instantiated, including dependencies. Details of the supplied test data and how to install and run the software should be detailed in the supplementary material.

Results – Examples of biological problems solved using the software, including results obtained with the deposited test data and associated parameters.

Availability and Future Directions – Where the software has been deposited. Any future work planned to be carried out by the authors, how others might extend the software.

A prerequisite for publication is that the results described in the paper must be reproducible when peer reviewers or editors choose to run the software on the deposited dataset using the parameters provided.

There is no minimal set of hardware platforms or operating systems that the software is required to run on provided they do not limit broad adoption of the software. Similarly, open-source software may depend on proprietary ancillary software, or external closed-source APIs, to run or to build. Such a dependency does not render software ineligible if it is otherwise open-source, but the dependency may be considered when judging the limits to broad adoption.

Authors are required to deposit the following with the journal upon submission for archival and review purposes.

1. Source Code, accompanied with the OSI-compliant license.
2. Binaries (optional).

3. Documentation. This must include instructions for building and installing the software from source, and any dependencies on both proprietary and freely available prerequisites. For software libraries, instructions for using the API are also required.
4. Test data and instructions. The test dataset(s) must enable complete demonstration of the capabilities of the software, and may either be available from a publicly accessible data archive via a provided URI, or may be specifically created and made available with the software. The results of running the software on the test data should be reproducible, and any external dependencies (such as external databases that may change) must be documented.

The archived software and associated test data, parameters and documentation are the copy of record that matches the associated published article. They must be deposited as supplemental information file of less than 100MB (ideally as .7z or .rar). This ensures that a version of the software is always available to the community. Recognizing that software is subject to on-going development we strongly encourage authors to deposit and manage their source code revisions in an established software version control repository (see Wikipedia for a list, and a feature comparison matrix). We further encourage software projects to provide active support and interact with their user communities via public mailing lists. Authors should provide links to the location of the software, mailing lists etc. as part of the article.

Once a positive response to a presubmission inquiry has been obtained, software articles and software should be submitted as a Research Article type and include a comment that this is a software article and should follow the Author Guidelines for manuscript, figure and table file formats. In addition, software articles should further follow the guidelines for software articles. For papers written in LaTeX, the LaTeX Guidelines for formatting the paper should be followed.

9. Materials Required for Manuscript Submission

Cover Letter

It is important that you include a cover letter with your manuscript. Please explain why this manuscript is suitable for publication in PLoS Computational Biology; why will your paper inspire the other members of your field, and how will it drive research forward? You are free to recommend a suitable Associate Editor to handle your submission; however, the editors reserve the right to contact an alternative—either from the board or a guest editor—if it is considered more appropriate. Please note that the cover letter will be available to the editors and to external peer reviewers as necessary, so be careful not to reveal anything of a confidential nature.

Author Status

It is the responsibility of the corresponding author to ensure that all authors are aware of and approve the submission of the manuscript, its content, authorship, and order of authorship. Confirmation of this action is required at submission of all manuscripts.

The involvement of any professional medical writer in publication must be declared. We encourage authors to consult the European Medical Writers' Association Guidelines on the role of medical writers. For all PLoS journals, the corresponding author must submit the manuscript, related files, and all required data and information. From the point of submission through to publication, all communication related to that manuscript will be directed to and received from the corresponding author only.

PLoS Computational Biology bases its criteria for authorship on those outlined in the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, which are summarized below. The contributions of all authors must be described. Contributions that fall short of authorship should be mentioned in the acknowledgments.

"Authorship credit should be based on

1. substantial contribution to conception and design, or acquisition of data, or analysis and interpretation of data;
2. drafting the article or revising it critically for important intellectual content;

3. and final approval of the version to be published

Authors should meet conditions 1, 2, and 3.

When a large, multi-center group has conducted the work, the group should identify the individuals who accept direct responsibility for the manuscript (3). These individuals should fully meet the criteria for authorship defined above and editors will ask these individuals to complete journal-specific author and competing interests disclosure forms. When submitting a group author manuscript, the corresponding author should clearly indicate the preferred citation and should clearly identify all individual authors as well as the group name.

Acquisition of funding, collection of data, or general supervision of the research group, alone, does not justify authorship. All persons designated as authors should qualify for authorship, and all those who qualify should be listed. Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content."

PLoS journals follow the COPE guidelines covering changes in authorship. Please note that if any changes to the list of authors of a manuscript are necessary after the initial submission of a manuscript to a PLoS journal but before its publication, the corresponding author may be asked to provide written confirmation that all authors consent to the change(s). The journal also reserves the right to request written confirmation from all authors (including those added, removed, or moved in the author order). Such written consent may be required before the revised submission is sent to the editors.

Financial Disclosure

This section should describe sources of funding that have supported the work. Please include relevant grant numbers and the URL of any funder's Web site. Please also include this sentence: "The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript." If this statement is not correct, you must describe the role of any sponsors or funders, and amend the aforementioned sentence as needed.

Competing Interests

The submitting author is asked at submission to declare, on behalf of all authors, whether there are any financial, personal, or professional interests that could be construed to have influenced the paper. The information entered here will appear in the published version, so please do not include the same in the manuscript file.

Reviewers are also asked to declare any interests that might interfere with their objective assessment of a manuscript. Any relevant competing interests of authors must be available to editors and reviewers during the review process and will be stated in published articles. Read more about the Public Library of Science's Competing Interests Policy.

Electronic Formats

Our submission system supports a limited range of formats for text and graphics. The following file formats/types and manuscript information are required before submission. If you are concerned about the suitability of your files, please contact us at [ploscompbiol\[at\]plos.org](mailto:ploscompbiol[at]plos.org).

Manuscript and Table Files

Articles can be submitted for review in DOC, DOCX, RTF, or PDF. Any articles that have been prepared in LaTeX will be accepted for review, but only in PDF format. After acceptance, only text files (RTF or DOC) of the revised manuscript and tables can be accepted for use in the production process.

Math Equations and DOCX

If your manuscript is or will be in DOCX and contains equations, you must follow the instructions below to make sure that your equations are editable when the file enters production.

If you have not yet composed your article, you can ensure that the equations in your DOCX file remain editable in DOC by enabling “Compatibility Mode” before you begin. To do this, open a new document and save as Word 97-2003 (*.doc). Several features of Word 2007/10 will now be inactive, including the built-in equation editing tool. You can insert equations in one of the two ways listed below.

If you have already composed your article as DOCX and used its built-in equation editing tool, your equations will become images when the file is saved down to DOC. To resolve this problem, re-key your equations in one of the two following ways.

1. Use MathType to create the equation. MathType is the recommended method for creating equations.
2. Go to Insert > Object > Microsoft Equation 3.0 and create the equation.

If, when saving your final document, you see a message saying “Equations will be converted to images,” your equations are no longer editable and PLoS will not be able to accept your file.

LaTeX

Articles prepared in LaTeX may be submitted in PDF format for use during the review process. After acceptance, however, .tex files and formatting information will be required as a zipped file. Please consult our LaTeX Guidelines for a list of what will be required.

Tables

Tables must conform to our Guidelines for Figure and Table Preparation and placed at the end of the article DOC or RTF file. Accepted LaTeX submissions only should have table files—which must also conform to these guidelines—uploaded individually into the online submission system.

Figure Files

For the article to be accepted for publication, the author will need to supply high-resolution versions of the figures in TIF or EPS format only. When preparing your figures, please ensure that the files conform to our Guidelines for Figure and Table Preparation. Please do not upload panels for a single figure separately (for example, Figure 1A, Figure 1B-1D, Figure 1E); each figure file should be a single montage of all panels. Queries can be sent to figures [at] plos.org.

If you are uploading your files in EPS format, please use the "create outlines" option under the type menu in Illustrator so that all text and fonts appear as intended in print. If you need additional help with figure preparation, please contact figures [at] plos.org.

PLoS does not accept vector EPS figures generated using LaTeX. We only accept LaTeX generated figures in TIFF format. Export your LaTeX files as PDFs, and then open them in GIMP or Photoshop and save as TIFF. In general, Figures must be generated in a standalone graphics application such as Adobe Illustrator, Inkscape, PyMol, MatLab, SAS, etc. Please see our Figure Guidelines for more information.

All figures will be published under a Creative Commons Attribution License, which allows them to be freely used, distributed, and built upon as long as proper attribution is given. Please do not submit any figures that have been previously copyrighted unless you have express written permission from the copyright holder to publish under the CCAL license.

Multimedia Files and Supporting Information

We encourage authors to submit essential supporting files and multimedia files along with their manuscripts. All supporting material will be subject to peer review, and should be smaller than 10 MB in size because of the difficulties that some users will experience in loading or downloading files of a greater size.

Supporting files should fall into one of the following categories: Dataset, Figure, Table, Text, Protocol, Audio, or Video. All supporting information should be referred to in the manuscript with a leading capital S (e.g., Figure S4 for the fourth supporting information figure). The numbered title and caption for each supporting information file should be included in the main article file, after the titles and captions for the main figures.

Supporting files will not be included in the typeset PDF, but will be referenced in the text and hosted online.

Supporting files may be submitted in a variety of formats, but should be publication-ready, as these files are not copyedited. Carefully consider whether your supporting information needs to be searchable and/or editable, and choose the most suitable format accordingly. See the Figure Guidelines for more detail about our requirements for multimedia files and the file formats we accept.

Ready to Submit Your Manuscript?

We have provided a Submission Checklist to help you prepare your materials for submission and to make the online submission process as straightforward as possible. Please take the time to look through the list before submitting your article.

Please login or register at our online submission system to begin the submission process. Files can be uploaded individually or together in a single ZIP file, and are automatically combined into a single PDF file, which must be approved by the author at the end of the submission process. This merged PDF is for internal and external peer review only. Original source files will be used to prepare accepted articles for publication.

10. Other Types of Articles

PLoS Computational Biology publishes a range of articles other than research articles. Contributions to the front section of the journal are subject to peer review. No publication charges apply. Prospective contributors are encouraged to review contributions in the respective sections of the journal before considering a submission.

Editorials — Editorials are written by PLoS Computational Biology editors. Guest editorials are invited by or at the discretion of the Editor-in-Chief. Editorials typically introduce new and changed features to the journal, discuss issues of professional development, and highlight developments in the field relevant to the readership.

Education articles — The goal of the Education section of PLoS Computational Biology is to provide both practical and background information on important computational methods and approaches used to investigate interesting biological problems. Contributions to the Education section can take several forms, including historical reviews and practical tutorials. Education articles should aim for 2000 and should not exceed 2500 words. The articles are generally invited, but unsolicited submissions will be considered and proceed at the discretion of the Education Editor.

Reviews — Reviews reflect rapidly advancing or topical areas in computational biology research that are of broad interest to the entire biology community and have not been covered significantly by other journals. A review should aim for 2500-3000 words and no more than 100 references and two or three figures or other display items. Reviews are received both by invitation and as unsolicited submissions and are handled by the Reviews Editors.

Perspectives — Perspectives in PLoS Computational Biology typically reflect an author's viewpoint on a particular development in science and how, based on current knowledge of the field and the progress in it, this development evidences or can lead to change in how science is conducted or interpreted. Perspectives are intended to be more prospective than retrospective but require sufficient background to place the points made in context. Perspectives are intended to invite debate and further comment as appropriate. The length is ideally around 2000 and limited to 2500 words. Suggestions for topics may be forwarded to ploscompbiol [at] plos.org and are usually handled by the Editor-in-Chief.

Messages from ISCB — As an official journal of the International Society for Computational Biology (ISCB), PLoS Computational Biology publishes in this section short informational articles invited by the ISCB Editor as well as announcements from the Society.

11. Overview of the Production Process

Before formal acceptance of the article for publication, the manuscript and all related files will be checked by PLoS staff to ensure that they comply with all essential formatting and manuscript preparation requirements. Please note that manuscripts are not subject to detailed copyediting. Therefore, please carefully review your manuscript, paying special attention to spelling, punctuation, and grammar, as well as scientific content. Authors who believe their manuscripts would benefit from in-depth professional copyediting are encouraged to use language-editing and copyediting services, such as the ones offered below (in alphabetical order):

American Journal Experts
Asia Science Editing
Bioedit Ltd
BiomEditor
BioScience Writers
Blue Pencil Science
Boston BioEdit
Carpe Diem Biomedical Writing and Editing
English Manager Science Editing
International Science Editing
Life Science Publishing
Online English
Professional Editing Services
Scienceditors.com
SciTechEdit International
Scitext Cambridge
Scribendi
Squirrel Scribe
Stallard Scientific Editing
Write Science Right

PLoS neither endorses nor takes responsibility for contracting with any of these individuals/companies, but we do recognize the value of the services they provide.

When an article has been accepted for publication, the manuscript files are transferred into our production system and will be published in PDF and HTML formats, with an XML download option. Articles are published online on a weekly schedule and archived in PubMed Central/PubMed within 5 to 10 days of publication.

12. Blogs, Wikis, Embargoes, and the Media

Authors are of course at liberty to present their findings at medical or scientific conferences ahead of publication. We recommend, however, that authors not contact the media or respond to such contact unless an article has been accepted for publication and an embargo date has been established. Respect for press embargoes will help to ensure that your work is reported accurately in the popular media. If a journalist has covered a piece of work ahead of publication, this will not affect consideration of the work for publication. See also our embargo guidelines for journalists.

Anexo III. Instrução para autores: revista *Current Protein & Peptide Science*



Instructions for Authors

ONLINE MANUSCRIPT SUBMISSION: An online submission and tracking service *via* Internet facilitates a speedy and cost-effective submission of manuscripts. The full manuscript has to be submitted online *via* Bentham's Content Management System (CMS) at <http://www.bentham-editorial.org/> (**View submission instructions online**).

Manuscripts must be submitted by one of the authors of the manuscript, and should not be submitted by anyone on their behalf. The principal/corresponding author will be required to submit a Covering Letter along with the manuscript, on behalf of all the co-authors (if any). The author(s) will confirm that the manuscript (or any part of it) has not been published previously or is not under consideration for publication elsewhere. Furthermore, any illustration, structure or table that has been published elsewhere must be reported, and copyright permission for reproduction must be obtained.

For all online submissions, please provide soft copies of all the materials (main text in MS Word or Tex/LaTeX), figures/illustrations in TIFF, PDF or JPEG, and chemical structures drawn in ChemDraw (CDX)/ISISDraw (TGF) as separate files, while a PDF version of the entire manuscript must also be included, embedded with all the figures/illustrations/tables/chemical structures etc.

It is imperative that before submission, authors should carefully proofread the files for special characters, mathematical symbols, Greek letters, equations, tables, references and images, to ensure that they appear in proper format.

References, figures, tables, chemical structures etc. should be referred to in the text at the appropriate place where they have been first discussed. Figure legends/captions should also be provided.

A successful electronic submission of a manuscript will be followed by a system-generated acknowledgement to the principal/corresponding author. Any queries therein should be addressed to ilyas@benthamscience.org

MANUSCRIPTS PUBLISHED: The Journal accepts mini- and full-length review articles written in English. Single topic/thematic issues may also be considered for publication.

Single Topic Issues: These issues will be restricted to invited review articles. Single Topic Special Editor will offer a short Perspective and coordinate the solicitation of manuscripts between 5-8 (for a mini-hot topic) to 8-12 (for full-length hot topic) from leading scientists. Authors interested in editing a single topic issue in an emerging topic of protein and peptide science may submit their proposal to the Editor-in-Chief at cpps@benthamscience.org for consideration. Visit <http://www.benthamscience.com/cpps/Special-Issue.htm> for viewing forthcoming issues.

MANUSCRIPT LENGTH:

Review Articles: The maximum page length limit for comprehensive Review articles is thirty journal pages. For Mini-Review articles, the maximum page length is nine journal pages. Each journal page is on average 900 words.

There is no restriction on the number of figures, tables or additional files e.g. video clips, animation and datasets, that can be included with each article online. Authors should include all relevant supporting data with each article (Refer to Supplementary Material section).

MANUSCRIPT PREPARATION:

The manuscript should be written in English in a clear, direct and active style. All pages must be numbered sequentially, facilitating in the reviewing and editing of the manuscript.

For further convenience, the customer support team available at **Bentham Publishing Services** (www.benthampublishingservices.com) can provide assistance to authors for the preparation of manuscripts.

MANUSCRIPT SECTIONS FOR PAPERS: Manuscripts may be divided into the following sections:

Covering letter

Title

Title page

Abstract

Keywords

Text organization

List of abbreviations (if any)

Conflict of interest (if any)

Acknowledgements (if any)

References

Appendices

Figures/illustrations (if any)

Chemical structures (if any)

Tables (if any)

Supportive/supplementary material (if any)

Covering Letter: It is mandatory that a signed covering letter also be submitted along with the manuscript by the author to whom correspondence is to be addressed, delineating the scope of the submitted article declaring the potential competing interests, acknowledging contributions from authors and funding agencies, and certifying that the paper is prepared according to the 'Instructions for Authors'. All inconsistencies in the text and in the reference section, and any typographical errors must be carefully checked and corrected before the submission of the manuscript. The article contains no such material or information that may be unlawful, defamatory, fabricated, plagiarized, or which would, if published, in any way whatsoever, violate the terms and conditions as laid down in the copyright agreement. The authors acknowledge that the publishers have the legal right to take appropriate action against the authors for any such violation of the terms and conditions as laid down in the copy right agreement. Covering letter can be downloaded from the journal's Web site. DOWNLOAD COVERING LETTER.

Title: The title of the article should be precise and brief and must not be more than 120 characters. Authors should avoid the use of non-standard abbreviations. The title must be written in title case except for articles, conjunctions and prepositions.

Authors should also provide a short 'running title'. Title, running title, byline, correspondent footnote and keywords should be written as presented in original manuscripts.

Title Page: Title page should include paper title, author(s) full name and affiliation, corresponding author(s) names complete affiliation/address, along with phone, fax and email.

Abstract: The abstract should not exceed 250 words for review papers summarizing the essential features of the article.

Keywords: Provide 6 to 8 keywords in alphabetical order.

Text Organization: The main text should begin on a separate page and should be divided into separate sections. The manuscript should be divided into title page, abstract and the main text. The text may be subdivided further according to the areas to be discussed, which should be followed by the Acknowledgement (if any) and Reference sections. The review article should mention any previous important recent and old reviews in the field and contain a comprehensive discussion starting with the general background of the field. It should then go on to discuss the salient features of recent developments. The authors should avoid presenting material which has already been published in a previous review. The authors are advised to present and discuss their observations in brief. The manuscript style must be uniform throughout the text and 10 pt Times New Roman fonts should be used. The full term for an abbreviation should precede its first appearance in the text unless it is a standard unit of measurement. The reference numbers should be given in square brackets in the text. Italics should be used for Binomial names of organisms (Genus and Species), for emphasis and for unfamiliar words or phrases. Non-assimilated words from Latin or other languages should also be italicized e.g. *per se*, *et al.* etc.

Greek Symbols and Special Characters: Greek symbols and special characters often undergo formatting changes and get corrupted or lost during preparation of manuscript for publication. To ensure that all special characters used are embedded in the text, these special characters should be inserted as a symbol but should not be a result of any format styling (*Symbol font face*) otherwise they will be lost during conversion to PDF/XML.

Authors are encouraged to consult reporting guidelines. These guidelines provide a set of recommendations comprising a list of items relevant to their specific research design. Chemical equations, chemical names, mathematical usage, unit of measurements, chemical and physical quantity & units must conform to SI and Chemical Abstracts or IUPAC.

All kinds of measurements should be reported only in International System of Units (SI).

List of Abbreviations: If abbreviations are used in the text either they should be defined in the text where first used, or a list of abbreviations can be provided.

Conflict of Interest: Financial contributions to the work being reported should be clearly acknowledged, as should any potential conflict of interest.

Acknowledgements: Please acknowledge anyone (individual/company/institution) who has contributed to the study by making substantial contributions to conception, design, acquisition of data, or analysis and interpretation of data, or who was involved in drafting the manuscript or revising it critically for important intellectual content. Please list the source(s) of funding for the study, for each author, and for the manuscript preparation in the acknowledgements section.

This journal complies with the International Committee of Medical Journal Editors' Uniform Requirements for Manuscripts Submitted to Biomedical Journals www.icmje.org and the FDA's Good Reprint Practices for the Distribution of Medical Journal Articles and Medical or Scientific Reference Publications on Unapproved New Uses of Approved Drugs and Approved or Cleared Medical Devices <http://www.fda.gov/oc/op/goodreprint.html>

References: References must be listed in the ACS Style only. All references should be numbered sequentially [in square brackets] in the text and listed in the same numerical order in the reference section. The reference numbers must be finalized and the bibliography must be fully formatted before submission.

See below few examples of references listed in the ACS Style:

Journal Reference:

[1] Bard, M.; Woods, R.A.; Bartón, D.H.; Corrie, J.E.; Widdowson, D.A. Sterol mutants of *Saccharomyces cerevisiae*: chromatographic analyses. *Lipids*, **1977**, 12(8), 645-654.

[2] Zhang, W.; Brombosz, S.M.; Mendoza, J.L.; Moore, J.S. A high-yield, one-step synthesis of o-phenylene ethynylene cyclic trimer via precipitation-driven alkyne metathesis. *J. Org. Chem.*, **2005**, 70, 10198-10201.

Book Reference:

[3] Crabtree, R.H. *The Organometallic Chemistry of the Transition Metals*, 3rd ed.; Wiley & Sons: New York, **2001**.

Book Chapter Reference:

[4] Wheeler, D.M.S.; Wheeler, M.M. In: *Studies in Natural Products Chemistry*; Atta-ur-Rahman, Ed.; Elsevier Science B. V: Amsterdam, **1994**; Vol. 14, pp. 3-46.

Conference Proceedings:

[5] Jakeman, D.L.; Withers, S.G.E. In: *Carbohydrate Bioengineering: Interdisciplinary Approaches*, Proceedings of the 4th Carbohydrate Bioengineering Meeting, Stockholm, Sweden, June 10-13, 2001; Teeri, T.T.; Svensson, B.; Gilbert, H.J.; Feizi, T., Eds.; Royal Society of Chemistry: Cambridge, UK, **2002**; pp. 3-8.

URL(WebPage):

[6] National Library of Medicine. Specialized Information Services: Toxicology and Environmental Health. <http://sis.nlm.nih.gov/Tox/ToxMain.html> (Accessed May 23, **2004**).

Patent:

[7] Hoch, J.A.; Huang, S. Screening methods for the identification of novel antibiotics. U.S. Patent 6,043,045, March 28, **2000**.

Thesis:

[8] Mackel, H. *Capturing the Spectra of Silicon Solar Cells*. PhD Thesis, The Australian National University: Canberra, December **2004**.

E-citations:

[9] Citations for articles/material published exclusively online or in open access (free-to-view) , must contain the exact Web

addresses (URLs) at the end of the reference(s), except those posted on an author's Web site unless editorially essential, e.g. 'Reference: Available from: URL'.

Some important points to remember:

- * All references must be complete and accurate.
- * All authors must be cited and there should be no use of the phrase *et al.*
- * Online citations should include the date of access.
- * Journal abbreviations should follow the *Index Medicus/MEDLINE*.
- * Take special care of the punctuation convention as described in the above-mentioned examples.
- * Superscript in the in-text citations and reference section should be avoided.
- * Abstracts, unpublished data and personal communications (which can only be included if prior permission has been obtained) should not be given in the reference section but they may be mentioned in the text and details provided as footnotes.
- * The authors are encouraged to use a recent version of EndNote (version 5 and above) or reference Manager (version 10) when formatting their reference list, as this allows references to be automatically extracted.

Appendices: In case there is a need to present lengthy, but essential methodological details, use appendices, which can be a part of the article. An appendix must not exceed three pages (Times New Roman, 12 point fonts, 900 max. words per page). The information should be provided in a condensed form, ruling out the need of full sentences. A single appendix should be titled APPENDIX, while more than one can be titled APPENDIX A, APPENDIX B, and so on.

Figures/Illustrations:

All authors must strictly follow the guidelines below for preparing illustrations for publication in **Current Protein & Peptide Science**. If the figures are found to be sub-standard, then the manuscripts will be rejected and the authors offered the option of figure improvement professionally by **Bentham Publishing Services**. The costs for such improvement will be charged to the authors

The authors should provide the illustrations as separate files, as well as embedded in the text file, numbered consecutively in the order of their appearance. Each figure should include a single illustration. Each figure should be closely cropped to minimize the amount of white space surrounding the illustration.

If a figure consists of separate parts, it is important that a single composite illustration file be submitted, containing all parts of the figure.

Photographs should be provided with a scale bar if appropriate, as well as high-resolution component files.

Scaling/Resolution:

For Line Art image type, which is generally an image based on lines and text and does not contain tonal or shaded areas, the preferred file format is TIFF or EPS, with colour mode being Monochrome 1-bit or RGB, in a resolution of 900-1200 dpi.

For Halftone image type, which is generally a continuous tone photograph and contains no text, the preferred file format is TIFF, with colour mode being RGB or Grayscale, in a resolution of 300 dpi.

For Combination image type, which is generally an image containing halftone in addition to text or line art elements, the preferred file format is TIFF, with colour mode being or RGB or Grayscale, in a resolution of 500-900 dpi.

Formats:

For illustrations, the following file formats are acceptable:

- **Illustrator**
- **EPS** (preferred format for diagrams)
- **PDF** (also especially suitable for diagrams)
- **PNG** (preferred format for photos or images)
- **Microsoft Word** (version 5 and above; figures must be a single page)
- **PowerPoint** (figures must be a single page)
- **TIFF**
- **JPEG** (conversion should be done using the original file)
- **BMP**
- **CDX** (ChemDraw)
- **TGF** (ISISDraw)

Bentham Science does not process figures submitted in GIF format.

If the large size of TIFF or EPS figures acts as an obstacle to online submission, authors may find that conversion to JPEG format before submission results in significantly reduced file size and upload time, while retaining acceptable quality. JPEG is a 'lossy' format, however in order to maintain acceptable image quality, it is recommended that JPEG files are saved at High or Maximum quality.

Files should not be compressed with tools such as Zipit or Stuffit prior to submission as these tools will in any case produce negligible file-size savings for JPEGs and TIFFs, which are already compressed.

Please do not:

1. Supply embedded graphics in your word processor (spreadsheet, presentation) document;
2. Supply files that are optimized for screen use (like GIF, BMP, PICT, WPG); the resolution is too low;
3. Supply files that are too low in resolution;
4. Submit graphics that are disproportionately large for the content.

Image Conversion Tools:

There are many software packages, many of them freeware or shareware, capable of converting to and from different graphics formats, including PNG.

Good general tools for image conversion include GraphicConverter on the Macintosh, PaintShop Pro, for Windows, and ImageMagick, which is available on Macintosh, Windows and UNIX platforms.

Note that bitmap images (e.g. screenshots) should not be converted to EPS, since this will result in a much larger file size

than the equivalent JPEG, TIFF, PNG or BMP, with no increase in quality. EPS should only be used for images produced by vector-drawing applications such as Adobe Illustrator or CorelDraw. Most vector-drawing applications can be saved in, or exported as, EPS format. In case the images have been originally prepared in an Office application, such as Word or PowerPoint, then the original Office files should be directly uploaded to the site, instead of being converted to JPEG or another format that may be of low quality.

Color Figures/Illustrations:

The cost for the first published page of color figures is **US\$ 940**; the second additional page will be for **US\$ 685** and each subsequent page for **US\$ 465**.

Chemical Structures:

Chemical structures MUST be prepared in ChemDraw/CDX and provided as separate file.

Structure Drawing Preferences:

[As according to the ACS style sheet]

Drawing Settings:

Chain angle	120°
Bond spacing	18% of width
Fixed length	14.4 pt (0.500cm, 0.2in)
Bold width	2.0 pt (0.071cm, 0.0278in)
Line width	0.6 pt (0.021cm, 0.0084in)
Margin width	1.6 pt (0.096cm)
Hash spacing	2.5 pt (0.088cm, 0.0347in)

Text settings:

Font	Times New Roman / Helvetica
Size	8 pt

Under the Preference Choose:

Units	points
Tolerances	3 pixels

Under Page Setup Use:

Paper	US letter
Scale	100%

Tables:

* Data Tables should be submitted in Microsoft Word table format.

* Each table should include a title/caption being explanatory in itself with respect to the details discussed in the table. Detailed legends may then follow.

* Table number in bold font i.e. Table **1**, should follow a title. The title should be in small case with the first letter in caps. A full stop should be placed at the end of the title.

* Tables should be embedded in the text exactly according to their appropriate placement in the submitted manuscript.

* Columns and rows of data should be made visibly distinct by ensuring that the borders of each cell are displayed as black lines.

* Tables should be numbered in Arabic numerals sequentially in order of their citation in the body of the text.

* If a reference is cited in both the table and text, please insert a lettered footnote in the table to refer to the numbered reference in the text.

* Tabular data provided as additional files can be submitted as an Excel spreadsheet.

Supportive/Supplementary Material: We do encourage to append supportive material, for example a PowerPoint file containing a talk about the study, a PowerPoint file containing additional screenshots, a Word, RTF, or PDF document showing the original instrument(s) used, a video, or the original data (SAS/SPSS files, Excel files, Access Db files etc.) provided it is inevitable or endorsed by the journal's Editor.

Supportive/Supplementary material intended for publication must be numbered and referred to in the manuscript but should not be a part of the submitted paper. In-text citations as well as a section with the heading "Supportive/Supplementary Material" before the "References" section should be provided. Here, list all Supportive/Supplementary Material and include a brief caption line for each file describing its contents.

Any additional files will be linked into the final published article in the form supplied by the author, but will not be displayed within the paper. They will be made available in exactly the same form as originally provided only on our Web site. Please also make sure that each additional file is a single table, figure or movie (please do not upload linked worksheets or PDF files larger than one sheet). Supportive/ Supplementary material must be provided in a single zipped file not larger than 4 MB.

Authors must clearly indicate if these files are not for publication but meant for the reviewers'/editors' perusal only.

PERMISSION FOR REPRODUCTION: Published/reproduced material should not be included unless you have obtained written permission from the copyright holder, which should be forwarded to the Editorial Office in case of acceptance of your article for publication.

For obtaining permission for reproducing any material published in an article by **Bentham Science Publishers**, please fill in the request **FORM** downloadable from the journal's Web site and send to cpps@benthamsience.org for consideration.

AUTHORS AND INSTITUTIONAL AFFILIATIONS: The author will be required to provide their full names, the institutional affiliations and the location, with an asterisk in front of the name of the principal/corresponding author. The corresponding author(s) should be designated and their complete address, business telephone and fax numbers and e-mail address must be stated to receive correspondence and galley proofs.

PAGE CHARGES: No page charges will be levied to authors for the publication of their review articles.

LANGUAGE AND EDITING: Manuscripts submitted containing many English typographical errors will not be published. Authors from non- English language countries are advised to use the services of our professional language editing department prior to submitting their manuscript to the Journal. Please contact **Bentham Publishing Services** www.benthampublishingservices.com for a language editing quote at e-mail: info@benthampublishingservices.com stating the total number of words of the article to be edited.

PROOF CORRECTIONS: Authors will receive page proofs of their accepted paper before publications. To avoid delays in publication, proofs should be checked immediately for typographical errors and returned **within 48 hours**. Major changes are not acceptable at the proof stage. If unable to send corrections **within 48 hours** due to some reason, the author(s) must at

least send an acknowledgement on receiving the galley proofs or the article will be published exactly as received and the publishers will not be responsible for any error occurring in the published manuscript in this regard.

The corresponding author will be solely responsible for ensuring that the revised version of the manuscript incorporating all the submitted corrections receives the approval of all the co-authors of the manuscript.

REPRINTS: Each first-named (corresponding) author will receive electronically five free e-reprints of their published article. Printed reprints may be ordered from the Publisher prior to publication of the article. First named authors may also order a personal print and online subscription of the journal at 50% off the normal subscription rate by contacting the subscription department at e-mail: subscriptions@benthamscience.org

OPEN ACCESS PLUS: Accepted articles can be published online for free open access for all to view, and be deposited by the Publishers in PubMed Central. Open access publishing provides the maximum dissemination of the article to the largest audience. Authors must pay for this service. All corresponding authors will be asked to indicate whether or not they wish to pay to have their paper made freely available on publication. If authors do not select the Open Access option, then their article will be published with standard subscription-based access at no charge.

Bentham Science offers authors the choice of open access publication of their articles at a fee of **US \$2,800** per published article which allows indefinite free-to-view online publication with Bentham Science.

Alternatively, authors may choose to publish their article with Bentham Science at a reduced fee for a limited open access period. Bentham Science is the first and only publisher to offer authors the choice at a reduced open access fee to have their article published for a limited open access period. For open access publication for a period of either twelve months or two months (limited open access option) the per article fee is US \$1,400 or US \$700, respectively.

For more information please contact us at e-mail: openaccess@benthamscience.org

REVIEWING AND PROMPTNESS OF PUBLICATION: All papers submitted for publication will be immediately subjected to editorial scrutiny, usually in consultation with members of the Editorial Advisory Board. Every effort will be made to assess the papers quickly. The papers will be typeset and the proofs dispatched to the authors normally within 4 weeks of their acceptance.

COPYRIGHT: Authors who publish in Bentham Science print & online journals will transfer copyright to their work to **Bentham Science Publishers**. Submission of a manuscript to the respective journals implies that all authors have read and agreed to the content of the Covering Letter or the Terms and Conditions. It is a condition of publication that manuscripts submitted to this journal have not been published and will not be simultaneously submitted or published elsewhere. Plagiarism is strictly forbidden, and by submitting the article for publication the authors agree that the publishers have the legal right to take appropriate action against the authors, if plagiarism or fabricated information is discovered. By submitting a manuscript the authors agree that the copyright of their article is transferred to the publishers if and when the article is accepted for publication. Once submitted to the journal, the author will not withdraw their manuscript at any stage prior to publication.

E-PUB AHEAD OF SCHEDULE: Bentham Science Publishers are pleased to offer electronic publication of accepted papers prior to scheduled publication. These peer-reviewed papers can be cited using the date of access and the unique DOI number. Any final changes in manuscripts will be made at the time of print publication and will be reflected in the final

electronic version of the issue. Articles ahead of schedule may be ordered by pay-per-view at the relevant links by each article stated *via* the E-Pub Ahead of Schedule (<http://www.benthamscience.com/cpps/E-Pub-Ahead-of-Schedule.htm>)

Disclaimer: Articles appearing in E-Pub Ahead-of-Schedule sections have been peer-reviewed and accepted for publication in this journal and posted online before scheduled publication. Articles appearing here may contain statements, opinions, and information that have errors in facts, figures, or interpretation. Accordingly, Bentham Science Publishers, the editors and authors and their respective employees are not responsible or liable for the use of any such inaccurate or misleading data, opinion or information contained of articles in the E-Pub Ahead-of- Schedule.

Anexo V. Súmula curricular



Nina da Mota Soares Cavalcanti

Possui graduação em Bacharelado em Ciências Biológicas pela Universidade Federal de Pernambuco (2003), Pós-Graduação em Gestão Ambiental (2009) pela Faculdade Frassinetti do Recife e Mestrado em Genética pela Universidade Federal de Pernambuco (2007). Doutora em Ciências Biológicas pelo Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal de Pernambuco. Tem experiência na área de Docência e pesquisa nas áreas de Bioinformática e Genética, com ênfase em Genética Vegetal, atuando principalmente nos projetos de sequenciamento da cana-de-açúcar, eucalipto, feijão-caupi e soja. Além disso, tem participado de treinamentos relacionados à qualidade e fiscalização ambiental. Também, tem participado da elaboração de RIMA e EIV e de implantação de SGA.

(Texto informado pelo autor)

Última atualização em 19/04/2012

Endereço para acessar este CV:
<http://lattes.cnpq.br/0405040565280319>

Dados Pessoais

Nome	Nina da Mota Soares Cavalcanti
Nome em citações bibliográficas	Soares-Cavalcanti, N M
Sexo	feminino
Filiação	Antonio Fernando Soares Cavalcanti e Ana Elizabete Simões da Mota Fernandes
Nascimento	07/06/1981 - Recife/PE - Brasil
Carteira de Identidade	5114846 SDS-PE - PE - 23/12/1999
CPF	00902819429
Endereço residencial	R. Manuel de Carvalho, 200 apt 601 Afritos - Recife 52050-360, PE - Brasil Telefone: 081 34275022
Endereço profissional	- - Brasil
Endereço eletrônico	e-mail para contato : ninamotasc@gmail.com e-mail alternativo : ninamotasc@hotmail.com

Formação Acadêmica/Titulação

- 2008 - 2012** Doutorado em Ciências Biológicas.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
Título: Análise in silico dos genes que participam das respostas aos estresses abióticos (seca/salinidade/congelamento) nos genomas do eucalipto, cana-de-açúcar e feijão-caupi, Ano de obtenção: 2012
Orientador: Ana Maria Benko Iseppon 
Palavras-chave: Estresse abiótico, Análise In Silico, Expressed sequence tag, Bioinformática
Áreas do conhecimento : Genética, Bioinformática, Genética Vegetal
Setores de atividade : Produtos e Processos Biotecnológicos Vinculados À Agricultura
- 2007 - 2009** Especialização em Pós-Graduação em Gestão Ambiental.
Faculdade Frassinetti do Recife, FAFIRE, Recife, Brasil
Título: Realidades da Sustentabilidade
Orientador: Prof. Dr. Ana Elizabete Fiuza Simões da Mota
- 2005 - 2007** Mestrado em Genética.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
Título: Estudo in silico de genes que codificam fatores de transcrição responsivos à seca, salinidade e congelamento nos genomas do eucalipto, cana e arroz, Ano de obtenção: 2007
Orientador: Profª Drª Ana Maria Benko Iseppon 
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
Palavras-chave: Análise In Silico, Bioinformática, Fatores de transcrição, Cana-de-açúcar, Eucalyptus, arroz
Áreas do conhecimento : Genética, Bioinformática, Genética Vegetal
Setores de atividade : Produtos e Processos Biotecnológicos Vinculados À Agricultura
- 2000 - 2003** Graduação em Bacharelado em ciências Biológicas.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
Título: Mineração da via SOS (Salt Overly-Sensitive) no genoma expresso da Cana-de-açúcar

Formação complementar

- 2011 - 2011** Curso de curta duração em A sustentabilidade e o Setor Industrial. Federação das Indústrias do Estado de Pernambuco, FIEPE, Recife, Brasil
- 2011 - 2011** Introdução ao Gerenciamento de Projetos. Ludo Cursos, LUDO, Brasil
- 2010 - 2010** ISO 14001:2008 - Gestão Ambiental - Como implement. SQS Consultores Associados, SQS, Brasil
- 2010 - 2010** Curso de curta duração em Aspectos regulatórios para empresas de biotecnolog. Serviço de Apoio às Micro e Pequenas Empresas de Recife, SEBRAE/PE, Recife, Brasil
- 2009 - 2009** FORMAÇÃO DE AUDITOR INTERNO ISO 14001:2004. SQS Consultores Associados, SQS, Brasil
Palavras-chave: ISO 14001:04, Qualidade Ambiental
- 2009 - 2009** Fiscalização Ambiental na Construção Civil. Instituto de Qualificação, IQ, Brasil
- 2004 - 2004** Curso de curta duração em Update em Bioinformática. Universidade Federal de Pernambuco, UFPE, Recife, Brasil
Palavras-chave: Bioinformática
- 2001 - 2001** Curso de curta duração em Polimorfismo de DNA: aspectos biológicos, éticos e. Universidade Católica de Pernambuco, UNICAP, Recife, Brasil
- 1994 - 1998** Curso de Inglês/BEC (Basic English Course). Associação Brasil-América, ABA, Brasil

Atuação profissional

1. Universidade Federal de Pernambuco - UFPE

Vínculo institucional

- 2005 - Atual** Vínculo: Colaborador , Enquadramento funcional: Estudante - Pesquisadora , Carga horária: 40, Regime: Integral
Outras informações:
O vínculo com a Universidade Federal de Pernambuco está relacionado com o doutoramento em andamento; tal vínculo será finalizado com o término do doutorado, em janeiro de 2012.
- 2000 - 2005** Vínculo: Estágio , Enquadramento funcional: pesquisa , Carga horária: 40, Regime: Integral

Atividades

- 09/2003 - 11/2004** Estágio, Centro de Ciências Biológicas, Departamento de Genética
Estágio:
Data mining junto aos projetos Sugarcane EST Genome Project (SUCEST) e Eucalyptus Sequencing Project Consortium (FOREST)
- 04/2002 - 08/2003** Estágio, Centro de Ciências Biológicas, Departamento de Genética
Estágio:
Análise citotaxinômica e evolutiva de Iridaceae (Laboratório de Genética e Biotecnologia Vegetal)
- 04/2002 - Atual** Pesquisa e Desenvolvimento, Centro de Ciências Biológicas, Departamento de Genética
Linhas de Pesquisa:
Bioinformática - Genômica e Proteômica
- 07/2000 - 11/2000** Estágio, Centro de Ciências Biológicas, Departamento de Zoologia
Estágio:
Análise de padrões de biodiversidade de coleóptera em ambientes de caatinga

2. Fundação de Hematologia e Hemoterapia de Pernambuco - HEMOPE

Vínculo institucional

- 2002 - 2002** Vínculo: Estágio , Enquadramento funcional: Pesquisa , Carga horária: 30, Regime: Parcial
Outras informações:
Caracterização de alterações estruturais no carótipo humano

Linhas de pesquisa

1. Bioinformática - Genômica e Proteômica

Objetivos: Data mining e análise estrutural de genes responsivos à estresses abióticos (seca, salinidade e congelamento) nos genomas da cana-de-açúcar, eucalipto, feijão caupi, arroz e soja
Palavras-chave: *Bioinformática, Análise In Silico, Estresse abiótico, angiosperma*
Áreas do conhecimento: *Genética, Genética Vegetal*
Setores de atividade: *Produtos e Processos Biotecnológicos Vinculados À Agricultura*

Áreas de atuação

1. BIOTECNOLOGIA
2. Genética Vegetal
3. Bioinformática
4. Qualidade Ambiental - Sistema de Gestão Ambiental
5. Gestão Ambiental
6. Biologia Molecular

Idiomas

Inglês Compreende Bem, Fala Bem, Escreve Bem, Lê Bem

Espanhol Compreende Bem, Fala Razoavelmente, Escreve Razoavelmente, Lê Bem

Produção em C, T&A

Produção bibliográfica

Artigos completos publicados em periódicos

1.  VIEIRA-MELLO, G. S., SANTOS, P. B., Soares-Cavalcanti, N M, Benko-Iseppon, A.M. Identification and expression of early nodulin in sugarcane transcriptome revealed by in silico analysis. *Lecture Notes in Computer Science.*, v.6685, p.72 - 85, 2011.
Referências adicionais: Português. Home page: [doi:10.1007/978-3-642-21946-7_6]
2.  SANTOS, P. B., Soares-Cavalcanti, N M, VIEIRA-MELLO, G. S., Benko-Iseppon, A.M. Osmoprotectants in the Sugarcane (*Saccharum* spp.) Transcriptome revealed by in silico evaluation. *Lecture Notes in Computer Science.*, v.6685, p.44 - 58, 2011.
Referências adicionais: Português. Home page: [doi:10.1007/978-3-642-21946-7_4]
3. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Belarmino, LC, Bezerra-Neto JP, AMORIM, L. L. B., FERREIRA NETO, J. R. C., PANDOLFI, V., Azevedo H M A, SILVA, R. L. O., Santos, M. G., Alves, M. V. S, Kido, EA
Prospecção de Genes de Resistência à Seca e à Salinidade em Plantas Nativas e Cultivadas. *Revista Brasileira de Geografia Física.*, v.06, p.1112 - 1134, 2011.
Palavras-chave: *Estresse abiótico, genes de resistência, angiosperma*
Áreas do conhecimento: *BIOTECNOLOGIA, Genética Vegetal*
Referências adicionais: Português. Meio de divulgação: *Impresso*
4. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Belarmino, LC, Barbosa-Silva, A., Kido, EA, MONTE, S. J. H., PANDOLFI, V., CALSA-JR, T., Benko-Iseppon, A.M.
In silico screening for pathogenesis related-2 gene candidates in *Vigna unguiculata* transcriptome. *Lecture Notes in Computer Science.*, v.6160, p.70 - 81, 2010.
Referências adicionais: Português.
5. Soares-Cavalcanti, N M, PANDOLFI, V., Kido, EA, Belarmino, LC, HOULLOU-KIDO, L. M., Benko-Iseppon, A.M.
PERFIL GERAL DE EXPRESSÃO ASSOCIADO À TOLERÂNCIA À SALINIDADE E SECA EM FEIJÃO-CAUPI (*Vigna unguiculata*). *Documentos - Embrapa Café.*, v.265, p.65 - 75, 2010.
Palavras-chave: *Estresse abiótico, Feijão Caupi, Genes Estresse-Induzidos*
Áreas do conhecimento: *Agronomia, Genética Vegetal, Melhoramento Genético*
Referências adicionais: Português. Meio de divulgação: *Impresso*
6. Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Belarmino, LC, BARROS, P.S., Benko-Iseppon, A.M.
Comparative In Silico Evaluation of MYB Transcription Factors in *Eucalyptus*, *Sugarcane* and *Rice* Transcriptomes. *Lecture Notes in Computer Science.*, v.5488, p.44 - 55, 2009.
Referências adicionais: Português.
7. BARROS, P.S., Soares-Cavalcanti, N M, VIEIRA-MELLO, G. S., Wanderley-Nogueira, AC, CALSA-JR, T., Benko-Iseppon, A.M.
In Silico Evaluation of Osmoprotectants in *Eucalyptus* Transcriptome. *Lecture Notes in Computer Science.*, v.5488, p.66 - 77, 2009.
Referências adicionais: Português. Meio de divulgação: *Impresso*
8. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Moraes, D. A. L., Belarmino, LC, Benko-Iseppon, A.M.
Abundance and Diversity of Resistance genes in sugarcane transcriptome.. *Genetics and Molecular Research.*, v.6, p.1 - 42, 2007.
Referências adicionais: Inglês. Meio de divulgação: *Impresso*
9.  Barbosa-Silva, A., Wanderley-Nogueira, AC, Silva, RRM, Belarmino, LC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
In silico survey of resistance (R) genes in *Eucalyptus* transcriptome. *Genetics and Molecular Biology.*, v.28, p.562 - 574, 2005.
Palavras-chave: *Serine-Threonine kinase, Nucleotide binding site, Leucine-rich repeats, Gene-for-gene interaction*
Áreas do conhecimento: *Bioinformática, Genética Vegetal*

Artigos aceitos para publicação

1. Belarmino, LC, Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, PANDOLFI, V., AMORIM, L. L. B., WINTER, P., Kahi, G., Benko-Iseppon, A.M.
Analysis and Annotation of Chickpea Bacterial Artificial Chromosome Sequences. Lecture Notes in Computer Science. , 2012.
Palavras-chave: Chickpea, BAC, Bioinformatic
Áreas do conhecimento : Genética, Bioinformática, Genética Vegetal
Referências adicionais : Português.
2. AMORIM, L. L. B., Belarmino, LC, PIRECK, B., Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, BRASILEIRO-VIDAL, A. C., Benko-Iseppon, A.M.
BIOINFORMATICS ANALYSIS OF CLASS II TRANSPOSABLE ELEMENTS IN THE COWPEA GENOME AS COMPARED WITH THE ACTUAL SOYBEAN KNOWLEDGE. Lecture Notes in Computer Science. , 2012.
Palavras-chave: cowpea, Bioinformatic
Áreas do conhecimento : Genética Vegetal, Bioinformática, BIOTECNOLOGIA
Referências adicionais : Português.
3. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Belarmino, LC, Bezerra-Neto JP, AMORIM, L. L. B., PANDOLFI, V., ARRUDA, H., Kido, EA
Prospecção de genes de resistência à seca e à salinidade em plantas nativas e cultivadas. Revista Brasileira de Geografia Física. , 2012.
Referências adicionais : Português.
4. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Belarmino, LC, Bezerra-Neto, JP, Benko-Iseppon, A.M.
An overall evaluation of the Resistance (R) and Pathogenesis Related (PR) superfamilies in soybean, medicago and arabidopsis. Genetics and Molecular Research. , 2011.
Palavras-chave: R Genes, PR Genes, Higher plants, Bioinformatic
Áreas do conhecimento : Bioinformática, Genética Vegetal, BIOTECNOLOGIA
Referências adicionais : Português.
5. Soares-Cavalcanti, N M, Belarmino, LC, Bezerra-Neto, JP, Wanderley-Nogueira, AC, LIRA, R. C., ARRUDA, H., Kido, EA, PANDOLFI, V., Benko-Iseppon, A.M.
DIFFERENTIAL GENE EXPRESSION IN THE PROSPECTION OF TOLERANCE GENES AGAINST DROUGHT AND SALINITY IN NATIVE AND CULTIVATED PLANTS. Revista Caatinga (UFERSA. Impresso). , 2011.
Palavras-chave: Abiotic stresses, Tolerance, Bioinformatic, Higher plants
Áreas do conhecimento : BIOTECNOLOGIA, Bioinformática, Genética Vegetal
Referências adicionais : Português.
6. VIEIRA-MELLO, G. S., SANTOS, P. B., Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
EARLY NODULINS IN THE SUGARCANE TRANSCRIPTOME: CHARACTERIZATION AND EXPRESSION PATTERN REVEALED BY IN SILICO ANALYSIS. Lecture Notes in Computer Science. , 2011.
Palavras-chave: Saccharum officinarum, Nodulinas, Transcriptoma
Referências adicionais : Inglês.
7. Soares-Cavalcanti, N M, Kido, EA, Belarmino, LC, Wanderley-Nogueira, AC, Bezerra-Neto JP, CAVALCANTI-LIRA, R., Benko-Iseppon, A.M.
Evaluation of the main gene families for drought and salinity in the GENOSOJA database, as compared with Medicago and Arabidopsis. Genetics and Molecular Biology (Impresso). , 2011.
Palavras-chave: Abiotic stresses, GENOSOJA, Medicago Trunculata, Stress-inducible genes
Referências adicionais : Inglês.
8. SANTOS, P. B., Soares-Cavalcanti, N M, VIEIRA-MELLO, G. S., Benko-Iseppon, A.M.
OSMOPROTECTANTS IN THE SUGARCANE (SACCHARUM SPP.) TRANSCRIPTOME REVEALED BY IN SILICO EVALUATION. Lecture Notes in Computer Science. , 2011.
Palavras-chave: Sugarcane, Abiotic stresses, Osmoprotectants
Referências adicionais : Inglês.
9. Soares-Cavalcanti, N M, Belarmino, LC, Kido, EA, PANDOLFI, V., MARCELINO-GUIMARAES, F. C., Rodrigues, F., PEREIRA, G.A.G., Benko-Iseppon, A.M.
Overall picture of expressed Heat Shock Factors in Glycine max, Lotus japonicus and Medicago truncatula. Genetics and Molecular Research. , 2011.
Palavras-chave: HSF, Fabaceae, Bioinformatic, Abiotic stresses, transcription factor
Áreas do conhecimento : Bioinformática
Referências adicionais : Inglês.
10. Soldatos TG, O'Donoghue SI, Satagopam VP, Barbosa-Silva, A., Pavlopoulos GA, Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Schneider R
Caipirini: using gene sets to rank literature. BMC Research Notes. , 2010.
Palavras-chave: Caipirini, Gene-Based Search, bioscience literature
Referências adicionais : Inglês.
11. Belarmino, LC, CAPRILES, P. V. S. Z., CABRAL, D. G. A., DARDENE, L. E., Barbosa-Silva, A., Soares-Cavalcanti, N M, CROVELLA, S., Benko-Iseppon, A.M.
In silico screening and comparative modeling of sugarcane defensins. Lecture Notes in Computer Science (LNCS). , 2010.
Referências adicionais : Inglês.
12. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Belarmino, LC, Barbosa-Silva, A., Kido, EA, MONTE, S. J. H., PANDOLFI, V., CALSA-JR, T., Benko-Iseppon, A.M.
In silico screening for pathogenesis related-2 (pr-2) gene candidates in Vigna unguiculata transcriptome. Lecture Notes in Computer Science (LNCS). , 2010.
Referências adicionais : Inglês.
13. ★ Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Belarmino, LC, BARROS, P.S., Benko-Iseppon, A.M.
Comparative in silico evaluation of MYB transcription factors in eucalyptus, sugarcane and rice transcriptomes. Lecture Notes in Computer Science (LNCS). , 2009.
Referências adicionais : Inglês.

14. BARROS, P.S., Soares-Cavalcanti, N M, VIEIRA-MELLO, G. S., Wanderley-Nogueira, AC, CALSA-JR, T., Benko-Iseppon, A.M. In silico evaluation of osmoprotectants in eucalyptus transcriptome. Lecture Notes in Computer Science (LNCS)., 2009.
Referências adicionais : Inglês.

Capítulos de livros publicados

1. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Belamino, LC, Kido, EA, Benko-Iseppon, A.M.
Insight on Pathogen Defense Mechanisms in the Sugarcane Transcriptome In: Special Issue (SI) of Functional Plant Science and Biotechnology on Sugarcane Pathology ed.UK : Global Science Book, 2011
Palavras-chave: Biotic stress, Pathogen, Sugarcane, SUCEST
Áreas do conhecimento : Bioinformática
Referências adicionais : Brasil/Português. Meio de divulgação: Impresso, ISBN: 999999999X
2. ★ Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Belamino, LC, Silva, RRM, Almeida, PML, Brunelli, KR, Kido, LMH, Kido, EA
Estresses ambientais: danos e benefícios em plantas In: Genes associados a estresses bióticos e abióticos em feijão-caupi *Vigna unguiculata* (L.) Walp. E outras angiospermas.1 ed.Recife/PE : MXM Gráfica e Editora, 2005, p. 350-359.
Palavras-chave: Estresses bióticos e Abióticos, Resistência à patógenos, Resistência à salinidades, Vigna unguiculata
Áreas do conhecimento : Genética Vegetal
Setores de atividade : Produção Vegetal
Referências adicionais : Brasil/Português. Meio de divulgação: Impresso, ISBN: 8587459201

Trabalhos publicados em anais de eventos (completo)

1. Wanderley-Nogueira, AC, Belamino, LC, Soares-Cavalcanti, N M, PANDOLFI, V., AMORIM, L. L. B., WINTER, P., KAHL, G., Benko-Iseppon, A.M.
Analysis and annotation of chickpea bacterial artificial chromosome sequences In: 8th Computational Intelligence Methods for Bioinformatics and Biostatistics., 2011, Gargnano, Itália.
Annals of the 8th Computational Intelligence Methods for Bioinformatics and Biostatistics. , 2011. v.8.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
2. AMORIM, L. L. B., Belamino, LC, PIERECK, B, Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, PANDOLFI, V., BRASILEIRO-VIDAL, A. C., Benko-Iseppon, A.M.
Bioinformatics analysis of class II transposable elements in the cowpea genome as compared with the actual soybean knowledge In: 8th Computational Intelligence Methods for Bioinformatics and Biostatistics., 2011, Gargnano, Itália.
Annals of the 8th Computational Intelligence Methods for Bioinformatics and Biostatistics. , 2011. v.8. p.1 - 12
Referências adicionais : Brasil/Português.
3. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Belamino, LC, Bezerra-Neto, JP, AMORIM, L. L. B., PANDOLFI, V., ARRUDA, H., Kido, EA
Prospecção de genes de resistência à seca e à salinidade em plantas nativas e cultivadas In: III Simpósio de Mudanças Climáticas, 2011, Petrolina, PE.
Anais do III Simpósio de Mudanças Climáticas. , 2011.
Referências adicionais : Brasil/Português.
4. VIEIRA-MELLO, G. S., BARROS, P.S., Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Early Nodulins in the Sugarcane Transcriptome: Characterization and Expression Pattern Revealed by In Silico Analysis In: CIBB 2010 (Computational Intelligence Methods for Bioinformatics and Biostatistics),, 2010, Salerno, Italy.
Annals of CIBB 2010 (Computational Intelligence Methods for Bioinformatics and Biostatistics),, 2010.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
5. BARROS, P.S., VIEIRA-MELLO, G. S., Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
in the Sugarcane (*Saccharum* spp.) Transcriptome revealed by in silico evaluation In: CIBB 2010 (Computational Intelligence Methods for Bioinformatics and Biostatistics),, 2010, Salerno, Italy.
Annals of CIBB 2010 (Computational Intelligence Methods for Bioinformatics and Biostatistics),, 2010.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
6. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Belamino, LC, Benko-Iseppon, A.M., Barbosa-Silva, A., Kido, EA, WINTER, P., PANDOLFI, V., MONTE, S. J. H., CALSA-JR, T.
In silico screening for Pathogenesis Related-2 gene candidates in *Vigna unguiculata* transcriptome. In: CIBB 2009 (Computational Intelligence Methods for Bioinformatics and Biostatistics In: CIBB 2009 (Computational Intelligence Methods for Bioinformatics and Biostatistics),, 2009, Genova Italy.
Annals of CIBB 2009 (Computational Intelligence Methods for Bioinformatics and Biostatistics),, 2009.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
7. Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Belamino, LC, BARROS, P.S., Benko-Iseppon, A.M.
Comparative In Silico Evaluation of MYB Transcription Factors in Eucalyptus, Sugarcane and Rice Transcriptomes In: CIBB 2008 (Computational Intelligence Methods for Bioinformatics and Biostatistics, 2008, Vietri Sur Mare, Italy.
Annals of CIBB 2008 (Computational Intelligence Methods for Bioinformatics and Biostatistics),, 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
8. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Moraes, D. A. L., Belamino, LC, Benko-Iseppon, A.M.
Abundance and Diversity of Resistance genes in sugarcane transcriptome In: X-meeting - 2 International Conference of AB3C, 2006
Genetics and Molecular Research. , 2006. v.6. p.1 - 42
Palavras-chave: genes de resistência, patógenos, Análise In Silico, Cana-de-açúcar
Áreas do conhecimento : Genética, Genética Vegetal
Setores de atividade : Produtos e Processos Biotecnológicos Vinculados À Agricultura
Referências adicionais : Brasil/Português. Meio de divulgação: Outro

Trabalhos publicados em anais de eventos (resumo)

1. Lima MO, Soares-Cavalcanti, N M, Belamino, LC, Benko-Iseppon, A.M.
Prospecção e análise in silico dos genes representantes da via SOS em *Ricinus communis* In: I Congresso Nacional de Ciências Biológicas / IV Simpósio de Ciências Biológicas, 2011, Recife.
Anais do I Congresso Nacional de Ciências Biológicas / IV Simpósio de Ciências Biológicas. Recife: UNICAP, 2011.
Referências adicionais : Brasil/Português. Meio de divulgação: Vários

Annals of 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy. , 2010.
Referências adicionais : Brasil/Bretão.

3. LIRA, R. C., Soares-Cavalcanti, N M, Bezerra-Neto, JP, Benko-Iseppon, A.M.
In Silico Characterization of the SOS (Salt Overly Sensitive) Pathway in the Expressed Genome of the Sugarcane In: 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy, 2010, Recife.
Annals of 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy. , 2010.
Referências adicionais : Brasil/Português.
4. LIRA, R. C., Bezerra-Neto, JP, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
In silico identification and expression analysis of vacuolar H⁺-ATPase and H⁺-pyrophosphatase in the transcriptome of soybean [Glycine max (L.) Merr.]. In: III Simpósio Brasileiro de Genética Molecular de Plantas, 2010, Ilheus, Bahia.
Anais do III Simpósio Brasileiro de Genética Molecular de Plantas. , 2010.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
5. Lima MO, Belarmino, LC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Inventário computacional da família de proteína relacionada à patogênese-2 (PR-2) em soja (Glycine max (L.) Merr.) In: XXXIII Reunião Nordestina de Botânica, 2010, Aracaju - SE.
Anais da XXXIII Reunião Nordestina de Botânica. , 2010.
Palavras-chave: Soja, Análise In Silico, PR-2
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
6. CABRAL, D. G. A., Lima MO, Soares-Cavalcanti, N M, Belarmino, LC, Benko-Iseppon, A.M.
Lipid Transfer Proteins: both fortifiers and soldiers on the frontlin. In: 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy, 2010, Recife.
Annals of 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy. , 2010.
Referências adicionais : Brasil/Bretão.
7. Belarmino, LC, Soares-Cavalcanti, N M, CAPRILES, P. V. S. Z., Benko-Iseppon, A.M.
secESTractor: a Workflow for Identification of Secretory Proteins from EST Datasets In: V Escola de Modelagem Molecular em Sistemas Biológicos do Laboratório Nacional de Computação Científica (LNCC), 2010, Petropolis, RJ.
Anais do V Escola de Modelagem Molecular em Sistemas Biológicos do Laboratório Nacional de Computação Científica (LNCC), , 2010.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
8. Lima MO, Belarmino, LC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
What are the targets of carbohydrate-active β -1-3-glucanases in Soybean (Glycine max (L.) Merr.)? In: 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy, 2010, Recife.
Annals of 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy. , 2010.
Referências adicionais : Brasil/Bretão.
9. ARAUJO, G. I. R., Soares-Cavalcanti, N M, Belarmino, LC, Benko-Iseppon, A.M.
Busca in silico da superfamília de tioninas no genoma expresso da soja (Glycine max) In: II SICBIO - Simpósio de Inovação em Ciências Biológicas., 2009, Recife.
Anais do II SICBIO - Simpósio de Inovação em Ciências Biológicas., 2009.
Referências adicionais : Brasil/Bretão. Meio de divulgação: Meio digital
10. Bezerra-Neto, JP, Belarmino, LC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Caracterização in silico de aquaporinas no transcriptoma de soja (Glycine max (L.) Merr.) In: XVII Congresso de Iniciação Científica CONIC-UFPE, 2009, Recife.
Anais do XVII Congresso de Iniciação Científica CONIC-UFPE. , 2009.
Referências adicionais : Brasil/Bretão. Meio de divulgação: Meio digital
11. Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Belarmino, LC, PANDOLFI, V., Benko-Iseppon, A.M.
In silico evaluation of NAC transcription factor associated to abiotic stress in the sugarcane transcriptome In: XII Congresso Brasileiro de Fisiologia Vegetal, 2009, Fortaleza-CE.
Anais do XII Congresso Brasileiro de Fisiologia Vegetal. , 2009.
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
12. Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Kido, EA, PANDOLFI, V., Barbosa-Silva, A., CALSA-JR, T., Benko-Iseppon, A.M.
Quantitative analysis of cowpea sage tags related to abiotic and biotic responses In: II Simposio Brasileiro de Genética Molecular de Plantas, 2009, Buzios-RJ.
Anais do II Simposio Brasileiro de Genética Molecular de Plantas. , 2009.
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
13. VIEIRA-MELLO, G. S., BARROS, P.S., Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, CASTELLETTI, C. H., CALSA-JR, T., Benko-Iseppon, A.M.
Análise computacional do gene DMI3 no genoma expresso do eucalipto In: Congresso Brasileiro de Genética, 2008, Salvador-BA.
Anais do Congresso Brasileiro de Genética. , 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
14. BARROS, P.S., Soares-Cavalcanti, N M, VIEIRA-MELLO, G. S., Benko-Iseppon, A.M.
Análise in silico dos genes DTPS1 e TTPB de Arabidopsis thaliana no genoma do feijão-caupi In: Congresso Brasileiro de Genética, 2008, Salvador-BA.
Anais do Congresso Brasileiro de Genética. , 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
15. BARROS, P.S., Soares-Cavalcanti, N M, VIEIRA-MELLO, G. S., CALSA-JR, T., Benko-Iseppon, A.M.
Análise in silico dos genes que participam da síntese da trealose em Arabidopsis thaliana no genoma do feijão-caupi. In: 55º Congresso Brasileiro de Genética, 2008, Salvador. BA.
Anais do 55º Congresso Brasileiro de Genética. , 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
16. Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Análise in silico dos genes que participam das respostas aos estresses abióticos (seca, salinidade e congelamento) nos genomas do eucalipto, cana-de-açúcar e feijão-caupi In: I Simposio de Inovação científica, 2008, Recife.
Anais do I Simposio de Inovação científica. , 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
17. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Data mining revela ortólogos ao gene PR1 de tabaco no transcriptoma do feijão-caupi In: 55º Congresso Brasileiro de Genética, 2008, Salvador, BA.
Anais do 55º Congresso Brasileiro de Genética. , 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital

18. VIEIRA-MELLO, G. S., Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, BARROS, P.S., Benko-Iseppon, A.M.
Estudo in silico do gene GmN70 de soja no genoma do feijão-caupi In: 59º Congresso Brasileiro de Botânica, 2008, Natal, RN.
Anais do 59º Congresso Brasileiro de Botânica. , 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
19. VIEIRA-MELLO, G. S., Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, SANTOS, P. B., Benko-Iseppon, A.M.
Estudo in silico do gene GM70 de soja no genoma do feijão-caupi In: 59. Congresso Brasileiro de Botânica, 4. Congresso Latinoamericano y del Caribe de cactáceas y Otras Suculentas, 30th Congress of International Organization for Succulent Plant Study, 2008, Natal-RN.
Anais do 59. Congresso Brasileiro de Botânica, 4. Congresso Latinoamericano y del Caribe de cactáceas y Otras Suculentas, 30th Congress of International Organization for Succulent Plant Study. , 2008.
Referências adicionais : Brasil/Português.
20. VIEIRA-MELLO, G. S., BARROS, P.S., Soares-Cavalcanti, N M, CALSA-JR, T., Benko-Iseppon, A.M.
In silico analysis of the CLAVATA1 (CLV1) gene in the sugarcane transcriptome In: Reunião anual da Sociedade de Bioquímica e Biologia molecular, 2008, Águas de Lindóia, SP.
Anais do Reunião anual da Sociedade de Bioquímica e Biologia molecular. , 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
21. VIEIRA-MELLO, G. S., BARROS, P.S., Soares-Cavalcanti, N M, CALSA-JR, T., Benko-Iseppon, A.M.
In silico analysis of the CLAVATA1 (CLV1) gene in the sugarcane transcriptome In: XXXVII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2008, Natal, RN.
Anais do XXXVII Reunião Anual da Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq. , 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
22. SANTOS, P. B., Soares-Cavalcanti, N M, VIEIRA-MELLO, G. S., Wanderley-Nogueira, AC, Benko-Iseppon, A.M.
Mineração revela ortólogos aos genes CMO e BADH no genoma do feijão-caupi In: 59. Congresso Brasileiro de Botânica, 4. Congresso Latinoamericano y del Caribe de cactáceas y Otras Suculentas, 30th Congress of International Organization for Succulent Plant Study, 2008, Natal-RN.
Anais do 59. Congresso Brasileiro de Botânica, 4. Congresso Latinoamericano y del Caribe de cactáceas y Otras Suculentas, 30th Congress of International Organization for Succulent Plant Study. , 2008.
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
23. BARROS, P.S., Soares-Cavalcanti, N M, VIEIRA-MELLO, G. S., Wanderley-Nogueira, AC, Benko-Iseppon, A.M.
Mineração revela ortólogos aos genes CMO e BADH no genoma do feijão-caupi In: 59º Congresso Brasileiro de Botânica, 2008, Natal, RN.
Anais do 59º Congresso Brasileiro de Botânica. , 2008.
Referências adicionais : Brasil/Português.
24. SANTOS, P. B., Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Análise computacional de genes que codificam glicina-betaína no genoma expresso da cana-de-açúcar e do eucalipto In: 52º Congresso Brasileiro de Genética, 2007, São Paulo.
Anais do 52º Congresso Brasileiro de Genética. , 2007.
Palavras-chave: Glicina-Betaina, Análise In Silico, SUCEST, FOREST
Áreas do conhecimento : Genética, Genética Vegetal
Setores de atividade : Produtos e Processos Biotecnológicos Vinculados À Agricultura
Referências adicionais : Brasil/Português.
25. BARROS, P.S., Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Análise computacional de genes que codificam glicina-betaína no genoma expresso da cana-de-açúcar e do eucalipto In: 53º Congresso Brasileiro de Genética, 2007, Águas de Lindóia, SP.
Anais do 53º Congresso Brasileiro de Genética. , 2007.
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
26. Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Benko-Iseppon, A.M.
ANÁLISE DOS GENES DA FAMÍLIA KUP/HAK/KT DE TRANSPORTADORES DE POTÁSSIO DE Arabidopsis thaliana NO TRANSCRIPTOMA DO EUCALIPTO In: Encontro de Genética do Nordeste, 2006, Recife-PE.
Anais do Encontro de Genética do Nordeste. , 2006.
Palavras-chave: Análise In Silico, Eucalyptus, Transportadores de potássio
Áreas do conhecimento : Bioinformática, Genética Vegetal
Setores de atividade : Outros
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
27. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Morais, D. A. L., Silva, RRM, Belarmino, LC
ANÁLISE IN SILICO DE GENES DE RESISTENCIA (R) NO GENOMA EXPRESSO DA CANA-DE-AÇÚCAR In: Encontro de Genética do Nordeste, 2006, Recife-PE.
Anais do Encontro de Genética do Nordeste. , 2006.
Palavras-chave: Estresse biótico, Genes R, Análise In Silico
Áreas do conhecimento : Genética Vegetal, Bioinformática
Setores de atividade : Outros
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
28. SANTOS, P. B., Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Análise in silico de osmoprotetores no genoma expresso da cana-de-açúcar e do eucalipto In: 57º Congresso Nacional de Botânica, 2006, Gramado.
Anais do 57º Congresso Nacional de Botânica. , 2006.
Palavras-chave: osmoprotetores, Estresse abiótico, Cana-de-açúcar, eucalipto, Análise In Silico
Áreas do conhecimento : Genética Vegetal, Genética
Setores de atividade : Produtos e Processos Biotecnológicos Vinculados À Agricultura
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
29. BARROS, P.S., Soares-Cavalcanti, N M, Belarmino, LC, Benko-Iseppon, A.M.
ANÁLISE IN SILICO DO GENE TPS1 (TREHALOSE-6-PHOSPHATE SYNTHASE) DE Arabidopsis thaliana NO GENOMA EXPRESSO DO EUCALIPTO In: XVII Encontro de Genética Regional., 2007, Recife.
Anais do XVII Encontro de Genética Regional. , 2006.
Referências adicionais : Brasil/Bretão.
30. SANTOS, P. B., Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
ANÁLISE IN SILICO DO GENE TPS1 (TREHALOSE-6-PHOSPHATE SYNTHASE) DE Arabidopsis thaliana NO GENOMA EXPRESSO DO EUCALIPTO In: Encontro de Genética do Nordeste, 2006, Recife-PE.
Anais do Encontro Regional de Genética. , 2006.
Palavras-chave: Estresse abiótico, Trehalose, Análise In Silico
Áreas do conhecimento : Bioinformática, Genética Vegetal

Setores de atividade : Outros

Referências adicionais : Brasil/Português. Meio de divulgação: Outro

31. Belamino, LC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
CARACTERIZAÇÃO E ANÁLISE DE MEMBROS DA FAMÍLIA MIP (PROTEÍNAS INTRÍNSECAS DE MEMBRANA) EM *Eucalyptus*. In: Encontro de Genética do Nordeste, 2006, Recife-PE.
Anais do Encontro Regional de Genética do Nordeste. , 2006.
Palavras-chave: Estresse abiótico, Aquaporinas, Análise In Silico, *Eucalyptus*
Áreas do conhecimento : Bioinformática, Genética Vegetal
Setores de atividade : Outros
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
32. Belamino, LC, Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Silva, RRM, Almeida, PML, Benko-Iseppon, A.M.
CATEGORIZAÇÃO DE GENES DO METABOLISMO DE SAIS (S) NO GENOMA EXPRESSO DA CANA-DE-AÇÚCAR E DO FEIJÃO-CAUPI In: Encontro de Genética do Nordeste, 2006, Recife-PE.
Anais do Encontro de Genética do Nordeste. , 2006.
Palavras-chave: Estresse abiótico, Genes S, Análise In Silico, *Eucalyptus*, Feijão Caupi
Áreas do conhecimento : Bioinformática
Setores de atividade : Outros
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
33. Belamino, LC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
E ANÁLISE DE MEMBROS DA FAMÍLIA MIP (PROTEÍNAS INTRÍNSECAS DE MEMBRANA) EM *Eucalyptus*.. In: XVII Encontro de Genética do Nordeste, 2006, Recife.
Anais do XVII Encontro de Genética do Nordeste. , 2006.
Referências adicionais : Brasil/Bretão. Meio de divulgação: Meio digital
34. Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, SANTOS, P. B., Benko-Iseppon, A.M.
Identificação de fatores de transcrição relacionados a estresse abiótico no transcriptoma do eucalipto In: 57º Congresso Nacional de Botânica, 2006, Gramado.
Anais do 57º Congresso Nacional de Botânica. , 2006.
Palavras-chave: Fatores de transcrição, Estresse abiótico, Bioinformática, eucalipto
Áreas do conhecimento : Genética, Genética Vegetal
Setores de atividade : Produtos e Processos Biotecnológicos Vinculados À Agricultura
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
35. Belamino, LC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Análise comparativa das aquaporinas no genoma expresso da cana-de-açúcar e outras duas monocotiledôneas In: X Congresso Brasileiro e Latino Americano de Fisiologia Vegetal, 2005, Recife/PE.
Anais do X Congresso Brasileiro e Latino Americano de Fisiologia Vegetal. , 2005.
Palavras-chave: Análise comparativa, In Silico, Aquaporinas, Genoma expresso, Cana-de-açúcar, Monocotiledôneas
Áreas do conhecimento : Bioinformática, Genética Vegetal
Setores de atividade : Produção Vegetal, Outros
Referências adicionais : Brasil/Português.
36. Soares-Cavalcanti, N M, Benko-Iseppon, A.M., Wanderley-Nogueira, AC
Análise in silico do fator de transcrição CBF1 em *eucalyptus* e outras plantas superiores In: 56º Congresso Nacional de Botânica, 2005, Curitiba.
Anais do 56º Congresso Nacional de Botânica. , 2005.
Palavras-chave: Salinidade, CBF1, *Eucalyptus*, Expressed sequence tag, Bioinformatic
Áreas do conhecimento : Bioinformática, Genética Vegetal
Setores de atividade : Produção Vegetal, Outros
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
37. Belamino, LC, Silva, RRM, Almeida, PML, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Aquaporinas e proteínas relacionadas no genoma expresso da cana-de-açúcar In: XXII Reunião Nordestina de Botânica, 2005, Teresina/PI.
Anais do XXII Reunião Nordestina de Botânica. , 2005.
Palavras-chave: Aquaporinas, Bioinformática, Cana-de-açúcar, Expressed sequence tag, Salinidade
Áreas do conhecimento : Genética Vegetal, Bioinformática
Setores de atividade : Produção Vegetal, Outros
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
38. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Caracterização e análise do gene codificante da glicoproteína Taumatina no genoma expresso da cana-de-açúcar In: 56º Congresso Nacional de Botânica, 2005, Curitiba.
Anais do 56º Congresso Nacional de Botânica. , 2005.
Palavras-chave: Glicoproteína Taumatina, Expressed sequence tag, Cana-de-açúcar, Análise In Silico
Setores de atividade : Outros, Produção Vegetal
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
39. Belamino, LC, Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Divisão e expressão diferencial de membros da família anexina em cana-de-açúcar In: 51º Congresso Brasileiro de Genética, 2005, Águas de Lindóia/SP.
Anais do 51º Congresso Brasileiro de Genética. , 2005.
Palavras-chave: Salinidade, Anexinas, Análise In Silico, Expressed sequence tag, Cana-de-açúcar, Expressão diferencial
Áreas do conhecimento : Bioinformática, Genética Vegetal
Setores de atividade : Produção Vegetal, Outros
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
40. Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Benko-Iseppon, A.M., Barbosa-Silva, A., Morais, D. A. L.
Evolução e perfil de expressão do gene de metabolismo de sais SOS1 em cana-de-açúcar e arroz In: 27 Reunião Nordestina de Botânica, 2004, Petrolina.
Anais do 27 Reunião Nordestina de Botânica. , 2004.
Palavras-chave: SOS1, Transcriptoma, Bioinformática, Cana-de-açúcar
Áreas do conhecimento : Bioinformática
Setores de atividade : Produtos e Processos Biotecnológicos Vinculados À Agricultura
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
41. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Identificação e análise estrutural do gene de resistência Xa21 no transcriptoma da cana-de-açúcar. In: 55 Congresso Nacional de genética, 2004, Viçosa.
Anal do 55 Congresso Nacional de Genética. , 2004.
Palavras-chave: Xa21, Transcriptoma, Cana-de-açúcar, Resistência à patógenos

Áreas do conhecimento : Genética Vegetal,Bioinformática,Fisiologia Vegetal
Setores de atividade : Produtos e Processos Biotecnológicos
Referências adicionais : Brasil/Português. Meio de divulgação: Outro

42. Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Benko-Iseppon, A.M.
Identification and IN SILICO evaluation of salinity resistance gene AtNHX1 from Arabidopsis thaliana in the eucalyptus expressed genome In: 55 Congresso Nacional de Botânica, 2004, Viçosa.
Anal do 55 Congresso Nacional de Botânica. , 2004.
Palavras-chave: AtNHX1 gene, Bioinformatic, Expressed sequence tag, Eucalyptus, Salinity resistance
Áreas do conhecimento : Genética Vegetal,Bioinformática,Fisiologia Vegetal
Setores de atividade : Produtos e Processos Biotecnológicos
Referências adicionais : Brasil/Inglês. Meio de divulgação: Outro
43. Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
IN SILICO ANALYSIS OF THE RPM1 RESISTANCE GENE IN EUCALYPTUS AND HIGHER PLANTS In: 50 Congresso Brasileiro de Genética, 2004, Costão do Santinho.
Anal do 50 Congresso Brasileiro de Genética. , 2004.
Palavras-chave: RPM1, Eucalyptus, Bioinformatic
Áreas do conhecimento : Bioinformática,Fisiologia Vegetal
Setores de atividade : Produtos e Processos Biotecnológicos
Referências adicionais : Brasil/Inglês. Meio de divulgação: Outro
44. Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Benko-Iseppon, A.M.
Salt Overly-Sensitive (SOS) genes in the sugarcane transcriptome as compared with Arabidopsis thaliana and other crops. In: 50 Congresso Brasileiro de genética, 2004, Costão do Santinho.
Anal do 50 Congresso Brasileiro de Genética. , 2004.
Palavras-chave: SOS genes, SOS pathway, Expressed sequence tag, Cana-de-açúcar, Bioinformatic
Áreas do conhecimento : Genética Vegetal,Bioinformática,Fisiologia Vegetal
Setores de atividade : Produtos e Processos Biotecnológicos
Referências adicionais : Brasil/Inglês. Meio de divulgação: Outro
45. ★ Soares-Cavalcanti, N M, Benko-Iseppon, A.M., Carvalho, R., Gitaí, J.
ESTUDOS CITOGENÉTICOS EM SETE ESPÉCIES DA FAMÍLIA IRIDACEAE In: 49 Congresso Nacional de Genética, 2003, Águas de Lindóia.
49 Congresso Nacional de genética. , 2003.
Palavras-chave: Iridaceae, Citogenética, Bandamento C, Coloração convencional
Áreas do conhecimento : Genética Vegetal
Setores de atividade : Desenvolvimento Rural, Produção Vegetal
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
46. ARAUJO, I. P., Wanderley-Nogueira, AC, Soares-Cavalcanti, N M, LOUREIRO, A. F., GUEIROS, R., MENDONÇA, T. M.
Transgênicos: domínio do conceito entre estudantes da UFPE In: 49 Congresso Brasileiro de Genética, 2003, Águas de Lindóia.
Anal do 49 Congresso Brasileiro de Genética. , 2003.
Palavras-chave: Transgênicos, Estudantes área saúde, Avaliação do Conhecimento
Áreas do conhecimento : Transgênico
Setores de atividade : Educação Média de Formação Geral
Referências adicionais : Brasil/Português. Meio de divulgação: Outro

Trabalhos publicados em anais de eventos (resumo expandido)

1. Soares-Cavalcanti, N M, PANDOLFI, V., Kido, EA, Belamino, LC, Kido, LMH, Benko-Iseppon, A.M.
PERFIL GERAL DE EXPRESSÃO ASSOCIADO À TOLERÂNCIA À SALINIDADE E SECA EM FEIJÃO-CAUPI (Vigna unguiculata) In: Simposio Sobre Tolerância à Deficiências Hídrica em Plantas. Adaptando as Culturas ao Clima do Futuro, 2010, Goiânia-GO.
Anais do Simposio Sobre Tolerância à Deficiências Hídrica em Plantas. Adaptando as Culturas ao Clima do Futuro. , 2010.
Palavras-chave: Estresse abiótico, Expressed sequence tag, Genes Estresse-Induzidos, Feijão Caupi
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
2. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Belamino, LC, AMORIM, L. L. B., Kido, EA, PANDOLFI, V.
Cowpea (Vigna unguiculata) Transcriptome and Differential Expression: The Project and Applications for Breeding Purposes In: XVII Plant and Animal genome Conference, 2009
Anals of XVII Plant and Animal genome Conference. , 2009. v.17. p.W304 -
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital

Apresentação de Trabalho

1. Lima MO, Soares-Cavalcanti, N M, Belamino, LC, Benko-Iseppon, A.M.
Prospecção e Análise estrutural in silico dos genes representantes da via SOS em Ricinus communis, 2011.
(Congresso,Apresentação de Trabalho)
Referências adicionais : Brasil/Português. Meio de divulgação: Outro; Local: UNICAP; Cidade: Recife; Evento: I Congresso Nacional de Ciências Biológicas/ IV Simpósio de Ciências Biológicas; Inst.promotora/financiadora: Universidade Católica de Pernambuco
2. VIEIRA-MELLO, G. S., BARROS, P.S., SANTOS, P. B., Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
Early Nodulins in the Sugarcane Transcriptome: Characterization and Expression Pattern Revealed by In Silico Analysis, 2010.
(Congresso,Apresentação de Trabalho)
Palavras-chave: Nodulinas, Saccharum officinarum, Transcriptoma
Referências adicionais : Itália/Inglês. Meio de divulgação: Outro; Local: Palazzo Comitini; Cidade: Palermo, Italy; Evento: SEVENTH INTERNATIONAL MEETING ON COMPUTATIONAL INTELLIGENCE METHODS FOR BIOINFORMATICS AND BIOSTATISTICS
3. LIRA, R. C., Soares-Cavalcanti, N M, Benko-Iseppon, A.M.
In Silico Characterization of the SOS (Salt Overly Sensitive) Pathway in the Expressed Genome of Sugar Cane, 2010.
(Congresso,Apresentação de Trabalho)
Referências adicionais : Brasil/Português. Meio de divulgação: Outro; Local: Centro de Ciencias Biologicas, UFPE; Cidade: Recife-PE; Evento: 1st Brazilian-Germany Meeting of Plant Systems Biology and Bioenergy; Inst.promotora/financiadora: UFPE, Camara Brasil-Alemanha, CCB
4. CABRAL, D. G. A., Lima MO, Soares-Cavalcanti, N M, Belamino, LC, Benko-Iseppon, A.M.
Lipid transfer proteins: both fortifiers and soldiers on the frontline, 2010. (Congresso,Apresentação de Trabalho)
Referências adicionais : Brasil/Português. Meio de divulgação: Outro; Local: Centro de Ciências Biológicas; Cidade: Recife-PE; Evento: 1st Brazilian-Germany Meeting of Plant Systems Biology and Bioenergy; Inst.promotora/financiadora: UFPE, Camara Brasil-Alemanha, CCB

5. SANTOS, P. B., Soares-Cavalcanti, N M, VIEIRA-MELLO, G. S., Benko-Iseppon, A.M. Osmoprotectants in the Sugarcane (*Saccharum spp.*) Transcriptome revealed by in silico evaluation, 2010. (Congresso, Apresentação de Trabalho)
Palavras-chave: osmoprotetores, Estresse abiótico, Cana-de-açúcar
Referências adicionais : Itália/Inglês. Meio de divulgação: Meio digital; Local: Palazzo Comitini; Cidade: Palermo, Italy; Evento: SEVENTH INTERNATIONAL MEETING ON COMPUTATIONAL INTELLIGENCE METHODS FOR BIOINFORMATICS AND BIOSTATISTICS
6. Lima MO, Belarmino, LC, Soares-Cavalcanti, N M, Benko-Iseppon, A.M. What are the targets of carbohydrate-active B-1-3-glucanase in soybean (*Glycine max* (L.) Merr.)?, 2010. (Congresso, Apresentação de Trabalho)
Referências adicionais : Brasil/Português. Meio de divulgação: Outro; Local: CCB, UFPE; Cidade: Recife-PE; Evento: 1st Brazilian-Germany Meeting of Plant Systems Biology and Bioenergy; Inst.promotora/financiadora: UFPE, Camara Brasil-Alemanha, CCB
7. ★ Benko-Iseppon, A.M., Kido, EA, MONTE, S. J. H., PANDOLFI, V., CALSA-JR, T., Kido, LMH, ROCHA, M. M., Brunelli, KR, BARBOSA P. K. A., WINTER, P., KAHL, G., BJORN, R., AMORIM, L. L. B., ONOFRE, A. V. C., FERREIRA-NETO, J. C., Vasconcelos A. T. R., Carvalho, R., Belarmino, LC, Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Barbosa-Silva, A., SANTOS, P. B., VIEIRA-MELLO, G. S., BRASILEIRO-VIDAL, A. C., BORTOLETI, K. C., PEDROSA-HARAND, A., GEPTS, P., FRANKLIN-DE-MELLO, N., ANDRADE, P. P., GRANJEIRO, T. B., CUNHA, O. L., MEIRELLES, D. B., ANDRADE, G. P., PIO-RIBEIRO, G., SITTOLIN, I. M., FREIRE-FILHO, F. R. Cowpea (*Vigna unguiculata*) Transcriptome And Differential Expression: The Project And Applications For Breeding Purposes, 2009. (Conferência ou palestra, Apresentação de Trabalho)
Referências adicionais : Estados Unidos/Inglês. Meio de divulgação: Meio digital-Home page: http://www.intl-pag.org/17/abstracts/W40_PAGXVII_304.html; Local: Town & Country Convention Center; Cidade: San Diego (California, USA); Evento: XVII International Plant & Animal Genome Conference; Inst.promotora/financiadora: Scherago International

Produção Técnica

Demais produções técnicas

1. Soares-Cavalcanti, N M, Belarmino, LC Apostila do Curso de Biossistemática Molecular Vegetal, 2010. (Desenvolvimento de material didático ou instrucional)
Palavras-chave: Bioinformática, Evolução, Planta
Referências adicionais : Brasil/Português. Meio de divulgação: Impresso
2. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Belarmino, LC Biossistemática na Evolução Vegetal, 2010. (Outro, Curso de curta duração ministrado)
Referências adicionais : Brasil/Português. 60 horas.
3. Soares-Cavalcanti, N M, Belarmino, LC Apostila do Curso de Introdução à Bioinformática, 2009. (Desenvolvimento de material didático ou instrucional)
Referências adicionais : Brasil/Português. Meio de divulgação: Impresso
4. Soares-Cavalcanti, N M, Belarmino, LC Curso Avançado em Bioinformática, 2009. (Outro, Curso de curta duração ministrado)
Referências adicionais : Brasil/Português. 60 horas. Meio de divulgação: Outro
5. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Wanderley-Nogueira, AC Genética na Agricultura, 2007. (Outro, Curso de curta duração ministrado)
Referências adicionais : Brasil/Português. 10 horas.
6. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Belarmino, LC Sistemática Molecular, 2007. (Outro, Curso de curta duração ministrado)
Referências adicionais : Brasil/Português. 24 horas.
7. Benko-Iseppon, A.M., Belarmino, LC, Barbosa-Silva, A., Brunelli, KR, CALSA-JR, T., Carvalho, R., CASTELLETTI, C. H., Kido, EA, Kido, LMH, PANDOLFI, V., Soares-Cavalcanti, N M, Wanderley-Nogueira, AC Atualização em Genética Molecular, 2006. (Outro, Curso de curta duração ministrado)
Referências adicionais : Brasil/Português. 2 semanas.
8. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Brunelli, KR, CALSA-JR, T., Carvalho, R., CASTELLETTI, C. H., Kido, EA, Kido, LMH Técnicas Moleculares, bioinformática e mapeamento aplicados ao melhoramento de plantas, 2005. (Outro, Curso de curta duração ministrado)
Referências adicionais : Brasil/Português. 2 semanas. Meio de divulgação: Outro

Orientações e Supervisões

Orientações e Supervisões concluídas

Trabalhos de conclusão de curso de graduação

1. Petra Barros. Petra Barros. 2007. Curso (Bacharelado em ciências Biológicas) - Universidade Federal de Pernambuco
Palavras-chave: osmoprotetores, Análise In Silico, Bioinformática, Eucalyptus
Áreas do conhecimento : Genética Vegetal, Genética
Setores de atividade : Produtos e Processos Biotecnológicos Vinculados À Agricultura
Referências adicionais : Brasil/Português.
No Trabalho de conclusão de curso de graduação da aluna Petra Barros atuei como co-orientadora.

Orientações e Supervisões em andamento

Iniciação científica

1. Rafaela Cavalcanti. Análise de genes que codificam transportadores de sódio responsivos à salinidade e seca no transcriptoma do feijão-caupi. 2009. Iniciação científica (Biomédicina) - Universidade Federal de Pernambuco
Referências adicionais : Brasil/Português.
No trabalho de Iniciação Científica da Aluna Rafaela Cavalcanti atuou como co-orientadora

Eventos

Participação em eventos

1. A sustentabilidade e o Setor Industrial, 2011. (Seminário)
2. Workshop crédito de carbono: perspectiva e oportunidades para a CHESF, 2011. (Outra)
3. Apresentação de Poster / Paineis no(a) II Simposio Brasileiro de Genética Molecular de Plantas, 2009. (Simpósio)
Quantitative Analysis of cowpea sage tags related to abiotic and biotic responses.
4. Apresentação de Poster / Paineis no(a) I Simposio de Inovação Científica, 2008. (Simpósio)
Análise in silico dos genes que participam das respostas aos estresses abióticos (seca, salinidade e congelamento) nos genomas do eucalipto, cana-de-açúcar e feijão-caupi.
5. Apresentação de Poster / Paineis no(a) XVII Encontro de Genética do Nordeste, 2006. (Congresso)
Análise dos genes da família KUP/HAK/KT de transportadores de potássio de *Arabidopsis thaliana* no transcriptoma do eucalipto.
6. Apresentação de Poster / Paineis no(a) 57 Congresso Nacional de Botânica, 2006. (Congresso)
Identificação de fatores de transcrição relacionados a estresse abiótico no transcriptoma do eucalipto.
7. Apresentação de Poster / Paineis no(a) 56 Congresso Nacional de Botânica, 2005. (Congresso)
Análise in silico do fator de transcrição CBF1 em *eucalyptus* e outras plantas superiores.
8. Apresentação de Poster / Paineis no(a) 27 Reunião Nordestina de Botânica, 2004. (Congresso)
Evolução e perfil de expressão do gene de metabolismo de sais SOS1 em cana-de-açúcar e arroz.
9. Apresentação de Poster / Paineis no(a) 55 Congresso Nacional de Botânica, 2004. (Congresso)
Identification and in silico evaluation of salinity resistance gene AtNHX1 from *Arabidopsis thaliana* in the *eucalyptus* expressed genome.
10. Apresentação de Poster / Paineis no(a) 50 Congresso Nacional de Genética., 2004. (Congresso)
Salt Overly-sensitive (SOS) genes in the sugarcane transcriptome as compared with *arabidopsis thaliana* and other crops.
11. Apresentação de Poster / Paineis no(a) 49 Congresso Nacional de Genética, 2003. (Congresso)
Estudos citogenéticos em sete espécies da família Iridaceae.
12. 47 Congresso Nacional de Genética., 2001. (Congresso)

Organização de evento

1. ABDELNOOR, R. V., Benko-Iseppon, A.M., PANDOLFI, V., Soares-Cavalcanti, N M, Belamino, LC, BARBOSA P. K. A.
II Reunião Científica do Consórcio Genosoma, 2010. (Outro, Organização de evento)
Palavras-chave: Glycine max, Sequenciamento do Genoma Expresso
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
2. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Belamino, LC, AMORIM, L. L. B., Bezerra-Neto, JP, ONOFRE, A. V. C., Wanderley-Nogueira, AC, ARRUDA, H., ALVES, G., SOTERO, D.
1st Brazilian-German Meeting Plant Systems Biology and Bioenergy, 2010. (Outro, Organização de evento)
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
3. Soares-Cavalcanti, N M
I Simpósio de Inovação em Ciências Biológicas, 2008. (Outro, Organização de evento)
Referências adicionais : Brasil/Português.
4. Benko-Iseppon, A.M., Kido, EA, Soares-Cavalcanti, N M, Wanderley-Nogueira, AC, Gitaí, J.
Curso "atualização em genética molecular", 2006. (Outro, Organização de evento)
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
5. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Gitaí, J.
IV Workshop Recife , 2006. (Outro, Organização de evento)
Referências adicionais : Brasil/Português. Meio de divulgação: Outro
6. Benko-Iseppon, A.M., Kido, EA, Soares-Cavalcanti, N M, Wanderley-Nogueira, AC
Curso "Técnicas moleculares, bioinformática e mapeamento aplicados ao melhoramento de plantas", 2005. (Outro, Organização de evento)
Referências adicionais : Brasil/Português. Meio de divulgação: Meio digital
7. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, Gitaí, J.
III Workshop Recife Megacity: open space project, 2005. (Outro, Organização de evento)
Referências adicionais : Brasil/Português. Meio de divulgação: Outro

Bancas

Participação em banca de trabalhos de conclusão

Graduação

1. Benko-Iseppon, A.M., Soares-Cavalcanti, N M, FERREIRA NETO, J. R. C., AMORIM, L. L. B.
Participação em banca de Marx Oliveira de Lima. Prospecção e análise in silico das esnaquinas no transcriptoma da soja (*Glycine max* (L.) Merr.), 2011
(Bacharelado em ciências Biológicas) Universidade Federal de Pernambuco
Palavras-chave: Bioinformática, Soja, esnaquinas
Áreas do conhecimento: Bioinformática, BIOTECNOLOGIA, Biologia Molecular
Referências adicionais: Brasil/Português.
2. Bezerra-Neto, JP, Benko-Iseppon, A.M., PANDOLFI, V., BARROS, P.S., Soares-Cavalcanti, N M
Participação em banca de João Pacifico Bezerra Neto. Caracterização in silico e análise evolutiva de aquaporinas no transcriptoma da soja (*Glycine max* (L.) Merr.), 2009
(Bacharelado em ciências Biológicas) Universidade Federal de Pernambuco
Palavras-chave: Aquaporinas, Bioinformática, Soja
Referências adicionais: Brasil/Português.

Totais de produção

Produção bibliográfica

Artigos completos publicados em periódico	9
Artigos aceitos para publicação	14
Capítulos de livros publicados	2
Trabalhos publicados em anais de eventos	56
Apresentações de Trabalhos (Conferência ou palestra)	1
Apresentações de Trabalhos (Congresso)	6

Produção Técnica

Curso de curta duração ministrado (outro)	6
Desenvolvimento de material didático ou instrucional	2

Orientações

Orientação concluída (trabalho de conclusão de curso de graduação)	1
Orientação em andamento (iniciação científica)	1

Eventos

Participações em eventos (congresso)	8
Participações em eventos (seminário)	1
Participações em eventos (simpósio)	2
Participações em eventos (outra)	1
Organização de evento (outro)	7
Participação em banca de trabalhos de conclusão (graduação)	2

Outras informações relevantes

- 1 - Participação da elaboração de Relatório de Impacto Ambiental (RIMA) referente à Linha de Transmissão da Usina Hidroelétrica Estreito (Consórcio entre as empresas Chesf, Construtora Queiroz Galvão, ENERGIMP S.A. e CNEC Engenharia S.A.) junto à empresa PROJETEC no ano de 2009. - Participação na equipe de elaboração de Estudo de Impacto de Vizinhança (EIV) do empreendimento Hotel Canarius de Gravata sob a coordenação da Arquiteta Lúcia de Fátima Soares Êscorel (CREA nº 8843-D/PE).



An overall evaluation of the Resistance (*R*) and Pathogenesis-Related (*PR*) superfamilies in soybean, as compared with *Medicago* and *Arabidopsis*

Ana C. Wanderley-Nogueira¹, Luis C. Belarmino¹, Nina da M. Soares-Cavalcanti¹, João P. Bezerra-Neto¹, Ederson A. Kido¹, Valesca Pandolfi¹, Ricardo V. Abdelnoor², Eliseu Bineck², Marcelo F. Carazzole^{3,4} and Ana M. Benko-Iseppon¹

¹Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Recife, PE, Brazil.

²Embrapa Soja, Londrina, PR, Brazil.

³Laboratório de Genômica e Expressão, Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.

⁴Centro Nacional de Processamento de Alto Desempenho em São Paulo, Universidade Estadual de Campinas, Campinas, SP, Brazil.

Abstract

Plants have the ability to recognize and respond to a multitude of pathogens, resulting in a massive reprogramming of the plant to activate defense responses including Resistance (*R*) and Pathogenesis-Related (*PR*) genes. Abiotic stresses can also activate *PR* genes and enhance pathogen resistance, representing valuable genes for breeding purposes. The present work offers an overview of soybean *R* and *PR* genes present in the GENOSOJA (Brazilian Soybean Genome Consortium) platform, regarding their structure, abundance, evolution and role in the plant-pathogen metabolic pathway, as compared with *Medicago* and *Arabidopsis*. Searches revealed 3,065 *R* candidates (756 in Soybean, 1,142 in *Medicago* and 1,167 in *Arabidopsis*), and *PR* candidates matching to 1,261 sequences (310, 585 and 366 for the three species, respectively). The identified transcripts were also evaluated regarding their expression pattern in 65 libraries, showing prevalence in seeds and developing tissues. Upon consulting the SuperSAGE libraries, 1,072 *R* and 481 *PR* tags were identified in association with the different libraries. Multiple alignments were generated for *Xa21* and *PR-2* genes, allowing inferences about their evolution. The results revealed interesting insights regarding the variability and complexity of defense genes in soybean, as compared with *Medicago* and *Arabidopsis*.

Key words: pathogen response, biotic stress, bioinformatics, *Glycine max*, *Medicago truncatula*.

Received: Accepted:

Introduction

In order to prevent the effects of pathogen attack, plants evolved the ability to recognize the threat and struggle against the invader as well as trigger an effective response (Bolton, 2009). One of the most important steps of this complex response lies in the detection of pathogen invaders by the plant, a step where *R* (Resistance) genes play a crucial role. This sensing involves the recognition of a pathogen gene product called avirulence (*avr*) factor by a correspondent *R* gene. The plant will be resistant and the pathogen growth and establishment will be impaired when both *avr* and *R* genes are compatible, leading to the so-

called Hypersensitive Response (HR) that triggers diverse responses, including local cell death to impair spreading of the pathogen (Bonas and Anckerveken, 1999). Besides this local reaction, the HR activates a signal cascade – including hormones and *PR* (Pathogen Related) genes, among others – that are able to establish resistance against a spectrum of different pathogen classes, this corroborating observations made at the beginning of the last century that plants, as well as animals (Benko-Iseppon *et al.*, 2010), may be immunized against the attack of a given pathogen after infection by another pathogen (Chester, 1933).

Besides a local reaction, plants may also display the Systemic Acquired Resistance (SAR). The SAR pathway is also common in many non-compatible plant-pathogen interactions (Nurnberg and Brunner, 2002). As soon as the pathogenic agent is detected, the plant induces a complex set of signal molecules able to activate defense proteins that

Send correspondence to Ana Maria Benko-Iseppon. Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego 1235, 50.670-420 Recife, PE, Brazil. E-mail: ana.iseppon@gmail.com.

may have a direct antimicrobial effect, as in the case of Pathogenesis-Related (*PR*) genes (Durrant and Dong, 2004). Alternatively, they may induce the production of secondary metabolites that impair pathogen movement or growth within the plant tissues (Sparla *et al.*, 2004).

Resistance genes are generally classified into five different groups or classes, defined according to their conserved domains (CD) (Bent, 1996; Hammond-Kosack and Jones, 1997; Ellis and Jones, 2000). The first class is represented by the *HMI* gene of maize that encodes a reductase able to inactivate toxins produced by the fungus *Helminthosporium carbonum* (Joahal and Briggs, 1992). It is the only *R* gene class where conserved domains are absent. A second class is represented by the *Pto* gene from tomato that confers resistance against the bacterium *Pseudomonas syringae* pv. *tomato*. It is characterized by a serine/threonine-kinase (ser/thre-kinase) domain, able to interact with the *avrPto* gene (Tang *et al.*, 1999). This gene was also identified in other plants, such as *Arabidopsis thaliana*, *Phaseolus vulgaris* (Melotto *et al.*, 2004), eucalyptus (Barbosa-da-Silva *et al.*, 2005) and sugarcane (Wanderley-Nogueira *et al.*, 2007).

The third class is represented by genes bearing two domains, viz. LRR (*Leucine Rich Repeats*) and NBS (*Nucleotide Binding Site*) (Liu *et al.*, 2004). This is the case of the *Rpm1* and *Rps2* genes from *A. thaliana*, the *N* gene from tobacco, *L6* from flax, *Prf* from tomato and *Rpg1* from soybean also found in common bean, and faba bean (Mindrinos *et al.*, 1994; Lawrence *et al.*, 1995; Salmeron *et al.*, 1996; Ashfield *et al.*, 2003). The fourth *R* gene class encodes a membrane-anchored protein composed of an extracellular LRR domain, a transmembrane region and a short intracellular tail in the C terminal. The *Cf* gene from tomato is an example of this class, conferring resistance against *Cladosporium fulvum* (Dixon *et al.*, 1996).

The *Xa21* gene from rice confers resistance to the bacteria *Xanthomonas oryzae* pv. *oryzae* and is a representative of the fifth class (Song *et al.*, 1995; Wang *et al.*, 1995). This gene encodes an extracellular LRR domain (similar to the *Cf* gene), as well as a ser/thre-kinase domain (similar to the *Pto* gene), suggesting an evolutionary connection among different classes in the genesis of plant *R* genes (Song *et al.*, 1997).

PR proteins comprise pathogen-induced proteins that are routinely classified into 17 families based on their biochemical and molecular biological properties, from *PR-1* to *PR-17* (van-Loon *et al.*, 2006). Similarities among sequences and serological or immunological properties form the basis of their classification (van-Loon *et al.*, 1999). Although most *PR* proteins are known to have antifungal activities, their active molecular mechanisms are not well understood except for *PR-2* (β -glucanases) and *PR3* (chitinases) (Kitajima and Sato, 1999). *PR1* is the most abundantly accumulated protein after pathogen infection and its genes have been cloned in many plants, such as to-

bacco (Gaffney *et al.*, 1993), *A. thaliana* (Metzler *et al.*, 1991), tomato (Tornero *et al.*, 1997) and apple. Although its phytochemical functions are unknown in all these species, this gene class is nonetheless considered to be a typical SAR marker (Bonasera *et al.*, 2006). *PR-5* is a thaumatin-like protein with high antifungal activity, being also expressed under cold stress in overwintering monocots where it exhibits antifreeze activities (Hon *et al.*, 1995; Atici and Nalbantolu, 2003; Griffith and Yaish, 2004). Other families like *PR-8* (Glycosyl hydrolase), *PR-9* (secretory peroxidase), *PR-14* (lipid transfer proteins), *PR-15* (oxalate oxydase) and *PR-17* (basic secretory proteins) (Nanda *et al.*, 2010) have been well studied and are believed to be involved in plant defense responses, although their molecular mechanisms have yet to be determined (Bolton, 2009). Most *PR* genes are expressed at a basal level under normal growth conditions, but are rapidly induced after pathogen infection. It is worthy of note that several *PR* genes are also regulated during development, leaf senescence and pollen maturation, as well as by environmental factors, such as osmotic, cold and light stress (Zeier *et al.*, 2004).

Soybean (*Glycine max*) is a globally important crop, providing oil and at least twice as much protein per acre as any other major grain (Libault *et al.*, 2010). Economically, soybean is the most valuable protein and edible oil crop in the world and serves as a model for seed and other developmental processes (Cannon *et al.*, 2009).

The present evaluation offers an overview of the main available sequences regarding plant-pathogen interaction of the *R* and *PR* classes in the soybean transcriptome, here compared with data available from *Arabidopsis* and *Medicago*, providing insights on the expression of such sequences in different tissues and inferring as to how these genes may have behaved over the course of evolution.

Material and Methods

Search and screening for *R* and *PR* genes in soybean, *Medicago* and *Arabidopsis* databases

For this purpose 59 proteins that play important roles in plant defense response were selected as seed sequences. The selected protein sequences were related to the 42 *R* and 17 *PR* gene classes described above. The *R* genes were previously compiled by Barbosa-da-Silva *et al.* (2005) and Wanderley-Nogueira *et al.* (2007), and *PR* seed sequences are available in Table S1 (Supplementary Material). All 59 seed sequences regarded full cDNAs that were obtained from the NCBI database and conceptually translated to improve search strategies.

For the identification of these gene analogs in soybean, *Medicago* and *Arabidopsis* transcriptomes, tBLASTx alignments were carried out against three platforms: GENOSOJA (The Brazilian Soybean Genome Consortium), TIGR (The Institute for Genomic Research) and

Insight on Pathogen Defense Mechanisms in the Sugarcane Transcriptome

Ana Carolina Wanderley-Nogueira¹, Ederson Akio Kido¹, Nina da Mota Soares-Cavalcanti¹, Luis Carlos Belarmino¹, João Pacífico Bezerra-Neto¹, William Lee Burnquist², Sabrina Moutinho Chabregas², José Ivo Baldani³, Ana Maria Benko-Iseppon^{1*}

¹ Genetics Department, Center of Biological Sciences, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego 1235, CEP 50.670-420, Recife, PE, Brazil

² Centro de Tecnologia Canavieira (CTC), Bairro Santo Antonio, CEP 13400-970, Piracicaba, SP, Brazil

³ Embrapa Agrobiologia (CNPAB), Rodovia BR 465 km 7, CEP 23890-000, Seropédica, RJ, Brazil

Corresponding author*: E-mail: ana.benko.iseppon@pq.cnpq.

ABSTRACT

Recognition of pathogen and activation of defense mechanisms is a common feature known from all multicellular organisms. Among higher plants Systemic Acquired Resistance (SAR) is known to activate Pathogen Related (PR) genes after recognition of the pathogen mediated by a Resistance (R) gene. Both gene classes (R and PR) represent the main mechanism against biotic and sometimes also abiotic stresses. Therefore, the identification of SAR metabolism components is an important concern regarding plant breeding. Sugarcane (*Saccharum officinarum*) is a major tropical and subtropical crop, actually grown in more than 120 countries, being especially important due to its value for the production of bioethanol, constituting a renewable energy source, besides the sugar production. The present work brings an overview on sugarcane Resistance and Pathogenesis-Related genes, regarding their structure, abundance and role in the plant-pathogen metabolic pathway and

also regarding abiotic stresses, as compared with rice and other higher plants. For this purpose a collection of 282,818 expressed sequence tags have been evaluated aiming to identify R and PR genes as well as important factors identified in the classic plant-pathogen metabolic KEGG pathway using rice full length cDNA as seed-sequences. The identified sugarcane genes have been used to screen four SuperSAGE libraries with 8,787,313 tags, allowing the identification of the main activated and repressed genes under abiotic stress (drought/salinity) conditions. The 1460 identified genes have been plotted on a rice virtual karyotype inferring about their distribution considering a putative synteny as a measure to infer about their relative position within rice and sugarcane chromosomes. The results revealed interesting insights on the variability and complexity of defense genes in sugarcane.

Key words: Bioinformatics, R-genes, PR-genes, biotic stress, abiotic stress

Abbreviations: **ABA**, Absciscic Acid; **AS**, Salicylic Acid; **avr**, avirulence; **BLAST**, Basic Local Alignment Tool; **EST**, Expressed Sequence Tag; **HR**, Hypersensitive Response; **LRR**, Leucine Rich Repeats; **NBS**, Nucleotide Binding Site; **PR**, Pathogenesis Related; **R**, Resistance; **ROS**, Reactive Oxygen Species; **SAR**, Systemic Acquired Resistance; **Ser/Thre-Kinase**, Serine/Threonine Kinase; **ST**, Signal Transduction; **SuperSAGE**, Super Serial Analysis of Gene

Short Running Title: **Pathogen Defense Mechanisms in Sugarcane**

INTRODUCTION

The prevalence of a disease constitutes an abnormal condition, affecting plant growth and impairing important physiological processes. As highlighted by FAO (2005), fungi, bacteria, virus and

IN SILICO SCREENING FOR PATHOGENESIS RELATED-2 GENE CANDIDATES IN *VIGNA UNGUICULATA* TRANSCRIPTOME

Ana Carolina Wanderley-Nogueira⁽¹⁾, Nina da Mota Soares-Cavalcanti⁽¹⁾, Luis Carlos Belarmino⁽¹⁾, Adriano Barbosa-Silva^(1,2), Ederson Akio Kido⁽¹⁾, Semiramis Jamil Hadad do Monte⁽³⁾, Valesca Pandolfi^(1,3), Tercilio Calsa-Junior⁽¹⁾, Ana Maria Benko-Iseppon⁽¹⁾

(1)Universidade Federal de Pernambuco
Center of Biological Sciences, Departament of Genetics, Laboratory of Plant Genetics and Biotechnology, R. Prof. Moraes Rêgo s/no., Recife, PE, Brazil
ana.benko.iseppon@pesquisador.cnpq.br

(2) Max-Delbrueck Center for Molecular Medicine
Computational Biology and Data Mining Group, Robert-Roessle-Str. 10, 13125, Berlin, Germany

(3) Universidade Federal do Piauí
Laboratory of Immunogenetics and Molecular Biology, Campus Petrônio Portela bloco 16, Teresina, PI, Brazil

Keywords: bioinformatics, pathogenesis-related, Beta-glucanases, biotic stress, cowpea.

abstract. Plants evolved diverse mechanisms to struggle against pathogen attack, for example the activity of Pathogenesis-Related (*PR*) genes. Within this category *PR*-2 encodes a Beta-glucanase able to degrade the polysaccharides present in the pathogen cell wall. The aim of this work was to screen the NordEST database to identify *PR*-2 members in cowpea transcriptome and analyze the structure of the identified sequences as compared with data from public databases. After search for *PR*-2 sequences in NordEST; CLUSTALx and MEGA4 were used to align *PR*-2 orthologs and generate a dendrogram. CLUSTER program revealed the expression pattern through differential display. A new tool was developed aiming to identify plant *PR*-2 proteins based in the HMMER analysis. Among results, a complete candidate from cowpea could be identified. Higher expression included all libraries submitted to biotic (cowpea severe mosaic virus, CPSMV) stress, as well as wounded and salinity stressed tissues, confirming *PR* expression under different kind of stresses. Dendrogram analysis showed two main clades, the outgroup and Magnoliopsida where monocots and dicots organisms were positioned as sister groups. The developed HMM model could identify *PR*-2 also in other important plant species, allowing the development of a bioinformatic routine that may help the identification not only of pathogenesis related genes but any other genes, classes that present similar conserved domains and motifs.

1 Introduction

In response to persistent challenge by a broad spectrum of microorganisms, plants have evolved diverse mechanisms to prevent infection [1]. The first one is the hypersensitive response (HR), that is immediate and starting with signal recognition from a pathogen elicitor by a host Resistance (*R*) gene leading to rapid cell death [2]; the second strategy is the systemic activation of genes encoding mitogen-activated protein kinases (MAPKs) and pathogenesis-related (*PR*) proteins which are directly or indirectly inhibitory towards pathogens and have been associated with the phenomenon of systemic acquired resistance (SAR) [3].

Identification and Expression of Early Nodulin in Sugarcane Transcriptome Revealed by in Silico Analysis

Gabriela Souto Vieira-de-Mello, Petra Barros dos Santos, Nina da Mota Soares-Cavalcanti, and Ana Maria Benko-Iseppon

Universidade Federal de Pernambuco, Center of Biological Sciences, Department of Genetics, Laboratory of Plant Genetics and Biotechnology, R. Prof. Moraes Régio 1235, CEP 50670-420, Recife, PE, Brazil
ana.benko.iseppon@pq.cnpq

Abstract. Nodulin genes have been defined as plant genes that are induced during nodule formation in legumes. Many studies, however, revealed a number of nodulins in non-legumes, including monocot plants, suggesting that these genes play additional roles besides nodulation. The presence and expression profile of the early nodulin genes (*Annezin*, *DMT3*, *NIN*, *NORK*, *CCS52A*, and *ENOD8*) was evaluated in the sugarcane transcriptome (237,954 ESTs) using in silico procedures. 129 sugarcane clusters were identified (out of 1,476 transcripts) and their expression profile was evaluated. Higher expression was observed in libraries of flowers, roots and normalized mix of tissues, confirming their multi-function character besides the plant-bacteria endophytic interaction in sugarcane. The multiple alignments revealed high homology among sugarcane sequences and respective proteins from other plants, mainly monocots, revealing a relatively conserved genetic structure among species, probably regarding ancient genetic processes.

Keywords: early nodulins, nitrogen fixation, expression pattern.

1 Introduction

Sugarcane is one of the most important sources of sugar and alcohol in the world and is cultivated in tropical and subtropical areas in more than 80 countries around the globe. Several sugarcane varieties have the ability to grow with low nitrogen fertilizer inputs, being selected for high yields with low inputs of inorganic nitrogen fertilizer [1]. This important crop establishes association with endophytic diazotrophic bacteria, including *Gluconacetobacter diazotrophicus*, *Herbaspirillum seropedicae* and *H. rubrisubalbicans*, showing unique features when compared with other nitrogen-fixing associations. Bacteria colonizes the intercellular spaces and vascular tissues of most organs of the infected symbiont promoting plant growth, without causing visible disease symptoms [1] [2]. It is still unclear which mechanisms are involved in the establishment of this particular type of interaction and what kind of molecules mediate signaling between plant and bacteria [3].

R. Rizzo and P. J. G. Lisboa (Eds.): CIBB 2010, LNBI 6685, pp. 72–85, 2011.
© Springer-Verlag Berlin Heidelberg 2011

Nodulins have been defined as plant genes that are exclusively induced during nodule formation in legume plants. Many studies, however, revealed a number of nodulin-related sequences in non-legumes, as *ENOD40* [3]. Moreover some non-leguminous plants, including rice, have the ability to perceive lipochitoooligosaccharide nodulation signal molecules (nod factors) produced by the rhizobia, suggesting that nodulation related processes are present in non legumes [4]. Recent evaluations indicated that the molecular communication between sugarcane and the microbes might involve lipopolysaccharides present in the outer membrane of these gram-negative bacteria [2]. In addition, some sugarcane genes involved in plant-bacteria signalization during the association and nitrogen metabolism are probably activated by the endophytic bacteria in the early steps of plant colonization, allowing the plant to assimilate the nitrogen fixed by the bacteria [5]. These genes also seem to act as nodule activators, once they present homology with some legume nodulins [1].

No previous evaluation of early nodulin genes was carried out in the sugarcane transcriptome, despite of the potential application of these genes in increasing the efficiency of the association between this plant and diazotrophic bacteria. In this context the present work aimed to perform an in silico identification and characterization of early nodulins in the sugarcane transcriptome, using known legume nodulin sequences as templates, evaluating also the expression profiles of nodulin-related sequences in this organism.

2 Material and Methods

For the annotation routine (Fig. 1) known full length cDNA sequences of early nodulin genes (*Annezin*, *DMT3*, *NIN*, *NORK*, *CCS52A* and *ENOD8*) from legumes (Tab. 1) were used as seed sequences against the SUCEST database with aid of a local tBLASTn search tool. Only sequences with e-value up to e^{-10} or less were used for a homology screening in Genbank using the BLASTx tool [6]. The cluster frame of the tBLASTn alignment was used to predict the Open Reading Frames (ORFs) for each selected cluster. Sugarcane clusters were translated using the ORF-finder tool at NCBI and screened for conserved motifs with aid of the RPS-BLAST CD-search tool.

The prevalence of sugarcane early nodulins was based on the number of reads that composed each cluster, followed by data normalization and calculation of the relative frequency (reads per library). A hierarchical clustering approach was applied using normalized data, from all identified nodulins transcripts, allowing the generation of a graphic representation with aid of the CLUSTER program [7]. The resulting dendrograms including both axes (using the weighted pair group for each gene class and library) were generated using the TreeView program for Windows [8]. Multiple alignments (CLUSTALx program) were generated using sugarcane complete sequences together with sequences from other organisms, searched NCBI. The phylogenetic analysis was performed using the MEGA program (Version 4 for Windows [9]) using maximum parsimony method, with bootstrap of 2,000 replications and pairwise deletion for the treatment of

Osmoprotectants in the Sugarcane (*Saccharum* spp.) Transcriptome Revealed by in Silico Evaluation

Petra Barros dos Santos, Nina da Mota Soares-Cavalcanti, Gabriela S. Vieira-de-Melo, and Ana Maria Benko-Iseppon

Universidade Federal de Pernambuco, Center of Biological Sciences, Department of Genetics, Laboratory of Plant Genetics and Biotechnology, R. Prof. Moraes Régio 1235, CEP 50670-420, Recife, PE, Brazil
ana.benko.iseppon@pq.cnpq

Abstract. Environmental stresses such as drought and salinity limit crop productivity in worldwide level. These stresses often lead to the accumulation of osmoprotectants in most organisms, including plants. In the present work, a search of known osmoprotectants (*P5CS*, *P5CR*, *INPSI*, *BADH*, *CMO*, *TPS*, *TPP*, *OASYL* and *SAI*) was carried out in the sugarcane transcriptome (237,954 expressed sequence tags) using *in silico* procedures. Alignments revealed that sugarcane presents a high number of osmoprotectant candidate genes, with 56 clusters found. *In silico* expression revealed higher expression in stressed callus tissues and those infected by *Herbaspirillum rubrisubtilicans* (HR), confirming the multi-function character of the osmoprotectants. As expected, the phylogenetic analysis revealed distinct groups among angiosperms, algae, animals, fungi and bacteria, in almost all dendrograms, with high degree of sequence conservation among angiosperms. As observed in comparative analysis between the ORFs of sugarcane and other organisms, the genic structure of these plants was relatively conserved suggesting that the accumulation of compatible solutes is an ancient metabolic adaptation.

Keywords: data mining, salt stress, drought stress, crop evolution, differential expression.

1 Introduction

Among several abiotic stresses, drought and salinity are the most important factors influencing the growth, survival, yield and natural distribution of plants worldwide [1]. Plants have developed mechanisms like stress avoidance to cope with low water content and the synthesis of compatible osmolytes, being one of the main mechanisms that organisms, including plants, have to prevent the harmful effects caused by abiotic stresses [2]. Osmoprotectants are also termed compatible solutes because they are accumulated by the plants without disturbance or interference in the cellular metabolism, besides their protective properties. Proline (Pro), glycinebetaine (GB), myo-inositol, trehalose and cysteine

(Cys) are important osmoprotectants that in a general way stabilize proteins and membranes against the denaturing effects of high salt concentrations and other harmful solutes, facilitating the retention of water in the cytoplasm, raising cellular osmotic pressure, and allowing the protection of membranes, protein complexes and cellular constituents [3].

A key to progress towards breeding better crops under abiotic stress has been to understand the changes in cellular, biochemical and molecular machinery of the plants. It is believed that osmoregulation would be the best strategy for abiotic stress tolerance, especially if osmoregulatory genes could be triggered in response to drought, salinity and high temperature [4]. To date, various approaches in genetic engineering have allowed the introduction of new pathways for the biosynthesis of various compatible solutes in plants [3].

The present work performed a data-mining based identification of osmoprotectant genes in the sugarcane database (SUCEST project) [5], using well known rice sequences as templates, and comparing the obtained clusters with sequences from public databases and literature data. Furthermore, the expression profile of the different osmoprotectant categories was evaluated, aiming to identify their role under different conditions in this crop. An *in silico* routine for this purpose is also presented, being applicable for the search of osmoprotectants in monocots in general. As little information is available about these molecules from sugarcane, that is cultivated mainly in tropical and subtropical areas and considered one of the world's most important crop plants, it is believed that this research may contribute for breeding purposes in pulse crops like sugarcane.

2 Material and Methods

Protein seed sequences of P5CS, P5CR, TPP, TPS, BADH, CMO and INPSI from rice were obtained at TIGR (The Institute for Genomic Research) database [6]. Each sequence was compared against the SUCEST database using tBLASTn tool and sequences with e-value equal to e^{-05} or less were annotated. After that a reverse alignment with BLASTx at NCBI was made to confirm the identity of the sequences. Sugarcane clusters were translated using the ORF Finder tool at NCBI and screened for conserved domains using RPS-BLAST CD-search tool [7]. Multiple alignments with proteins from sugarcane and other organisms that presented complete domains were generated at the CLUSTALx program [8]. To avoid influence of sequence sizes in the alignments, the non aligned 5' and 3' extremities were excluded, as well as autapomorphic, non informative internal sequence regions.

A phylogenetic maximum parsimony analysis using bootstrap function (2,000 replications) was performed, generating a consensus tree with a cut-off of 50 using the program MEGA Version 4 for Windows [9]. For each cladogram the most basal organisms have been manually settled as outgroup, in all cases non angiosperms.

The prevalence of sugarcane clusters were verified by direct counting of the reads that composed each cluster, followed by data normalization (considering

ANALYSIS AND ANNOTATION OF CHICKPEA BACTERIAL ARTIFICIAL CHROMOSOME SEQUENCES

Luis C. Belarmino⁽¹⁾, Ana C. Wanderley-Nogueira⁽¹⁾, Nina M. Soares-Cavalcanti⁽¹⁾, Valesca Pandolfi⁽¹⁾, Lidiane L. B. Amorim⁽¹⁾, Peter Winter⁽²⁾, Guenter Kahl⁽²⁾, Ana M. Benko-Iseppon^{(1)*}

(1) Universidade Federal de Pernambuco

Center of Biological Sciences, Department of Genetics, Laboratory of Plant Genetics and Biotechnology, R. Prof. Moraes Rêgo 1235, CEP 50670-901, Recife-PE, Brazil.

(2) Plant Molecular Biology

Biocenter, University of Frankfurt, Marie-Curiestr. 9, D-60439, Frankfurt/Main, Germany.

*Corresponding author, email: ana.benko.iseppon@pq.cnpq.br

Keywords: *Ab initio* gene prediction, stress related genes, SSR, molecular markers, transposable elements.

Abstract. Despite of the huge amount of genomic sequence data for some crop species, few *in silico* analysis have been carried in order to generate information important for breeding purposes. This is the case of chickpea (*Cicer arietinum* L.), a crop that presents a narrow genetic background making it difficult for breeders to produce cultivars with durable resistance to major biotic and abiotic stresses. Molecular markers have been used in chickpea mapping to pinpoint agronomical features like tolerance/resistance to help the development of new cultivars. Here, we report an in-depth analysis of six bacterial artificial chromosomes (comprising 228,5 kbp) generated for chickpea. We have used an EST-cross-alignment strategy in order to take advantages of the huge amount of data available from legume species. Nineteen genes were identified among the BACs that might play in some extent a role in pathogen resistance. A region of 59 kb showed intense synteny with *Medicago truncatula*. Moreover, we have investigated the content of repetitive DNA in these BACs and found that this region has been extensively shaped by transposition events, aside of bearing a set of microsatellites sequences with a thorough potential for linkage map analysis. The *in silico* procedures used are recommended for further analysis regarding genomic sequences in crop species aiming to help the conversion of sequence data in applicable molecular markers.

1 Introduction

Chickpea (*Cicer arietinum* L.) is an annual, self-pollinated diploid species ($2n = 16$) with a relatively small genome (740 Mbp) [1], presenting a total production of 9.7 Mt and an average productivity of 882 kg/ha, being the world's third most widely grown food legume [2]. However, the growth and productivity of chickpea are adversely affected by biotic and abiotic stresses [3]. Increasing the combined tolerance/resistance to biotic and abiotic stresses are major aims of chickpea breeders around the world. In this context, progress in the development of novel genetic tools such as DNA molecular markers, dense genetic maps, and whole-genome transcription profiling techniques have been developed to assist chickpea improvement [4, 5, 6, 7].

The generation of Bacterial Artificial Chromosomes (BACs) is essential for the construction of a physical map, generating a valuable resource for map-based cloning, identification of gene-controlling regions and comparative genomics [8]. BAC libraries representing several fold genome coverage have been produced for the two model legume crops *Medicago truncatula* and *Lotus japonicus* L., and in other agronomic crops as

IN SILICO ANALYSIS OF CLASS II TRANSPOSABLE ELEMENTS IN THE COWPEA AS COMPARED WITH THE ACTUAL SOYBEAN KNOWLEDGE

Lidiane L. B. Amorim⁽¹⁾, Luis C. Belarmino⁽¹⁾, Bruna Piereck⁽¹⁾, Nina M. Soares-Cavalcanti⁽¹⁾, Ana C. Wanderley-Nogueira⁽¹⁾, Valesca Pandolfi⁽¹⁾, Ana C. Brasileiro-Vidal⁽¹⁾, Ana M. Benko-Iseppon^{(1)*}

(1) Universidade Federal de Pernambuco

Center of Biological Sciences, Department of Genetics, Laboratory of Plant Genetics and Biotechnology, R. Prof. Moraes Rêgo 1235, CEP 50670-901, Recife-PE, Brazil.

*Corresponding author, email: ana.benko.iseppon@pq.cnpq.br

Keywords: In silico analysis, mobile DNA sequences, legume evolution, *Vigna unguiculata*, *Glycine max*.

Abstract. The completion of plant genome sequences has enabled unprecedented systematic analysis of coding and non-coding DNA fractions. Transposable elements (TEs) are important genome components being still poorly investigated in cowpea. The recent publication of the cowpea genespace sequences opened the possibility for an in depth analysis of its transposon content. Using known TE from soybean as seed sequences we analyzed 153,476 sequences of the cowpea genespace database and found 1,602 TE elements belonging to six of the previously recognized superfamilies of class II elements: *Mutator* (736 sequences), *PIF/Harbinger* (376), *CACTA* (367), *Helitron* (77), *Tc1/Mariner* (37) and *hAT* (9). An association analysis permitted the identification of numerous TEs in the vicinity of genes (especially kinases, resistance genes and transcription factors) indicating their importance in the evolution of those regions. Our analysis provides a preview regarding the abundance and diversity of the transposable elements in cowpea.

1 Introduction

The availability of plant genome sequences marked a decisive turn in our understanding of angiosperm genome content and evolution [1]. These sequences confirm that transposable elements (TEs) are among the most abundant components of the genome in higher plants, acting as recombination hot spots, allowing the acquisition of specific cell functions or still influencing protein-coding regions [2, 3]. Therefore, the precise and comprehensive detection of TE sequences is a prerequisite for fully understanding their function in eukaryotic genomes [4].

There are two functional classes of transposable elements according to their transposition mechanisms: class I (retrotransposons), which move via a RNA intermediate in a 'copy-and-paste' fashion, and class II (transposons), which move via a DNA intermediate by a 'cut-and-paste' mechanism. Class I members can be divided into five orders: long terminal repeat (LTR), dispersed nuclear element (DIRS), penelope-like element (PLE), short interspersed nuclear element (SINE) and long interspersed nuclear element (LINE). On the other hand, class II elements can be divided into two major subclasses: subclass 1 comprises classical 'cut-and-paste' TEs of the order TIR and subclass 2 holds DNA TEs that undergo a transposition process that entails replication without double-stranded cleavage, the rolling-circle *Helitrons* and the self-replicating *Maverick* (or *Polinton*) elements [5, 6]. Transposable elements subclasses can be further split into superfamilies based on structural features and phylogenetic clustering. In

In Silico Evaluation of Osmoprotectants in *Eucalyptus* Transcriptome

Petra dos Santos Barros, Nina M. Soares-Cavalcanti, Gabriela S. Vieira-Mello,
Ana C. Wanderley-Nogueira, Tercílio Calsa-Junior,
and Ana M. Benko-Iseppon*

Universidade Federal de Pernambuco, Center for Biological Sciences (CCB),
Genetics Department, Laboratory for Plant Genetics and Biotechnology,
Av. Prof. Moraes Rêgo s/no. Recife, PE, Brazil
Tel.: +55-81-2126-8520
ana.benko.iseppon@pesquisador.cnpq.br

Abstract. The synthesis of osmoprotectants is one of the main mechanisms to dribble harmful effects caused by abiotic stresses. In the present work, osmoprotectant genes were selected and used in a search for orthologs in the FORESTs database using *in silico* procedures. The studied genes were here analyzed for the first time in *Eucalyptus*, including 51 identified orthologs from nine gene families ($P5CS=3$, $P5CR=4$, $TPS1=10$, $TPPB=8$, $SAT=6$, $OASTL=13$, $CMO=1$, $BADH=3$ and $INPS1=3$). In general, the most represented genes were those involved in the trehalose and cysteine synthesis, probably due to their vital importance in the maintenance of cellular homeostasis. Overall expression pattern revealed significant amounts of reads in most libraries. Dendrograms based on analysis of selected genes from other flowering plants revealed no clear separation between mono and dicots, since diversity regarding the evolution and diversity in gene signatures may be more related to fitness and adaptation, than evolutionary patterns.

Keywords: Salinity and Drought Stress, Osmolytes, Data Mining, Crop Evolution.

1 Introduction

Despite of significant information regarding response to abiotic stresses in herbaceous plants, as rice and arabidopsis, scarce information is available for understanding such mechanisms in woody plants, principal components of arid and semi-arid regions of the word [1]. Physiological characters widely associated with drought resistance in eucalyptus include greater osmotic adjustment, tissue elasticity, higher leaf water content and higher ability to resist desiccation [2]. Drought and salt stresses are often interconnected and cause cellular damages, such as osmotic and oxidative stresses. To cope with these environmental influences, plants have developed arrays of physiological and biochemical strategies

* Corresponding author.

Simpósio sobre Tolerância à Deficiência Hídrica em Plantas: Adaptando as Cul- turas ao Clima do Futuro



PERFIL GERAL DE EXPRESSÃO ASSOCIADO À TOLERÂNCIA À SALINIDADE E SECA EM FEIJÃO-CAUPI (*Vigna unguiculata*)

Nina da Mota Soares-Cavalcanti¹; Valesca Pandolfi¹; Ederson Akio Kido¹; Luis Carlos Belarmino¹; Laureen Michelle Houllou Kido²; Ana Maria Benko-Iseppon^{1,2}

Introdução

Dentre as plantas cultivadas e amplamente utilizadas pela população nordestina, nenhuma tem tantas características vantajosas como o feijão-de-corda (*Vigna unguiculata* (L.) Walp). Conhecido também como feijão-caupi, inclui o popular feijão-de-corda, fradinho ou macassar e o feijão-verde. Também é chamado de “*southern pea*” ou ervilha do sul nos Estados Unidos, país no qual constitui produto para exportação e consumo interno, na forma de enlatados (do tipo seleta de legumes). Algumas variedades e espécies próximas são usadas na obtenção de *moyashi* (broto-de-feijão), uma cara iguaria da cozinha asiática, já difundida no mundo todo, que também é comercializada como alimento enlatado ou congelado. Além de produzir os grãos citados, trata-se de uma das plantas mais versáteis que existem, sendo muito valiosa como forrageira, com uma capacidade ímpar de fertilizar solos pobres, mesmo em condições de calor, seca e salinidade. A capacidade do feijão-caupi de crescer sob condições de solo e ambiente estressantes tem sido reconhecida por muitos cientistas. No Brasil, trata-se do único feijão capaz de sobreviver com sucesso na região Norte (alta umidade, muita chuva, solo argiloso) e no Nordeste (seca, solo arenoso, por vezes salino, muito sol), crescendo também em locais menos estressantes e com clima temperado, como no estado de Santa Catarina e o estado norte-americano da Califórnia (Benko-Iseppon, 2001).

O feijão-caupi também se apresenta muito promissor como fonte de importantes genes para o melhoramento de outras leguminosas, como a soja, o feijão comum, a ervilha, a lentilha, o amendoim, o grão-de-bico, etc. (Benko-Iseppon, 2001). Comparado a essas leguminosas o caupi apresenta características superiores para a recuperação de áreas degradadas, uma vez que em associação com rizóbio nativo torna-se eficiente fixador de nitrogênio sob as mais adversas condições, o que o torna um excelente candidato para uso em biorremediação (recuperação de ambientes degradados), transformando regiões em processo de desertificação em áreas novamente férteis e produtivas (Benko-Iseppon, 2001).

Estresse abiótico em plantas

O crescimento e desenvolvimento vegetal podem ser severamente afetados por estresses abióticos, como seca, alta salinidade e temperaturas baixas

¹UFPE - Universidade Federal de Pernambuco. Recife, PE.

²CETENE - Centro de Tecnologias Estratégicas do Nordeste. Recife, PE. E-mail: ana.benko.iseppon@pq.cnpq.br