UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

RESPOSTAS TRANSCRICIONAIS E ESTRESSE-INDUZIDAS EM GENÓTIPOS CONTRASTANTES DE *Glycine max* (SOJA) E *Vigna unguiculata* (FEIJÃO-CAUPI).

RECIFE

FEVEREIRO/2016

JOÃO PACIFICO BEZERRA NETO

RESPOSTAS TRANSCRICIONAIS E ESTRESSE-INDUZIDAS EM GENÓTIPOS CONTRASTANTES DE *Glycine max* (SOJA) E *Vigna unguiculata* (FEIJÃO-CAUPI).

Tese de doutorado apresentada à coordenação do Programa de Pós-Graduação em Ciências Biológicas como parte dos requisitos para obtenção do título de Doutor em Ciências Biológicas, na área de concentração Biotecnologia/ Biologia Celular e Molecular.

Orientadora: Drª Ana Maria Benko Iseppon

Coorientadora: Drª Valesca Pandolfi

RECIFE

FEVEREIRO/2016

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

Bezerra Neto, João Pacifico
Repostas transcricionais e estresse-induzidas em genótipos constrastantes de *Glycine max* (Soja) e *Vigna unguiculata* (Feijão-caupi) / João Pacifico Bezerra Neto. – Recife: O Autor, 2016.

206 f.: il.
Orientadora: Ana Maria Benko Iseppon, Valesca Pandolfi
Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Ciências Biológicas, 2016.
Inclui referências e apêndices

1. Leguminosa 2. Soja 3. Feijão-caupi 4. Bioinformática 5. Regulação da expressão gênica I. Iseppon, Ana Maria Benko (orient.) II. Pandolfi, Valesca (coorient.) Título.

583.74 CDD (22.ed.) UFPE/CCB-2017-124

RESPOSTAS TRANSCRICIONAIS E ESTRESSE-INDUZIDAS EM GENÓTIPOS CONTRASTANTES DE *Glycine max* (SOJA) E *Vigna unguiculata* (FEIJÃO-CAUPI).

Tese apresentada ao program de Pós-graduação em Ciências Biológicas da Universidade Federal de Pernambuco, como requisito final exigido para a obtenção do título de Doutor em Ciências Biológicas, área de concentração: **Biotecnologia**.

Data da Apresentação: 12 / 02 / 2016.

_____

Drª. Ana Maria Benko Iseppon (Orientadora)
Departamento de Genética – UFPE

_____

Dr. David Anderson de Lima Morais
Sherbrooke University – Canadá

_____

Dr. Antonio Mauro Rezende
Centro de Pesquisas Aggeu Magalhães – Recife/PE

_____

Dr. Ederson Akio Kido
Departamento de Genética – UFPE

_____

Drª Valesca Pandolfi
Departamento de Genética – UFPE

Aos meus pais (Ivonete Pacifico de Souza e Adalberto Vanderlei de Souza Filho) e ao meu irmão (Ivanberto Pacifico de Souza), ***dedico****.*

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, que me faz ver além do agora, me faz confiar no que não conheço, confiar no que já está planejado para minha vida.

Agradeço à minha mãe, Ivonete Pacifico de Souza, por ser para mim um exemplo de determinação e perseverança, por ter me ensinado que pelos filhos todo sacrifício é válido, por me apoiar em cada decisão difícil, sendo motivação para seguir em frente.

Agradeço também ao meu pai, Adalberto Vanderlei de Souza Filho, por ter feito o possível e impossível para permitir minha formação acadêmica e pessoal, por me mostrar que podemos ir além dos limites impostos.

Ao meu irmão, Ivanberto Pacifico, por ter me apoiado e incentivado nesta jornada. Aos meus familiares, em especial à minha tia Ivone Pacifico, minha segunda mãe, por todo o carinho e confiança em mim depositados.

Agradeço enormemente à Profª Drª Ana Maria Benko Iseppon, por ter me orientado desde minha graduação, por ter confiado em mim ao longo desses anos, por ser um exemplo de pessoa e profissional.

Agradeço a Drª Valesca Pandolfi (Vale), por sua coorientação, por ser amiga nas horas que mais necessitei, por me fazer rir nos momentos certos.

Agradeço ao Dr Adriano Barbosa da Silva, por ter me recebido em Luxemburgo e me orientado por novos caminhos na bioinformática.

Estendo meus agradecimentos aos amigos de laboratório, cada um com seu espaço e devida importância na minha vida. Aos antigos: Marx Lima, Bruna Piereck, Stephani, Sheyla Silva, Mitalle e Flávia, e aos novos: Carlos, Lívia, Dulce e Carol, que compartilham comigo grandes momentos, por tornarem meu dia-a-dia no laboratório mais animado.

Agradeço a Rebeca Rivas, que se fez presente nos momentos mais importantes, ainda que a distância física existisse.

Agradeço aos amigos que fiz no doutorado, mas que levarei para toda vida, Silvany Araújo e Hévila, vocês marcaram minha vida para sempre, aprendi com vocês que os vínculos não precisam ser só profissionais.

Agradeço à Vanessa Souza pelo auxílio na dissolução dos inúmeros problemas que teimavam em surgir, por sua amizade e orientações essenciais.

Agradeço a Suane e Suziany por me lembrarem que a amizade não tem prazo de validade.

Agradeço aos amigos que se fizeram presentes, mesmo que estando do outro lado do oceano, em especial à Jéssica Soares, que me ensinou que a qualidade de uma amizade não se mede em tempo de convivência.

Agradeço aos amigos que torceram por mim a cada momento, Paulo e Alice, Gabriel e Rayane, Willames, Rogério, Wanessa, e em especial à Waléria Kelly, a minha louca predileta. Aos que entenderam minha ausência.

Agradeço o suporte financeiro oferecidos pelas agências de fomento à pesquisa, CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), BNB (Banco do Nordeste do Brasil), FINEP (Financiadora de Estudos e Projetos) e FACEPE (Fundação de Amparo à Ciência e Tecnologia de Pernambuco).

E a todos que de alguma forma contribuíram para a realização e conclusão deste trabalho.

Muito obrigado.

“Ninguém nasce feito, é experimentando-nos no mundo que nós nos fazemos. ”

(Paulo Freire)

RESUMO

As plantas evoluíram para sobreviver em ambientes onde muitas vezes são impostas condições adversas, tais como estresses abióticos (temperatura, luz, seca, salinidade, frio), ou bióticos (vírus, bactérias, fungos e nematoides). Para sua sobrevivência, desenvolveram inúmeros mecanismos que permitem a detecção de mudanças ambientais, bem como a indução de respostas específicas às condições estressantes impostas, minimizando as perdas. Existem genes-chave nos mecanismos de adaptação, especialmente os relacionados à desintoxicação, à homeostase e à reprogramação dos padrões de expressão gênica, envolvendo mudanças em nível fisiológico. A identificação de genes-candidatos promissores para o melhoramento de espécies cultivadas com relação aos principais estresses ainda está aquém das necessidades. Assim, a identificação e caracterização de genes relacionados com a resposta vegetal a estresses foi realizada para as culturas da soja e do feijão-caupi, em seus respectivos transcriptomas, por métodos computacionais. Quando estão sob estresse, as plantas podem ativar respostas celulares, incluindo a produção de proteínas antioxidantes, com o intuito de minimizar os danos e evitar a ação tóxica de ROS (Espécies Reativas de Oxigênio) nas células vegetais. Neste contexto, foram identificados 1.273 transcritos em feijão-caupi e 451 transcritos em soja, distribuídos em 15 categorias de genes ROS que desempenham papéis importantes no estresse oxidativo. Estes genes compõem um grupo de enzimas antioxidantes que trabalham em conjunto para manter um nível de estado estacionário intracelular, promovendo o crescimento da planta, desenvolvimento, ciclo celular, a sinalização hormonal, reforçando respostas aos estressores ambientais abióticos e bióticos, semelhante ao observado em outras espécies de plantas. Além das ROS fatores de transcrição (FTs) representam um papel crucial, como os principais reguladores da tolerância vegetal ao estresse. Nesse contexto, foi realizada uma identificação das famílias de FTs, presentes no transcriptoma de duas variedades contrastantes de feijão-caupi (sensível e tolerante a seca). Foram identificados 4.822 transcritos, classificados em 64 famílias, com expressão diferencial nos diferentes tempos de exposição ao estresse e cultivares, exibindo indução da expressão em condições de estresse. As interações entre as famílias gênicas reguladoras e os genes regulados permitiram a criação de modelos computacionais para a compreensão da arquitetura e funcionamento da rede de regulação gênica vegetal frente ao estresse, permitindo a identificação eficiente de candidatos para o melhoramento vegetal e fins biotecnológicos.

Palavras-chave: Estresses Ambientais, Leguminosas, Bioinformática, Expressão Gênica, Genes de defesa.

ABSTRACT

Plants evolved to survive in environments that often impose adverse conditions, such as abiotic (temperature, light, drought, salinity, cold) and biotic stresses (viruses, bacteria, fungi and nematodes). For survival, they developed several mechanisms that enable the detection of environmental changes, as well as induction of specific responses to the imposed stress conditions, minimizing losses. There are key genes associated to adaptation mechanisms; especially those related to detoxification, homeostasis and gene expression patterns reprogramming, involving changes at physiological level. The identification of promising candidate genes for cultivated species improvement related to main stresses is still far from the necessities. Thus, the identification and characterization of genes related to plant response to stress was carried out for soybean and cowpea in their transcriptome by computational methods. When under stress, plants can activate cellular responses, including production of antioxidant proteins, in order to minimize damage and avoid toxic action of ROS (Reactive Oxygen Species) in plant cells. In this context, 1,273 transcripts were identified in cowpea and 451 transcripts in soybean, distributed into 15 ROS gene categories that play important roles in oxidative stress. These genes form a group of antioxidant enzymes that work in concert to maintain a steady intracellular level state, promoting plant growth, development, cell cycle, and hormonal signaling, reinforcing responses to biotic and abiotic environmental stressors, like observed in other plant species. Apart from ROS, transcription factors (TFs) play a crucial role as the primary regulators of plant stress tolerance. In this context, the identification of TF families was conducted for transcriptome data for two contrasting cowpea varieties (sensitive and tolerant to drought). 4,822 transcripts were identified, being classified into 64 families, differentially expressed in different times of exposure to stress and cultivars, exhibiting induction of expression under stress conditions. The interactions between regulatory gene families and regulated genes allowed the creation of computational models for understanding the architecture and operation of the plant against stress gene regulation network, allowing efficient identification of candidates for plant breeding and biotechnological purposes.

Keywords: Environmental Stresses, Legume, Bioinformatics, Gene Expression, Genes of Defense.

LISTA DE FIGURAS

LISTA DE TABELAS

values: up-regulated transcripts or tags; Negative values: down-regulatedtranscripts or tags. (Disponível no link: Material Suplementar)

CAPÍTULO III

Table1. Cowpea and soybean contigs matching results for each procured ROS family, showing the best match for each gene family in both species, e-value, score, size in nucleotide (nt) and amino-acid (aa), presence and kind of conserved domains and number other matches in cowpea (EST and RNAseq). Abbreviations: PO: Pingo de Ouro; SI: Santo Inácio; [c]: complete; [i]: incomplete.

LISTA DE ABREVIATURAS

•OH	Hydroxyl radical
1O2	Singlet oxygen
aa	Aminoacid; aminoácido
ABA	Ácido abscísico
ABFs	Abre-binding factors
ABRE	Aba-responsive element
AOX	Alternative oxidase
AP2	Apetala2
AP2/ERF	Apetala2/ethylene response factor
AP-MS	Affinity purification and mass spectrometry
APX	Ascorbate peroxidase
AS	Salicylic acid
ATX1	Arabidopsis homolog of trthorax
BAC	Cromossomo artificial de bacteria
BCP	Blue copper protein
bHLH	Basic helix-loop-helix
BiFC	Bimolecular fluorescence complementation
BLAST	Basic local alignment search tool
bZIP	Basic leucine zipper
CAT	Catalase
CBF/DREB	C-repeat binding factor/drought-responsive element binding factor
CDPKs	Calmodulin-dependent protein kinases
CGI	Cowpea genomics initiative
CGKB	Cowpea genespace/genomics knowledge base
DDBJ	Dna data bank of japan
DHAR	Dehydroascorbate reductase
DREB	Dehydration-responsive element binding factor
EMBL-EBI	European molecular biology laboratory - european bioinformatics institute

ERD1	Early responsive to dehydration stress 1
ERF	Ethylene responsive factor
EST	Expressed sequence tag
ET	Ethylene
FPKM	Fragments per kilobase million
FT	Fatores e transcrição
GLR	Glutaredoxin
GPX	Glutathione peroxidase
GR	Glutathione reductase
H2O2	Hydrogen peroxide
HMM	Hidden markov model
INSDC	International nucleotide sequence database collaboration
JÁ	Jasmonic acid
JGI	Joint genome institute
MAPKs	Mitogen-activated protein kinases
MDAR	Monodehydroascorbate reductase
NAC	No apical meristem
NADPHox	Nadph oxidase
NADPHoxLike	Nadph oxidase-like
NCBI	National center for biotechnology information
Nt	Nucleotide; nucleotídeo
O2•−	Superoxide anion
ORF	Open reading frame
PO	Pingo de ouro
PPIs	Protein-protein interaction
PrxR	Peroxiredoxin
PTM	Post-translational modifications
RCD1	Radical induced cell death 1
RNAseq	Sequenciamento de rna
ROS	Reactive oxygen species

RPKs	Receptor protein kinases
RT-qPCR	Real-time pcr analysis
SBML	System biology markup language
SI	Santo inácio
SIB1	Sigma factor-interacting protein 1
SOD	Superoxide dismutase
SRO1	Similar to rcd one
STIFDB	Stress-responsive transcription factor database
TF	Transcription factor
Trx	Thioredoxins
UNIPROT	Universal protein resource
Y2H	Yeast two-hybrid

SUMÁRIO

INTRODUÇÃO

As plantas são frequentemente expostas a diferentes adversidades ambientais, incluindo estresses abióticos (como temperaturas extremas, seca, salinidade, estresse oxidativo) e bióticos (por exemplo, ataque de insetos, de herbívoros e de patógenos). Tais fatores afetam negativamente o crescimento vegetal, impedindo que as plantas alcancem todo seu potencial genético (CRAMER et al., 2011). Como forma de adaptação, os vegetais desenvolveram mecanismos fisiológicos e moleculares de adaptação a tais estresses. Entretanto, o custo metabólico e energético sofrido pelas plantas frente às condições adversas muitas vezes resulta em impactos negativos para a produção e produtividade agrícola (WANG et al., 2001). Por exemplo, estresses ambientais – tais como temperaturas extremamente elevadas, excesso ou falta de chuvas, ventos fortes e geadas, dentre outros – têm ocasionado efeitos negativos nos sistemas de produção, nas mais diversas regiões do planeta, contribuindo para perdas na produtividade superiores a 50 % nas principais culturas agrícolas (WANG et al., 2003; MITTLER; BLUMWALD, 2010). Tais adversidades, aliadas ao atual cenário de crescimento da população mundial têm levado à necessidade de desenvolvimento de novas tecnologias que visem uma maior produção agrícola, de forma sustentável, a fim de atender à crescente demanda por alimentos.

No Brasil, a crescente expansão da soja [*Glycine max* (L.) Merr.] para novas fronteiras agrícolas com diferentes condições ambientais tem aumentado a demanda por novas estratégias que deem sustentabilidade à produção dessa cultura nos diferentes locais, cujo rendimento ainda é bastante afetado por diferentes fatores de estresse (POPP et al., 2003; ASSAD et al., 2007). Por sua vez, o feijão-caupi [*Vigna unguiculata* L. (Walp.)] constitui a segunda maior cultura agrícola do Nordeste brasileiro. Embora nessa região existam acessos com características agronômicas requeridas pelo mercado consumidor, há uma demanda pela aquisição de novas fontes de resistência / tolerância aos principais estresses, em especial contra seca e salinidade (no âmbito abiótico) e o ataque por vírus (no âmbito biótico), que se destacam como fatores limítrofes da produtividade do feijão-caupi (FREIRE-FILHO et al., 2011; EMBRAPA, 2014).

Em plantas superiores, a reprogramação dos padrões de expressão gênica figura entre os mecanismos-chave na adaptação a condições ambientais limitantes. As plantas se adaptaram para responder a essas tensões tanto em nível molecular, quanto fisiológico e bioquímico, sendo controladas principalmente por fatores de transcrição (FTs) induzidos ou reprimidos por uma variedade de estresses (BARTELS; SUNKAR, 2005). Considera-se que genes estresse-induzidos incluam moléculas indutoras de tolerância, assim como reguladores da expressão

ativos na transdução de sinais de diversas vias metabólicas (NAKASHIMA; YAMAGUCHI-SHINOZAKI, 2009). O conjunto de todas as interações entre FTs e seus genes-alvo em uma célula pode ser representado como uma rede, a qual fornece uma estrutura poderosa para elucidar os princípios gerais de organização da regulação da transcrição (CHEN; ZHU, 2004).

Para entender a regulação e os mecanismos de adaptação aos estresses ambientais, torna-se fundamental a identificação dos componentes que compõem estes sistemas regulatórios, incluindo tanto os genes codificantes de FTs, como aqueles codificantes de seus efetores. Com base no conhecimento destas interações é possível a criação de modelos computacionais que auxiliem na compreensão e elucidação da arquitetura de uma determinada rede em células quando sob condições de estresse (CHEN; ZHU, 2004; LAI et al., 2011).

Estudos de transcriptomas em grande escala têm revelado muitos genes "desconhecidos" ou sem função predita, potencialmente envolvidos na resposta ao estresse em plantas. Abordagens sistemáticas, tais como localização de genes no genoma e análise global da expressão, combinadas com métodos computacionais, compreendem um importante passo na investigação da biologia do estresse. Nesse contexto, a manipulação genética da expressão de genes codificantes de FT já se tornou uma ferramenta útil na engenharia metabólica, considerando-se que uma abordagem semelhante usando novos candidatos pode levar a uma maior tolerância a estresses ambientais em diferentes culturas (CHEN; ZHU, 2004).

Desta forma, a identificação de novos genes ou genes com novas características em culturas de importância agrícola, a exemplo da soja e do feijão-caupi (em especial para o Nordeste brasileiro), poderá trazer grandes avanços no melhoramento destas e de outras leguminosas cultivadas. Consequentemente, para lidar com estresses torna-se necessário um melhor entendimento dos mecanismos de respostas das plantas em nível celular e molecular frente a tais condições ambientais.

O presente estudo objetivou a identificação de genes relacionados com a resposta vegetal frente estresses ambientais (bióticos e abióticos) e suas vias de interação/regulação nas culturas da soja e feijão-caupi, a partir de dados disponíveis em seus respectivos bancos de dados, com ênfase em genes codificantes para fatores de transcrição, favorecendo assim um maior entendimento sobre os mecanismos de tolerância e vias relacionadas à sobrevivência vegetal sob condições adversas.

1. Estresses ambientais em plantas superiores

As plantas evoluíram para viver em ambientes onde, muitas vezes, são expostas a fatores de estresse, ou até mesmo a uma combinação destes, impedindo que alcancem seu potencial genético de crescimento e reprodução (ROCKSTROM; FALKENMARK, 2000; BRAY, 2004). Para lidar com estas adversidades, os vegetais desenvolveram mecanismos específicos que lhes permitem detectar mudanças ambientais, desencadeando respostas moleculares (Figura 1) visando minimizar danos e conservar recursos valiosos para seu crescimento e reprodução (RIZHSKY et al., 2004; MITTLER; BLUMWALD, 2010). Desta forma, torna-se fundamental a compreensão da natureza das múltiplas respostas vegetais aos estresses, para o delineamento de estratégias mais eficientes para o desenvolvimento de plantas resistentes a múltiplas adversidades, com altos padrões de rendimento.

Ao longo dos anos, estresses abióticos (calor, frio, seca, salinidade, nutrientes) têm promovido impactos negativos sobre a agricultura mundial, principalmente em termos de perdas na produtividade, observando-se perdas de mais de 50 %, nas principais culturas agriculturáveis (WANG et al., 2003). Além destes fatores, as plantas precisam defender-se do ataque de uma vasta gama de pragas e patógenos, incluindo fungos, bactérias, vírus, nematoides e insetos herbívoros (HAMMOND-KOSACK; JONES, 2000).

Sabe-se que cada tipo de estresse (biótico ou abiótico) promove o desencadeamento de respostas complexas em níveis celulares e moleculares, a fim de minimizar ou evitar danos maiores, garantindo a sobrevivência do indivíduo, ainda que se reduzam seus níveis de crescimento e produtividade (HERMS; MATTSON, 1992).

Os diferentes fatores de estresse que ocorrem de modo combinado podem ser considerados aditivos ou interativos (NIINEMETS, 2010). Quando uma tensão inicial, ou anterior, altera a resposta normal de uma planta frente a um segundo estresse, os fatores estressantes podem ser considerados interativos. Os efeitos interativos deletérios de dois ou mais estresses abióticos sobre uma determinada cultura vegetal vêm sendo amplamente estudados (MITTLER; BLUMWALD, 2010). Por exemplo, calor e estresse hídrico, em particular, podem causar danos bem mais intensos às culturas, quando comparados a qualquer outro tipo de estresse individual (MITTLER, 2006; BARNABAS et al., 2008).

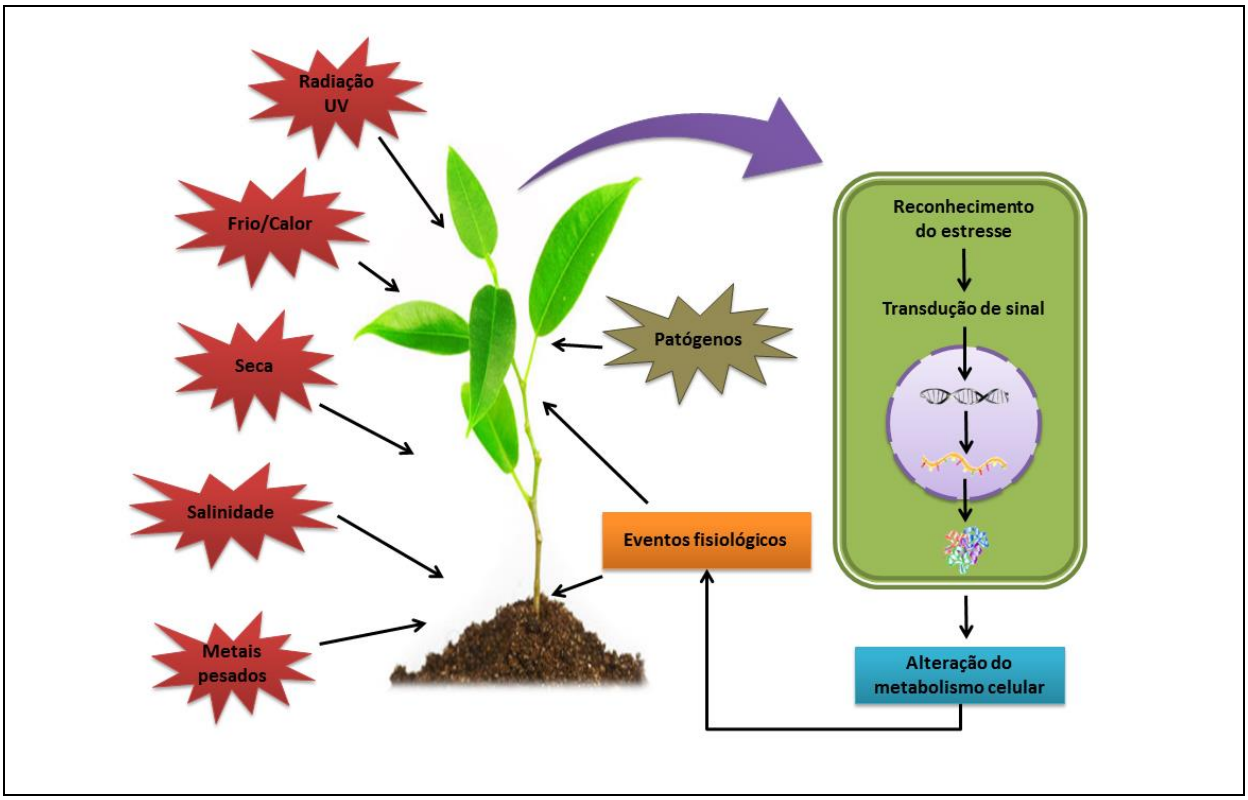


Figura 1. Esquema das interações planta - estresses ambientais, percepção do estresse e ativação de cascatas metabólicas para geração de respostas fisiológicas (Figura do autor).

Na interação entre estresses bióticos e abióticos, observou-se que o aumento da temperatura pode criar um efeito interativo negativo, diminuindo a resistência a patógenos bacterianos, virais, fúngicos e por nematoides. Por exemplo, temperaturas médias mais altas em experimentos controlados por um período de seis anos correlacionaram o aumento da temperatura a uma maior susceptibilidade da cultura de trigo (*Triticum aestivum*) ao fungo *Cochliobolus sativus* (SHARMA et al., 2007). Já em tabaco (*Nicotiana tabacum*) e em *Arabidopsis thaliana* a resposta de hipersensibilidade (*Hypersensitive Response*, HR) - e as respostas de defesa (*Resistance*, R) mediadas por genes na defesa contra *Pseudomonas syringae* e vírus tornam-se comprometidas em altas temperaturas, permitindo um aumento da infecção por tais organismos patogênicos (WANG et al., 2009). Da mesma forma, a exposição de *A. thaliana* ao estresse hídrico também resultou em maiores níveis de infecção por *P. syringae* (MOHR; CAHILL, 2003).

Curiosamente, os estresses abióticos podem também interagir de forma positiva com o estresse causado por patógenos. Segundo Wiese et al. (2004) um aumento do estresse osmótico (induzido pelo excesso de salinidade) resultou numa maior resistência da cevada (*Hordeum vulgare* L.) ao fungo *Blumeria graminis*, agente causador do oídio, uma das principais doenças desta cultura. Outro exemplo desta interação positiva foi observado em tomate (*Lycopersicon*

esculentum) no qual a incidência de estresse de seca culminou em maior resistência ao fungo *Botrytis cinerea* (ACHUO et al., 2006). Agentes patogênicos podem também interferir ativamente nas relações hídricas das plantas, apresentando de forma secundária um efeito positivo sobre a tolerância das plantas contra determinado estresse abiótico. Por exemplo, com a finalidade de aumentar sua patogenicidade, observa-se que alguns patógenos induzem ao fechamento estomático, reduzindo assim a perda de água a partir dos tecidos infectados (BEATTIE, 2011).

Modelos de previsão climática atuais indicam que nos próximos 50-100 anos as temperaturas médias aumentarão em 3-5°C, afetando drasticamente os sistemas agrícolas mundiais (IPCC, 2014), através do aumento dos períodos e da frequência de secas ou inundações, além da ocorrência de ondas de calor (MITTLER; BLUMWALD, 2010; IPCC, 2014). Estas mudanças, aliadas à pressão crescente sobre a produtividade global de alimentos devido ao aumento populacional, resultam numa demanda elevada (e urgente) por variedades com maior tolerância contra a seca, o calor e a salinidade (TAKEDA; MATSUOKA, 2008; NEWTON et al., 2011). Desta forma, compreender os mecanismos de resposta das plantas a estresses múltiplos (e possivelmente simultâneos) é fundamental para o desenvolvimento de culturas tolerantes aos diferentes tipos de estresses ambientais.

1.1 Controle da expressão gênica frente a estresses

Juntos, os estresses bióticos e abióticos são responsáveis por desencadear uma série de modificações morfológicas, fisiológicas, bioquímicas, moleculares e metabólicas, muitas das quais funcionando como respostas das plantas para obter tolerância a tais estresses (BENKO-ISEPPON et al., 2011). Frequentemente, em culturas de importância agrícola, tais mecanismos de resposta são acompanhados de um custo metabólico e energético alto, refletindo, muitas vezes, em baixa produtividade (WANG et al., 2001; NOGUEIRA et al., 2005).

Vários estresses resultam em efeitos gerais e específicos sobre o crescimento e desenvolvimento da planta. Observa-se, por exemplo, um crescimento reduzido da planta quando esta é submetida à seca, devido à diminuição da fotossíntese imposta pelo estresse osmótico e/ou pela indisponibilidade de nutrientes no solo, devido à falta de água. Sob efeito de seca, o crescimento da planta pode ser ainda mais comprometido em decorrência da salinidade do solo, gravando o efeito da seca em decorrência da toxidez provocada pelo excesso

de íons (ZHU, 2002). Estes e muitos outros exemplos apontam para respostas comuns das plantas diante de diferente fatores de estresse.

A prevenção do estresse osmótico, causado pela seca, por exemplo, parece depender da minimização da abertura estomática e da perda de água cuticular, bem como da maximização da absorção de água, através do crescimento da raiz e ajustamento osmótico. Durante o estresse salino, o ajuste osmótico também parece desempenhar um papel importante na manutenção da homeostase osmótica, enquanto a sobrevivência induzida por congelamento durante o estresse osmótico pode depender da prevenção ou retardamento da formação de núcleos de gelo. Como parte da resposta ao estresse, a regulação da expressão gênica envolve tanto mudanças universais e únicas em níveis de transcrição de determinados genes (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2000). Com base na presença destes mecanismos de tolerância parece prudente esperar que as plantas tenham múltiplas vias de percepção de estresse e transdução de sinais, as quais parecem apresentar interligações (*cross-talk*) em diferentes pontos das diferentes vias (CHINNUSAMY et al., 2003).

A reprogramação dos padrões de expressão gênica é um dos mecanismos-chave na adaptação (tanto em nível molecular, quanto fisiológico e bioquímico) das plantas superiores às condições ambientais limitantes (BARTELS; SUNKAR, 2005), sendo os produtos dos genes estresse-induzidos considerados os promotores de tolerância e/ou reguladores da expressão por meio de transdução de sinais nas diversas vias (NAKASHIMA; YAMAGUCHI-SHINOZAKI, 2009). Muitos estresses abióticos alteram vias gênicas e cascatas de sinalização visando à restauração da homeostase celular, compreendendo genes e vias que estão na base da resistência ao estresse abiótico, sendo parte integrante da regulação gênica global. Em geral genomas vegetais possuem cerca de 7 % de suas sequências codificantes para fatores de transcrição (FTs), demonstrando a complexidade da regulação da transcrição (UDVARDI et al., 2007). No genoma de *A. thaliana*, por exemplo, são conhecidos cerca de 1.500 FTs, muitos relacionados à expressão gênica em resposta a diferentes tipos de estresse (RIECHMANN et al., 2000).

2. Controle da expressão gênica em eucariotos

Os organismos eucariotos multicelulares geralmente apresentam um genoma maior do que os procariotos, sendo organizado em vários cromossomos e com maior complexidade em termos de sequências. Muitas espécies eucarióticas possuem os mesmos genes, com as mesmas sequências que outras espécies de plantas e animais. Além disso, as mesmas sequências de

DNA (apesar de não apresentarem as mesmas proteínas) são encontradas em todas as células nucleadas diploides, mesmo que estas células formem tecidos diferentes, com propriedades e funções distintas. Estas diferenças resultam das diferentes formas com que os genes são regulados (ativados ou reprimidos) em resposta a estímulos ambientais, ou seja, as funções específicas de diferentes tipos celulares são resultado da regulação gênica diferencial (WATSON, 2006; PHILLIPS; HOOPES, 2008; ALBERTS, 2010).

De modo geral, a regulação da expressão gênica ocorre em dois níveis. Primeiro, a transcrição é controlada limitando a quantidade de mRNA que é produzido a partir de um gene em particular. O segundo nível de controle ocorre por meio de eventos pós-transcricionais, ou seja, após a síntese proteica, através de alterações em sua estrutura, o que pode resultar na mudança e/ou inibição da atividade funcional (WATSON, 2006).

A regulação da transcrição em eucariotos é então de forma genérica o resultado dos efeitos combinados das propriedades estruturais (enovelamento do DNA) e as interações das proteínas chamadas fatores de transcrição. Para uma melhor compreensão, as principais diferenças em termos de estrutura da cromatina e expressão entre eucariotos e procariotos estão sumarizadas na Tabela 1 (PHILLIPS; HOOPES, 2008; ALBERTS, 2010).

Tabela 1. Principais diferenças entre eucariotos e procariotos em termos de estrutura e complexidade do genoma e sua influência na transcrição.

	Procariotos	**Eucariotos**
Estrutura do Genoma	Simples, geralmente circular, por vezes acompanhado por pequenos segmentos de DNA acessórios (plasmídeos).	Genoma arranjado em cromossomos, com nucleossomos limitando o acesso ao DNA.
Tamanho do Genoma	Relativamente pequeno.	Relativamente grande
Localização da transcrição e tradução	Acoplado devido à inexistência de barreiras nucleares (estrutura da célula procariótica).	Transcrição nuclear e tradução citoplasmática.
Organização dos genes	Operons, com genes de funções similares agrupados próximos.	Operons geralmente não encontrados, com cada gene possuindo sua própria região promotora.
Padrão da transcrição	Ativa	Inativa
Estrutura do DNA	DNA altamente enovelado com algumas proteínas associadas	Cromatina supercondensada, associada a histonas em nucleossomos.

2.1 Fatores de transcrição (FTs) e controle da expressão

FTs são proteínas reguladoras cuja função é ativar ou (mais raramente) inibir a transcrição através da ligação de sequências específicas de DNA. Os FTs apresentam grande afinidade (~106 vezes maior) ao seu sítio de ligação na molécula do DNA, do que com qualquer outra região da cadeia. Tais regiões são altamente conservadas e têm sido utilizadas para categorizar os FTs em "famílias" (ex. proteínas MADS-box). FTs também podem ser classificados pela estrutura tridimensional de suas proteínas, incluindo os modelos de hélice-volta-hélice (*Helix-Turn-Helix*) básico, hélice-alça-hélice (*Helix-Loop-Helix*) e as proteínas dedo de zinco (*Zinc Finger Protein*) (Figura 2). Estes motivos estruturais diferentes resultam em FTs específicos para as sequências de consenso nas quais se ligam (REMENYI et al., 2004; WATSON, 2006).

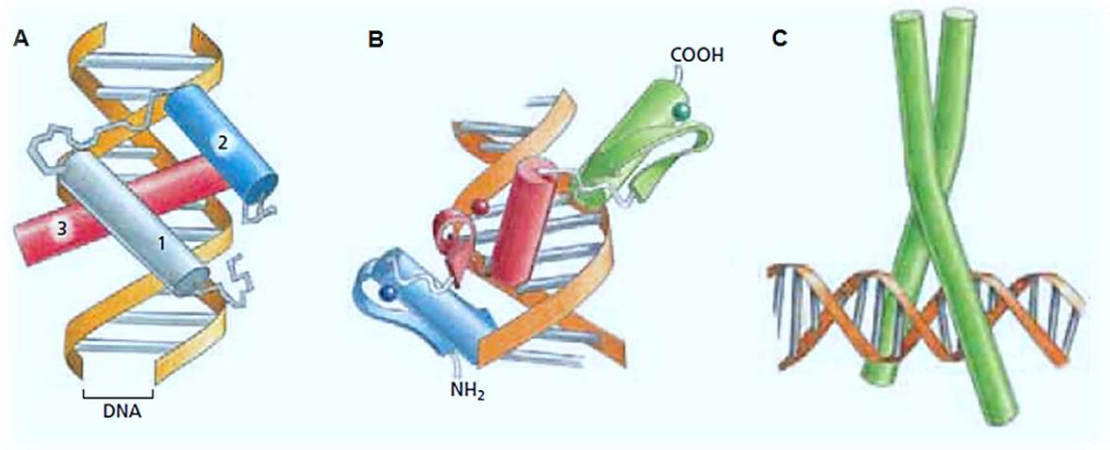


Figura 2. Principais motivos de ligação de Fatores de transcrição (FTs) ao DNA. (A) Homeodomínio – um motivo estrutural encontrado em muitas proteínas de ligação de DNA em eucariotos; (B) Motivo Dedo de Zinco – construído a partir de uma alfa-hélice e uma folha-beta (o último mostrado como uma seta torcida) mantidas juntas por uma molécula de zinco (indicada pelas esferas coloridas); (C) Ziper de Leucina – formado por duas alfa-hélices, cada uma constituída por uma molécula de proteína diferente (Adaptado de Alberts, 2010).

Os FTs são considerados as mais importantes moléculas em mecanismos de regulação gênica procariotas e eucariotas (PULVERER, 2005). Em eucariotos, a regulação da expressão por FTs é chamada de combinatória, uma vez que envolve interações coordenadas de múltiplas proteínas; em contraste com os procariotos, onde uma única proteína é, usualmente, necessária. Em procariotos, a RNA-Polimerase pode acessar praticamente qualquer promotor numa cadeia de DNA, sem a presença de ativadores ou repressores. Assim, o "estado fundamental" de expressão nos procariotos é dito “não restritivo”. Já em eucariotos, o estado fundamental de

expressão é “restritivo”, onde, apesar de existirem promotores, eles encontram-se inativos na ausência de FTs. Por exemplo, a RNA-polimerase II, não pode se ligar a promotores de DNA em eucariotos sem a ajuda dos FTs (STRUHL, 1999).

Em muitos organismos eucarióticos o promotor contém uma sequência TATA-box conservada. Além dessa, várias outras sequências consenso também existem e são reconhecidos pelas diferentes famílias de FTs. A transcrição se inicia quando as regiões promotoras se ligam a um FT, promovendo uma série de interações entre múltiplas proteínas (ativadores, reguladores e repressores) no mesmo local, ou em outro promotor. Em última análise, um complexo de transcrição é formado no promotor que facilita a ligação e transcrição pela RNA polimerase (Figura 3; PHILLIPS; HOOPES, 2008; ALBERTS, 2010).

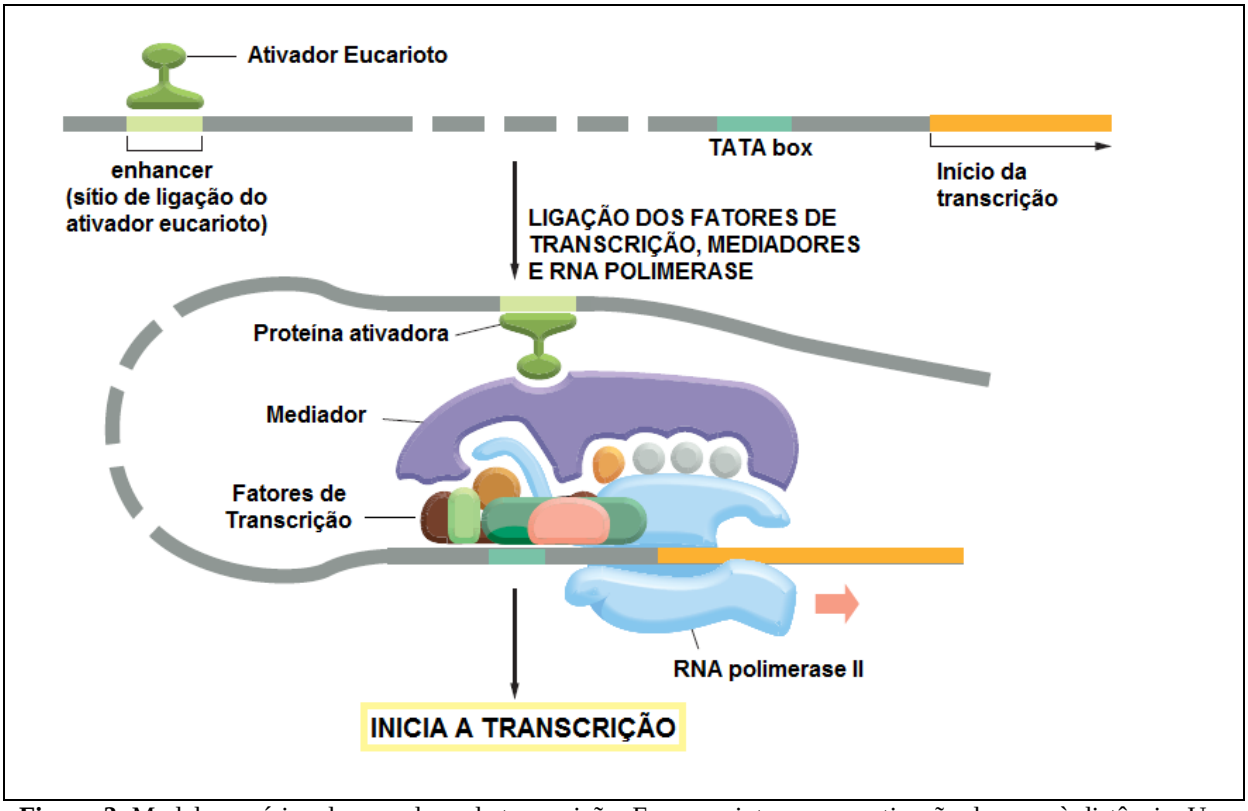


Figura 3. Modelo genérico do complexo de transcrição. Em eucariotos ocorre ativação do gene à distância. Um ativador de proteína ligado ao DNA atrai a RNA polimerase e os fatores de transcrição específicos para a região promotora. O dobramento do DNA permite o contato entre o ativador (*enhancer*) e o complexo de transcrição ligado ao promotor. No caso ilustrado na figura acima, um grande complexo de proteínas denominado de “Mediador” atua como um intermediário. A região de DNA segmentada representa uma variação no comprimento entre o intensificador e o começo da transcrição, podendo chegar a uma extensão de dezenas de milhares de pares de nucleotídeos (Adaptado de Alberts, 2010).

Tal como em procariontes, moléculas repressoras eucarióticas podem, por vezes, se ligar a elementos silenciadores localizados próximos a um gene e inibir a ligação, montagem ou a

atividade do complexo de transcrição, modificando, assim, a expressão gênica. Considerando o estado fundamental de transcrição como restritivo, a regulação positiva é a forma predominante de controle na maioria dos sistemas caracterizados. Muitos FTs encontram-se ligados ao DNA até serem removidos por uma molécula de sinalização, enquanto outros só podem se ligar ao DNA quando induzidos por uma molécula sinalizadora. A ligação de um tipo de FT também pode influenciar a ligação de outros. Assim, a expressão de genes em eucariotos é altamente variável, dependendo do tipo de ativadores envolvidos, e da presença dos sinais que controlam a ligação (BABU et al., 2004; PHILLIPS; HOOPES, 2008).

2.2 Mecanismos de controle da expressão

Mesmo que uma célula apresente FTs, a transcrição de determinados genes pode não ocorrer, uma vez que estes FTs podem não estar ligados aos seus sítios específicos na molécula do DNA. A associação da molécula de DNA com as proteínas é o primeiro passo no seu silenciamento. O complexo formado pelas proteínas de DNA e histonas é conhecido como cromatina, sendo tal complexo firmemente ligado ao DNA. Portanto, o estado da cromatina é fundamental no acesso dos FTs e RNA-polimerase aos promotores, contribuindo para o estado restritivo da expressão gênica. Assim, para que a transcrição do gene ocorra, a estrutura da cromatina deve ser "desenrolada" ou ao menos 'afrouxada', para que as regiões promotoras sejam reconhecidas (PHILLIPS; HOOPES, 2008; CHENG; BLUMENTHAL, 2010).

A estrutura da cromatina também contribui para os diferentes níveis de complexidade na regulação da expressão gênica, permitindo a regulação simultânea de genes funcional ou estruturalmente relacionados, os quais se apresentam em grupos espaçados ou em domínios de DNA específicos (SPROUL et al., 2005). A interação da cromatina com ativadores e repressores pode resultar em domínios de cromatina que estão abertos, fechados ou prontos para a ativação. Estas variações permitem a existência de fenômenos específicos nos eucariotos como a transcrição de genes específicos e distintos em diferentes fases de desenvolvimento, bem como ao longo da memória epigenética dos indivíduos. Eles também permitem a manutenção de estados celulares diferenciados, o que é crucial para a sobrevivência dos organismos multicelulares (STRUHL, 1999).

No entanto, FTs não são os únicos responsáveis pela regulação de genes, uma vez que os eucariotos contam ainda com mecanismos epigenéticos, incluindo a metilação do DNA, o *imprinting*, a presença de RNAs não-codificantes, modificações pós-traducionais, dentre outros

mecanismos que podem influenciar a transcrição de genes específicos, determinando se um gene será "ligado" ou "desligado" em um determinado organismo (PHILLIPS; HOOPES, 2008).

O DNA de todas as células eucarióticas apresenta-se estreitamente associado às histonas, formando o nucleossomo, uma estrutura globular composta por 146 pares de bases (pb) de DNA envolvidos em torno de duas moléculas de cada uma das histonas H2A, H2B, H3, e H4, com uma molécula de histona H1 ligado o DNA ao complexo. Este complexo DNA/Histonas tem consequências importantes para a transcrição, já que a estrutura da cromatina é um aspecto crítico na expressão de genes em células eucarióticas, uma vez que os ativadores e repressores da expressão não só interagem com os FTs gerais e outros componentes transcricionais, mas também atuam na indução de alterações na estrutura da cromatina (ALBERTS, 2010; CHENG; BLUMENTHAL, 2010). Tal relação (entre a estrutura da cromatina e a transcrição) é evidente, por exemplo, quando visualizamos os cromossomos politênicos de espécies de *Drosophila,* onde as regiões cromossômicas descondensadas do genoma representam as regiões ativamente engajadas na síntese de RNA. Do mesmo modo, os genes ativamente transcritos em células de vertebrados estão presentes numa fração descondensada da cromatina, a qual se torna mais acessível aos FTs, do que o resto do genoma (COOPER, 2000).

Entretanto, mesmo descondensados, os genes ativamente transcritos permanecem ligados às histonas e complexados em nucleossomos impedindo, assim, a interação dos FTs e da RNA-polimerase com a fita de DNA. Somente após o processo de acetilação de histonas e ligação de proteínas não-histônicas é que ocorre a remodelação dos nucleossomos resultando, assim, na ligação dos FTs. A acetilação de histonas está correlacionada à cromatina transcricionalmente ativa numa grande variedade de tipos celulares. O processo envolve a redução da carga líquida positiva das histonas, enfraquecendo a sua ligação ao DNA, bem como alterando suas interações com outras proteínas. Inversamente, as histonas desacetilases (que removem os grupos acetil a partir de caudas de histonas) estão geralmente associadas com repressores da transcrição em levedura e em células de mamíferos. Assim, as histonas são reguladas tanto para ativação quanto para a repressão de genes, desempenhando um papel essencial na expressão de genes eucarióticos (CHENG; BLUMENTHAL, 2010).

Outro mecanismo pelo qual o controle da transcrição relaciona-se com a estrutura da cromatina refere-se à metilação do DNA, onde resíduos de citosinas (C) - que antecedem guaninas (G) na cadeia de DNA - podem ser modificados pela adição de grupos metil na posição

carbono-5. Esta metilação promove a redução da atividade transcricional de genes que contêm altas frequências de dinucleotídeos CpG na vizinhança dos seus promotores. A metilação inibe a transcrição destes genes por meio da ação de uma proteína, MeCP2 (*Methyl Cpg Binding Protein 2*), que se liga especificamente ao DNA metilado e reprime a transcrição. Curiosamente, as funções da proteína MeCP2, associada a histona desacetilases, interligam a metilação do DNA com as alterações na estrutura dos nucleossomos (Messerschmidt et al., 2014). Um papel regulador importante da metilação do DNA foi estabelecido no fenômeno conhecido por impressão genômica (*imprinting*), que controla a expressão de alguns genes envolvidos no desenvolvimento de embriões de mamífero. Na maioria dos casos, ambos os alelos (paternos e maternos) de um gene são expressos em células diploides. No entanto, existem alguns genes 'impressos', cuja expressão é parental-depende (se herdados da mãe ou do pai). Em alguns casos, apenas o alelo paterno de um gene é expresso, sendo o alelo materno transcricionalmente inativo. Para outros genes marcados, o alelo materno é expresso e o alelo paterno é inativo (CREWS et al., 2007; TANG; HO, 2007).

Nos últimos anos tem sido relatado que alguns RNAs "não codificantes" podem participar da regulação da transcrição, atuando como um cofator no recrutamento alvo-dirigido de modificações no DNA e cromatina em loci específicos. Micro RNAs (miRNA), como a palavra diz, são RNAs pequenos (de, aproximadamente 22 nucleotídeos) não codificantes, que atuam na regulação da expressão gênica através da interação com o RNA mensageiro (RNAm) (KANDURI, 2011).

Os miRNA são transcritos por uma RNA Polimerase II, gerando um miRNA primário (pri-miRNA) longo. No núcleo, a enzima Drosha forma um complexo com a proteína ligadora de RNA dupla fita DGCR8, e o pri-miRNA é processado em um miRNA precursor de ~70 nt (pré- miRNA), no formato de um grampo (*hairpin*). Drosha é uma RNase III, essencial para o processamento de miRNAs capazes de serem exportados do núcleo. O pré-miRNA é exportado para o citosol pelo complexo Exportina 5 dependente de GTP (Guanosina trifosfato), onde é processado pela enzima Dicer, gerando um RNA dupla fita (dsRNA; *double - stranded RNA*) de, aproximadamente 22 nt. Este RNA é incorporado ao complexo silenciador induzido por RNA (RISC). Neste complexo, apenas uma fita do dsRNA é considerada para o miRNA maduro, o qual é ligado ao RNAm alvo, regulando, assim, a expressão gênica tanto em nível transcricional, traducional ou ambos (Ketting; Plasterk, 2004; Meola et al., 2009). A fita restante do dsRNA (não incorporada ao complexo RISC) é denominada miRNA* e, portanto, degradada pela célula (MEOLA et al., 2009).

Outra classe de RNAs são os pequenos sRNA (50 à 250 nt) os quais são altamente estruturados e também não-codificantes. Estes controlam a expressão de diversos genes, incluindo genes de virulência de agentes patogênicos, sendo importantes candidatos em inferências biotecnológicas na defesa contra microrganismos resistentes a drogas convencionais (HOWDEN et al., 2013).

2.3. Interação de genes no controle da expressão

Muitos FTs atuam no controle da expressão de uma variedade de regiões promotoras diferentes, enquanto outros são apenas ativadores para regiões promotoras específicas (LYER et al., 2001). Diversas moléculas de sinalização podem influenciar na forma como os FTs se ligam às regiões promotoras, seja pela ligação destas moléculas com os fatores, seja pela modificação dos domínios funcionais. É possível ainda, que um FT possa responder a um sinal físico, tal como a luz vermelha ou vermelha distante, mas o sinal deve ser percebido e traduzido para a forma de um ativador modificado quimicamente, que interage diretamente com o FT (WATSON, 2006; PHILLIPS; HOOPES, 2008).

Portanto, a complexidade da regulação da expressão nos organismos eucariotos mostra ser o resultado da combinação de diversos fatores, que englobam a estrutura e remodelagem da cromatina, a percepção de sinais que ativam ou inibem a ação dos fatores. Assim, o controle da transcrição depende das interações entre todos os FTs e se eles atraem a RNA polimerase ou bloqueiam-na antes de iniciar a transcrição. Além disso, vários FTs podem sobrepor-se, criando uma massa do tamanho de um ribossomo. Uma vez unidos, alterações nos domínios funcionais de um FT e/ou interações covalentes com outros fatores pode transformar os níveis de transcrição, ligando ou desligando um gene específico, dependendo se eles permitem ou inibem o recrutamento da RNA polimerase (PHILLIPS; HOOPES, 2008; ALBERTS, 2010).

Um ‘potenciador’ (*enhancer*) típico pode ser de até 500 pb, podendo conter vários locais de ligação para pelo menos dois ou três de FTs (LEVINE; TIJAN, 2003). Dois FTs ligados em locais próximos entre si na cadeia de DNA podem se combinar para formar um dímero e dobrar o DNA, o que se acredita ser o mecanismo chave no processo de ativação gênica. A estrutura da cromatina permite a associação de *enhancers* mesmo quando eles são ligados a sequências de DNA centenas de pares de bases de distância entre si. Alguns FTs atuam como elementos de ligação entre os *enhancers* distantes e promotores, formando conexões com outras proteínas (HINNEBUSCH, 2006).

3. Principais vias relacionadas às respostas vegetais ao estresse

Frio, seca e alta salinidade são condições de estresse comuns que afetam negativamente tanto o crescimento da planta, como sua produção. Compreender os mecanismos pelos quais as plantas transmitem os sinais para a maquinaria celular para ativar respostas adaptativas é de fundamental importância para desenvolver culturas mais tolerantes, melhorando a eficiência da produção. Neste contexto, as vias de transdução de sinal são o "elo" entre os mecanismos de detecção/percepção do estresse e a resposta genética. Para lidar com as mudanças ambientais, as plantas tendem a ativar cascatas de transdução de sinais que controlam e coordenam as respostas fisiológicas e bioquímicas necessárias para a sua adaptação (HASEGAWA et al., 2000; ZHU, 2002).

As plantas devem produzir uma resposta adequada às várias condições específicas de estresse biótico e abiótico, visto que em muitos casos, os diferentes tipos de pressões ambientais provocam reações diferenciadas. Por exemplo, o estresse térmico geralmente faz com que as plantas abram os estômatos para resfriar as folhas, mas sob condições de seca este mecanismo seria desvantajoso, pois mais água seria perdida (RIZHSKY et al., 2004). Da mesma forma, o aumento da transpiração causada pelo estresse de calor pode aumentar a absorção de sal ou metais pesados, aumentando o dano provocado pela toxicidade desses fatores (MITTLER; BLUMWALD, 2010). Assim, as plantas devem equilibrar de forma eficiente a alocação de recursos entre crescimento e defesa frente aos estresses, já que a resposta ao estresse pode ser de alto custo em detrimento do crescimento e produção vegetal (BECHTOLD et al., 2010). Como exemplo, pode-se citar a expressão constitutiva de uma proteína rica em prolina em *Arabidopsis* que confere tolerância ao estresse osmótico, salino e ao calor, mas produz plantas raquíticas (PRIYANKA et al., 2010). O custo da defesa frente estresses pode ser reduzido se o gene em questão atuar em diferentes tipos de estresses ambientais, o que pode explicar a sobreposição de várias vias de resposta (ASSELBERGH et al., 2008). Isso pode ser observado para certas vias moleculares de sinalização, onde os FTs e algumas proteínas efetoras são ativados tanto por estresses bióticos como abióticos (DUBOS et al., 2010; ATKINSON et al., 2011).

Os diferentes estresses abióticos quase sempre induzem a ativação de genes de vias de sinalização celular (SHINOZAKI; YAMAGUCHI-SHINOZAKI, 2000; ZHU, 2002) e respostas celulares semelhantes, como a produção de proteínas de estresse e a superexpressão de substâncias antioxidantes (CUSHMAN; BOHNERT, 2000).

De uma forma geral, os estresses abióticos iniciam com a percepção do estresse, desencadeando uma cascata de eventos moleculares, levando à transmissão e o processamento de sinais para geração de respostas mediadas por fitormônios, mensageiros secundários, FTs, genes e ou proteínas sinalizadoras (Figura 4) (PANDEY, 2008).

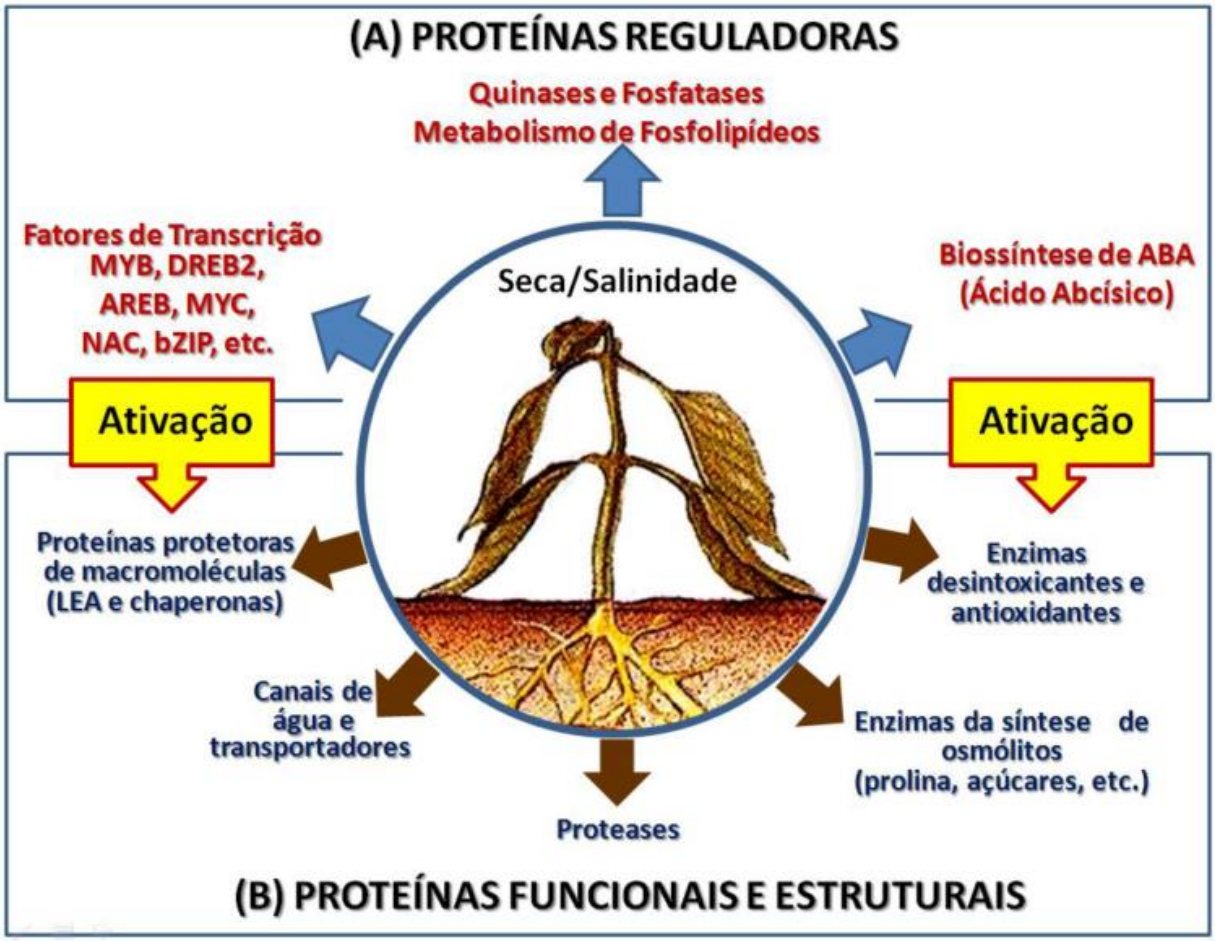


Figura 4. Esquema geral da resposta vegetal ao estresse abiótico, em especial, seca e salinidade. As células vegetais recebem os vários sinais através de uma variedade de sensores (muitos ainda desconhecidos), sendo os sinais traduzidos para as diferentes vias por meio de hormônios vegetais, transdutores de sinal e reguladores de transcrição. Genes estresse-induzidos são regulados por diversos sinais, e muitos regulados por fatores de transcrição (TFs) induzidos por estresse (uma cascata transcricional). Muitos destes genes codificam para proteínas diretamente envolvidas na tolerância ao estresse (Fonte: Benko-Iseppon et al., 2011).

A ocorrência de estresses de forma simultânea apresenta um maior grau de complexidade, visto que as respostas a estes são, em grande parte, controladas por diferentes vias de sinalização hormonal, que podem interagir e inibir-se mutuamente (ANDERSON et al., 2004; ASSELBERGH et al., 2008). Muitas vezes, a exposição a um patógeno ou praga aumenta os efeitos de um estresse abiótico, tais como a deficiência de água (AUDEBERT et al., 2000), enquanto que o estresse abiótico de longa duração pode enfraquecer as defesas das plantas e aumentar a suscetibilidade a diversos tipos de patógeno (MITTLER; BLUMWALD, 2010).

As respostas moleculares das plantas para estresses envolvem interações e *crosstalk* com muitas vias moleculares (TAKAHASHI et al., 2004). Em muitos casos, a biologia de sistemas tem sido eficientemente usada para elucidar algumas das vias regulatórias importantes nas respostas das plantas ao estresse. Um dos primeiros sinais envolve a produção de espécies reativas de oxigênio (ROS - *Reactive Oxygen Species*), que modificam a atividade enzimática e a regulação gênica (MITTLER et al., 2011; MOLASSIOTIS; FOTOPOULOS, 2011).

A incidência de ROS nas células vegetais incluem o ânion superóxido (O^2), o radical hidroxila (•OH) e o peróxido de hidrogênio (H_2O_2). Eles são produzidos como bioprodutos normais no metabolismo aeróbico, em diversos compartimentos celulares, principalmente cloroplastos, mitocôndria, peroxissomos, retículo endoplasmático, parede celular e apoplasto (GILL; TUTEJA, 2010; SHARMA et al., 2012). Em todos os organismos aeróbicos os níveis de ROS são regulados por vias que metabolizam as espécies reativas por meio de genes conhecidos como recicladores (BHATTACHARJEE, 2012).

O balanço entre produção e eliminação das ROS pode ser perturbado por uma variedade de fatores ambientais, muitas vezes favorecendo o aumento dos níveis intracelulares dos agentes oxidantes, que podem ocasionar dano oxidativo. Par evitar os danos causados por ROS, as plantas superiores fazem usos dos diferentes antioxidantes que compõem o sistema de defesa oxidativa, como o superóxido dismutase (SOD), catalase (CAT), ascorbato peroxidase (APX), monodehidroascorbato redutase (MDHAR), glutationa redutase (GR), dentre outros (MILLER et al., 2010; SHAMA et al., 2010; KRISHNAMURTHY; RATHINASABAPATHI, 2013). Estas enzimas atuam em diversos compartimentos celulares atuando quando as plantas são submetidas ao estresse oxidativo. A interação destes agentes antioxidantes nos diferentes compartimentos faz com que eles estejam associados com diferentes vias de defesa, influenciando o crescimento vegetal e desenvolvimento, tanto nos processos de divisão, quanto de morte celular (PINTO; GARA, 2004; GILL et al., 2013)

A sinalização via ROS em resposta a estresses (em especial ao abiótico) e suas interações com os hormônios tem sido foco de diversos estudos. Dentre os hormônios reguladores de resposta ao estresse destacam-se o ácido abscísico (ABA) e etileno (ET). O ABA é um regulador central de muitas respostas das plantas a estresses ambientais, envolvendo particularmente aqueles que têm efeito osmótico (CRAMER, 2010; HUBBARD et al., 2010). Sua sinalização pode ser rápida, sem envolver atividade transcricional, como no caso do controle da abertura estomática através da regulação do transporte de íons e água. Há também

respostas mais lentas, que envolvem respostas transcricionais para controle do crescimento, germinação e mecanismos de proteção (KIM et al., 2010).

O atual modelo de sinalização ABA inclui três componentes principais, receptores (PIR/Pyl/RCAR), proteínas fosfatases (PP2C) e proteínas quinases (SnRK2/OST1) (MA et al., 2009; PARK et al., 2009). Estudos sobre a regulação da transcrição para estresse de seca e salinidade revelaram que ambas as vias são dependentes e ao mesmo tempo independentes de ABA (Figura 5). A desidratação induz um aumento dos níveis endógenos de ABA que ativam os genes-alvo codificantes para fatores de sinalização, FTs, enzimas metabólicas, dentre outros (YAMAGUCHI-SHINOZAKI; SHINOZAKI, 2006). Na fase vegetativa, a expressão de genes responsivos ao ABA é regulada principalmente por FTs da classe bZIP, dependente de ABA (FUJITA et al., 2005; YOSHIDA et al., 2010). A ativação de cascatas de sinalização ABA resulta num aumento da tolerância das plantas ao estresse por desidratação. Em contraste outros elementos, tais como DREB (*Dehydration Response Element B*), estão relacionados com a resposta para desidratação independentemente do ABA, assim como vias relacionadas ao estresse térmico (MARUYAMA et al., 2004; SAKUMA et al., 2006).

O etileno (ET) também está envolvido em muitas respostas ao estresse (YOO et al., 2009), incluindo a seca, inundações, calor, frio, injúrias e radiação UV (GODA et al., 2008). Existem interações conhecidas entre ET e ABA durante a seca, amadurecimento de frutos, e de dormência dos brotos. Todas essas interações fazem a resposta da planta ao estresse muito complexa (SUN et al., 2010; WILKINSON; DAVIES, 2010).

A investigação das respostas moleculares contra estresses múltiplos, frequentemente enfoca a sobreposição de padrões de transcrição. Para este efeito, vários estudos têm sido conduzidos, nos quais plantas de diferentes grupos são expostas a um estresse ou outro, em paralelo, sendo seus padrões de expressão gênica comparados (SEKI et al., 2002). Assim, os conjuntos sobrepostos de genes que são regulados por ambas as tensões são então identificados e propostos como componentes de uma resposta comum, ou pontos de *cross-talk* entre as vias de sinalização. Muitos destes genes-chave são tidos como bons alvos para melhorar a tolerância ao estresse em plantas cultivadas (DENBY; GEHRING, 2005; SWINDELL, 2006).

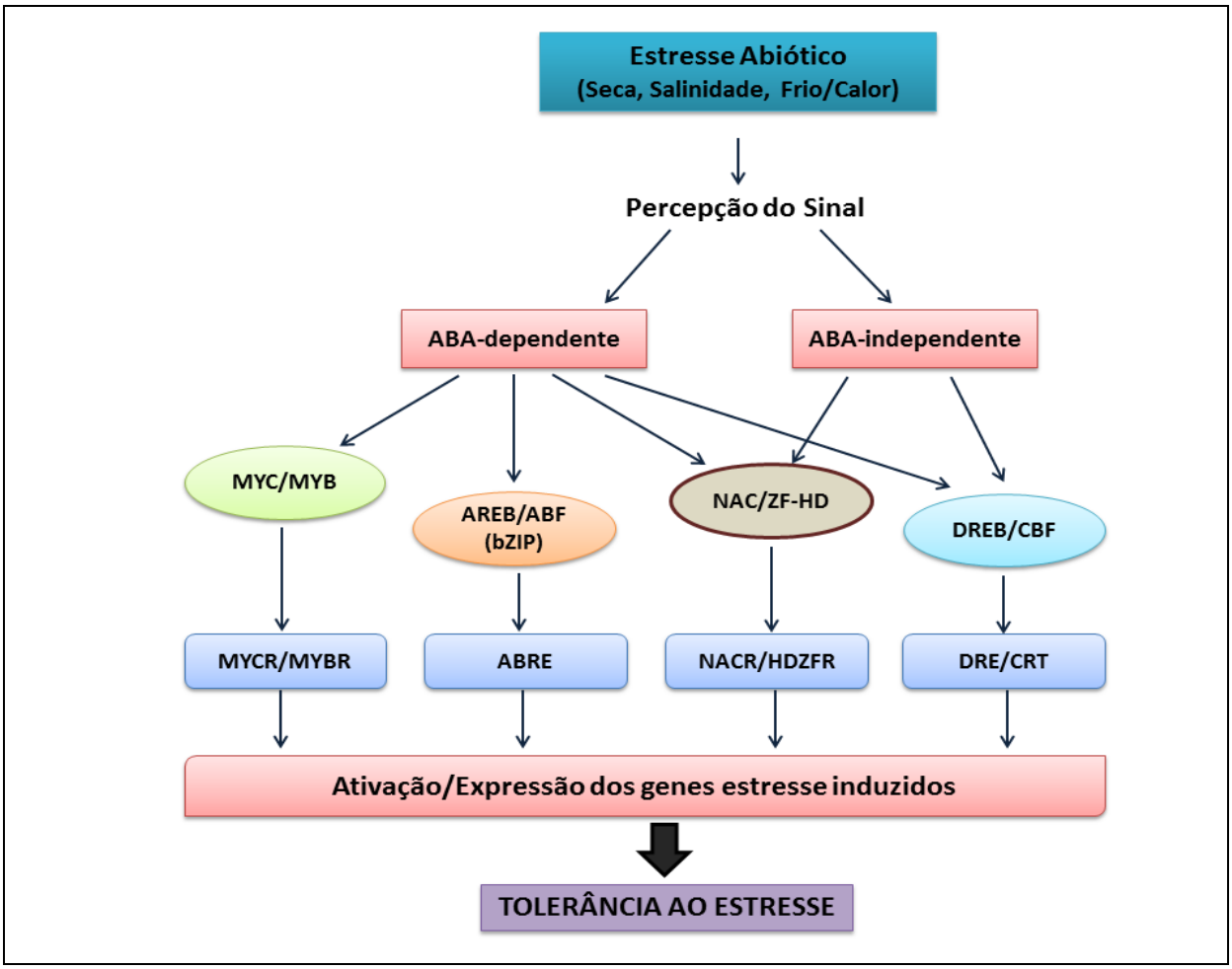


Figura 5. Modelo esquemático da sinalização dependente e independente de ABA. Cascata de sinalização associada aos principais fatores de transcrição responsivos ao estresse abiótico (Figura do autor).

4. Bioinformática e sua contribuição para biotecnologia vegetal

A identificação dos genes, suas funções e suas relações para regulação nos diferentes organismos são fundamentais para a compreensão dos processos biológicos, incluindo diferenciação celular, morfogênese, determinação fenotípica, resistência a patógenos e adaptabilidade. Neste sentido, a grande quantidade de informações geradas para espécies de plantas como *A. thaliana*, arroz (*Oryza sativa*), *Populus tricocharpa* (álamo) e soja deram início à era das "ômicas" e da biologia de sistemas em plantas. Esta nova era na genética molecular compreende o estudo de transcritos (transcriptômica), de proteínas (proteômica) (CHEN; HARMON, 2006), dos metabólitos celulares (metabolômica) (BHALLA et al., 2005), cujo principal objetivo é traçar perfis funcionais da célula e/ou tecido, fornecendo uma visão holística da interação entre os genes e seus efeitos sobre o fenótipo.

Um dos aspectos fundamentais para avanços nessas investigações foi o desenvolvimento e aplicação de técnicas e ferramentas computacionais, cada vez mais eficientes para armazenamento e interpretação dos resultados obtidos (PROSDOCINI et al., 2002). Neste contexto, surge a bioinformática como um campo multidisciplinar, formado pela convergência de tecnologias biológicas, biomédicas e da informação, visando responder a questões biológicas através de análises com ferramentas computacionais (BAXEVANIS, 2001). Por meio da bioinformática é possível, portanto, manipular uma grande quantidade e diversidade de dados biológicos; sendo os programas e algoritmos desenvolvidos capazes de armazenar, processar, analisar, decifrar estruturas, traçar relações entre moléculas e vias e interpretar grande quantidade de informações (BORÉM; SANTOS, 2001). À medida que as informações foram sendo geradas, novos avanços nas tecnologias relacionadas às "ômicas" foram sendo obtidos, os quais têm permitido, por exemplo a identificação de mecanismos de sinalização transcricional, traducional e pós-traducional, que regulam as respostas vegetais ao estresse (AHUDA et al. 2010). Aliados à bioinformática, estudos com plantas-modelo, como *A. thaliana* e *Medicago truncatula* têm sido fundamentais na elucidação das bases genéticas e bioquímicas sob tais situações estressantes, permitindo a integração das informações relacionadas às várias "ômicas" (HIRAYAMA; SHINOZAKI, 2010).

Atualmente centenas de bancos de dados vêm disponibilizando informações específicas para determinadas espécies ou grupos de organismos, incluindo sequências nucleotídicas ou proteicas, de grupos de moléculas ou de elementos gênicos ou genômicos (FERNANDEZ-SUAREZ; GALPERIN, 2012).

Dentre os bancos de dados públicos que armazenam grande quantidade de dados de sequências de nucleotídeos e de proteínas, destacam-se o GenBank, do NCBI (National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/); o DDBJ (DNA Data Bank of Japan; http://www.ddbj.nig.ac.jp) e o EMBL-EBI (European Molecular Biology Laboratory - European Bioinformatics Institute; http:// www.ebi.ac.uk), que juntos, formam o INSDC (International Nucleotide Sequence Database collaboration; http://www.insdc.org/). Estes bancos trabalham em cooperação para estabelecer formatos de dados, metadados e protocolos que facilitem a submissão segura dos dados, além de proporcionar o intercâmbio contínuo de dados em todo o mundo. Estas propriedades fazem do INSDC uma das alianças globais de arquivos de dados biológicos mais duradores (NAKAMURA et al., 2012).

Por sua vez, o UNIPROT (Universal Protein Resource; http://www.uniprot.org) é um banco de dados que integra, interpreta e padroniza dados de inúmeras fontes para formar o

maior e mais completo catálogo de sequências proteicas e de anotação funcional. Assim, o UNIPROT disponibiliza uma base de conhecimento de sequências proteicas com livre acesso, estável e abrangente, sendo essas sequências totalmente classificadas e ricamente anotadas e curadas (THE UNIPROT CONSORTIUM, 2012).

Outro banco de destaque é o PLAZA 2.5 (http://bioinformatics.psb.ugent.be /plaza/), um banco de dados *'online'* que disponibiliza diversas ferramentas de análise e visualização interativa, permitindo simultaneamente a exploração de informações genômicas de 25 espécies vegetais, cobrindo uma ampla variedade taxonômica. Os dados disponíveis consistem em anotações estruturais e funcionais de genes, famílias de genes homólogos, múltiplos alinhamentos de sequência, árvores filogenéticas e regiões colineares dentro e entre espécies (VAN BEL et al., 2011).

Paralelamente à criação dos bancos de dados, várias ferramentas que permitem o acesso aos dados gerados foram desenvolvidas. Tais ferramentas permitiram, por exemplo, uma análise de predição *in silico* de sequências desconhecidas e novos genes, por meio de análises de alinhamentos comparativos, bem como relações de regulação entre diferentes genes de uma via ou processo (KENT et al., 2002; SAEYS et al., 2006). A Tabela 2 compila algumas das ferramentas mais comumente usadas na área de bioinformática.

Diversas áreas vêm emergindo a partir da bioinformática nos últimos anos, com destaque para a mineração de texto e biologia de sistemas (KEMPER et al., 2010). Além destas, algumas áreas emergentes adicionais, tais como análise de imagem, computação em grade, evolução dirigida, design racional de proteínas (LOOGER et al., 2003; BIASINI et al., 2014), bioinformática relacionados ao microRNA (BROWN; SANSEAU, 2005; LIU et al., 2014) e modelagem em epigenômica (FAZZARI; GREALLY, 2004; LIM et al., 2010) merecem menção.

Sem a mineração de texto (*Text Mining*), a informação relacionada a interações biomoleculares e biológicas disponível na literatura, apesar de acessível, seria pouco explorada na prática. Uma área importante na mineração de texto é a obtenção de relações entre entidades, tais como genes e proteínas. A atual geração de sistemas combina métodos estatísticos com a aprendizagem de máquina para capturar o conhecimento disponível na literatura de como genes e proteínas se associam, de forma a criar sistemas utilizáveis com alta precisão (RHEE et al., 2006). Assim, a abordagem combinatória utilizando múltiplas plataformas e a integração dos seus resultados é uma estratégia eficaz para o esclarecimento de sistemas moleculares, bem como para o desenvolvimento de estratégias que visem melhorar a produtividade vegetal. Além

disso, a genômica comparativa entre as plantas modelo nos permite compreender as propriedades biológicas de cada espécie, acelerando a descoberta de genes e suas funções (EDWARDS; BATLEY, 2004; MOCHIDA; SHINOZAKI, 2010).

Tabela 2. Ferramentas de bioinformática para análise de sequências de nucleotídeos e proteínas, alinhamentos, modelagem e estatística. Um levantamento amostral de ferramentas (dados compilados pelo autor).

Ferramentas	Descrição
Análise de Sequências	
EMBOSS	Um pacote de software gratuito e de código aberto para a biologia molecular de alta qualidade.
G-Language	Fornece pacotes para análise de genomas.
CodonCode Aligner	Programa para montagem e edição de sequências, formação de contig e detecção de mutações.
Estrutura de Proteínas	
Amber7	Conjunto de programas que permitem aos usuários realizar simulações de dinâmica molecular, em particular sobre biomoléculas.
Charmm	Simulações macromoleculares, incluindo minimização de energia.
Cn3d; Rasmol	Visualização de estruturas tridimensionais.
EsyPred3D	Predição de estruturas proteicas 3D
Modeller	Modelagem tridimensional de proteínas.
Alinhamento de Sequências	
BLAST	Alinhamento de sequências de nucleotídeos e proteínas.
Clustal, Muscle, T-Coffee	Alinhamento múltiplo de sequências
MEGA	Ferramenta integrada para alinhamento de sequências, construção de árvores, estimativa de taxas de evolução molecular, teste de hipóteses evolutivas, etc.
ModelTest	Teste de modelos de alinhamento.
Phylogeny.fr	Ferramentas on-line de alinhamento e criação de árvores
Análises Metagenômicas	
Orphelia	Predição de genes que codificam para proteínas com origem filogenética desconhecida.
MetaGene Mark	Predição gênica para metagenomas.
Glimmer-MG	Sistema para a identificação de genes em sequências de DNA de metagenômica
Análises Estatísticas	
Gromacs	Pacote para realização de dinâmica molecular para sistemas com centenas de milhões de partículas.
R	Estatística computacional.

5. Biologia de sistemas: interação gênica em vias metabólicas e de resposta ao estresse

A análise de dados genômicos e transcriptômicos em associação com modelos de redes metabólicas vem auxiliando na exploração de dados de uma forma global, revelando genes relacionados, por exemplo, com doenças e seus pontos de convergência de regulação (GALHARDO et al., 2013). Em eucariotos, sabe-se que o controle da expressão se dá através de interações entre os vários FTs, sendo mais complexas em termos de elucidação (CHEN; ZHU, 2004).

As plantas utilizam redes de regulação gênica, compostas por vários componentes, tais como DNA, RNA, proteínas e pequenas moléculas, para a regulação de vários processos biológicos, permitindo assim, que elas se adaptem a um determinado ambiente em mudança ou para respostas comuns ao seu desenvolvimento. A disponibilidade de métodos experimentais de alto rendimento permite a determinação dos níveis de expressão de milhares de genes, bem como de interações proteína-proteína ou proteína-DNA. Neste contexto, abordagens de biologia de sistemas permitem a integração destas informações, contribuindo para a compreensão das propriedades destes sistemas biológicos em células ou tecidos específicos, tornando-se úteis na elucidação das propriedades das redes de regulação transcricional nas plantas (PU; BRADY, 2010).

A Biologia de Sistemas é uma área multidisciplinar que visa integrar os dados relativos a genes e proteínas individuais e investigar o comportamento e relações entre os diversos elementos de um sistema biológico para explicar o seu funcionamento (KITANO, 2002); ou seja, descreve uma abordagem para a compreensão da biologia que dá ênfase às interações entre os componentes conhecidos através de biologia molecular, a despeito das propriedades isoladas dos próprios componentes. A abordagem da biologia de sistemas para doenças complexas, por exemplo, reconhece um número potencialmente grande de elementos envolvidos na progressão da doença. Assim, a reconstrução da rede nestes casos pode permitir a identificação de inúmeras associações genéticas (FUNKE et al., 2013).

Embora vários estudos forneçam uma visão geral sobre os princípios gerais de redes biológicas, a ligação entre a organização de tais redes e a dinâmica do genoma ao nível de genes, mRNAs, proteínas e seu papel no comportamento dos sistemas permanecem obscuros (LIN et al., 2012). Vários algoritmos computacionais foram desenvolvidos para reconstruir redes regulatórias utilizando combinações de diferentes tipos de dados, incluindo dados de

perfil de expressão. Alguns destes algoritmos têm sido usados com sucesso para reconstruir redes de ordem superior, apresentando uma visão mais abrangente da organização em rede. Dada a crescente acumulação de dados sobre plantas, estes métodos computacionais têm potencial para produzir conhecimentos valiosos sobre a regulação da transcrição em resposta a pressões ambientais em vegetais (CHEN; ZHU, 2004).

5.1. Redes regulatórias / modelagem de mapas de interação

A célula de um organismo (seja ela animal ou vegetal) é um sistema complexo, cujas características são definidas através da atividade de muitos componentes que se comunicam entre si através de interações par-a-par. Tais componentes podem ser representados por uma série de 'nós', conectados por ligações, que representam as interações. Os diversos nós e ligações entre estes formam uma rede (BARABASI; OLTIVAI, 2004).

Existem diversos tipos de redes biológicas, ou seja, aquelas que representam vias metabólicas e redes que representam interações entre genes e os FTs que os regulam, as quais podem ser de dois tipos: direcionais e não-direcionais (Figura 6). As redes direcionais são aquelas onde a informação representada apresenta uma direção, como redes em que FTs regulam diferentes genes. Já as redes não-direcionais, a informação representada não possui nenhuma direção, como exemplo, as redes de interação proteicas, em que se representam as interações físicas entre proteínas distintas (BARABASI; OLTIVAI, 2004).

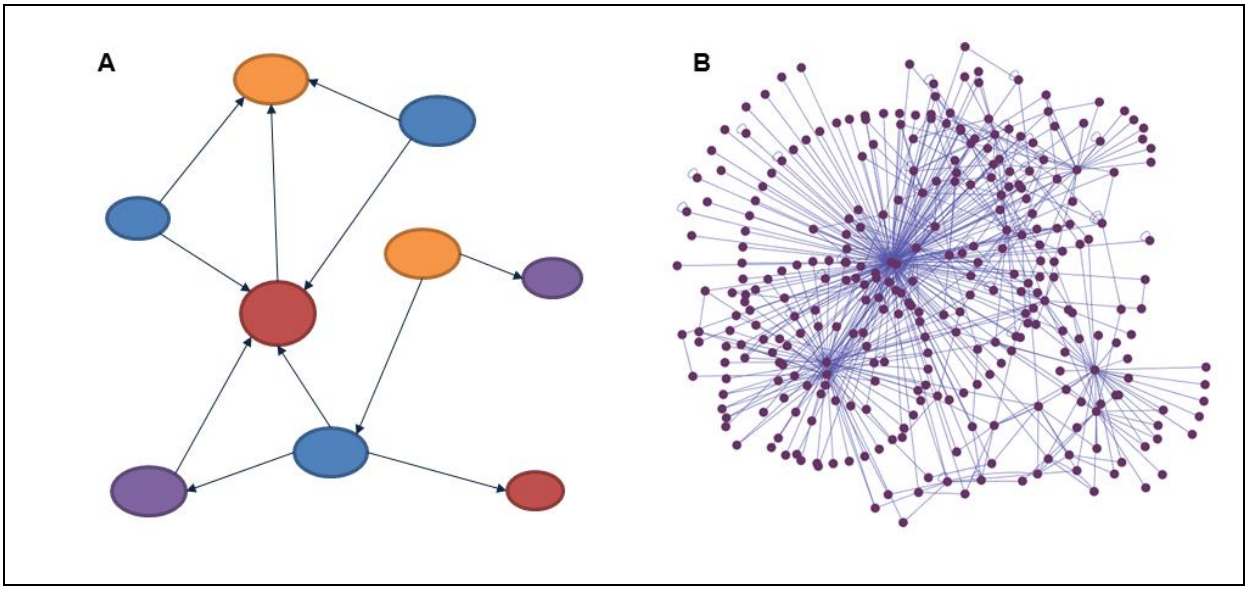


Figura 6. Modelos hipotéticos de redes regulatórias. (A) Rede direcional, as interações entre os nós possuem direção e (B) Rede não-direcional, não existe direção entre as interações dos nós (Imagem do autor).

Devido ao elevado número de genes envolvidos nessas redes, ativando e reprimindo *feedbacks*, sua dinâmica é extremamente complexa e difícil de inferir. Com o desenvolvimento de tecnologias de análise de expressão gênica em larga escala, como RNA-Seq, tornou-se possível a análise de milhares de genes simultaneamente, permitindo a construção de redes regulatórias, baseadas nos dados de expressão. Neste sentido, diversas redes genéticas têm sido propostas, a exemplo das Redes Bayesianas (FRIEDMAN, 2004; DOJER et al., 2006), Equações Estruturais (XIONG et al., 2004), Redes Booleanas Probabilísticas (PAL et al., 2005) e Equações Diferenciais (MESTL et al.,1995).

Entretanto, apesar desses métodos modelarem bem redes regulatórias onde existe informação biológica disponível, a determinação de todas as relações entre os componentes de uma rede ainda é uma tarefa difícil. O ponto fundamental desta abordagem é que ela não apresenta um teste estatístico, sendo arbitrária a escolha do limiar de informação, tornando mais complicada a interpretação e identificação das verdadeiras interações da rede. Assim, a escolha do limiar para se determinar a veracidade de uma interação se torna subjetiva. Uma solução seria a construção da rede em módulos, onde cada módulo seria composto por vários genes e, só depois é que se construiria a rede integrando-se módulo-a-módulo (XU et al., 2004; YAMAGUCHI et al., 2007). Os módulos funcionais consistem em subconjuntos de nós altamente interconectadas dentro das redes, os quais compreendem nós fortemente conectados que, no caso de redes de coexpressão, representam grupos de genes coexpressos que podem agir em conjunto dentro de um processo biológico com o qual estão associados.

Uma variedade de algoritmos de agrupamento de rede vem sendo desenvolvida para identificar os módulos em redes, incluindo, por exemplo, o algoritmo de agrupamento Markov (ENRIGHT et al., 2002), MCLUST (FRALEY; RAFTERY, 2003), biclustering (PRELIĆ et al., 2006), e o algoritmo de agrupamento heurístico (MUTWIL et al., 2010). Apesar de eficientes, alguns desses métodos podem não ser realistas em um contexto biológico, onde genes individuais são capazes de atuar em múltiplas vias. Além disso, a variabilidade na detecção do módulo também pode ser agravada pela geração de diferentes redes por meio de métodos diferentes (AOKI et al., 2007).

O mapeamento e análise de sistemas de redes de regulação em grande escala, permite a descoberta de propriedades emergentes, como a interação de fatores regulatórios e seus alvos, tal como os FTs e seu efeito regulatório em seus alvos gênicos (LONG et al., 2008). O grande desafio será validar as interações geradas pelas várias redes e incorporar essas interações e suas consequências funcionais de forma ‘espaço-temporal’. Este ponto é importante para decifrar

como essas redes são interligadas entre si, permitindo a compreensão de como o organismo se desenvolve, cresce e se reproduz. Estudos transcricionais têm sido aplicados com sucesso na exploração dos processos de regulação gênica e da dinâmica da transcrição (DINNENY et al., 2008; WEINHOFER et al., 2010).

5.1.1 Principais ferramentas e contribuições para o entendimento dos sistemas biológicos

Qualquer estudo em biologia de sistemas depende de *softwares* e dados adequados, como informações disponíveis em bancos de dados *on-line*, muitos com ferramentas de análise (Tabela 3). Alguns bancos, inclusive, oferecem acesso a dados de transcriptoma, onde ferramentas como o Genevestigator e o ATTED II são úteis para a busca de genes ou proteínas com perfis de expressão particulares ou coexpressão com genes avos (TOUFIGHI et al., 2005; OBAYASHI et al., 2011). A análise de conjunto de genes deve ser apoiada por evidências estatísticas, pois associações de genes exclusivamente com base na coexpressão nem sempre são informativas, apesar de aumentarem o poder preditivo quando incluem outro tipo de dados.

Iniciativas como o IntAct (KERRIEN et al., 2006), o *Molecular Interaction database* (CHATR-ARYAMONTRI et al., 2007), *Database of Interacting Proteins* (SALWINSKI et al., 2004), *Biomolecular Interaction Network Database* (ALFARANO et al., 2005) e o BioGRID (STARK et al., 2006) foram estabelecidas para o acúmulo sistemático e organização dos dados de interação. Um dos recursos mais abrangentes para a localização de proteínas é o banco de dados de localização subcelular de proteínas de *Arabidopsis*, que armazena as localizações de proteínas subcelulares determinadas experimentalmente e preditas em *Arabidopsis* (HEAZLEWOOD et al., 2007). Outras ferramentas amplamente utilizadas estão o AraNet, GeneMania e STRING, conhecidas pela facilidade de utilização (WARDE-FARLEY et al., 2010; SZKLARCZYK et al., 2011).

Tradicionalmente, a biologia de sistemas, muitas vezes se concentra em espécies-modelo. No entanto, muito pode ser aprendido através da exploração de ferramentas de genômica comparativa, o que pode ajudar a atribuir funções aos membros de famílias multigênicas, definindo conjuntos de ortólogos e identificando motivos regulatórios de genes relevantes. A base de dados Plaza, por exemplo, associa os dados genéticos de 25 espécies de plantas disponíveis (VAN BEL et al., 2011).

Tabela 3. Bases de dados de acesso livre e ferramentas para análise de redes regulatórias gênicas (Adaptado de Bassel et al., 2012).

Aplicações	Ferramentas
Bancos de Dados e Repositórios	
Expressão Gênica	BAR, Genevestigator, ATTED, AT-TAX, TileViz
Expressão de Proteínas	Pep2pro
Interação Proteína-Proteína	ANAP, BAR, PAIR, IntAct, AtPIN, AtPID, AthPPI, Associomics, PRIN (rice), BioGRID
Localização de Proteínas	SUBA, PREDOTAR, TargetP
Análise Conjunta de Genes	BAR, AMIGO, DAVID, FuncAssociate, BiNGO, ATCOECIS, MapMan, AraNet, AgriGO
Comparação de Genomas	Plaza, OrthologID, DoOP
Regulação Gênica e Arquitetura de Promotores	
Detecção de Motivos	BAR, ATCOECIS, TAIR patmatch, DoOP, STAMP, Weeder
Análise de Sítios de Ligação	PLACE, AthaMap, AGRIS
Integração de dados e Inferência de Redes	LeMoNe, ENIGMA, CORNET, Ondex, AraNet, GeneMania
Computação genérica e Modelagem de Ambientes	R, MatLab, Systems Biology Workbench
Visualização de Dados	AraCyc, PlantCyc, Cytoscape, MapMan, REACTOME

De modo geral, a elucidação da regulação da transcrição envolve a identificação de ambos os elementos de regulação. Vários algoritmos e ferramentas estão disponíveis para tais análises (LADUNGA, 2010), como o Lemone e o ENIGMA, que são pacotes de *software* para a construção de redes em módulos a partir de dados de expressão gênica (MAERE et al., 2008). A inferência de redes gênicas regulatórias compreende um campo ativo e ainda em expansão, onde novos e poderosos algoritmos estão sendo continuamente desenvolvidos, oferecendo ótimas opções de modelagem para sistemas biológicos com necessidades mínimas de *script* e extremamente eficientes para análises de modelos de interação (SAURO et al., 2003).

5.1.2 Biologia de sistemas em leguminosas

Muitos métodos computacionais foram desenvolvidos para prever a atividade de genes reguladores e seus alvos em *A. thaliana* (NEEDHAM et al., 2009; YAO et al., 2011), embora a predição de redes regulatórias gênicas em espécies vegetais ainda esteja em fase inicial, seja pela ausência de ferramentas adequadas, seja pela ausência de integração de métodos computacionais com dados de expressão e outras fontes de dados para o estudo de genes coexpressos. Especificamente, muito pouco tem sido feito para a construção de redes de regulação gênicas para leguminosas, com exceção das espécies-modelo economicamente relevantes como a soja, apesar da grande quantidade de dados de expressão gênica acumulados durante os últimos anos (BENKO-ISEPPON et al., 2012).

No entanto, em casos específicos, a informação experimental é escassa. Assim, construir redes regulatórias gênicas a partir de diferentes abordagens não é uma tarefa fácil. Rodriguez-Llorente et al. (2009), utilizando novos métodos computacionais do *'Symbiosis Interactoma*', determinaram a dinâmica de interações entre proteínas envolvidas na simbiose da bactéria modelo *Sinorhizobium meliloti* com suas plantas hospedeiras. O método foi utilizado para guiar as técnicas experimentais que tentam validar novas proteínas envolvidas em diferentes estágios de simbiose, trazendo à tona uma metodologia para o estudo de interações planta-microorganismo, fornecendo suporte para validações experimentais, e também, aplicações no estudo de outros processos complexos, tais como doenças causadas por patógenos.

Em contrapartida, com a disponibilidade do genoma completo da soja (SCHMUTZ et al., 2010) e numerosas anotações de genes e proteínas – por exemplo, SoyDB, um banco de dados anotado para função de FTs (WANG et al., 2010) – torna-se importante desenvolver e integrar os métodos de bioinformática para a construção de redes gênicas confiáveis, seja por meio dos dados de expressão, seja com os dados de genômica funcional. Nesse sentido Zhu et al. (2012), projetaram um protocolo para integrar diversos métodos de bioinformática, permitindo a partir de dados de expressão de larga escala da soja, elucidar os mecanismos e processos biológicos por meio de redes regulatórias no desenvolvimento e crescimento celular sob condições de estresse. Esta metodologia pode ser extrapolada para outras leguminosas, reduzindo a problemática de anotação de dados e sendo robusta o suficiente para a geração das redes a partir dos dados do transcriptoma.

Correlacionar expressão entre genes (análise de coexpressão) é uma abordagem poderosa para analisar grandes conjuntos de dados, principalmente para genes envolvidos nas

mesmas vias biológicas (USADEL et al., 2009; BASSEL et al., 2011). Verdier et al. (2013) investigaram um modelo para a leguminosa *Medicago truncatula*, visando à construção de uma rede que descrevesse as últimas três semanas de desenvolvimento da semente antes da maturação. A rede revelou módulos coexpressos distintos, relacionados com a aquisição de tolerância à dissecação, longevidade e abertura da vagem. A aquisição de tolerância à dissecação e a ocorrência de dormência foi associada a genes de resposta ao estresse abióticos e a genes tardios de embriogênese.

Tendo em vista a necessidade de um quadro de bioinformática mais integrado para a previsão de redes de regulação gênicas a partir de dados de expressão de soja, Zhu et al. (2013) desenvolveram um protocolo integrado para análise de expressão diferencial, clusterização gênica, identificação de motivos de ligação ao DNA e predição de função a partir de dados de RNA-Seq. Embora a estrutura básica da construção da rede de regulação seja a mesma que o método anterior (ZHU et al., 2012), o pré-processamento e a normalização dos dados de RNA-Seq, a filtragem de genes expressos diferencialmente, bem como a aplicação biológica foram distintas. Utilizando a soja, eles determinaram a rede para o processo de nodulação da planta, sendo verificado que este processo apresenta três fases, permitindo a identificação de diferentes genes para cada uma das fases. Este tipo de abordagem transpõe os dados de RNAseq a um novo patamar de importância para a construção de redes em um contexto biológico.

6. Genômica funcional de soja e de feijão-caupi

Devido à crescente necessidade de pesquisas que atendessem as demandadas agrícolas, industriais e de alimentação, visando incrementar o processo produtivo da soja e do feijão-caupi, foram criados os programas para estudo do genoma expresso destas leguminosas. Em 2007, o governo brasileiro iniciou o consórcio GENOSOJA (Consórcio Nacional para Estudos para Genoma da Soja) com o objetivo de aperfeiçoar o processo de produção deste grão, com ênfase para fatores que afetam a produção nacional, como a ocorrência de secas, pragas e patógenos (NASCIMENTO, 2009).

Atualmente, o genoma da soja está sendo investigado por vários grupos em diferentes países, incluindo estudos de genômica estrutural e funcional, transcriptoma e proteoma. O projeto também tem como objetivo estudar a estrutura e função dos genes e proteínas, bem como gerar o perfil de expressão de soja sob estresses bióticos e abióticos, como a ferrugem asiática, nematoides, seca e fixação de nitrogênio. Os dados gerados vêm sendo depositados em

bancos de dados públicos (GenBank) e de acesso restrito no site do consórcio GENOSOJA (ABDELNOOR et al., 2009; BENKO-ISEPPON et al., 2012).

Com relação ao feijão-caupi, a Rede Nordestina de Biotecnologia (ReNorBio) realizou, sob coordenação da Universidade Federal de Pernambuco, a análise genômica funcional, estrutural e comparativa em *V. unguiculata* - NordEST, objetivando o sequenciamento de etiquetas de sequências expressas (ESTs - *Expressed Sequence Tags*) para a identificação de genes de interesse nessa cultura para fins de melhoramento. O projeto integra 10 instituições e 11 Laboratórios, onde foram obtidas 27.453 sequências de EST e mais de 21 milhões de sequências de SuperSAGE (*Super Serial Analysis of Genome Expression*) relacionadas à expressão de genes de resistência a fatores bióticos e abióticos importantes para o melhoramento do feijão-caupi (BENKO-ISEPPON et al., 2010).

Essa geração massiva de dados sobre a expressão de genes em diversas condições pode tornar as culturas (como o feijão-caupi e a soja) altamente produtivas e rentáveis, sendo sua anotação um dos requisitos para o entendimento da genômica funcional, ou seja, a chave para a interpretação biológica dos dados (CONESA et al., 2005; SHIVASHANKAR et al., 2006). Esta abordagem é especialmente promissora na anotação de sequências diferencialmente expressas, por exemplo, para a análise das respostas em plantas submetidas a estresses abióticos, fornecendo uma ferramenta poderosa para a genômica funcional (CHEN et al., 2002).

6.1. Importância socioeconômica da soja e do feijão-caupi

A soja (*G. max*) compreende uma das culturas mais importantes do mundo, sendo usualmente comercializada em grão, farelo e óleo bruto e refinado, constituindo importante fonte de proteína, além de apresentar grande potencial na fabricação de plásticos e biodiesel (SEAB, 2013). No mundo, o agronegócio da soja movimenta aproximadamente 230 bilhões de dólares por ano. Estima-se que são produzidas por ano cerca de 267,61 milhões de toneladas anuais deste grão (USDA, 2014). No Brasil, destaca-se como a espécie vegetal cultivada de maior importância econômica e social. A expansão de seu cultivo teve início no final dos anos 60, tornando-se, em curto período de tempo, um dos principais produtos da exploração agrícola e da economia brasileira. Com isso, a soja colocou o Brasil em posição de grande destaque no agronegócio mundial, tornando-se o maior produtor mundial desta oleaginosa (USDA, 2014).

Atualmente, a cultura de soja ocupa um lugar de destaque no mercado de "commodities" do Brasil, com uma produção superior a 91 milhões de toneladas, na previsão da safra de

2014/2015 (CONAB, 2014; USDA, 2014). Entre as culturas agrícolas de destaque mundial, a soja tem sido objeto de um grande número de estudos que buscam compreender e quantificar as relações ecofisiológicas que interferem na produção da cultura devido aos desafios encontrados para sua adaptação nos diferentes ambientes (POPP et al., 2003). Porém, a crescente expansão da soja para novas fronteiras agrícolas brasileiras, com diferentes condições edafoclimáticas, tem exigido um aumento na demanda por tecnologias que deem sustentabilidade à produção dessa cultura em locais onde seu rendimento é afetado por fatores bióticos e abióticos (ASSAD et al., 2007).

O Brasil produz e consome diversos tipos de feijão, destacando-se os do grupo I, que inclui feijão-comum (*Phaseolus vulgaris*), com aproximadamente 60 % da produção (principalmente, o grão tipo comercial carioca – mais consumido no Brasil) e o grupo II: feijão-caupi (*V. unguiculata*), com média de 42 % da produção, entre os anos de 2005 e 2009 (FREIRE-FILHO et al., 2011; EMBRAPA, 2014).

O feijão-caupi, popularmente conhecido como "feijão-macassar", feijão verde ou "feijão-de-corda" é uma Fabaceae herbácea, nativa da África e bastante cultivada nas regiões tropicais dos continentes africano, asiático e americano (EHLERS; HALL, 1997), com destaque na economia Nordestina, constituindo o principal alimento proteico e energético das comunidades rurais (TEÓFILO et al., 2001). Apresenta alta rusticidade e adaptabilidade às condições de estiagem prolongadas e capacidade de se desenvolver em solos de baixa fertilidade. Estas características, tais como as de precocidade (ISHIYAKU et al., 2005), tolerância à seca (PIMENTEL et al., 2002) e à salinidade (WANG et al., 2003) fazem do feijão-caupi uma cultura promissora como fonte de importantes genes para o melhoramento de outras leguminosas, como a soja, o feijão comum, a ervilha, a lentilha, o amendoim e o grão-de-bico, entre outras (BENKO-ISEPPON, 2001).

Mais de 5,4 milhões de toneladas de feijão-caupi são produzidas em todo o mundo, sendo no Brasil o feijão-caupi cultivado predominantemente no sertão semi-árido da região Nordeste, onde é a segunda maior cultura agrícola, estando atrás apenas da mandioca (*Manihot esculenta*). De acordo com a FAO (Organização para a Alimentação e Agricultura das Nações Unidas), desde 2012, o rendimento médio do caupi foi de 483 kg/ha no continente africano, que está 50 % abaixo do rendimento potencial da cultura. Em alguns métodos de cultivo mais tradicionais, o rendimento pode ser muito baixo, não ultrapassando os 100 kg/ha. Apesar da sua ampla variabilidade genética para praticamente todos os caracteres de interesse agronômico,

com relação a outras culturas, feijão-caupi ainda é pouco melhorado (ROCHA et al., 2003). Atualmente os projetos de melhoramento da cultura visam ao desenvolvimento de cultivares com alta qualidade de grão, resistência à seca e à salinidade, resistência múltipla a viroses, a doenças fúngicas e bacterianas, porte mais compacto e mais ereto, que possibilitem a colheita mecânica e características para processamento industrial (FREIRE-FILHO et al., 2005; 2011).

Considerando que seu consumo médio é de 18 Kg ano por pessoa, ele abastece a mesa de 28,2 milhões de nordestinos e gera 1,1 milhões de empregos. Apresenta-se como uma excelente fonte de proteínas (23-25 % em média), carboidratos (62 %, em média), vitaminas e minerais, além de possuir grande quantidade de fibras e não conter colesterol. Representa alimento básico para as populações de baixa renda do Nordeste brasileiro. Por toda sua importância socioeconômica, o feijão-caupi necessita de uma maior atenção, tanto com ênfase nas políticas de abastecimento, quanto para a pesquisa que foquem o melhoramento (FREIRE-FILHO et al. 2011).

Esforços têm sido feitos pelos segmentos que trabalham com biotecnologia em várias instituições no sentido de criar nova cultivares com características diferenciadas, que contribuam para minimizar perdas na sua produção, melhorando o entendimento da base genética de respostas adaptativas a situações de desafio e explorando esse conhecimento com a finalidade de melhorar, por exemplo, a tolerância à seca nessas plantas (TALAMÈ et al., 2007). A busca pelo conhecimento e pelo aprimoramento de novas técnicas nessa área está sendo de grande importância para o avanço tecnológico e o conhecimento dos países desenvolvidos em relação a essas leguminosas (BAJAJ et al., 2000).

6.2 Bancos de dados relacionados as culturas da soja e do feijão-caupi

A soja apresenta um genoma com tamanho aproximado de 1,1 Gb (ARUMUGANATHAN; EARLE, 1991), sendo considerado complexo devido aos diversos ciclos de duplicações ocorridos nos últimos 45 milhões de anos, possuindo entre 40 e 60 % de sequências repetitivas (SCHMUTZ et al., 2010).

Em janeiro de 2010, o consórcio internacional de pesquisas em soja liderado pelo JGI (Joint Genome Institute) concluiu o sequenciamento da cultivar Williams 82, a mais plantada nos Estados Unidos. A montagem final foi realizada com 13 milhões de sequências, com cobertura de aproximadamente 7,2 vezes seu genoma, obtendo um comprimento final de 950

Mb, distribuídos em 20 cromossomos. Um total de 66.153 genes foram preditos a partir das sequências genômicas, sendo 46.430 deles com alta confiabilidade (SCHMUTZ et al., 2010). Os dados estão disponíveis na base de dados Phytozome (http://www.phytozome.net/soybean) e foram os primeiros a serem armazenados no servidor do projeto.

A quantidade total de dados disponíveis publicamente para soja no GenBank (NCBI) inclui mais de 120 mil sequências de nucleotídeos (principalmente mRNA), ~ 1460000 ESTs, ~ 368 mil sequências genômicas, ~ 80.000 proteínas, com pelo menos 118 estruturas tridimensionais depositadas e mais de 6,2 milhões de SNPs (BENKO-ISEPPON et al., 2012).

Com relação a transcritos, um banco de dados foi desenvolvido para o depósito de sequências, aprofundando o nível de informações sobre a transcrição (LIBAULT et al., 2010). Além disso, o banco de dados de SoyDB (http://casp.rnet.missouri.edu/soydb/) é especificamente curado para FTs de soja (WANG et al., 2010). Vários conjuntos de dados gerados usando vários tecidos ou diferentes estágios de desenvolvimento já foram depositados. Por exemplo, pode-se acessar os dados gerados a partir de soja submetidas a infecção por patógenos, tal como a *Pseudomonas syringae* (ZABALA et al., 2006). Outro banco de destaque é a Base de Soja Conhecimento (SoyKB), um recurso web-based abrangente para pesquisa relacionada a cultura da soja, projetado para lidar com o armazenamento e integração de informações genômicas, transcriptômicas (EST e microarray, por exemplo), proteômicas, metabolômica e incluindo informações a cerca do fenótipo (TRUPTI et al., 2012).

Outras bases encontram-se disponíveis para soja e outras leguminosas importantes, como o SoyBase (http://soybase.org) (GRANT et al., 2009; DU et al., 2010), o Soy-TFKB (Soybean Transcription Factor Knowledge Base; http://www.igece.org/SoybeanTF), o SGMD (The Soybean Genomics and Microarray Database; http://bioinformatics.towson.edu/SGMD) (ALKHAROUF; MATTHEWS, 2004) e o LegumeTFDB (http: // legumetfdb .psc.riken.jp / index.pl) (MOCHIDA et al., 2010). Esse número de bancos e dados disponíveis para soja mostra que os estudos de ômicas da cultura representam como um caminho promissor para o entendimento de características importantes para o melhoramento da cultura.

No âmbito nacional, foi criado em 2007/2008 o consórcio GENOSOJA (http://bioinfo.cnpso.embrapa.br/genosoja/), com o objetivo de estudar o genoma da soja a partir de sua organização em nível estrutural, buscando caracterizar e regiões genômicas importantes e seus produtos, contribuindo assim para a identificação de genes utilizando metodologias de transcriptômica e proteômica, especialmente considerando-se a resposta das

plantas frente a estresses bióticos e abióticos. Desta forma, a rede GENOSOJA tem por objetivo abordar não apenas se um gene é induzido ou reprimido sob uma determinada condição, mas também para determinar o contexto em que se expressa, incluindo as interações com outros genes, sua localização física e produtos, permitindo inferências sobre genes importantes e suas vias metabólicas, vitais para o desenvolvimento e estudo de plantas tolerantes a situações adversas. Para isto foi desenvolvido o banco de dados GENOSOJA (http://bioinfo03.ibi.unicamp.br/soja/), que inclui um conjunto de ferramentas que integram todos os dados do projeto, em comparação com sequências disponíveis em outros bancos de dados públicos (BENKO-ISEPPON et al., 2012; NASCIMENTO et al., 2012).

Para o feijão-caupi, que apresenta um genoma nuclear estimado em 620 Mb (um dos menores genomas dentro das leguminosas e plantas vasculares), são poucas as iniciativas disponíveis de genômica funcional, quando comparado com a soja (ARUMUGANATHAN; EARLE, 1991). Isto se deve provavelmente ao fato das diferentes cultivares estarem sendo analisadas no âmbito do desenvolvimento de marcadores moleculares, deixando os estudos genômicos aquém das outras grandes culturas de interesse econômico. No entanto, existem pesquisas e desenvolvimento de bancos voltados para estudos do genoma de feijão-caupi. A iniciativa genômica para o caupi (CGI - *Cowpea Genomics Initiative*) realizou o sequenciamento do feijão-caupi, fornecendo uma anotação e análise dos dados gerados. O 'Cowpea Genespace/Genomics Knowledge Base' (CGKB) foi criado visando o gerenciamento, a análise e a disseminação da informação obtida a partir do sequenciamento do caupi. Assim, o CGKB (http://cowpeagenomics. med.virginia.edu/CGKB) foi desenvolvido com vistas à anotação funcional dos genes, determinação de vias metabólicas, homologia baseada em HMM (*Hiden Markov Model*), dentre outras funções (CHEN et al., 2007).

Além deste, encontram-se disponíveis um banco de dados para o mapa físico (http://phymap.ucdavis.edu:8080/cowpea) que compreende a preparação de um mapa genético de consenso contendo mais do que mil marcadores SNP para gerar um mapa físico do feijão-caupi baseado em BACs (Cromossomos Artificias de Bactérias). Merece menção também o HarvEST, um banco para dados de ESTs (http://harvest.ucr.edu), que é apoiado pelo 'USDA Plant Genome program' e pelo 'CGIAR Generation Challenge Program' (KAGA et al., 2014).

No Brasil, foi estabelecida, em 2007, a rede NordEST, agrupando 11 laboratórios com objetivo de gerar dados de genômica funcional úteis para o melhoramento da cultura e programas de biotecnologia. Assim, foram gerados e analisados 500.179.927 transcritos,

incluindo 453.952.833 RNAseq (100 bp Illumina HiSeq TruSeq); 46.582.833 HT-SuperSAGE tags (26 bp, Solexa-Illumina); 298.119 SuperSAGE tags (26 bp, 454 Life Sciences/Roche); 32.084 LongSAGE tags (19-21 bp); e 314.765 ESTs (Expressed Sequence Tags) (49,820 gerados pela rede NordEST) (BENKO-ISEPPON et al., 2013). Ainda no escopo da iniciativa, está sendo conduzido o mapeamento físico dos cromossomos, incluindo amostras de microssatélites, rDNA e BACs, o que torna viável a rápida transferência de dados entre culturas visando o melhoramento. Mais recentemente mais de quatro milhões de sequências de RNAseq foram geradas e montadas no banco de dados NordEST, que conta com um conjunto de ferramentas que integram os dados do projeto, e permitem a anotação de diversos genes importantes para o melhoramento e biotecnologia da cultura (BENKO-ISEPPON et al., 2013).

REFERÊNCIAS BIBLIOGRÁFICAS

Abdelnoor, R.V. et al. GENOSOJA – A Brazilian Soybean Genome Consortium. In: Plant & Animal Genome XVII, San Diego. **Plant & Animal Genome Conference XVII** Abstracts (www.intl-pag.org/17/abstracts/P01_ PAGXVII_034.html), 2009.

Achuo, E.A.; Prinsen, E.; Hofte, M. Influence of drought, salt stress and abscisic acid on the resistance of tomato to *Botrytis cinerea* and *Oidium neolycopersici*. **Plant Pathology**, v. 55, p. 178–186, 2006.

Ahuja, I.; De Vos, R.C.H.; Bones, A.M. et al. Plant molecular stress responses face climate change. **Trends in Plant Science**, v. 15, p. 664–674, 2010.

Alberts, B. Biologia molecular da célula. 5. ed. Porto Alegre: **Artmed**, 2010. xxxv, 1268, 40, 50p. ISBN 9788536320663 (enc.).

Alfarano, C. et al. The Biomolecular Interaction Network Database and related tools 2005 update. **Nucleic Acids Research**, v. 33, p. D418–D424, 2005.

Alkharouf, N.W.; Matthews, B.F. SGMD: the soybean genomics and microarray database. **Nucleic Acids Research**, v. 32, p. D398–D400, 2004.

Anderson, J.P.; Badruzsaufari, E.; Schenk, P.M.; Manners, J.M.; Desmond, O.J.; Ehlert, C.; Maclean, D.J.; Ebert, P.R.; Kazan, K. Antagonistic interaction between abscisic acid and jasmonate– ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis. **The Plant Cell**, v. 16, p. 3460–3479, 2004.

Aoki, K.; Ogata, Y.; Shibata, D. Approaches for extracting practical information from gene co-expression networks in plant biology. **Plant Cell Physiol**, v. 48, p. 381–390, 2007.

Arumuganathan, K.; Earle, E.D. Nuclear DNA content of some important plant species. **Plant Molecular Biology Reporter**, v. 9, p. 208-218, 1991.

Assad, E. D.; Marin, F. R.; Evangelista, S. R. et al. Forecast system of soybean crop yield for Brazil. **Pesquisa Agropecuária Brasileira**, v. 42, n. 5, p. 615-625, 2007.

Asselbergh, B.; De Vieesschauwer, D.; Hofte, M. Global switches and fine-tuning—ABA modulates plant pathogen defense. **Molecular Plant-Microbe Interactions**, v. 21, p. 709–719, 2008.

Atkinson, N.J.; Dew, T.P.; Orfila, C.; Urwin, P.E. Influence of combined biotic and abiotic stress on nutritional quality parameters in tomato (*Solanum lycopersicum*). **Journal of Agricultural and Food Chemistry,** v. 59, p. 9673–9682, 2011.

Audebert, A.; Coyne, D.L.; Dingkuhn, M.; Plowright, R.A. The influence of cyst nematodes (*Heterodera sacchari*) and drought on water relations and growth of upland rice in Cote d'Ivoire. **Plant and Soil**, v. 220, p. 235–242, 2000.

Babu, M.M.; Luscombe, N.M.; Aravind, L.; Gerstein, M.; Teichmann, S.A. Structure and evolution of transcriptional regulatory networks. **Current Opinion in Structural Biology**, v. 14, p. 283–291, 2004.

Bajaj, S.; Targolli, J.; Liu, L-F.; Ho, T-H.D.; Wu, R. Transgenic approaches to increase dehydration-stress tolerance in plants. **Molecular Breeding**, v. 5, p. 493–503, 2000.

Barabasi, A.L.; Oltvai, Z.N. Network biology: Understanding the cell's functional organization. **Nature Reviews Genetics**, v. 5, n. 2, p. 101-?-U15, 2004.

Barnabas, B.; Jager, K.; Feher, A. The effect of drought and heat stress on reproductive processes in cereals. **Plant, Cell and Environment**, v. 31, p. 11–38, 2008.

Bartels, D.; Sunkar, R. Drought and salt tolerance in plants. **Critical Reviews in Plant Sciences**, v. 24, p. 23–58, 2005.

Bassel, G.W.; Gaudinier, A.; Brady, S.M.; Hennig, L.; Rhee, S.Y.; De Smet, I. Systems Analysis of Plant Functional , Transcriptional , Physical Interaction , and Metabolic Networks. **Plant Cell**, v. 24, p. 3859–75, 2012.

Bassel, G.W.; Lan, H.; Glaab, E.; Gibbs, D.J.; Gerjets, T.; Krasnogor, N.; Bonner, A.J.; Holdsworth, M.J.; Provart, N.J. Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. **Proceedings of the National Academy of Sciences USA**, v. 108, p. 9709–9714, 2011.

Baxevanis, A.D. Bioinformatics and the internet. **Methods Biochemistry Anal**, v.43, p. 1-17. 2001.

Beattie, G.A. Water relations in the interaction of foliar bacterial pathogens with plants. **Annual Review of Phytopathology**, v. 49, p. 533–555, 2011.

Bechtold, U.; Lawson, T.; Mejia-Carranza, J.; Meyer, R.C.; Brown, I.R.; Altmann, T.; Ton, J.; Mullineaux, P.M. Constitutive salicylic acid defences do not compromise seed yield, drought tolerance and water productivity in the Arabidopsis accession C24. **Plant, Cell and Environment**, v. 33, p. 1959–1973, 2010.

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

Benko-Iseppon, A.M.; Kido, E.A.; Pandolfi, V. et al. Brazilian cowpea transcriptome project: over 20 million expressed sequence tags to analyze and breed salinity and virus resistance. **Resumos do III Congresso Brasileiro de Biotecnologia**. Fortaleza, CE, p. 97-98, 2010.

Benko-Iseppon, A.M.; Soares-Cavalcanti, N.M.; Belarmino, L.C; Bezerra-Neto J.P.; 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, E.A. Prospecção de genes de resistência à seca e à salinidade em plantas nativas e cultivadas. **Revista Brasileira de Geografia Física**, v. 6, p. 1112-1134, 2011.

Benko-Iseppon, A.M.; Nepomuceno, A.L.; Abdelnoor, R.V. GENOSOJA - The Brazilian Soybean Genome Consortium: High Throughput Omics and Beyond. **Genetics and Molecular Biology**, v. 35, p. i–iv, 2012.

Benko-Iseppon, A.M. Brazilian Cowpea Genome Consortium: understanding responses against biotic and abiotic stresses. In: **Annual Meeting of the West African Cowpea Consortium (WACC 2013)**. Organized by Kirkhouse Trust, Lomé, Togo, November/2013.

Bhalla, R.; Narasimhan, K.; Swarup, S. Metabolomics and its role in understanding cellular responses in plants. **Plant Cell Reports**, v. 24, n. 10, p. 562-71, 2005.

Bhattacharjee, S. Reactive oxygen species and oxidative burst: roles in stress, senescence and signal transduction in plants. **Current Science**, v. 89, n. 7, p. 1113–1121, 2005.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic Acids Research**, v. 42, n. W252–W258, 2014.

Borém, A.; Santos, F.R. Biotecnologia simplificada. **Editora Suprema**. Viçosa, MG, 2001.

Bray E.A. Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. **Journal of Experimental Botany**, v. 55, p. 2331–2341, 2004.

Brown, J.R.; Sanseau, P. A computational view of microRNAs and their targets. **Drug Discovery Today**, v. 10, p. 595–601, 2005.

Chatr-aryamontri, A.; Ceol, A.; Palazzi, L.M.; Nardelli, G.; Schneider, M.V.; Castagnoli, L.; Cesareni, G. MINT: the Molecular INTeraction database. **Nucleic Acids Research**, v. 35, p. D572–D574, 2007.

Chen, J.; Lee, S.; Zhou, G. et al. High-throughput GLGI procedure for converting a large number of serial analysis of gene expression tag sequences into 3' complementary DNAs. **Genes Chromosomes Cancer**, v. 33, p. 252-61, 2002.

Chen, S.; Harmon , A.C. Advances in plant proteomics. **Proteomics**, v. 6, p. 5504 – 5516, 2006.

Chen, W.J.; Zhu, T. Networks of transcription factors with roles in environmental stress response. **Trends Plant Science**, v. 9, n. 12, p. 591–596, 2004.

Chen, X.; Laudeman, T.W.; Rushton, P.J.; Spraggins, T.; Timko, M.P. CGKB: An Annotation Knowledge Base for Cowpea (*Vigna unguiculata* L.) Methylation Filtered Genomic Genespace Sequences. **BMC Bioinformatics**, v. 8, p. 129, 2007.

Cheng, X.; Blumenthal, R.M. Coordinated chromatin control: structural and functional linkage of DNA and histone methylation. **Biochemistry**, v. 49, n. 14, p. 2999-3008, 2010.

Chinnusamy, V.; Schumaker, K.; Zhu, J.K. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. **Journal of Experimental Botany**, v. 55, n. 395, p. 225–236, 2003.

CONAB. Companhia Nacional de Abastecimento. Disponível em <http://www.conab.gov.br>. Acesso em agosto/2014.

Conesa, A.; Götz, S.; García-Gómes, J.M. et al. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v.21, n. 18, p. 3674-3676, 2005.

Cooper, G.M. The Cell—A Molecular Approach, 2nd ed., **Sinauer Association**, Sunderland, MA, 2000.

Cramer, G.R. Abiotic stress & plant responses from the whole vine to the genes. **Australian Journal of Grape and Wine Research**, v. 16, p. 86-93, 2010.

Cramer, G. R.; Urano, K.; Delrot, S.; Pezzotti, M.; Shinozaki, K. Effects of abiotic stress on plants: a systems biology perspective. **BMC Plant Biology**, v. 11, n. 1, p. 163, 2011.

Crews, D.; Gore, A.C.; Hsu, T.S.; Dangleben, N.L.; Spinetta, M.; Schallert, T.; Anway, M.D.; Skinner, M.K. Transgenerational epigenetic imprints on mate preference. **Proceedings National Academy Science USA**, v.104, n. 14, p. 5942 – 5946, 2007.

Cushman, J.C.; Bohnert, H.J. Genomic approaches to plant stress tolerance. **Current Opinion in plant Biology**, v. 3, p. 117-124, 2000.

Denby, K.; Gehring, C. Engineering drought and salinity tolerance in plants: lessons from genome-wide expression profiling in *Arabidopsis*. **Trends in Biotechnology**, v. 23, p. 547–552, 2005.

Dinneny, J.R.; Long, T.A.; Wang, J.Y.; Jung, J.W.; Mace, D.; Pointer, S.; Barron, C.; Brady, S.M.; Schiefelbein, J.; Benfey, P.N. Cell identity mediates the response of Arabidopsis roots to abiotic stress. **Science**, v. 320 p. 942–945, 2008.

Dojer, N.; Gambin, A.; Mizera, A.; Wilczynski, B.; Tiuryn, J. Applying dynamic Bayesian networks to perturbed gene expression data. **BMC Bioinformatics**, v. 7, p.249, 2006.

Du, J.; Grant, D.; Tianetal, Z. SoyTEdb: acomprehensive database of transposable elements in the soybean genome. **BMC Genomics**, v. 11, n. 1, article 113, 2010.

Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. MYB transcription factors in Arabidopsis. **Trends in Plant Science**, v. 15, p. 573–581, 2010.

Edwards, D.; Batley, J. Plant bioinformatics: from genome to Phenome. **Trends in Biotechnology**, v. 22, n. 5, p. 232-237, 2004.

Ehlers, J.D.; Hall A.E. Cowpea (*Vigna unguiculata* (L.) Walp). **Field Crops Research**, v.53, p.187-204, 1997.

EMBRAPA, Empresa Brasileria de Pesquisa Agropecuária. Disponível em: <http://www.cnpso. EMBRAPA.br>. Acesso em: 05 de agosto de 2014.

Enright, A.J.; Van Dongen, S.; Ouzounis, C.A. An effi- cient algorithm for large-scale detection of protein families. **Nucleic Acids Research**, v. 30, p. 1575–1584, 2002.

Fazzari, M.J.; Greally, J.M. Epigenomics: beyond CpG islands. **National Review of Genetics**, v. 5, p. 446– 55, 2004.

Fernandez-Suarez, X.M.; Galperin, M.Y. The 2013 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. **Nucleic Acids Research**, v. 41, n. D1, p. D1–D7, 2012.

Fraley, C.; Raftery, A.E. Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. **Journal of Classification**, v. 20, p. 263–286, 2003.

Freire-Filho, F.R.; Rocha, M.M.; Ribeiro, V.Q. *et al*. Adaptabilidade e estabilidade produtiva de feijão-caupi. **Ciência Rural**, v. 35, p. 24-30, 2005.

Freire Filho, F.R.; Ribeiro, V.Q.; Rocha, M. De M.; Silva, K.J.D. E; Nogueira, M.S.R.; Rodrigues, E.V. Feijão-caupi no Brasil: produção, melhoramento genético, avanços e desafios. **Teresina: Embrapa Meio-Norte**, 2011. 84p.

Friedman, N. Inferring cellular networks using probabilistic graphical models. **Science**, v. 303, p. 799-805, 2004.

Fujita, Y.; Fujita, M.; Satoh, R.; Maruyama, K.; Parvez, M.M.; Seki, M.; Hiratsu, K.; Ohme-Takagi, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. **Plant Cell**, v. 17, n. 12, p. 3470-3488, 2005.

Funke, C.; Schneider, S.; Berg, D.; Kell, D.B. Genetics and iron in the systems biology of Parkinson's disease and some related disorders. **Neurochemistry**, v. 62, n. 5, p. 637–652, 2013.

Galhardo, M.; Sinkkonen, L.; Berninger, P.; Lin, J.; Sauter, T.; Heinäniemi, M. Integrated analysis of transcript-level regulation of metabolism reveals disease-relevant nodes of the human metabolic network. **Nucleic Acids Research**, v. 42, n. 3, p. 1474-1496, 2013.

Gill, S.S.; Anjum, N.A.; Hasanuzzaman, M.; Gill, R.; Trivedi, D.K.; Ahmad, I.; Pereira, E.; Tuteja, N. Glutathione and glutathione reductase: a boon in disguise for plant abiotic stress defense operations. **Plant Physiology and Biochemistry**, v. 70, p. 204–212, 2013.

Gill, S.S.; Tuteja, N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. **Plant Physiology and Biochemistry**, v. 48, p. 909-930, 2010.

Goda, H.; Sasaki, E.; Akiyama, K.; Maruyama-Nakashita, A.; Nakabayashi, K.; Li, W.; Ogawa, M.; Yamauchi, Y.; Preston, J.; Aokim, K.; Kiba, T.; Takatsuto, S.; Fujioka, S.; Asami, T.; Nakano, T.; Kato, H.; Mizuno, T.; Sakakibara, H.; Yamaguchi, S.; Nambara, E.;

Kamiya, Y.; Takahashi, H.; Hirai, M.Y.; Sakurai, T.; Shinozaki, K.; Saito, K.; Yoshida, S.; Shimada, Y. The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. **Plant Journal**, v. 55, n. 3, p. 526-542, 2008.

Grant, D.; Nelson, R.T.; Cannon, S.B.; Shoemaker, R.C. SoyBase, the USDA-ARS soybean genetics and genomics database. **Nucleic Acids Research**, v. 38, n. 1, p. D843–D846, 2009.

Hammond-Kosack, K.E.; Jones, J.D.G. Response to plant pathogens. In: Buchannan B, Gruissem W, Jones R, eds. Biochemistry and molecular biology of plants. **American Society of Plant Physiologists**, 1102–1157, 2000.

Hasegawa, P.M.; Bressan, R.A.; Zhu, J.K.; Bohnert, H.J. Plant cellular and molecular responses to high salinity. **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 51, p. 463–499, 2000.

Heazlewood, J.L.; Verboom, R.E.; Tonti-Filippini, J.; Small, I.; Millar, A.H. SUBA: The Arabidopsis Subcellular Database. **Nucleic Acids Research**, v. 35, p. D213–D218, 2007.

Herms, D.A.; Mattson, W.J. The dilemma of plants—to grow or defend. **Quarterly Review of Biology**, v. 67, p. 283–335, 1992.

Hinnebusch, A.G. eIF3: a versatile scaffold for translation initiation complexes. **Trends in Biochemical Sciences**, v. 31, p. 553-562, 2006.

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

Howden, B.P.; Beaume, M.; Harrison, P.F.; Hernandez, D.; Schrenzel, J. et al. Analysis of the small RNA transcriptional response in multidrug-resistant *Staphylococcus aureus* after antimicrobial exposure. **Antimicrobial Agents and Chemotherapy**, v. 57, p. 3864–3874, 2013.

Hubbard, K.E.; Nishimura, N.; Hitomi, K.; Getzoff, E.D.; Schroeder, J.I. Early abscisic acid signal transduction mechanisms: newly discovered components and newly emerging questions. **Genes Development**, v. 24, n. 16, p. 1695-1708, 2010.

IPCC, 2014: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Ishiyaku, M.F.; Singh, B.B.; Craufurd, P.Q. Inheritance of time to flowering in cowpea [*Vigna unguiculata* (L.) Walp.]. **Euphytica**, v.142, p.291–300, 2005.

Kaga, A.; Isemura, T.; Shimizu, T.; Somta, P.; Srinives, P. Asian Vigna Genome Research. p. 33–40, 2014.

Kanduri, C. Kcnq1ot1: a chromatin regulatory RNA. **Seminars in Cell & Developmental Biology**, v. 22, p. 343–350, 2011.

Kemper, B.; Matsuzaki, T.; Matsuoka, Y.; Tsuruoka, Y.; Kitano, H.; Ananiadou, S.; Tsujii, J. PathText: a text mining integrator for biological pathway visualizations. **Bioinformatics**, v. 26, p. i374–81, 2010.

Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M. et al. The human genome browser at UCSC. **Genome Research**, v. 12, p. 996-1006, 2002.

Kerrien, S. et al. IntAct - open source resource for molecular interaction data. **Nucleic Acids Research**, v. 35, p. D561–D565, 2007.

Ketting, R.F.; Plasterk, R.H. What's new about RNAi? Meeting on siRNAs and miRNAs. **EMBO Reports**, v. 5, n. 8, p. 762-765, 2004.

Kim, T.H.; Bohmer, M.; Hu, H.; Nishimura, N.; Schroeder, J.I. Guard cell signal transduction network: advances in understanding abscisic acid, CO2, and Ca2+ signaling. **Annual Review of Plant Biology**, v. 61, p. 561-591, 2010.

Kitano, H. Systems Biology: a brief overview. **Science**, v. 295, p. 1662-1664, 2002.

Krishnamurthy, A.; Rathinasabapathi, B. Novel interplay between auxin and reactive oxygen species signaling Oxidative stress tolerance in plants. **Plant Signaling and Behavior**, v. 8, n. 10, p. 1–5, 2013.

Ladunga, I. An overview of the computational analyses and discovery of transcription factor binding sites. **Methods in Molecular Biology**, v. 674, p. 1–22, 2010.

Lai, Z.; Li, Y.; Wang, F.; Cheng, Y.; Fan, B.; Yu, J.-Q.; Chen, Z. Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. **Plant Cell**, v. 23, p. 3824–3841, 2011.

Levine, M.; Tjian, R. Transcription regulation and animal diversity. **Nature**, v. 424, p. 147–151, 2003.

Libault, M.; Farmer, A.; Joshi, T. et al. An integrated transcriptome atlas of the crop model *Glycine max*, and tissues in comparative analyses in plants. **Plant Journal**, v. 63, n. 1, p. 86–99, 2010.

Lim, S.J.; Tan, T.W.; Tong, J.C. Bioinformation Computational Epigenetics : the new scientific paradigm. **Bioinformation**, v. 4, n. 7, p. 331-337, 2010.

Lin, C-C.; Chen, Y-J.; Chen, C-Y.; Oyang, Y-J.; Juan, H-F.; Huang, H-C. Crosstalk between transcription factors and microRNAs in human protein interaction network. **BMC Systems Biology**, v. 6, p.18, 2012.

Liu, B.; Li, J.; Cairns, M.J. Identifying miRNAs, targets and functions. **Briefings in Bioinformatics**, v. 15, p. 1–19, 2014.

Long, T.A.; Brady, S.M.; BenfeY, P.N. Systems approaches to identifying gene regulatory networks in plants. **Annual Review of Cell and Developmental Biology**, v. 24, p. 81–103, 2008.

Looger, L.L.; Dwyer, M.A.; Smith, J.J.; Hellinga, H.W. Computational design of receptor and sensor proteins with novel functions. **Nature**, v. 423, p. 185–190, 2003.

Lyer, V.R.; Horak, C.E.; Scafe, C.S.; Botstein, D.; Snyder, M.; Brown, P.O. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. **Nature**, v. 409, n. 6819, p. 533-538, 2001.

Ma, Y.; Szostkiewicz, I.; Korte, A.; Moes, D.; Yang, Y.; Christmann, A.; Grill, E. Regulators of PP2C phosphatase activity function as abscisic acid sensors. **Science**, v. 324, n. 5930, p. 1064-1068, 2009.

Maruyama, K.; Sakuma, Y.; Kasuga, M.; Ito, Y.; Seki, M.; Goda, H.; Shimada,Y.; Yoshida, S.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Identification of cold-inducible downstream genes of the Arabidopsis DREB1A/CBF3 transcriptional factor using two microarray systems. **Plant Journal**, v. 38, n. 6, p. 982-993, 2004.

Meola, N.; Gennarino, V.A.; Banfi, S. microRNAs and genetic diseases. **Pathogenetics**, v. 2, n. 1, p. 7, 2009.

Messerschmidt, D.M.; Knowles, B.B.; Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. **Genes Development**, v. 28, n. 8, p. 812-828, 2014.

Mestl, T.; Plahte, E.; Omholt, S.W. A mathematical framework for describing and analyzing gene regulatory networks. **Journal of Theorical Biology**, v. 176, p. 291-300, 1995.

Miller, G.; Suzuki, N.; Ciftci-Yilmaz, S.; Mittler, R. Reactive oxygen species homeostasis and signalling during drought and salinity stresses. **Plant Cell Environmental**, v. 33, p. 453–67, 2010.

Mittler, R. Abiotic stress, the field environment and stress combination. **Trends in Plant Science**, v. 11, p. 15–19, 2006.

Mittler, R.; Blumwald, E. Genetic engineering for modern agriculture: challenges and perspectives. **Annual Review of Plant Biology**, v. 61, p. 443–462, 2010.

Mittler, R.; Vanderauwera, S.; Suzuki, N.; Miller, G.; Tognetti, V.B.; Vandepoele, K.; Gollery, M.; Shulaev, V.; Van, B.F. ROS signaling: the new wave? **Trends in Plant Science**, v. 16, n. 6, p. 300-309, 2011.

Mochida, K.; Shinozaki, K. Genomics and Bioinformatics Resources for Crop Improvement. **Plant & Cell Physiology**, v. 51, n. 4, p. 497–523, 2010.

Mochida, K.; Yoshida, T.; Sakurai, T.; Yamaguchi-Shinozaki, K.; Shinozaki, K.; Tran, L.-S.P. LegumeTFDB: an integrative database of *Glycine max*, *Lotus japonicus* and *Medicago truncatula* transcription factors. **Bioinformatics**, v. 26, n. 2, p. 290–291, 2010.

Mohr, P.G.; Cahill, D.M. Abscisic acid influences the susceptibility of *Arabidopsis thaliana* to *Pseudomonas syringae* pv. tomato and *Peronospora parasitica*. **Functional Plant Biology**, v. 30, p. 461–469, 2003.

Molassiotis, A.; Fotopoulos, V. Oxidative and nitrosative signaling in plants: two branches in the same tree? **Plant Signaling and Behavior**, v. 6, n. 2, p. 210-214, 2011.

Mutwil, M.; Usadel, B.; Schütte, M.; Loraine, A.; Ebenhöh, O.; Persson, S. Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. **Plant Physiology**, v. 152, p. 29–43, 2010.

Nakamura, Y.; Cochrane, G.; Karsch-Mizrachi, I. ON BEHALF OF THE INTERNATIONAL NUCLEOTIDE SEQUENCE DATABASE COLLABORATION. The International Nucleotide Sequence Database Collaboration. **Nucleic Acids Research**, v. 41, n. D1, p. D21–D24, 2012.

Nakashima, K.; Ito, Y.; Yamaguchi-Shinozaki, K. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. **Plant physiology**, v. 149, n. 1, p.88-95, 2009.

Nascimento, L.C. et al. Bioinformatics Analysis Applied to GENOSOJA Project. In: 5th International Conference of the Brazilian Association for Bioinformatics and Computational Biology. **Anais da 5th International Conference of the Brazilian Association for Bioinformatics and Computational Biology**, Rio de Janeiro, 2009.

Nascimento, L.C.; Costa, G.G.L.; Binneck, E.; Pereira, G.A.G.; Carazzolle, M.F. A Web-Based Bioinformatics Interface Applied to the GENOSOJA Project: Databases and Pipelines. **Genetics and Molecular Biology**, v. 35, p. 203–11, 2012.

Needham, C.J.; Manfield, I.W.; Bulpitt, A.J.; Gilmartin, P.M.; Westhead, D.R. From gene expression to gene regulatory networks in *Arabidopsis thaliana*. **BMC System Biology**, v. 3, 85, 2009.

Newton, A.C.; Johnson, S.N.; Gregory, P.J. Implications of climate change for diseases, crop yields and food security. **Euphytica**, v. 179, p. 3–18, 2011.

Niinemets, U. Responses of forest trees to single and multiple environmental stresses from seedlings to mature plants: past stress history, stress interactions, tolerance and acclimation. **Forest Ecology and Management**, v. 260, p. 1623–1639, 2010.

Nogueira, R.M.C.; Araújo, E.L.; Willadino, L.G.; Cavalcante, U.M.T. Estresses ambientais: danos e benefícios em plantas. **Imprensa Universitária**, Recife, 2005.

Obayashi, T.; Nishida, K.; Kasahara, K.; Kinoshita, K. ATTED-II updates: Condition-specificgenecoexpression to extend coexpression analyses and applications to a broad range of flowering plants. **Plant Cell Physiology**, v. 52, p. 213–219, 2011.

Pal, R.; Datta, A.; Bittner, M.L; Dougherty, E.R. Intervention in context-sensitive probabilistic Boolean networks. **Bioinformatics**, v. 21, p. 1211-1218, 2005.

Pandey, G.K. Emergence of a novel calcium signaling pathway in plants: CBL-CIPK signaling network. **Physiology and Molecular Biology of Plants**, v. 14, p. 51- 68, 2008.

Park, S.Y.; Fung, P.; Nishimura, N.; Jensen, D.R.; Fujii, H.; Zhao, Y.; Lumba, S.; Santiago, J.; Rodrigues, A.; Chow, T.F.; Alfred, S.E.; Bonetta, D.; Finkelstein, R.; Provart, N.J.; Desveaux, D.; Rodriguez, P.L.; McCourt, P.; Zhu, J.K.; Schroeder, J.I.; Volkman, B.F. Cutler SR: Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. **Science**, v. 324, n. 5930, p. 1068-1071, 2009.

Phillips, T.; Hoopes, L. Transcription factors and transcriptional control in eukaryotic cells. **Nature Education**, v. 1, n. 1, p. 119, 2008.

Pimentel, C.; Abboud, A.C.S.; Roy-Macauly, H. et al. Leaf protoplasmic tolerance to drought, in two cowpea genotypes cultivated at the field. **Revista Universidade Rural**, Série Ciências da Vida, v. 22, n.1, p.07–14, 2002.

Pinto, M.C.; Gara, L. Changes in the ascor- bate metabolism of apoplastic and symplastic spaces are associated with cell differentiation. **Journal of Experimental Botany**, v. 55, n. 408, p. 2559–2569, 2004.

Popp, M.P.; Dillon, C.R.; Keisling, T.C. Economic and weather influences on soybean planting strategies on heavy soils. **Agricultural Systems**, v. 76, p. 969-984, 2003.

Preli, C.A.; Bleuler, S.; Zimmermann, P.; Wille, A.; Bühlmann, P.; Gruissem, W.; Hennig, L.; Thiele, L.; Zitzler, E. A systematic comparison and evaluation of biclustering methods for gene expression data. **Bioinformatics**, v. 22, p. 1122–1129, 2006.

Priyanka, B.; Sekhar, K.; Reddy, V.D.; Rao, K.V. Expression of pigeonpea hybrid-proline-rich protein encoding gene (CcHyPRP) in yeast and Arabidopsis affords multiple abiotic stress tolerance. **Plant Biotechnology Journal**, v. 8, p. 76–87, 2010.

Prosdocini, F.; Cerqueira, G.C.; Binneck, E.; Silva, A.F. et al. Bioinformática: manual do usuário. **Biotecnologia Ciência e Desenvolvimento**, v. 29, n. 12-25, 2002.

Pu, L.; Brady, S. Systems biology update: cell type-specific transcriptional regulatory networks. **Plant Physiology**, v. 152, n. 2, p. 411–419, 2010.

Pulverer, B. Sequence-specific DNA-binding transcription factors. **Nature Milestones**, 2005. doi: 10.1038/nrm1800

Remenyi, A.; Scholer, H.; Wilmanns, M. Combinatorial control of gene expression. **Nature Structural and Molecular Biology**, v. 11, p. 812–815, 2004.

Rhee, S.Y.; Dickerson, J.; Xu, D. Bioinformatics and Its Applications in Plant Biology. **Annual Review Plant Biology**, v. 57, p. 335–60, 2006.

Riechmann, J.L.; Heard, J.; Martin, G. et al. Arabidopsis transcription factors: genome wide comparative analysis among eukaryotes. **Science**, v. 290, p. 2105–2110, 2000.

Rizhsky, L.; Liang, H.J.; Shuman, J.; Shulaev, V.; Davletova, S.; Mittler, R. When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. **Plant Physiology,** v. 134, p. 1683–1696, 2004.

Rocha, M.M.; Lima, J.A.A.; Freire-Filho F.R.M. et al. Resistência de genótipos de caupi (*Vigna unguiculata* L. Walp.) de tegumento branco a isolados de vírus das famílias Bromoviridae, Comoviridae e Potyviridae. **Revista Ciência Rural**, v. 8, p. 85-92, 2003.

Rockstrom, J.; Falkenmark, M. Semiarid crop production from a hydrological perspective: gap between potential and actual yields. Critical Reviews in Plant Sciences, 19:319–346, 2000.

Rodriguez-llorente, I.; Caviedes, M.A.; Dary, M.; Palomares, A.J.; Cánovas, F.M.; Peregrín-Alvarez, J.M. The Symbiosis Interactome: A Computational Approach Reveals Novel Components, Functional Interactions and Modules in *Sinorhizobium Meliloti*. **BMC Systems Biology**, v. 3, n. 63, p. 1–18, 2009.

Saeys, Y.; Inza, I.; Larrañaga, P. A review of feature selection techniques in bioinformatics. **Bioinformatics**, v. 23, p. 2507–1257, 2007.

Sakuma, Y.; Maruyama, K.; Osakabe, Y.; Qin, F.; Seki, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Functional analysis of an Arabidopsis transcription factor, DREB2A, involved in drought-responsive gene expression. **Plant Cell**, v. 18, n. 5, p. 1292-1309, 2006.

Salwinski, L.; Miller, C.S.; Smith, A.J.; Pettit, F.K.; Bowie, J.U.; Eisenberg, D. The Database of Interacting Proteins: 2004 update. **Nucleic Acids Research**, v. 32, p. D449–D451, 2004.

Sauro, H.M.; Hucka, M.; Finney, A.; Wellock, C.; Bolouri, H.; Doyle, J.; Kitano, H. Next generation simulation tools: The Systems Biology Workbench and BioSPICE integration. **OMICS**, v. 7, p. 355–372, 2003.

Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J. et al. Genome sequence of the palaeopolyploid soybean. **Nature**, v. 463, n. 7278, p. 178–183, 2010.

SEAB, Secretaria de Estado de Agricultura e Abastecimento. Soja: análise da conjuntura agropecuária. 2013. Disponível em: <http://www.agricultura.pr.gov.br/>. Acesso em: 05 ago. 2014.

Seki, M.; Narusaka, M.; Ishida, J.; Nanjo, T.; Fujita, M.; Oono, Y.; Kamiya, A.; Nakajima, M.; Enju, A.; Sakurai, T.; Satou, M.; Akiyama, K.; Taji, T.; Yamaguchi-Shinozaki, K.; Carninci, P.; Kawai, J.; Hayashizaki, Y.; Shinozaki, K. Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold, and high-salinity stresses using a full-length cDNA microarray. **Plant Journal**, v. 31, p. 279-292, 2002.

Sharma, P.; Dubey, R.S.; Pessarakli, M. Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. **Journal of Botany**, ID: 217037, 2012.

Sharma P, Jha AB, Dubey RS (2010) Oxidative stress and antioxidative defense systemin plants growing under abiotic Stresses. In: Handbook of Plant and Crop Stress, Pessarakli M, Eds. 89–138, **CRC Press**, Taylor and Francis Publishing Company, Fla, USA, 3rd edition.

Sharma, R.C.; Duveiller, E.; Ortiz-Ferrara, G. Progress and challenge towards reducing wheat spot blotch threat in the Eastern Gangetic Plains of South Asia: is climate change already taking its toll? **Field Crops Research**, v. 103, p. 109–118, 2007.

Shinozaki, K.; Yamaguchi-Shinozaki, K. Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. **Current Opinion of Plant Biology**, v. 3, p. 217-223, 2000.

Shivashankar, H.N.; Gasser, R.B.; Ranganathan, S. A hitchhiker's guide to expressed sequence tag (EST) analysis. **Briefings in Bioinformatics**, v.8, n.1, p.6-21, 2006.

Sproul, D.; Gilbert, N.; Bickmore, W.A. The role of chromatin structure in regulating the expression of clustered genes. **Nature Reviews Genetics**, v. 6, p. 775–781, 2005.

Stark, C.; Breitkreutz, B.J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M. BioGRID: a general repository for interaction datasets. **Nucleic Acids Research**, v. 34, p. D535–D539, 2006.

Struhl, K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. **Cell**, v. 98, p. 1–4, 1999.

Sun, L.; Zhang, M.; Ren, J.; Qi, J.; Zhang, G.; Leng, P. Reciprocity between abscisic acid and ethylene at the onset of berry ripening and after harvest. **BMC Plant Biology**, v. 10, p. 257, 2010.

Swindell, W.R. The Association Among Gene Expression Responses to Nine Abiotic Stress Treatments in *Arabidopsis thaliana*. **Genetics**, v. 174, n. 4, p. 1811–1824, 2006.

Szklarczyk, D.; Franceschini, A.; Kuhn, M.; Simonovic, M.; Roth, A.; Minguez, P.; Doerks, T.; Stark, M.; Muller, J.; Bork, P.; Jensen, L.J.; von Mering, C. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. **Nucleic Acids Research**, v. 39, p. D561– D568, 2011.

Takahashi, S.; Seki, M.; Ishida, J.; Satou, M.; Sakurai, T.; Narusaka, M.; Kamiya, A.; Nakajima, M.; Enju, A.; Akiyama, K.; Yamaguchi-Shinozaki, K.; Shinozaki, K. Monitoring the expression profiles of genes induced by hyperosmotic, high salinity, and oxidative stress and abscisic acid treatment in Arabidopsis cell culture using a full-length cDNA microarray. **Plant Mol Biology**, v. 56, n. 1, p. 29-55, 2004.

Takeda, S.; Matsuoka, M. Genetic approaches to crop improvement: responding to environmental and population changes. **Nature Reviews Genetics**, v. 9, p. 444–457, 2008.

Talamè, V.; Ozturk, N. Z.; Bohnert, H.J. et al. Barley transcript proles under dehydration shock and drought stress treatments: a comparative analysis. **Journal of Experimental Botany**, v. 58, n. 2, p. 229–240, 2007.

Tang, W.Y.; Ho, S.M. Epigenetic reprogramming and imprinting in origins of disease. **Reviews in Endocrine and Metabolic Disorders**, v. 8, p. 173-82, 2007.

Teófilo, E.M.; Paiva, J.B.; Filho, S.M. Polinização Artificial em Feijão Caupi (Vigna unguiculata (L.) Walp). **Ciência e Agrotecnologia**, v. 25, p. 220-223, 2001.

THE UNIPROT CONSORTIUM. Update on activities at the Universal Protein Resource (UniProt) in 2013. **Nucleic Acids Research**, v. 41, n. D1, p. D43– D47, 2012.

Toufighi, K.; Brady, S.M.; Austin, R.; Ly, E.; Provart, N.J. The Botany array resource: e-Northerns, expression angling, and promoter analyses. **Plant Journal**, v. 43, p. 153–163, 2005.

Trupti, J.; Kapil, P.; Michael, R. et al., Soybean knowledge base (SoyKB): a web resource for soybean translational genomics. **BMC Genomics**, v. 13, p. S15, 2012.

Udvardi, M.K.; Kakar, K.; Wandrey, M. et al. Legume transcription factors: global regulators of plant development and response to the environment. **Plant Physiology**, v. 144, p. 538-549, 2007.

UNITED STATES DEPARTMENT OF AGRICULTURE. World agricultural supply and demand estimates. 2013. Disponível em: <http://www.usda.gov/oce/commodity/wasde/latest.pdf>. Acesso em: 05 ago. 2014.

Usadel, B.; Obayashi, T.; Mutwil, M.; Giorgi, F.M.; Bassel, G.W.; Tanimoto, M.; Chow, A.; Steinhauser, D.; Persson, S.; Provart, N.J. Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. **Plant Cell Environmental**, v. 32, p. 1633–1651, 2009.

Van Bel, M.; Proost, S.; Wischnitzki, E.; Movahedi, S.; Scheerlinck, C.; Van De Peer, Y.; Vandepoele, K. Dissecting Plant Genomes with the PLAZA Comparative Genomics Platform. **Plant Physiology**, v. 158, n. 2, p. 590–600, 2011.

Verdier, J.; Lalanne, D.; Pelletier, S.; Torres-jerez, I.; Righetti, K.; Bandyopadhyay, K.; Leprince, O. et al. A Regulatory Network-Based Approach Dissects Late Maturation Processes Related to the Acquisition of Desiccation Tolerance and Longevity of Medicago. **Plant Physiology**, v. 163, p. 757–74, 2013.

Wang, W.; Vinocur, B.; Altman, A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. **Planta**, v.218, p.01–14, 2003.

Wang, W.X.; Vinocur, B.; Shoseyov, O.; Altman, A. Biotechnology of plant osmotic stress tolerance: physiological and molecular considerations. **Acta Horticulturae**, v. 560, p. 285-292, 2001.

Wang, Y.; Bao, Z.L.; Zhu, Y.; Hua, J. Analysis of temperature modulation of plant defense against biotrophic microbes. **Molecular Plant-Microbe Interactions**, v. 22, p. 498–506, 2009.

Wang, Z.; Libault, M.; Joshi, T.; Valliyodan, B.; Nguyen, H.T.; Xu, D.; Stacey, G.; Cheng, J. SoyDB: a knowledge database of soybean transcription factors. **BMC Plant Biology**, v. 10, p. 14, 2010.

Warde-Farley, D. et al. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. **Nucleic Acids Research**, v. 38(Web Server issue): W214– W220, 2010.

Watson, James D. et al. Biologia molecular do gene. 5. ed. Porto Alegre: **Artmed**, 2006. xxxii, 728 p. + 1 CD-ROM ISBN 9788536306841 (enc.).

Weinhofer, I.; Hehenberger, E.; Roszak, P.; Hennig, L.; Köhler, C. H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. **PLoS Genetics**, v. 6, p. e1001152, 2010.

Wiese, J.; Kranz, T.; Schubert, S. Induction of pathogen resistance in barley by abiotic stress. **Plant Biology**, v. 6, p. 529–536, 2004.

Wilkinson, S.; Davies, W.J. Drought, ozone, ABA and ethylene: new insights from cell to plant to community. **Plant Cell Environ**, v. 33, n. 4, p. 510-525 2010.

Xiong, M.; Li, J.; Fang, X. Identification of genetic networks. **Genetics**. v. 166, p. 1037-1052, 2004.

Xu, X.; Wang, L.; Ding, D. Learning module networks from genome-wide location and expression data. **FEBS Letter**, v. 578, p. 297-304, 2004.

Yamaguchi, R.; Yoshida, R.; Imoto, S.; Higuchi, T.; Miyano, S. Finding module-based gene networks in time-course gene expression data with state space models. **IEEE Signal processing magazine**, v. 24, p. 37-46, 2007.

Yamaguchi-Shinozaki, K.; Shinozaki, K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. **Annual Review of Plant Biology**, v. 57, p. 781-803, 2006.

Yao, C.W.; Hsu, B.D.; Chen, B.S. Constructing gene regulatory networks for long term photosynthetic light acclimation in *Arabidopsis thaliana*. **BMC Bioinformatics**, v. 12, p. 335, 2011.

Yoo, S.D.; Cho, Y.; Sheen, J. Emerging connections in the ethylene signaling network. **Trends of Plant Science**, v. 14, n. 5, p. 270-279, 2009.

Yoshida, T.; Fujita, Y.; Sayama, H.; Maruyama, K.; Mizoi, J.; Shinozaki, K.; Yamaguchi-Shinozaki, K. AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. **Plant Journal**, v. 61, n. 4, p. 672-685, 2010.

Zabala, G.; Zou, J.; Tuteja, J.; Gonzalez, D.O.; Clough, S.J.; Vodkin, L.O. Transcriptome changes in the phenylpropanoid pathway of *Glycine max* in response to *Pseudomonas syringae* infection. **BMC Plant Biology**, v.6, article 26, 2006.

Zhu, J.K. Salt and drought stress signal transduction in plants. **Annual Review of Plant Biology**, v. 53, p. 247-273, 2002.

Zhu, M.; Dahmen, J.L.; Stacey, G.; Cheng, J. Predicting Gene Regulatory Networks of Soybean Nodulation from RNA-Seq Transcriptome Data. **BMC Bioinformatics**, v. 14, n. 1, 2013.

Zhu, M.; Xin, D.; Trupti, J.; Dong, X.; Stacey, G.; Cheng, J. Reconstructing Differentially Co-Expressed Gene Modules and Regulatory Networks of Soybean Cells. **BMC Genomics**, v. 13, n. 1, 2012.

Capítulo de livro aceito:

Dr. David J. Burritt, Dr. Md. Anwar Hossain, Dr. Shabir Hussain Wani and Dr. Soumen Bhattacharjee (Editors). "Drought Tolerance in Plants, Vol 2: Molecular and Genetic Perspectives". Springer International Publishing, Heidelberg, Germany (2016).

Drought Tolerance in Plants, Vol 2: Molecular and Genetic Perspectives: Transcription factors involved in plant drought tolerance regulation

Transcription factors involved in plant drought tolerance regulation

Lidiane L. Barbosa Amorim[1,2], João Pacífico Bezerra-Neto[1], Rômulo da Fonseca do Santos[1], José Ribamar Ferreira da Costa Neto[1], Ederson Akio Kido[1], Mitalle Matos[1] & Ana Maria Benko-Iseppon[1]

[1]Department of Genetics, Laboratory of Plant Genetics and Biotechnology, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego, 1235¸50.670-423, Recife, PE, Brazil
[2]Federal Institute of Education, Science and Technology of Piauí, Rua Projetada, s/n, CEP 64500-000, Oeiras, PI, Brazil

*Corresponding Author's e-mail: ana.iseppon@gmail.com

Abstract

Drought stress induces the expression of a significant number of transcription factor (TF) genes that are involved in stress tolerance and response. Nowadays, novel "omics" and next generation sequencing provides a global picture of drought-stress responses and a comprehensive understanding of TF regulatory networks that modulate this response. Several TFs have been shown to be involved in the regulation of stress-responsive genes under drought condition, including *Arabidopsis thaliana*, rice (*Oryza sativa*, Poaceae), soybean (*Glycine max*, Fabaceae) and other plants. Most data have been obtained from transcriptome analyzes, studies using small RNA molecules, chromatin modulation, and genomic DNA modifications. Based on recent advances in genomic technologies, improvements in crop drought stress tolerance have been attained from marker-assisted breeding and gene transfer. In this chapter, we summarize recent progress on the role of TFs in response to drought in plants. Studies on crops have shown that facing drought requires ET and ABA signaling pathways through selectively interacting with TFs, once again highlighting a close relationship between different signaling pathways referred in this chapter.

Keywords: Drought Tolerance, Plant Transcription Factor, Protein-protein Interaction, Epigenetics.

Contents

1 Introduction

The complex genetic basis related to drought tolerance requires a deeper understanding of molecular and physiological mechanisms, also concerning regulation of networks involved in related responses such as water deficit, osmotic, oxidative and heat stress. The recognition of these mechanisms is fundamental for the development of adapted cultivars to each type of abiotic stress (Vicente-Serrano et al. 2013; Rabara et al. 2014). When exposed to drought, plants induce a series of changes at physiological, biochemical and molecular level. Such modifications may affect cell viability due to the production of reactive oxygen species (ROS) responsible for the oxidation of multicellular components, such as proteins, lipids and nucleic acids (Jaspers e Kangasjärvi, 2010; Zhang et al. 2014).

Drought is also responsible for inhibition of respiration, stomatal closure, stimulation of root growth, changes in photosynthesis and assimilation of nutrients, among others (Lata et al. 2011; Pinheiro e Chaves, 2011; Siddiqui et al. 2015). In addition to the mentioned effects, water deficit stimulates the production of abscisic acid (ABA) a central regulator of the stress response, which confers drought tolerance, increasing expression of many stress-responsive genes by inducing signaling cascades (Cutler et al. 2010; Kim et al. 2015).

In addition to regulatory events mediated by ABA, other groups of genes activated by water deficit confer tolerance to the plant, including transcription factors (TFs) that play a role as regulators and central molecular switches in the control of gene expression. Thus, any change in TF activity would result in modulation of a genetic network of the plant, making them potential targets for induction of adaptive responses to abiotic stresses such as drought (Tripathi et al. 2014).

2 Response to different abiotic stresses share common responsive TFs

Major abiotic stresses – including high salinity, drought, cold, and heat – influence the entire plant metabolism and adversely affect growth, development, fertility and productivity of crops (Krasensky and Jonak 2012). After detection of stress, plants stimulate signal transduction by various signaling pathways inducing a range of physiological and biochemical responses at cellular and whole-organism levels to enable them to maintain homeostasis against these stresses (Hirayma and Shinozaki 2010; Krasensky and Jonak 2012). Many protein kinases participate in abiotic stress signal transduction, including mitogen-activated protein kinases

(MAPKs), calmodulin-dependent protein kinases (CDPKs), receptor protein kinases (RPKs), and ribosomal protein kinases (Shu et al. 2015).

Membrane disorganization, protein denaturation, metabolic toxicity, repression of cell growth and photosynthesis are some of the consequences of abiotic stress on plants. Also, plants alter their metabolism by the accumulation of osmolytes and generation of second messengers such as calcium and ROS (Farooq et al. 2009; Pinheiro and Chaves 2011). Modulation of intracellular Ca^{2+} levels is sensed by calcium-binding protein, initiating a protein phosphorylation cascade that targets TFs, which convey these signaling cues and activate or repress expression of stress-inducible genes and enhances stress tolerance (Xiong et al. 2002; Harrison 2012). Stress-induced changes may participate in the generation of phytohormones including abscisic acid (ABA), ethylene (ET), jasmonic acid (JA), and salicylic acid (SA). Various interactions can take place between phytohormone and TFs to amplify the initial signal and induce a second round of signaling that can follow the same pathway or use altogether different signaling pathway components (Atkison and Urwin 2012).

The ultimate goal of signal transduction pathways is to elicit a specific well-timed biological response and to modify gene transcription. Many TF-families have been suggested to play a major role for transcriptional reprogramming associated with plant stress response (Nuruzzaman et al. 2013). These TF genes are capable of controlling the expression of a broad range of elements through sequence-specific interactions with *cis*-regulatory DNA elements in the promoters of target stress-related genes containing TF binding sites, specifically activating or repressing the expression of these target genes, directing the expression in a synchronized manner (Gordan et al. 2011).

Among all described phytohormones, ABA is an important player in the regulation of responses during abiotic stresses, especially regarding drought and salt stresses. It promotes rapid changes in rates of transcription, transcript processing, and stability, also modifying conformational states of regulatory molecules that control RNA processing like TFs (Cutler et al. 2010).

Molecular analysis of promoters of ABA-responsive genes led to the identification of a conserved *cis*-acting element, designated as ABRE (ABA-responsive element) (Choi et al. 2000; Zhu et al. 2002). ABREs can be recognized by ABRE-binding proteins (AREBs) or ABRE-binding factors (ABFs), which belongs to group A of basic leucine zipper (bZIP) TFs (Uno et al. 2000). In *Arabidopsis*, approximately 75 distinct bZIP-type TFs have been identified (Jakoby et al. 2002). Recent genome-wide analysis based on RNA-Seq, EST (expressed sequence tag) and microarray analyses have revealed different numbers of bZIP

superfamily members in plants, with 125 representatives in maize (Wei et al. 2012), 45 in castor bean (Jin et al. 2014), 55 in grapevine (Liu et al. 2014a), and 64 in cucumber (Baloglu et al. 2014). Hence, TFs are important regulators that control gene expression under abiotic stress. Among them, NAC (No Apical Meristem, ATAF1/2), AP2/ERF (Apetala2/ethylene response factor), MYB, C2H2 zinc finger, and WRKY superfamilies (Hirayama et al. 2010; Nakashima et al. 2014; Osakabe et al. 2014; Krannich et al. 2015) deserve mentioning. This section exemplarily describes functional aspects of some important TF families with a functional role in plant response to abiotic stress (Table 1).

2.1 bZIP-type transcription factor Family

bZIP proteins can be identified based on the presence of a bZIP domain, characterized by 60 to 80 amino acids in length, a basic region (BR) and a leucine zipper (LZ) that is functionally distinct (Hurst 1994). In the model plant *A. thaliana* bZIP family is divided into ten groups (A, B, C, D, E, F, G, H, I and S), considering amino acid sequence similarities of bZIP domains and protein structure. For *Arabidopsis,* it has been proposed that members of A, B and S groups play crucial roles in ABA-signaling and abiotic stress response (Jakoby et al. 2002; Lindemose et al. 2013).

AREB/ABFs belongs to the A-group bZIP comprising nine homologs in *Arabidopsis*, and can bind to ABRE, a *cis*-regulatory element found in promoters of many ABA- and stress-responsive genes. This element plays a pivotal role in the regulation of abiotic stress response including drought, salt, and cold (Fujita et al. 2011; 2013). Among the nine members of AREB/ABF identified in *Arabidopsis,* AREB1/ABF2, AREB2/ABF4 and ABF3 are induced by ABA treatment during vegetative growth, dehydration and high salinity, (Choi et al. 2000; Uno et al. 2000). In *Arabidopsis* overexpression of ABF3 or ABF4 resulted in ABA hypersensitivity and enhanced drought tolerance with changes in expression levels of ABA- and other stress-regulated genes.

Fujita et al. (2005) have reported that overexpression of *AREB1* in transgenic *Arabidopsis* plants leads to ABA hypersensitivity and enhanced drought tolerance while overexpression of *ABF2* increased plant tolerance to high salinity, drought, heat and oxidative stresses (Kim et al. 2004; Li et al. 2013). AtbZIP37 (ABF3) and AtbZIP38 (ABF4/AREB2) were up-regulated in response to ABA signal, dehydration, and salinity. Overexpression of ABF3 and ABF4 also resulted in ABA hypersensitivity and several other ABA/stress-associated phenotypes, including enhanced drought tolerance (Kang et al. 2002; Oh et al. 2005). Similarly, overexpression of *PtABF* gene of *Poncirus trifoliata (Rutaceae)* enhanced dehydration and drought tolerance in tobacco (Huang et al. 2010).

Table 1. A list of transcription factors (TFs) described in this paper.

Gene function	Organisms	Gene name	Inducers	Function in abiotic stress	Reference
bZIP	*Arabidopsis thaliana*	AREB1/ABF2	ABA, drought, salt stress	Regulator of ABA signaling under drought stress	Fujita et al. 2005
bZIP	*Arabidopsis thaliana*	AtbZIP37 (ABF3)	ABA, drought, salt stress	Regulator of ABA signaling under drought stress	Kang et al. 2002
bZIP	*Arabidopsis thaliana*	AtbZIP38 (ABF4/AREB2)	ABA, drought, cold, salt stress	Mediates drought stress response at the molecular level	Kang et al. 2002
bZIP	*Poncirus trifoliata*	PtABF	ABA, drought, cold	Regulation of expression of genes involved in stress response and tolerance	Huang et al. 2010
bZIP	*Oryza sativa*	TRAB1	ABA, drought, salt stress	Regulation of gene expression via phosphotylation in response to ABA	Hobo et al. 1999; Kagaya et al. 2002
bZIP	*Oryza sativa*	OsABI5	ABA, salt stress	Regulation of adaptive stress response and plant fertility	Zou et al. 2008
bZIP	*Oryza sativa*	OsbZIP23	Drought, salt	Regulation of expression of genes involved in stress response and tolerance	Xiang et al. 2008
bZIP	*Oryza sativa*	OsbZIP72	Drought, ABA	Regulation of expression of ABA-stress-related genes and enhanced drought tolerance	Lu et al. 2009
bZIP	*Oryza sativa*	OsbZIP46	Drought, ABA and heat	Regulation of expression of ABA-stress-related genes	Tang et al. 2012
bZIP	*Tamarix hispida*	ThbZIP1	ABA, drought, salt stress	Mediates ABA and salt stress response at the molecular level	Ji et al. 2013
bZIP	*Capsicum annuum*	CaBZ1	Drought, salt, and ABA	Increased the expression of ABA, regulation of stress-related genes	Moon et al. 2015
bZIP	*Glycine max*	GmbZIP1	Drought, salt, cold, ABA, low temperature	Regulation of expression of ABA-stress-related genes	Gao et al. 2011
bZIP	*Vitis vinífera*	VvbZIP23	Drought, salt, cold, ABA	Regulation of genes involved in abiotic response	Tak and Mhatre 2013
bZIP	*Vitis vinífera*	VvbZIP45	Drought, ABA, heat	Not reported	Liu et al. 2014b Nicolas et al. 2014
bZIP	*Solanum lycopersicum*	SIAREB1	Drought, salt	Regulation of genes involved in abiotic response	Orellana et al. 2010
bZIP	*Solanum lycopersicum*	SIAREB2	Drought, salt	Regulation of genes involved in abiotic response	Orellana et al. 2010
WRKY	*Arabidopsis thaliana*	AtWRKY33	Salt, cold, oxidative stress and UV radiation	Control of cellular ROS levels in abiotic stress signaling	Jiand and Deyholos 2013
WRKY	*Arabidopsis thaliana*	AtWRKY25	Cold an heat	Regulation of heat inducible genes	Jiand and Deyholos 2013
WRKY	*Arabidopsis thaliana*	AtWRKY39	Heat	Regulator of SA-dependent heat stress defense pathways	Li et al. 2010
WRKY	*Arabidopsis thaliana*	AtWRKY36	ABA and drought	Regulation of expression of ABA-stress-related genes	Ren et al. 2010
WRKY	*Oryza sativa*	OsWRKY45	ABA, drought, salt	Regulation of genes involved in abiotic response	Ren et al. 2010
WRKY	*Oryza sativa*	OsWRKY11	Heat and drought	Regulation of genes involved in desiccation tolerance	Wu et al. 2009

WRKY	*Glycine max*	GmWRKY20	ABA, salt, drought, cold	Mediates ABA signaling and regulates the expression of wax biosynthetic genes	Lou et al. 2013
WRKY	*Glycine max*	GmWRKY54	Salt, drought	Regulation of genes involved in abiotic response	Zhou et al. 2008
WRKY	*Gossypium hirsutum*	GhWRKY25	Salt, drought, Cold, ABA, AS	Regulation of the expression of ROS-related genes	Liu et al. 2015
WRKY	*Gossypium hirsutum*	GhWRKY17	Salt, drought, H_2O_2	Regulation of expression of ABA-stress-related genes	Yan et al. 2014
WRKY	*Triticum aestivum*	TaWRKY10	Salt, cold, H_2O_2	Regulation of osmotic balance, ROS scavenging and transcription of stress related genes	Wang et al. 2013
WRKY	*Triticum aestivum*	TaWRKY44	Salt, cold, drought, H_2O_2	ROS elimination and activation of stressed-associated genes	Wang et al. 2015
WRKY	*Dendrathema grandiflorum*	DgWRKY3	Salt, drought	Regulation of genes involved in osmotic adjustment, membrane protection, and oxidative stress response	Liu et al. 2013
WRKY	*Tamarix hispida*	ThWRKY14	ABA, salt, drought	Regulation of genes involved in abiotic response	Zheng et al. 2013
NAC	*Arabidopsis thaliana*	ANAC019, ANAC055, ANAC072	ABA, drought, salt	Regulation of genes involved in abiotic response	Tran et al 2003
NAC	*Oryza sativa*	OsNAC6/SNAC2	Salt, cold, drought	Regulation of abiotic stress responsive gene expression	Nakashima et al. 2007 Hu et al. 2008
NAC	*Oryza sativa*	OsNAC5	Salt, cold, drought	Regulation of OsLEA3 expression	Takasaki et al. 2010
NAC	*Oryza sativa*	OsNAC10	Salt, cold, drought	Enhancement of grain yield under drought conditions	Jeong et al. 2010
NAC	*Oryza sativa*	ONAC063	Salt, cold	Regulation of abiotic stress responsive gene expression	Yokotani et al. 2009
NAC	*Oryza sativa*	SNAC1	Salt, cold, drought, ABA	Increase of stomatal closure and expression of genes involved in stress tolerance	Hu et al. 2006 Liu et al.2014
NAC	*Glycine max*	GmNAC2	Salt, cold, drought	Downregulation of the expression. of ROS-related genes	Pinheiro et al. 2009 Tran et al. 2009

Several members of rice bZIP TFs were able to withstand various abiotic stresses with the ability to improve drought and salt tolerance in transgenic rice. The first member of rice is *TRAB1*, a TF-induced by ABA under drought and salt stress. TRAB1 can be activated via ABA-dependent phosphorylation (Hobo et al. 1999). The second member identified was *OsABI5*, *a homolog of TRAB1 that* encodes a protein that can bind to ABRE and was induced by ABA and high salinity, being down-regulated in seedlings grown under drought and cold conditions. Overexpression of *OsABI5* in rice provided high sensitivity to salinity stress (Zhou

et al. 2007; 2008). Expression of OsbZIP23 also was induced by drought, salinity and ABA treatment. Transgenic rice overexpression OsbZIP23 exhibited ABA hypersensitivity and improved tolerance to drought and salinity (Xiang et al. 2008). Overexpression of another member of bZIP, OsbZIP72, also exhibited hypersensitivity to ABA and led to drought tolerance in *transgenic* rice plants (Lu et al. 2009), while OsbZIP46 increased ABA sensitivity but had no positive effect on drought resistance (Tang et al. 2012). Recently, Liu et al. (2014b) showed that OsbZIP71 was strongly induced by drought and ABA treatments and repressed under salt stress. Real-time PCR analysis (RT-qPCR) revealed that abiotic stress-related genes were up-regulated in overexpressing plants. Thus, these reports indicate that OsbZIP factors have significant roles in the control of abiotic stress tolerance, with potential biotechnological applications for improving abiotic stress tolerance of plants.

Similar results were obtained with transgenic *Arabidopsis* plants overexpressing ThbZIP1 from *Tamarix hispida* (Tamaricaceae). The overexpression led to significant tolerance to drought and salt stress, but plants were sensitive to ABA treatment. Microarray analysis showed that many ROS scavenging genes were up-regulated by *ThbZIP1* under salt stress conditions (Ji et al. 2013). Also, Zhang et al. (2012) showed that ABP9, a maize *bZIP* gene, conferred drought, salt, and cold tolerance in transgenic plants, but plants overexpressing *ABP9* also displayed significantly sensitivity to ABA. Transgenic plants overexpressing *ABP9* presented improved salt stress tolerance by the modulation of ROS levels. In addition, both *ABP9* and *ThbZIP1* transformed plants showed improved water-retention capacity (Zhang et al. 2012; Ji et al. 2013).

A novel bZIP TF from hot pepper, CaBZ1, presented homology to other bZIP genes of *Arabidopsis* group S, and its expression was strongly induced by multiple stress stimuli, such as ABA, cold and salinity (Moon et al. 2015). Overexpression of CaBZ1 reduced water loss, increased ABA-induced stomatal closure, and altered expression levels of stress-inducible and TF genes, such as NAC and CBF (Moon et al. 2015). Consistently, a novel bZIP from soybean named GmbIZP1 also conferred stress tolerance not only to *drought* but also to salt and cold stresses (Gao et al. 2011).

In grape (*Vitis vinifera*, Vitaceae), a bZIP member named VvbZIP23 was isolated from cultivar ‘Mangoo’ and was identified as an important plant regulator of abiotic stress responses. Its expression was found to be strongly induced by a wide spectrum of abiotic stresses, including drought, salt, cold, and application of abscisic, salicylic, and jasmonic acid (Tak and Mhatre 2013). Some evidences indicated that VvbZIP45 transcripts were also involved in

abiotic stress response with upregulation by ABA, including drought stress (Liu et al. 2014a; Nicolas et al. 2014).

In cultivated tomato (*Solanum lycopersicum*, Solanaceae) the expression of two AREB/ABFs proteins (SlAREB1 and SlAREB2) was induced by both drought and salinity, although expression of SlAREB1 was strongly affected. Tomato mutants overexpressing SlAREB1 showed increased tolerance to salinity and drought stresses compared with wild-type plants. Notably, microarray analyzes revealed that many defense genes associated with abiotic and biotic stress were upregulated (Orellana et al. 2010). More importantly, the results suggested that SlAREB1 TF was involved in ABA response to abiotic stress and possibly also in response to pathogens during plant defense, mediating crosstalk between abiotic and biotic responses in tomato plants, similar to CAbZIP1 from pepper expressed in *Arabidopsis* (Lee et al. 2006).

Considering the multiple functions of bZIP TFs in abiotic stress response these potential candidate genes have received far less attention for application in the improvement of drought tolerance in crops. Such an approach would be interesting towards a better understanding of the network regulation associated with bZIP TFs.

2.2 WRKY TFs

WRKY family is defined by the presence of highly conserved WRKY-DBD of 60 amino acids in length, which contains the almost invariant WRKYGQK sequence motif at the N-terminal and a zinc-binding motif with features of C-C-H-H (C-X_{4-5}-C-X_{22-23}-H-X-H) or C-C-H-C (C-X_7-C-X_{23}-H-X-C) at the C-terminal region (Eulgem et al. 2000). WRKY proteins have been found to modulate gene expression in plants under biotic and abiotic stresses. Besides roles in response to such stresses, WRKY proteins are involved in a whole range of physiological processes that have profound effects on growth and development, such as senescence, dormancy, morphogenesis of trichomes and embryos, and metabolism (Rushton et al. 2010). WRKY acts via protein-protein interaction and even cross- and autoregulation (Pan and Jiang 2014).

Specific WRKY TFs, which help in the expression of a cluster of stress-responsive genes, are being targeted (Banerjee and Roychoudhury 2015). For example, in *Arabidopsis* AtWRKY33 downstream target genes with functions in detoxification of ROS (such as glutathione *S*-transferase *GSTU11*, peroxidases, and lipoxygenase *LOX1*), increasing salinity tolerance (Jiang and Deyholos 2009). AtWRKY33 also affected tolerance to heat through

modulation of transcriptional reprogramming (Li et al. 2011), indicating a crossover of regulatory roles within various stress responses. Cross-regulation among AtWRKY25, AtWRKY26 and AtWRKY33 is essential in promoting tolerance against high-temperature stress by positively regulation of the cooperation between the heat-shock protein and ethylene-activated signaling pathways (Li et al. 2011). Furthermore, AtWRKY39 is heat stress-induced and acts as a positive regulator of SA-dependent heat stress defense pathways (Li et al. 2010). Another good example in rice is *OsWRKY11* gene, whose overexpression under the control of heat shock protein HSP101 promoter led to enhanced heat and drought tolerance in transgenic rice (Wu et al. 2009).

WRKY genes may mediate crosstalk between plant abiotic tolerance and ABA-related signaling. For instance, overexpression of GhWRKY17 from cotton (*Gossypium hirsutum*, Malvaceae) in *Nicotiana benthamiana* (Solanaceae) responded to drought and salt stress through ABA signaling (Yan et al. 2014). Transcription levels of ABA-inducible genes, including AREB, DREB, and LEA, were repressed under drought and salt stress conditions (Yan et al. 2014). Similar results were obtained for *T. hispida* after overexpression of ThWRKY4, conferring tolerance to salt stress. Expression of ThWRKY4 alone was not sufficient to activate some stress related genes (bZIP, DOF, bHLH), but needed to be activated by ABA signals (Zheng et al. 2013). Results showed that *OsWRKY45* also play a role in ABA signaling and drought tolerance in rice (Tao et al. 2011). Moreover, overexpression of OsWRKY45 resulted in constitutive expression of ABA-induced responses and abiotic-related stress factors, also markedly enhancing drought resistance (Qiu and Yu 2009). Interestingly, AtWRKY63 (ABO3) mutants exhibited ABA-induced stomatal closure resulting in lower drought tolerance and higher ABA sensitivity (Ren et al. 2010). Consistent with these findings, expression of wild soybean GmWRKY20 gene in *Arabidopsis* enhanced drought tolerance and also regulated ABA signaling. GmWRKY20 overexpression lines were more sensitive to ABA during stomatal closure when compared with the wild type (Luo et al. 2013). On the other hand, transgenic plants overexpressing *GmWRKY13* exhibited decreased sensitivity to ABA, and exhibited less tolerance towards high salt and mannitol treatment in comparison to the wild types (Zhou et al. 2008). Taken together, these data demonstrate that WRKY TFs play key roles in ABA-dependent pathway and drought-responsive signaling networks (Luo et al. 2013).

Involvement of some soybean WRKY genes with salt and cold tolerance was experimentally demonstrated. For example, *GmWRKY21* transgenic plants were tolerant to cold stress, whereas overexpression of GmWRKY54 in *Arabidopsis* conferred drought and salt tolerance (Chen et al. 2007; Zhou et al. 2008). High levels of GmWRKY54 are thought to

induce expression of *salt tolerance Zn finger* (*STZ/Zat10*) and *DREB2A* (Zhou et al. 2008). In *Arabidopsis* AtWRKY8 was implicated in the modulation of salinity stress and shows high up-regulation upon NaCl treatment (Hu et al. 2013) while overexpression of AtWRKY18 and AtWRKY60 increased plant sensitivity to salt and osmotic stresses (Chen et al. 2010). GhWRKY25 from cotton also lead to improved tolerance to salt stress, but reduced plant tolerance to drought stress (Liu et al. 2015), whereas a reverse phenomenon was reported for DgWRKY3 from chrysanthemum (*Dendranthema grandiflorum*, Asteraceae). Heterologous expression of DgWRKY3 in tobacco revealed that transgene expression was lower under drought when compared to salinity stress (Liu et al. 2013).

In wheat, WRKY confers multiple abiotic stress tolerance. Exposure to salt, cold, drought, and H_2O_2 induced the expression of TaWRKY10, TaWRKY2, TaWRKY19, and TaWRKY44. Overexpression of TaWRKY10 and TaWRKY44 enhanced drought and salt tolerance by direct or indirect activation of stress-related genes, ROS scavenging, and osmotic balance (Wang et al. 2013). TaWRKY19 and TaWRKY2 overexpression conferred higher tolerance to salinity, drought and low temperature in *Arabidopsis* (Niu et al. 2012). Together, these examples illustrate how stress-responsive WRKY TFs may represent convergence points between pathways with different functions, also indicating them as a promising target for applied studies in crop species.

2.3 NAC TFs for abiotic genes responsive expression

NAC TF gene family is one of the largest and most diversified in plants with about 110 genes in *Arabidopsis* (Jensen et al. 2010), 150 genes in rice (Nuruzzaman et al. 2010) and 101 genes in soybean (Pinheiro et al. 2009). Transcriptomic studies indicate that a large number of NAC TFs are associated with abiotic stress, such as drought, salinity, and cold, as well as ABA phytohormone.Involvement of NAC TFs in abiotic stress is regulated through binding to regulatory promoter regions at the transcriptional level, alternative splicing at post-transcriptional level, association to ubiquitins, and dimerization and/or interaction with other non-NAC proteins at post-translational level (Puranik et al. 2012).

Some stress-responsive NAC genes are grouped into the SNAC category which includes three subgroups (Nuruzzaman et al. 2010). In *Arabidopsis* members of subgroup III-3 (ANAC019, ANAC055, and RD26 or ANAC072) were induced by high salinity, drought, JA, and under control of central ABA perception and signaling network (Tran et al. 2004; Jensen et al. 2010). Overexpressing of these three NAC TFs conferred drought tolerance in

transgenic *Arabidopsis by up-regulation of several stress-inducible genes (Tran et al. 2003).* These TFs bind to ERD1 (EARLY RESPONSIVE TO DEHYDRATION STRESS 1) promoter, which is induced by dehydration (Tran et al. 2004; Naika et al. 2013). Interestingly, ERD1 up-regulation depends on co-overexpression of two *cis*-elements (ZF homeodomain transcriptional activator ZFHD1 and NAC TFs), suggesting cooperative regulation of stress responses via members of different TF families (Lindemose et al. 2013). ANAC019 and RD26 also conferred ABA-hypersensitivity, suggesting positive regulation of ABA signaling (Tran et al. 2004; Jensen et al. 2010).

Several members of NAC family in rice have been identified with potential association with abiotic stress response or signaling. For example, SNAC1 overexpression improved drought tolerance under field condition, also conferring strong tolerance to salinity (Hu et al. 2006). OsNAC6/SNAC2 was also induced by drought, cold, and high salinity. OsNAC6 overexpression in rice transgenic plants improved drought and high-salt stress through modification of expression of a significant number of stress-responsive genes with several functions such as detoxification, redox homeostasis, and proteolytic degradation (Chung et al. 2009). Overexpression of OsNAC5 also affected plant tolerance to salinity by up-regulating the expression of stress-inducible genes, such as OsLEA3 (Takasaki et al. 2010). Song et al. (2011) showed that transgenic rice plants with reduced OsNAC5 expression by RNA interference (RNAi) resulted in less tolerant individuals to abiotic stresses than control plants. Recent results demonstrated that overexpression of OsNAC5 enhanced drought tolerance and grain yield under field conditions (Jeong et al. 2013). Furthermore, overexpression of OsNAC10, OSNAC45 and ONAC063 also enhanced abiotic stress tolerance of rice playing an important role in inducing responses to high-salinity stress (Jeong et al. 2010; Yokotani et al. 2009).

NAC genes also improved stress tolerance in other crops. In *Brassica napus* (Brassicaceae) nine BnNAC were induced in response to cold, ABA treatment and drought (Hegedus et al. 2003). In soybean, NAC showed to be induced by abiotic stress, participating in stress tolerance. GmNAC2 acts as a negative stress regulator and its overexpression lead to hypersensitivity to drought, salinity and cold (Jin et al. 2013).

3 Epigenetic control of TF response

The tridimensional structure of the DNA in eukaryotes is dynamic, directly influencing gene expression. During the cell cycle, due to the association of DNA with proteins called

histones, given regions may have a lower or a higher level of folding. These topological alterations reflect, respectively, permissive or inhibiting states regarding the access of the transcriptional machinery (TFs, effector proteins, etc.). Certain genomic regions have constitutively lower (called chromatin) or higher levels of folding (constitutive heterochromatin). Additionally, there are those regions whose topology may be variable (facultative heterochromatin) during the life cycle of the organism. This occurs due to histone modifications in N-terminal tails, which act as substrate to chemical associations that change the chromatin structure and/or signaling for the recruitment of inhibitors/activators of gene expression (Jenuwein and Allis 2001). There is a number of these alterations in such proteins (for a review see Chinnusamy and Zhu 2010; Bannister and Kouzaides 2011). The set of modifications orchestrated by combinations involving these proteins is named 'histone code', being recognized by other proteins and resulting in downstream events (Strahl and Allis 2000; Jenuwein and Allis, 2001).

Another type of chemical alteration of chromatin, also associated with the transcriptional regulation, is represented by the methylation of cytosines in target genes. In plants methylation of promoter sequences is generally associated with inhibition of transcription. Otherwise, methylation on coding regions usually has discrete effects during the referred process (Stam et al. 1998; Zilberman et al. 2007). As exposed, it is clear that both chemical histone modifications and methylation of cytokines alter expression of genes without causing changes in their nucleotide sequences. Additionally, such modifications represent no mutations and may be transferred (inherited) from one generation to the next (Molinier et al. 2006; Hauser et al. 2011). Thus, such processes are representative of so-called epigenetic phenomena. Russo (1997) conceptualizes this term as "any inheritable modification of gene expression coming from the alteration on the accessibility of the transcriptional machinery to target genes, and not of mutation on its sequences". Additionally, there are recent adjustments of this concept, suggesting that the requirement of heritability is falling into disuse and that epigenetic may be defined as "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" (Bird 2007). Thus, there is no consensus or concrete definition to the term epigenetic, a condition also highlighted by Richards et al. (2010).

Once the DNA is on a permissive configuration for the access of the transcriptional machinery, TFs interact with specific promoter regions and execute their function, helping the RNA polymerase II to initiate the transcription. However, recent findings revealed that expression of TFs may also be directly regulated by epigenetic modifications. This was initially

observed in *A. thaliana* for the gene ATX1 (ARABIDOPSIS HOMOLOG OF TRTHORAX). ATX1 carries a SET [Su(var)3-9, E(z) and Trithorax] domain that is highly conserved and presents histone methylase activity (Rea et al. 2000). ATX1 is an epigenetic regulator of histone H3K4 methyltransferase activity (i.e., it methylates lysine four of histone H3; Alvarez-Venegas 2003), a modification associated with gene activation (Schneider et al. 2004). It is also known that ATX1 does not act on global methylation of K4 residues in H3, acting only in part of them (Alvarez-Venegas and Zoya Avramova, 2005). However, an analysis revealed that a high amount of genes (around 1,600) exhibited expression changes on *atx1* mutants, with approximately half of the genes being inducted and the other half-suppressed (Alvarez-Venegas et al. 2006). Aiming to explain the occurrence of such drastic and opposite alterations (induction versus suppression) in the *Arabidopsis* transcriptome after knocking-out of a single gene, target genes with altered expression have been prospected. As an outcome, 60 TF coding genes were identified, 42 of them activated and 18 repressed (Alvarez-Venegas et al. 2006). Once there was a direct correlation between epigenetic modification and expression of TFs, Alvarez-Venegas et al. (2007) conducted a series of analysis to determine how the action of ATX1 influenced the activity of TFs and also of a high number of additional genes. Since TF WRKY70 was suppressed 7.2 times in control samples during the previous test, this gene was chosen as study model (Alvarez-Venegas et al. 2006). This particular TF is important because it represents a convergence point between the jasmonate signaling pathway (acting as a suppressor) and salicylic acid pathway (acting as an activator), being an important component of the biotic stress response (**Figures 1A-B**). The authors observed that WRKY70 is the primary target of ATX1, which acts on the nucleosomes of this TF, methylating [specifically, trimethylating (me3)] K4 residues of H3 (H3K4me3) histones. Such epigenetic modification was responsible for the activation of WRKY70. An alteration of the expression of ATX1 lead, consequently, to a change in expression of WRKY70 (**Figures 1A-B**) and all genes regulated by this specific TF (induced of and suppressed). This explained, at least partially, how the alteration of one single gene caused an abrupt change in the expression of a myriad of other genes.

In relation to TFs associated with abiotic stress response, Alvarez-Venegas et al. (2006) also observed that an *atx1* mutant influenced the activity of ZAT10 (suppression of 2.7 times) and R26 (suppression of 3.72 times), among other genes. The first gene encodes a TF that induces the expression of genes associated with ROS response. Some evidences indicate that ZAT10 enhance tolerance of plants to stresses such as high salinity, heat and osmotic stresses (Mittler et al. 2006). The second gene (R26) encodes an NAM TF, which mediates regulation

between ABA and JA signaling pathways during responses to drought or wounding stresses (Fujita et al. 2004). These data suggested that epigenetic mechanisms may regulate TFs associated with drought stress, a presumption confirmed by recent experiments carried out by Verkest et al. (2015). These authors analyzed a population of isogenic lines (identic genotypes) of rapeseed (*B. napus*) to select individuals with best indicators regarding drought tolerance (high NAD(P)H content and low respiration) when compared to other lines. Since the population used consisted of identic genotypes, any variation in this index could be attributed to epigenetic factors. After several selection steps, two lines were selected: PEG1 and PEG2. The first presented significant tolerance to drought when compared to its control and PEG2. ChIP-seq (chromatin immunoprecipitation sequencing) data revealed that many genomic regions of PEG1 were enriched for H3K4 being absent in its control. The contrast of these regions enriched with epigenetic markers when compared to transcriptome data has shown that a significant portion of genes associated with H3K4 was differentially expressed, including many TFs as AP2/ EREBP, WRKY, NAC, GRAS, C3H, C2H2, HSF, bHLH, and C2C2.

Besides acting directly in the regulation of TF expression, epigenetic phenomena act indirectly in the regulation of responses during periods of unfavorable conditions for plant development. Every epigenetic modification may influence the access of TFs (Jenuwein and Allis 2001) or their recruitment (Watt and Molloy 1998), and several reports exist on alteration in plant epigenomes submitted to drought and associated stresses, including osmotic and heat stress, among other (Table 1). Despite advances in gene regulation towards developmental processes and stress response, further studies are mandatory for a better understanding of epigenetic mechanisms, also considering the application of this knowledge for plant breeding purposes.

4 Bioinformatic studies of TFs involved in abiotic stress

The advent of experimental high-performance platforms, in particular, next-generation sequencing (NGS), optimized assay systems and advanced bioinformatics approaches have enabled comprehensive and maximized studies of plant genomes in a quick and economically viable manner. Transcriptome studies may be used for quantitative analysis of thousands expressed genes related to germination, growth and development, flowering, and conditions of biotic and abiotic stresses, allowing to understand plant mechanisms of the stress response (Oktem et al. 2008). The identification of the expression of regulatory elements in the huge amount of new data has been facilitated by bioinformatic methods and by the availability of several online repositories.

Algorithms designed for data interpretation allow robust biological discoveries, such as prediction of possible TFs, their location in the genome and evaluation of their regulation in stress conditions, as compared to non-stressed controls. TF identification can be divided into two simplified approaches; the first one consists in detecting these elements in databases, by performing an analysis of these repositories, while the second approach is based on pattern recognition in sequences, allowing their grouping in different plant TF families previously described.

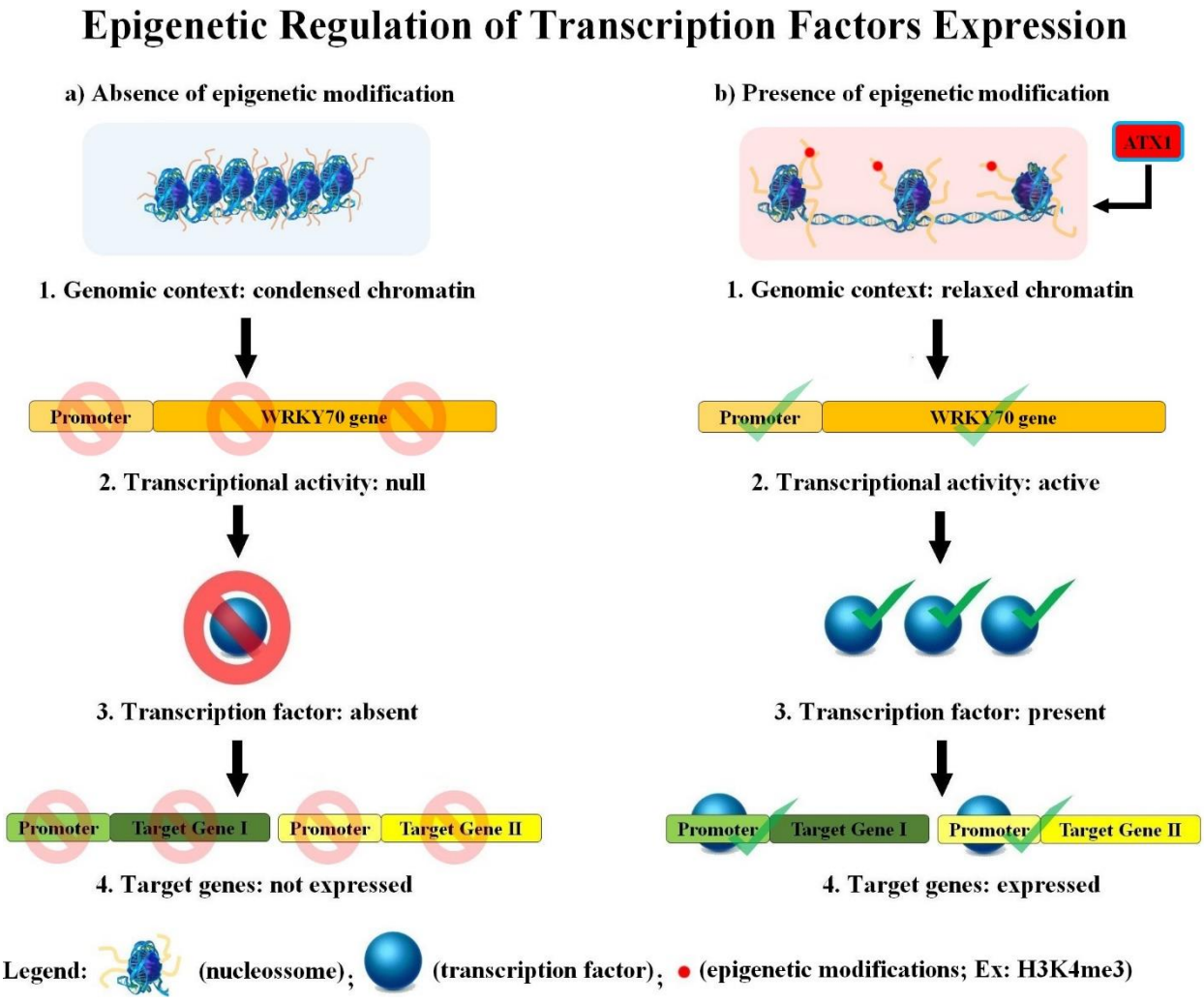


Figure 1. Schematic representation of epigenetic regulation mediated by the expression of WRKY70 transcription factor, by the methylase of histones ATX1m and the developments in the activity of target genes composing the SA (salicylic acid) signaling pathway. **(a)** Situation highlighting the absence of expression of WRKY70 due to the high level of chromatin condensation (absence of ATX1 activity) and, as a result, the absence of expression of its target genes; **(b)** Situation highlighting the induction of WRKY expression due to the occurrence of epigenetic modification (performed by ATX1).

Table 2. Selection of reports concerning epigenetic alterations in response to drought or associated stresses (osmotic stress, heat, etc.) in plants.

Analyzed epigenetic markers	Species	Target gene	Analyzed tissue	Epigenetic state during stress	Tolerance to studied stress	Reference
H3K27me3 and H3K27me2	*Solanum lycopersicum*	*Asr2*	Root	Decrease in the content of H3K27me2	Not analyzed	Gonzales et al. (2013)
H3T3ph	*Arabidopsis thaliana*	Genome wide	Leaf	Decrease in the phosphorylation	hypersensitivity	Wang et al. (2015)
H3K4me3, H3K9Ac and H3K27Ac	*Physcomitrella patens*	Genome wide	Gametophores	Increase in the content of H3K4me3, H3K9Ac and H3K27Ac.	not analyzed	Widiez et al. (2014)
Cytosine methylation	*Oryza sativa*	Genome wide	Leaf	Hypomethylation in the tolerant accession and hypermethylation in sensible accession.	not analyzed	Gayacharan and Joel (2013)

Stress response mechanisms involve complex regulation of multiple genes and TFs. The first step involves the identification of these factors to unravel mechanisms associated with their regulation (Babu et al. 2004; Dubos et al. 2010; Atkinson et al. 2011). The main computational methods developed for identifying TFs on a genome-wide scale involved *in silico* approaches to evaluate the presence or absence of TFs and definition of their characteristics (DNA binding domains, auxiliary areas, and lost domains) (Harrison, 1991). The first TF analysis at genomic level was performed by Riechmann et al. (2000) for *A. thaliana*.

Several available databases focused initially on identification and annotation of TFs, like PlnTFDB (Pérez-Rodríguez et al. 2010), PlantTFDB (Jin et al. 2013), LegumeTFDB (Mochida et al. 2010) and TreeTFDB (Mochida et al. 2013). Databases for specific organisms were also created, like SoyDB (Wang et al. 2010), DATF (Guo et al. 2005) and DPTF (Zhu et al. 2007), all for plants such as soybean (*Glycine max*), *Arabidopsis* and *Populus trichocarpa* (Salicaceae), which may be used for evolutionary studies of TF in plants, as for prediction of TFs in new sequenced genomes. Public databases such as PlantTFDB (http://planttfdb.cbi.pku.edu.cn) and PlnTFDB (http://plntfdb.bio.uni-potsdam.de/v3.0) have cataloged and predicted TFs for more than 50 plant species of different taxonomic categories, including algae, bryophytes, gymnosperms and angiosperms. Additionally, in 2009 the STIFDB (Stress-responsive Transcription Factor Database) was created to summarize information from abiotic stress responsive genes and TFs associated with these genes in *A. thaliana* (Shameer et al. 2009). Its assembly is the result of data clustering

from analytical profiles of stress responses in plants (Sanghera et al. 2011; Babitha et al. 2013) and computational studies of stress-induced gene regulation (Georgii et al. 2012). The new version (STIFDB2 database; http://caps.ncbs.res.in/stifdb2) added data regarding two rice varieties (*O. sativa* subsp. *japonica* and *O. sativa* subsp. *indica*). The data was compiled from an analysis of 15 different stress types, including cold, osmotic, dehydration, heat, salinity, radiation and oxidative stress, among others. Different families and subfamilies of TFs, their DNA binding domains and reference data from the literature related to his prediction/description are also listed in the new version. Its construction was based on mining genomic data and identification of 5,984 unique genes related to stress, with more than 38,500 associations signaling and/or responding to stress. Briefly, the database can be used to identify potential stress-induced TFs, their gene regulation and may also be extrapolated to determine protein-protein interactions between TFs. With the inclusion of orthologs from other species, the database can be used to study evolutionary conservation and carry out comparative analyzes between TFs responsive to abiotic stress in different plant species (Naika et al. 2013).

PlnTFDB is another useful database, consisting of a public repository, which aims to identify and catalog all genes involved in transcriptional control in plants, currently containing 28,190 protein models and 26,180 sequences of different proteins, organized into 84 families, with classification based on domain structure. The classification of TFs is based on domain structures previously built in Pfam database, with prediction via hmmpfam, an alignment tool included in the HMMER package, which is based on the identification of protein motifs via HMM (Hidden Markov Model) (Eddy 1996; Pérez-Rodriguez et al. 2010). Despite its functionality, few tools are available in PlnTFDB, allowing analysis of sequences just like a TF repository of information. Many domains are included in other databases/repositories like PROSITE (http://prosite.expasy.org), a database that includes protein domains, families, and functional sites; and SMART (http://smart.embl-heidelberg.de) allowing gene identification, annotation of domains, and sequence architecture.

Additionally, DBD (http://www.transcriptionfactor.org) database deserves mentioning with tools for TF prediction in complete sequenced genomes, using resources provided by Pfam (http://pfam.xfam.org) as well as information on domain organization, also systematizing information for gene families and genomes (Kummerfeld and Teichm 2005). DBD provides for users the possibility to search DNA binding domains in a given protein sequence. Despite its functionality, this database only focuses on TFs, not including transcription regulators (TRs).

Regarding the identification of TFs and TRs by comparing functional domains, some tools and scripts are available, like the unpublished tool ITAK (http://bioinfo.bti.cornell.edu/cgi-

bin/itak/index.cgi). This tool adopts the same domain classification of PlnTFDB, based on data cataloged in Pfam to predict genes coding for TFs and TRs. This tool is available in two versions (online and standalone). The online version, even being friendly to the user, analyzes only 50 protein sequences at a time, making its use impractical to evaluate sequence analysis on a genomic scale. In contrast, the standalone version, is capable of realizing more robust analyzes, but may be a challenge for most biologists, since the implementation is based on Linux command lines, often requiring optimization on a cluster computer system.

Another tool that provides a functional protein analysis, classifying families by prediction of domains and important target sites in sequences is InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan5) which allows the alignment of sequences against a database of protein signatures. InterProScan comprises 14 search programs for domain signature or domain pattern, being useful for prediction of TFs and TRs (Hunter et al. 2009). This tool has been widely used for large-scale analysis, as in the case of *Medicago truncatula* (Fabaceae), which had its factors and transcription regulators identified in different gene families by presence of known binding DNA domains (Kakar et al . 2008). For soybean, Wang et al. (2010) developed a protocol for TF and TR prediction using InterProScan and a previous mapping in species database. More recently, a group of online tools called MEME (http://meme-suite.org) was developed, which includes 13 tools for discovery and enrichment of motifs as well as motif pattern comparison (motif-motif comparisons). Since motifs are responsible for many biological functions, their detection and characterization comprise necessary steps in the analysis of molecular cell interactions, including regulation of gene expression (Bailey et al. 2015).

In recent years, probabilistic models have been used for performing sequence alignments, especially using HMM-based softwares. One of the most widely known software, HMMER, implements in their algorithm a probabilistic inference to build complex models of specific positions in a given sequence by HMM. This software is a powerful tool for homology search, based on probabilistic inference methods (Eddy 2009; Johnson et al. 2010; Finn et al. 2011). Today HMMER is a major tool for domain search and TF classification, implemented in most plant TF databases, being widely used to find TFs related to abiotic stress.

In microalgae, a basal group of Viridiplantae, computational identification of TFs has been carried out for *Chlamydomonas reinhardtii*, *Volvox carteri, Galdieria sulfuraria* and *Nannochloropsis oceanica* (Eustigmatophyceae) (Pérez-Rodriguez et al. 2010; Vieler et al. 2012). Relationships and differences between TFs of this ancient group with higher plants remain uncertain. Hu et al. (2014) analyzed three different genomes [*N. oceanica* (IMET1 strain), *N. oceanica* (CCMP1779 strain), and *N. gaditana* (CCMP526 strain)] by comparison of conserved

domains, using the PlantTFDB and its implemented pipeline (Jin et al. 2013). The analysis uncovered 125, 119 and 85 TFs in the three genomes respectively, corresponding to 1.26%, 0.99% and 0.94% of their proteomes. The predicted TFs were then classified into different families based on DNA binding domain. Considering the three analyzes, 26 families were shared, with 19 TFs common for all three species, whereas MYB, bZIP, MYB-related and NF-YC were the most abundant TF families, representing more than 40% of the factors found.

In higher plants, a total of 1533 TF-genes was identified for *Arabidopsis*, being classified in 34 families (Riechmann et al. 2000), while 1611 TFs were predicted for rice, classified in 37 families of genes (Xiong et al. 2005). In turn, for soybean 51 HMM models deposited in Pfam and 11 models created via HMMER tool were used to predict existing TFs in the soybean genome. It was possible to identify 5035 protein sequences corresponding to 4342 different loci from 61 TF families. From these sequences a database for specific soybean TFs was created, the SoybeanTFDB (http://soybeantfdb.psc.riken.jp), grouping basic and relevant information about the motifs found, promoter regions, genome distribution, as well as the alignment of members of the 61 TF families found. Focusing in abiotic stress response, promoter regions of TFs were analyzed to identify *cis* elements in PLACE database (http://www.dna.affrc.go.jp/PLACE) a repository of identified *cis-motifs*, including those that respond to biotic stresses (Higo et al. 1999; Mochida et al. 2010).

This type of analysis focused on specific families is available for several plant species, but only some important TF groups for specific stresses and growth regulation had been studied. This often happens because most TF families are vast and complex in higher plants. This is the case of WRKY, NAC, and bZIP families, which are highly diverse and active in plant expression reprogramming under stress conditions (Nakashima et al. 2012; Rinerson et al. 2015; Seo and Choe 2015). For example, Ooka et al. (2003) identified in their work 75 isoforms of NAC in rice and 105 for *Arabidopsis*, while at least 157 orthologues in rice were identified by Lu et al. (2015). In soybean at least 101 members of NAC were predicted (Pinheiro et al. 2009), while 163 complete sequences were predicted to *P. trichocarpa*, being grouped into 18 different subfamilies (Hu et al. 2010). For WRKY at least 74 members have been characterized in *Arabidopsis*, and more than 100 members have been reported for rice (Wu et al. 2005; Zhang and Wang 2005). Similarly high numbers were found in grape (*Vitis vinifera*, Vitaceae) and tomato (*S. lycopersicum*) with 98 and 84 members, respectively (Wang et al. 2014; Chen et al. 2015). Members of bZIP TFs have also been predicted in several organisms, such as *Cucumis sativus* (Curcubitaceae), with 64 isoforms identified (Baloglu et al. 2014), while 75 members were predicted for *Arabidopsis* (Jakoby et al.

2002), 92 for Sorghum bicolor (Poaceae, Wang et al. 2011) and 96 for Brachypodium distachyon (Poaceae, Liu and Chu, 2015).

The identification based on computational methods is faster when compared to traditional molecular biological methods. Its efficiency is directly related to data integration from public databases, associated with sensitive prediction algorithms, providing a new approach for understanding main plant molecular responses related to stress response, adaptation, and tolerance (Pérez-Rodríguez et al. 2010; Jin et al. 2013; Naika et al. 2013). Until now, there is no universal pipeline for identification and classification of plant TFs that can be used by the scientific community in a comprehensive way. However, prediction of TFs in plant genomes can be made without great difficulty when appropriate tools and repositories are integrated into a workflow. Here, we present two workflows for TFs prediction, using tools and resources available in the previously mentioned databases (Figure 2).

Predictions based on sequence alignment/motifs (HMM) (Fig. 2A), uses TF databases currently available. Initially, it is necessary to select protein families of sequences of interest in the respective database (e.g. PlantTFDB, PlnTFDB, etc.) to be used as seed sequences to generate the alignment and build the motif pattern using hmmbuild tool implemented in HMMER 3.0 package. The motif generated for a given TF family is then confronted via hmmsearch tool (also implemented in the package) against the protein or translated nucleotide sequences for the chosen species, using a cut-off value of e-05 (or lower). This step will select candidates for TFs in a selected organism that show similarity with predicted TFs available. Sequences obtained in this step shall be annotated against GenBank to confirm their function, as well as the similarity level with homologous sequences previously characterized. This analysis is simple and can be applied to all TF families available in repositories. The disadvantage of this workflow resides in the fact that this is not a comprehensive analysis, requiring a longer time to increase the number of TF families analyzed.

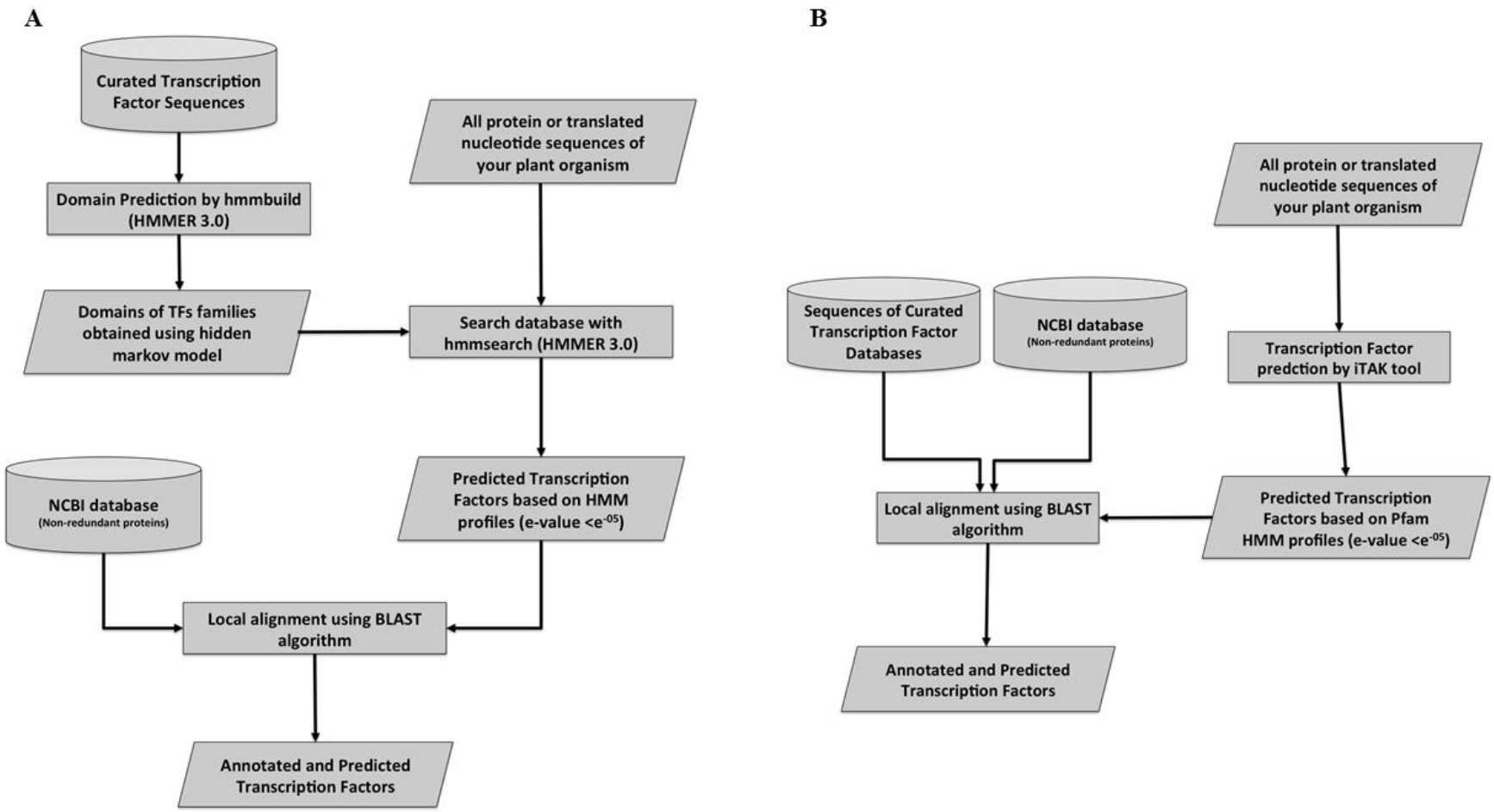


Figure 2. Exemplary workflows for *in silico* transcription factors prediction, using available TF databases. (A) Scheme for TF prediction using HMMER 3.0 via alignment of sequences against HMM models obtained via hmmbuild (aligning sequences obtained in public databases); (B) Predicting transcription factors in the genome, based on ITAK tool (offline version) and annotation against NCBI GenBank.

In contrast, the workflow outlined in Figure 2B presents a pipeline for a global analysis of TFs. Prediction, in this case, is based on HMM alignment of motifs of known families via ITAK tool. The script implemented in the program will identify all TF candidates, generating files for alignments, classification, and translation of nucleotide sequences for the chosen genome, using a recommended cut-off of e^{-05} (or lower). Also here, these candidates can then be aligned via BLAST against public TF databases and GenBank to identify their function and confirm similarity to orthologs from other species. The disadvantage of this method is the requirement of the command line to install packages and run scripts, a limitation for many biologists, besides the necessity of computing power, often associated with routine implementation on computer clusters, paralleling the process to accelerate the analysis.

5 Interaction network

Modeling metabolic pathways is fundamental to a system biological analysis of complex processes, facilitating studies about signaling mechanisms including abiotic and biotic stresses and allowing identification of molecular targets for biotechnological applications in crop plants. Generation of interaction networks also contributes to the understanding of metabolic, physiologic and cellular mechanisms involved in such processes. Among regulators, many TFs figure as hubs, acting in the expression control of several stress-inducible genes, forming gene networks and/or signaling cascades. Regulation of gene expression is complex and controlled by an interaction network between different regulatory protein and *cis*-regulatory sequences present in promoters of their target genes (Harbison et al. 2004). Promoter regions act as molecular "switches" in gene expression, but also as signal transduction termination points in signaling processes (Yamaguchi-Shinozaki and Shinozaki 2005; Yamaguchi-Shinozaki et al. 2006; Tran et al. 2007; Nakashima et al. 2009). The interaction between TFs and their DNA binding sites are an integral part of regulatory networks (Bulyk 2007).

Changes in expression patterns occur after perception of a given environmental perturbation by the plant, with the molecular machinery reprogramming TFs activity, allowing the creation of an effective defensive state. Many of these genes are multifunctional, able to induce tolerance to more than one stress situation (Xiong et al. 2002; Cramer et al. 2011; Chul-Lee and Luan 2012; Miura and Tada, 2014; Rejeb et al. 2014). The interaction between abiotic and biotic

stresses, for example, induces complex responses to different stressors. Under stress, accumulation of certain metabolites may modify plant response to both biotic/abiotic stresses, favoring cross-talking response, protecting the organism to more than one type of stress pressure (Rasmussen et al. 2013). Thus, changes in ion flux, increases ROS and hormone synthesis, as the first responses after stress perception, whereas the resulting signal transduction triggers metabolic defense reprogramming (Bartoli et al. 2013). For example, Tsutsui and coworkers (2009) reported that DREB TF may regulate stress response between abiotic and biotic stress in *A. thaliana*, increasing plant resistance to cold and pathogens simultaneously.

Available sequences of reference genomes revolutionized the knowledge of plant genetics and its interactome networks, especially considering protein-protein interaction (PPIs) networks, allowing a comprehensive identification of genes involved in the response to abiotic and biotic stresses. Plant protein-protein interactome networks can be generated either using '*in vivo'* or '*in silico'* methods. The most reliable experimental methods for plant interactome mapping are yeast two-hybrid (Y2H, Y3H), bimolecular fluorescence complementation (BiFC) with affinity purification and mass spectrometry (AP-MS) (Zhang et al. 2010). Considering *in silico* methods in PPI detection, there are sequence-based approaches, structure-based approaches, chromosome proximity, gene fusion, ortholog-based sequence approach, phylogenetic profile analysis, gene co-expression profiles, sequence co-evolution, synthetic lethality data, *in silico* two-hybrid systems, gene cluster versus gene neighbor analysis, protein domain information, protein interface analysis, and protein docking methods, among others (Smith 2002; Aytuna et al. 2005; Skrabanek et al. 2008; Rao et al. 2014). Another *in silico* method used to understand main mechanisms associated with signaling and gene response regard pathway model representations, build using computational tools. Softwares that allow description of molecular interactions are based on 'System Biology Markup Language' (SBML; http://www.sbml. org), that allow consistent and intuitive graphical construction of front pathways related, for example, to evaluate plant environmental stress responses. The maps cited here are traditionally created via CellDesigner v. 4.4 (http://www.celldesigner.org/), a modeling tool for the design of biochemical networks with graphical user interface (Hucka et al. 2003; Funahashi et al. 2003; Kitano et al. 2005).

In recent years, several approaches were designed to predict plant interactomes in species of agronomic importance, like soybean, *Citrus sinensis* (Rutaceae) and rice, providing insights into mechanisms and pathways involved mainly with disease resistance (Musungu et al. 2015).

Currently, some plant repositories and databases for whole genome scale, for transcriptomic and proteomic data are available and some act as repositories of plant protein-protein interaction platforms. Many of these repositories are available online for searches and free downloads, such as STRING, IntAct, MINT, TAIR and BioGRID (Stark *et al.* 2006; Breitkreutz et al. 2008; Swarbreck et al. 2008; Aranda et al. 2010).

For *Arabidopsis* and some other plants, stresses including drought, osmotic, salinity, oxidative, wounding, and heavy metals were found to induce expression of TF proteins that modulate specific genes with defined binding regions (Figure 3; Takuhara et al. 2011). Several reviews focus on specific TF roles under abiotic stress, with the main emphasis on families that present key roles in stresses as drought, salinity, high temperature and hormones, being responsible for the modulation of several genes. Due to the fine-tuned relationship between TFs and other genes, it is clear that the understanding of such processes is directly related to the deciphering of regulatory networks, especially in interactions between members and the dynamic interaction among these, as well as analysis of each member that act in more than one signaling pathway. This type of analysis has been generated for several TF families, with emphasis on basic leucine zipper (bZIP), AP2/ERF, NAM/ATAF1/CUC2 (NAC), WRKY, MYB, and basic helix-loop-helix (bHLH). Many members of these TF families were characterized with respect to their role in abiotic stress pathways, resulting in increase of plant tolerance/resistance against stresses (Ciftci-Yilmaz and Mittler 2008; Chen et al. 2012; Dubos et al. 2010; Fujita et al. 2011; Mizoi et al. 2012;. Puranik et al. 2012;. Rushton et al. 2012; Nuruzzaman et al, 2013; Danquah et al. 2014).

TFs function through interaction with a broad range of proteins, including other TFs, cofactors and chromatin modifiers (Grove and Walhout 2008; Chi et al. 2013). Tolerance to abiotic stresses occurs in plants when TFs interacted with multiple proteins. For many TFs, the formation of dimers is necessary prior to DNA binding, including homo- or heterodimers (Funnell and Crossley 2012). Dimerization can be a mechanism of either positive or negative transcription control (Schulman 2013). Examples of TFs that require dimerization for binding include bZIP and bHLH (Amoutzias et al. 2007). Group-A subfamily of bZIP-type includes ABA response elements factors (ABF1-ABF4) and ABA-responsive element binding proteins (AREB1-AREB3), which interact with ABA-responsive elements (ABREs) under water deficit condition in *Arabidopsis* (Fujita et al. 2009). AREB proteins bind to ABA-responsive elements, which are major *cis*-acting elements in ABA-dependent pathways (Yoshida et al. 2010). Overexpression of AREB1 in rice

and soybean enhances drought tolerance (Oh et al. 2005; Barbosa et al. 2013). Overexpression of AREB2/ABF3 and ABF4 in *Arabidopsis* also confers drought tolerance (Kim et al. 2004), whereas overexpression of an active form of AREB1/ABF2 exhibited enhanced drought tolerance (Fujita et al. 2005). These three AREB/ABF TFs (from both rice and *Arabidopsis*) need ABA for full target gene activation. In turn, *areb1 areb2 abf3* triple mutant presented reduced tolerance to drought stress as compared to single and double *AREB/ABF* knockout mutants, indicating cooperative action between these three TFs (Uno et al. 2000; Yoshida et al. 2010).

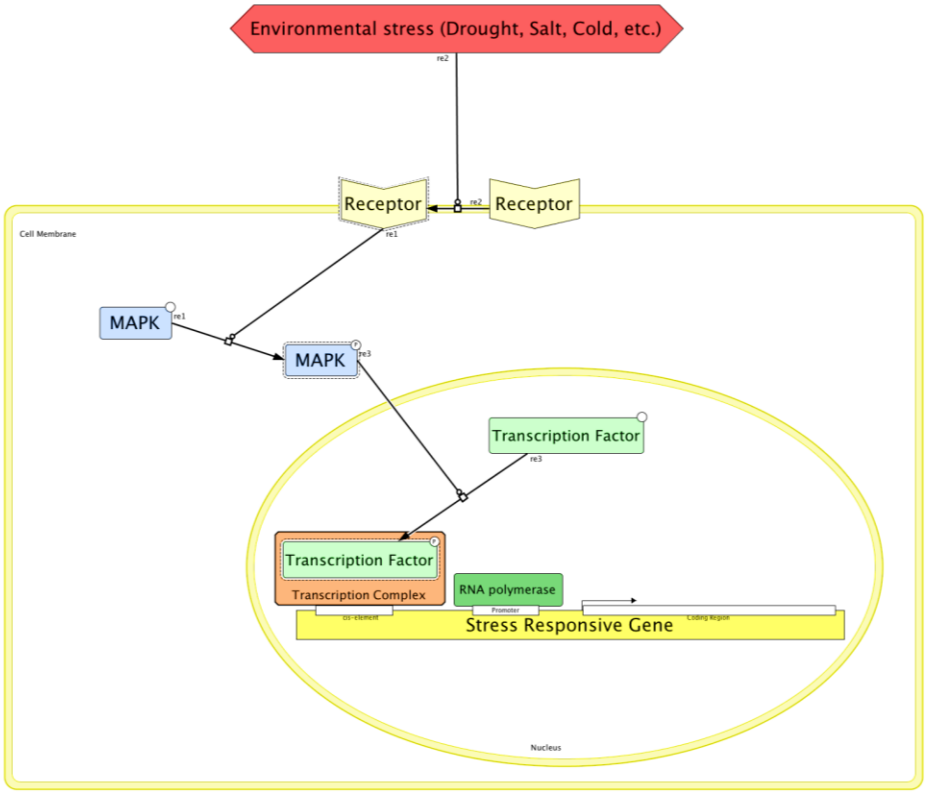


Figure 3. A simplified model of abiotic transcriptional activation and gene expression regulation created using CellDesigner ver.4.4. The graphical representation is based on system biology graphical notation (SBGN). The model shows proteins (transcription factors, general MAPK), general receptors, genes and protein complex. The active state of the molecules is indicated by a dashed line surrounding the molecule. The phosphorylated state is indicated by letter 'P' inside protein residue symbol. The frame in yellow represents the cellular membrane and compartments.

Analysis of the promoter region of other dehydration-induced genes led to the discovery of a group within AP2 TF subfamily named DREB/CBF (dehydration responsive element binding/C-repeat binding factor), which is unique to plant species and implicated in ABA-independent regulation (Figure 4).

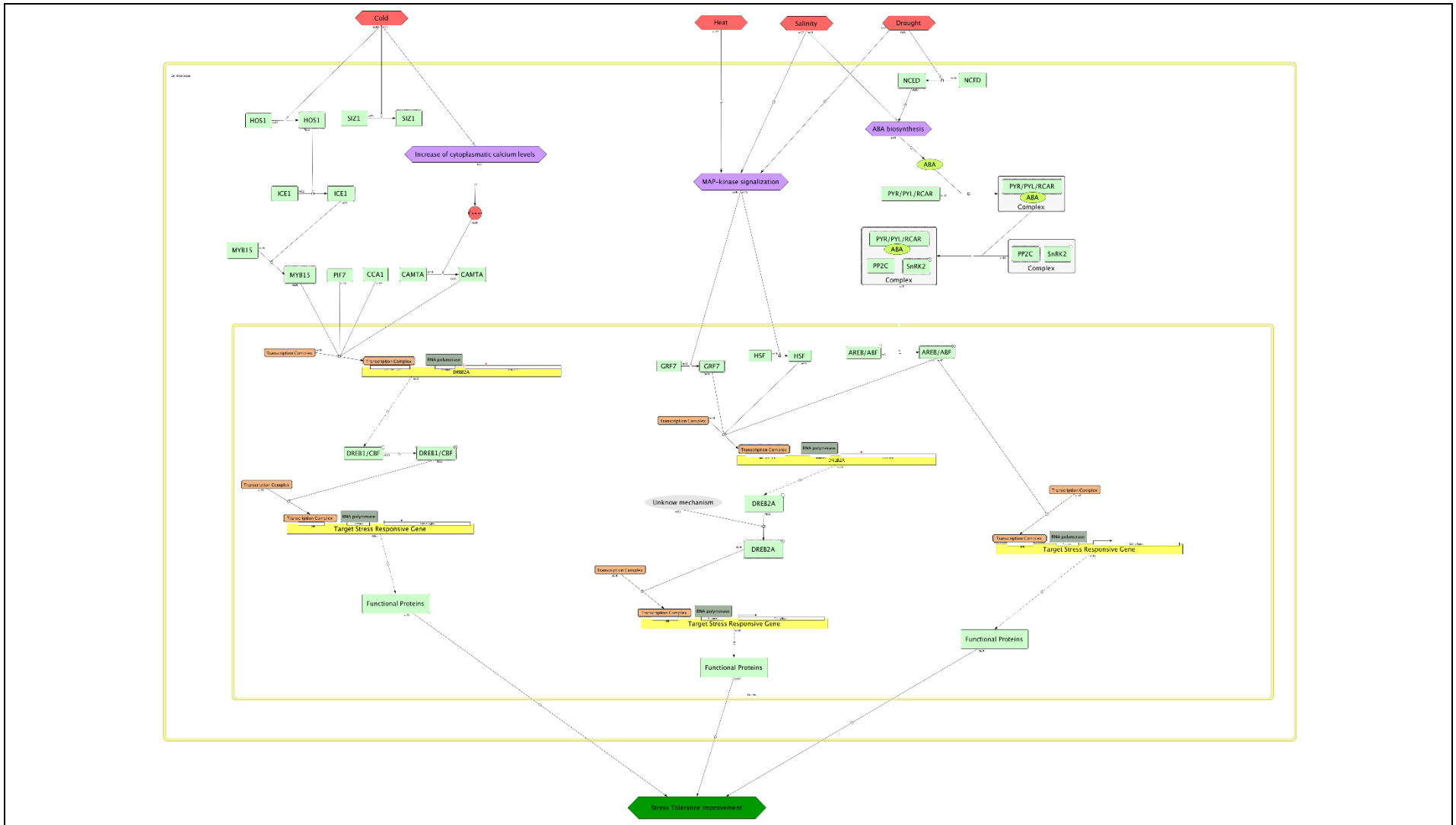


Figure 4. Representation of abiotic stress signaling networks mediated by AREB, DREB1, and DREB2-type transcription factors. Model created using CellDesigner ver.4.4, with a graphical representation of plant abiotic stress responses in abscisic acid (ABA)-dependent and ABA-independent gene expression. Drought, salt, high and low-temperature module the level and activity of the TF group and their target genes. The upper part of the figure shows transcription cascades involved in rapid responses to abiotic stress, driven by Kinases (MAPK) or not. Lower parts of the figure show transcription cascades involved in gene activation. The pathways suggest crosstalk between the stress types. The active state of the molecules is indicated by a dashed line surrounding the molecule. The phosphorylated state is indicated by letter 'P' inside protein residue symbol. The frame in yellow represents the cellular membrane and compartments.

Proteins of DREB subfamily activate C-repeat or dehydration response element (DRE/CRT) located in the promoters of drought-inducible genes and regulate expression of genes under water deficit (Liu et al. 1998). In turn, members of DREB1/CBF subgroup (DREB1A/CBF3, DREB1B/CBF1, and DREB1/CBF2) are cold inducible and known to be involved in cold stress response, while DREB2 group (DREB2A and DREB2B) plays a role in dehydration and heat stress response (Mizoi et al. 2012). The physical interaction between AREB/ABFs and DREB1A/CBF3, DREB2A, and DREB2C has been described (Lee et al. 2010). Lee et al. (2010) showed that the domains of AP2 proteins (DREB1A and DREB2A) interact with ABF2 and that additional ABF family members (ABF3 and ABF4) interact with DREB2C. The interaction between DREB2C, ABF2, and ABF4 can be observed in *Arabidopsis* predicted network (Figure 5A), indicating crosstalk between elements of ABA-dependent and ABA-independent response pathways. Subsequently, Kim et al. (2011) showed that an ABRE promoter sequence and AREB/ABF were involved in DREB2A expression, suggesting a complex interaction between AREB and DREB regulons at gene expression and also at the protein level.

Yeast two-hybrid assays involving DREB2A, DREB2B, and DREB2C, showed their interaction with the regulatory hub protein RCD1 (Radical Induced Cell Death 1) and its closest paralogue SRO1 (SIMILAR TO RCD ONE) (Jaspers et al. 2009), which is also observed in *Arabidopsis* DREB2C predicted interaction (Figure 5A). RCD1 is degraded during heat shock stress and RCD1-interaction deficient *dreb2a* splice variant (DREB1a.2), which lacks RCD1-interacting region and is accumulated during heat shock. This observation suggests that removal of RCD1 protein or loss of interaction with DREB2A is necessary for DREB2A function under abiotic stress and that RCD1 may mediate REB2A degradation (Vainonen et al. 2012). Also, predicted interaction analysis have connected DREB2C to RCF3, which encodes a KH-domain containing putative RNA-binding protein. RCF3 is an important upstream regulator of heat stress-responsive genes expression and thermotolerance in *Arabidopsis*, indicating that interaction between DREB2 and RCF3 cooperates in regulating heat tolerance in *Arabidopsis* (Guan et al. 2013).

Recently expression of an NAC member, ATAF1, was also induced by different abiotic stresses. Jensen et al. (2013) reported that *Arabidopsis* ATF1 may regulate ABA-dependent gene expression of ABRE regulons. Interestingly, cooperative action of ANAC096 and AREB/ABF factors (ABF2/AREB1 and ABF4/AREB2) was necessary for dehydration and osmotic stress

response (Xu et al. 2013). In soybean, GmNAC20 may participate in the regulation of stress tolerance through activation of DREB/CBF-COR pathway (Hao et al. 2011). The *Arabidopsis* ATAF1 predicted interaction allowed the identification of an interaction between ATAF1 and DREB2A, probably playing a significant role in drought response (figure 5B). ATAF1 also interacts with ERD14 protein, which is ABA-regulated. EDR14 is a member of the dehydrin family that accumulates in response to abiotic stresses in *Arabidopsis*, including drought, low temperature, and salinity (Alsheikh et al. 2003). Functional analyzes of ABA-regulated ERD14 showed prevention of heat-induced aggregation and/or inactivation of distinct enzyme substrates (Kovacs et al. 2008). Thus, ERD proteins act protecting the cell metabolism during stress and the induction of these genes by ABA during heat shock may help to prevent denaturing effects.

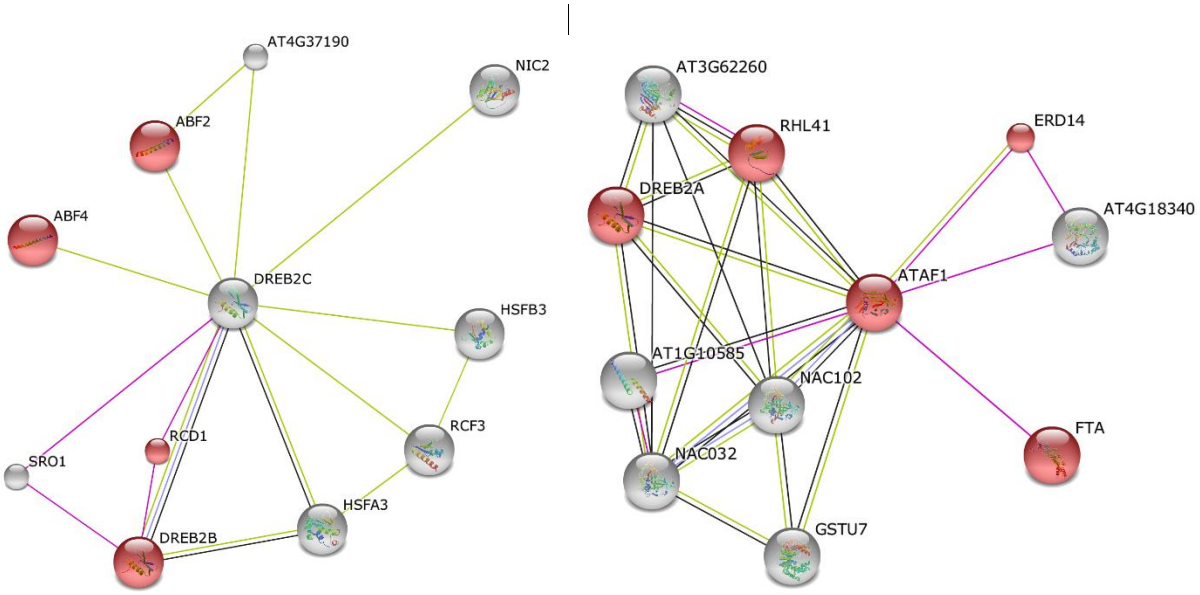


Figure 5. Schematic representation of the transcription factor (A) DREB2C and (B) AtATAF1 from *Arabidopsis thaliana* interaction with other proteins predicted in silico by STRING database program version 9.0 (Szklarczyk et al. 2011). Red colors indicate GO Biological Process response to water deprivation.

Other examples of TFs that form dimers are WRKY proteins (AtWRKY18, AtWRKY40, and AtWRKY60) from *Arabidopsis*. They can interact with themselves and with each other to form homo- and hetero-complexes through the leucine zipper present at the N-termini of WRKY proteins (Xu et al. 2006). For instance, yeast two-hybrid assays revealed that AtWRKY40 and

AtWRKY60 interact with AtWRKY36 and AtWRKY38 (Arabidopsis Interactome Mapping Consortium 2011). Subsequently, yeast two-hybrid system also revealed a significant interaction of AtWRKY30 with AtWRKY53, AtWRKY54, and AtWRKY70 (Besseau et al. 2012). Diverse WRKY genes are under direct positive or negative control by WRKY TF-factors via a specific feedback mechanism (auto/cross-regulation) (Pandey and Somssich 2009).

WRKY family is one the best-studied TF regarding plant protein interactome networks. AtWRKY33 has been shown to be induced by oxidative stress, and its overexpression in *Arabidopsis* improved oxidative and drought stresses (Jiang and Deyholos 2009). To illustrate the strong interaction between WRKY33 and other proteins, we exemplarily used the database view of STRING (Figure 6).

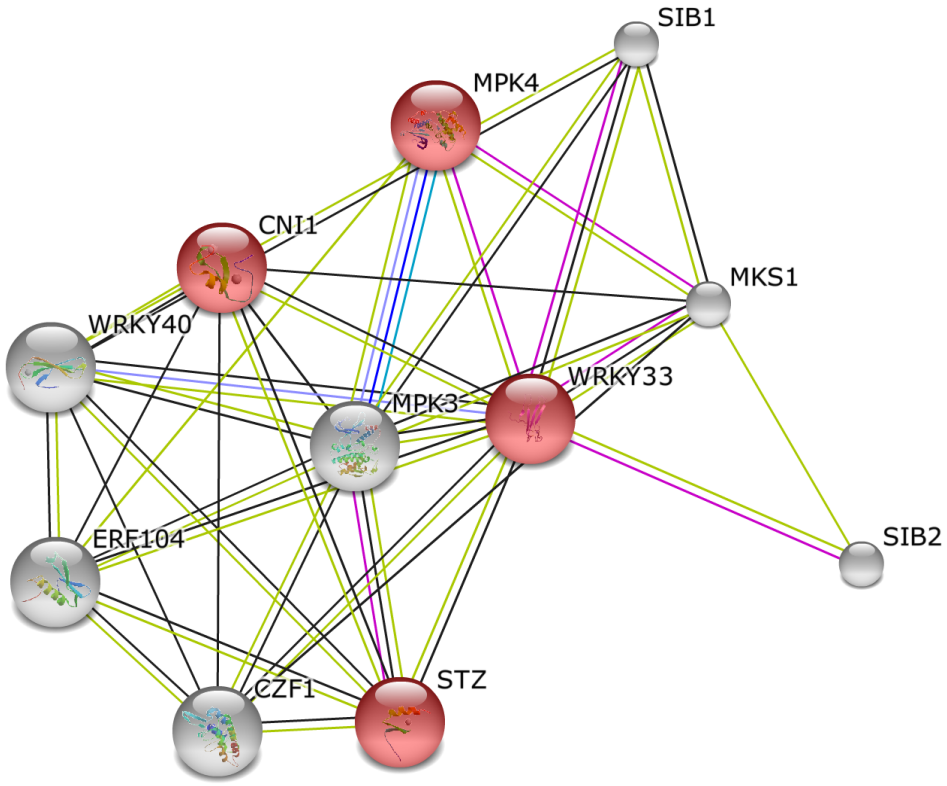


Figure 6. Schematic representation of the AtWRKY33 interaction with other proteins predicted in silico by STRING database program version 9.0 (Szklarczyk et al. 2011).

The output shows that *At*WRKY33 interacted with VG protein that contains a conserved FXXVQXLTG region, named VQ domain, like SIB1 (SIGMA FACTOR-INTERACTING PROTEIN 1). These proteins play diverse roles in plant defense response, drought and salt tolerance, and also growth and development (Jing and Lin 2015). Based on recent reports, Kim et al. (2013) suggested that the interaction of WRKY TF with VQ domain proteins regulates the DNA binding activity of the TF either as positive or negative cofactors of WRKY TFs. Lai et al. (2011) revealed that the VQ motif of SIB1 is necessary for its interaction with WRKY33 in plant defense against necrotrophic pathogens. In turn, Hu et al. (2013) showed that VQ9 protein interacted with WRKY8 in an experiment using yeast two-hybrid system. Mutation of VQ9 increased tolerance to salt stress. Therefore, VQ9 protein probably acts as a transcriptional repressor that antagonizes WRKY8 to maintain a proper balance of WRKY8-mediated signaling pathways to allow salinity stress tolerance (Hu et al. 2013). In rice, proteins bearing VQ domain are co-regulators of WRKY TFs in response to disease resistance and response to environmental stress (Kim et al. 2013). Also in grapevines, VQ14 and V17 may cooperate with VvWRKY25 and VvWRKY27 to mediate abiotic stress responses (Wang et al. 2015).

WRKY protein contain clustered proline-directed serines (SP clusters) as potential phosphorylation sites of MPKs (Ishihama and Yoshioka, 2012). Further, phosphorylation of WRKY by MPKs could be an important mean to transduce the signal to the nucleus, a fact also pointed by previous evaluations with MPKs that implicated in abiotic tolerance. In *Arabidopsis*, MPK4 and MPK6 are activated by cold, salt and drought stress. During *salt* and *cold* stresses, MPK6 (and *MPK4*) was *activated* by MKK2 (Teige et al. 2004), while MPK3/6 activated *WRKY33* expression. *In vivo,* WRKY33 protein autoregulates their expressions, via a positive feedback loop by binding to their own promoter (Mao et al. 2011). Upon drought stress, rice OsWRKY30 can be phosphorylated by MPK3 and MPK6, which then lead to an enhanced drought tolerance (Shen et al. 2012). According to Banerjee and Roychoudhury (2015), the MAPK cascades involved in phosphorylating WRKYs associated with abiotic stress are less studied in comparison to the biotic counterparts. Another significant interaction of WRKY33 was observed regarding STZ, a zinc-finger protein that exhibits a repressor domain activated under severe salt stress to control survival mechanisms, leading to increased salt stress tolerance (Sakamoto et al. 2004; Mittler et al. 2006). Transgenic lines of soybean overexpressing GmWRKY54 showed enhanced salt and drought tolerance, maybe through regulation of STZ/Zat10 (Zhou et al. 2012).

6 Molecular modification of TFs

As a consequence of external stimuli and intracellular signals, TFs can be targets of covalent modifications (post-translational modifications, PTM), usually of enzymatic nature, during or after their respective biosynthesis (Khidekel and Hsieh-Wilson 2004). Such PTMs add functional groups to TFs, such as ethyl, methyl groups, causing changes in their features, including protein stability, binding specificity, transcriptional activity, subcellular distribution and even regarding interactions with other proteins (Filtz et al. 2014) as well as epigenetic modifications (Benayoun and Veitia 2009).

So, every functional aspect of TF may be influenced by these changes, which can act in isolated forms or interconnected (Filtz et al. 2014). Thus, cross or associated answers from various PTMs have been reported (Yang and Seto 2008). For example, acetylation of some H3 lysine (K) residues in a particular position can mask the occurrence of ubiquitination (another PTM, which also occurs in K residues), possibly signaling for protein degradation via proteasome 26S (Freiman and Tjian 2003; Tootle and Rebay 2005). Considering TFs as regulators of multiple processes, even a potential PTM code, such as the proposed histone code, has been proposed (Benayoun and Veitia 2009), highlighting the importance of these modifications.

Among several databases currently available on the web, a bank that covers different PTM proteins is available (69 types; http://ptmcode.embl.de; Minguez et al. 2013), including phosphorylation, acetylation, methylation, adenylation, glycosylation, ubiquitination, hydroxylation, ADP-ribosylation, amidation, carboxylation, sulfation and prenylation (Minguez et al. 2013). Some of these PTM deserve mentioning, as most frequently cited in relation to TFs, as phosphorylation, acetylation, ubiquitination, and sumoylation.

Protein phosphorylations are the most frequent PTMs, being also the better studied (Cieśla et al. 2011; Khoury et al. 2011). They regard reversible modifications that result in conformational changes in protein structure, generally associated with biological activation of proteins, turning them on or off. In turn, dephosphorylation, promoted by a phosphatase enzyme, reverses the situation and may lead to inactivation of the protein. The addition of one or more phosphate groups by a kinase at an amino acid of a target protein chain – usually Serine (S), Threonine (T) or, to a lesser extent, Tyrosine (Y) – increases the negative charge around these modifications. Such additions expose other amino acids located more centrally in the protein, in a way that a non-polar

and hydrophobic region of a protein can become polar and hydrophilic, leading to a conformational change in the protein structure and interaction with other residues (Kitchen et al. 2008).

TFs are generally phosphorylated at multiple sites, and this phosphorylation process may regulate the function of a given TF (Komeili and Shea 1999). Many studies aiming at the understanding of how phosphorylation/dephosphorylation regulates TFs function are available and the process is well documented (e.g. Jackson 1992; Whitmarsh and Davis 2000; Peck 2003).

An important initiative generated a database covering triplets of genes, including the ability of a TF to regulate a second gene (the target gene) which is dependent on one or more PTMs, that are catalyzed by a third gene, usually a modifying enzyme (Everett et al. 2009). Given that TFs are targets of different kinases and phosphatases, phosphorylation can integrate a set of information from extracellular stimuli that induce signal transduction pathways that provide a prompt response with the required flexibility in gene regulation (Jackson 1992).

Another PTM regards protein acetylation, in which an acetyl group is incorporated by replacing a hydrogen atom in a given amino acid. This modification is reversible (deacetylation) and can neutralize the positive charge (generally of a K) what may change the function of a given protein (Ren and Gorovsky 2001). These changes can involve features such as DNA recognition, protein-protein interaction, and protein stability, among others. Acetylase enzymes can act on DNA-binding proteins, such as histones and TFs, or even on extra-nuclear proteins such as tubulin (Kouzarides 2000). Histone acetylation is closely related to chromatin remodeling. In this process, certain regulatory elements (which are complexed with histones) are important for the expression of certain genes in eukaryotes.

On the other hand, acetylation of non-histone proteins has been associated with the modulation of cell signaling in multiple levels and this participation would be analogous to that involving phosphorylation and dephosphorylation. In the excellent review of Spange et al. (2009) examples of proteins affected by acetylation are compiled, including TFs, highlighting biological implications brought about by each modification. Exemplary cases of TF are also presented by Polevoda and Sherman (2002), among a diversity of acetylated proteins.

In a study with human cells, Martínez-Balbás et al. (2002) demonstrated that acetylation of the DNA binding domain E2F1 TF influenced the ability of this TF to bind to DNA, also influencing its transcriptional activation capacity and the half-life of the respective protein. E2F is an important TF family active in the cell cycle, coordinating the transcription of genes necessary for the progression of the cell into the S phase. Expression of some genes of this family has been

also studied in other plants like rice and tobacco (Kosugi and Ohashi 2002). In *A. thaliana*, E2F1 was described by Jager et al. (2001) as a member of a multigene family with different activities.

It should be kept in mind that changes in the ability to bind to DNA depend on where acetylation occurred in the protein chain and that acetylation will not always stimulate transcription (Polevoda and Sherman 2002). According to Khidekel and Hsieh-Wilson (2004), acetylation may decrease TF binding to promoters, mitigate transcription, stop TF-coactivator ligation or still cause nuclear kidnapping of TFs. A web resource called AceK, which aims to locate acetylated sites in histones and non-histone proteins, is available at http://csb.cse.yzu.edu.tw/AceK/ (Cheng-Tsung et al. 2014).

Another PTM type that involves TFs regards ubiquitination. In this process an ubiquitin (small eukaryotic regulatory protein with 8.5 kDa) binds to the last amino acid of the chain (Glycine 76) with a K amino acid in the target protein. This binding may signalize for degradation of the target via proteasome (ubiquitin 26S proteasome system, UPS) or may change its cellular localization, its activity or even prevent or promote other protein interactions (Chan and Hill, 2001). This PTM may involve a single ubiquitin (monoubiquitination) or a chain of ubiquitin (polyubiquitination), but only the polyubiquitination of specific lysine residues (K48 and K29) signalize the protein degradation via UPS. UPS pathway influences some processes associated with plant development, including response to biotic and abiotic stresses (Lyzenga and Stone 2012). Other types of polyubiquitination have been related to diverse events, such as endocytosis, intracellular trafficking and protein activation (Pickart and Fushman, 2004). In turn, monoubiquitination has been implicated in events such as DNA repair and gene silencing (Mukhopadhyay and Riezman 2007). The process related to this PTM involves different steps: activation, conjugation, and connection that are commonly performed by enzymes as a ubiquitin-activating enzyme (UBA; E1), a ubiquitin-conjugating enzyme (UBC; E2), and ubiquitin ligase (E3). The completion of these steps culminates by the conjugation of ubiquitin to a K residue of the substrate protein (Chan and Hill 2001).

The relation of E3 ubiquitin ligases with their target proteins and plant innate immunity has been also studied. E3 ubiquitin (Ub)-ligase protein – whose importance is reflected by the presence of over 1,400 coding genes in *A. thaliana* (Mazzucotelli et al. 2006) – are classified in different families, according to their structural and functional characteristics, reflecting their specificity to given substrates. Duplan and Rivas (2014) reported a positive action of an RING-type Ub-ligase in the basal resistance activation (PTI-PAMP-triggered immunity; Jones and Dangl 2006) of *Vitis pseudoreticulata* after fungal infection, by inducing proteolysis of the target-TF (*WRKY11*) via

UPS. In this case, the target TF has been considered to be a negative regulator of plant defense response (Yu et al. 2013).

Another PTM regards sumoylation, characterized by a labile covalent bond of a c-terminal glycine residue of SUMO (small ubiquitin-like modifier protein) to a K acceptor of a target protein (Hay 2005). Thus, SUMO also competes with other PTM by K substrate. This reversible association may change the protein activity, its nuclear localization, also affecting protein-protein interaction (Verger et al. 2003). One explanation for these changes would be that the presence of multiple sites of mono-sumoylation or formation of poly-sumoylation of a protein greatly increases its mass and surface area, affecting protein interaction with other proteins and also its activity. The importance of this conformational change in the regulation of transcription is not entirely understood. However, TF sumoylation has been associated with repression of the activity of certain targets. In this regard, some mechanisms have been proposed. One is that sumoylation promotes interaction of TFs with co-repressors (as histone deacetylase co-repressors), suggesting a complex crosstalk between acetylation and sumoylation, affecting gene regulation (Gill 2015).

In *A. thaliana*, TF MYB30 is sumoylated in K283, and this has been associated with ABA (abscisic acid phytohormone) signaling, in response to abiotic stress (Zheng et al. 2012). The replacement of K283R blocks sumoylation in *A. thaliana* protoplasts. In the same way, bZIP TF ABI5 also suffers sumoylation (mainly in K391) what results in a negative regulation of ABA signaling pathway (Miura et al. 2009). In both cases, sumoylation serves to balance expression of genes that are responsive to ABA (Zheng et al. 2012). A list of these genes with variations in their expression can be accessed in Zheng et al. (2012).

It is evident that TFs and their amino acid residues are subject to various PTMs. Furthermore, some residues of target proteins, such as K, are subject to different PTMs such as acetylation, sumoylation, and ubiquitination, among others, allowing various combinations and influencing general and specific responses involving many different pathways.

7 Concluding remarks and perspectives

Many studies showed that TFs are key regulators of both ABA-dependent and ABA-independent abiotic stress response. Several examples demonstrate how TFs can be co-regulated under different abiotic stress and might be part of the same signal transduction networks.

In the course of the last 15 years, significant efforts have focused on the generation and improvement of high-throughput analytical technologies for investigating the roles and mechanisms associated with TFs. Many actual findings considering responses to drought and related abiotic stresses rely on the development of omics inferences, especially genomics, transcriptomics, and proteomics, helping to uncover complex roles of TF networks.

Next Generation Sequencing is getting more reliable and cheaper, and shall bring additional evidence to a better understanding of TF function in non-model plants and groups scarcely studied, as most woody species. The availability of a higher sequencing coverage in transcriptomes is necessary, considering that TFs present discrete expression but can induce drastic changes under stress, as exemplified in this review.

Here we selected some key TFs with proposed similar functions in distinct plant species, uncovering their critical role in the induction of signaling cascades, leading to cell reprogramming after stress perception. However, many other deserve attention and possibly the combination of different TFs will be advisable in biotechnological inferences. Additional databanks and tools for bioinformatics and system biology analysis are being developed and will allow the recognition of new interactions in different species that can be compared for identification of new mechanisms considering these important molecules. Further, chromatin modeling and mechanisms of post-transcriptional regulation of TFs regard frontiers to be unveiled, whereas the possible manipulation of these processes for breeding purposes is still in its infancy.

Acknowledgements

We are highly thankful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for fellowships and financial support.

References

Alsheikh MK, Heyen BJ, Randall SK (2003) Ion binding properties of the dehydrin ERD14 are dependent upon phosphorylation. J Biol Chem 278(42):40882-9

Alvarez-Venegas R, Pien S, Sadder M, Witmer X, Grossniklaus U, Avramova Z (2003) ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. Current Biology 138(1):627–637

Alvarez-Venegas R, Avramova Z (2005) Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive Arabidopsis genes and in atx1 mutants. Nucleic Acids Res 33:5199–5207

Alvarez-Venegas R, Abdallat A Al, Guo M, Alfano JR, Avramova Z (2007) Epigenetic control of a transcription factor at the cross section of two antagonistic pathways. Epigenetics 2:106–113

Alvarez-Venegas R, Sadder M, Hlavacka A, Baluska F, Xia Y, Lu G, Firsov A, Sarath G, Moriyama H, Dubrovsky JG, Avramova Z (2006) The Arabidopsis homolog of trithorax, ATX1, binds phosphatidylinositol 5-phosphate, and the two regulate a common set of target genes. Proc Natl Acad Sci USA 103:6049–6054

Amoutzias GD, Veron AS, Weiner J 3rd, Robinson-Rechavi M, Bornberg-Bauer E, Olivr SG, Robertson DL (2007) One Billion Years of bZIP Transcription Factor Evolution: Conservation and Change in Dimerization and DNA-Binding Site Specificity. Mol Biol Evol 24:827–835

Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an *Arabidopsis* interactome map. Science 333:601-607

*Aranda B, Achuthan P, Alam-Faruque Y, Armean, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerreien S, Khadake J, Kerssemakers J, Leroy C, Menden M, Michaut M, Montecchi-Palazzi L, Neuhauser SN, Orchard S, Perreau V, Roechert B, van Eijk K, Hermjakob H (*2010*)* The IntAct molecular interaction database in 2010.*Nucleic Acids Res* 38*:*D525–D531

Atkinson NJ, Dew TP, Orfila C, Urwin PE (2011) Influence of combined biotic and abiotic stress on nutritional quality parameters in tomato (Solanum lycopersicum). J Agric Food Chem 59:9673–9682

Atkinson NJ, Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. J Exp Bot 63(10):3523–3543

Aytuna AS, Gursoy A, Keskin O (2005) Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. Bioinformatics 21:2850–2855

Babitha KC, Ramu SV, Pruthvi V, Mahesh P, Nataraja KN, Udayakumar M (2013) Co-expression of AtbHLH17 and AtWRKY28 confers resistance to abiotic stress in Arabidopsis. Transgenic Res 22(2):327-341

Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14: 83–291

Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME Suite. Nucleic Acids Res 1, 43(W1):W39-49

Baloglu MC, Eldem V, Hajyzadeh M, Unver T (2014) Genome-wide analysis of the bZIP transcription factors in cucumber. PLoS One 9: e96014

Banerjee A, Roychoudhury A (2015). WRKY Proteins: Signaling and regulation of expression during abiotic stress responses. Sci World J Article ID 807560

Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. Cell Res 21:381–395

Barbosa EGG, Leite JP, Marin SRR, Marinho JP, Corrêa Carvalho FJ, Fuganti-Pagliarini R, Farias JRB, Neumaier N, Marcelino-Guimarães FC, Yamaguchi-Shinozaki K, Nakashima K, Maruyama K, Kanamori N, Fujita Y, Yoshida T, Nepomuceno AL (2013) Overexpression of the ABA-dependent *AREB1* transcription factor from *Arabidopsis thaliana* improves soybean tolerance to water deficit. Plant Mol Biol Rep 31:719–730

Bartoli CG, Casalongué CA, Simontacchi M., Marquez-Garcia B, Foyer CH (2013) Interactions between hormone and redox signaling pathways in the control of growth and cross-tolerance to stress. Environ Exp Bot 94:73–88

Benayoun BA,Veitia RA (2009) A post-translational modification code for transcription factors: sorting through a sea of signals. Trends Cell Biol 19(5):189-197

Besseau S, Li J, Palva ET (2012) WRKY54 and WRKY70 co-operate as negative regulators of leaf senescence in *Arabidopsis thaliana*. J Exp Bot 63:2667-2679

Bird A (2007) Perceptions of epigenetics. Nature 447:396–398

Breitkreutz B-J, Stark C, Reguly T, Boucher L, Breitkreutz A, Livstone M, Oughtred R, Lackner DH, Bähler J, Wood V,Dolinski K, Tyers M (2008) BioGRID Interaction Database: 2008 update Nucleic Acids Res 36:D637–D640

Bulyk M L (2007) Protein Binding microarrays for the characterization of protein-DNA interaction. Adv Biochem Eng Biotechnol 104:65-85

Chan NL, Hill CP (2001) Defining polyubiquitin chain topology. Nat Struct Biol 8(8):650-2

Chen M, Wang QY, Cheng XG, Xu ZS, Li LC, Ye XG, Xia LQ, Ma YZ (2007) GmDREB2, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. Biochem Biophys Res Commun 353**:**299–305

Chen L, Song Y, Li S, Zhang L, Zou C, and Yu D (2012) The role of WRKY transcription factors in plant abiotic stresses. Biochim Biophys Acta 1819:120–128

Chen Y-Y, Li M-Y., Wu X-J, Huang Y, Ma J, Xiong A-S (2015) Genome-wide analysis of basic helix-loop-helix family transcription factors and their role in responses to abiotic stress in carrot. Mol Breed. 35(5):1-12

Cheng-Tsung Lu, Tzong-Yi Lee, Yu-Ju Chen, and Yi-Ju Chen (2014) An Intelligent System for Identifying Acetylated Lysine on Histones and Nonhistone Proteins. BioMed Research International 2014: ID 528650

Chinnusamy V, Zhu J (2010) Epigenetic regulation: chromatin modeling and small RNAs, in Abiotic Stress Adaptation in Plants: Physiological, Molecular and Genomic Foundation (eds Pareek AA, Sopory SK and Bohnert HJ), Springer, Dordrecht, pp. 217–241

Choi H, Hong J, Ha J, Kang J, Kim SY (2000) ABFs, a family of ABA-responsive element binding factors. J Biol Chem 275:1723-30

Chung PJ, Kim YS, Jeong JS, Park SH, Nahm BH, Kim JK (2009) The histone deacetylase OsHDAC1 epigenetically regulates the OsNAC6 gene that controls seedling root growth in rice. Plant J 5:764–776

Cieśla J, Frączyk T, Rode W (2011) Phosphorylation of basic amino acid residues in proteins: important but easily missed. Acta Biochim Pol 58(2):137-48

Ciftci-Yilmaz S, Mittler R (2008) The zinc finger network of plants. Cell Mol Life Sci 65:1150–1160

Cramer GR, Urano K, Delrot S, Pezzotti M, Shinozaki K.(2011) Effects of abiotic stress on plants: A systems biology perspective. BMC Plant Biol 11:163–177

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic acid: Emergence of a core signaling network. Annu Rev Plant Biol 61:651–679

Danquah A, de Zelicourt A, Colcombet J, Hirt H (2014) The role of ABA and MAPK signaling pathways in plant abiotic stress responses. Biotechnol Adv 32(1):40-52

Jager SM, Menges M, Bauer UM, Murra JA (2001) Arabidopsis E2F1 binds a sequence present in the promoter of S-phase-regulated gene AtCDC6 and is a member of a multigene family with differential activities. Plant Mol Biol 47(4):555-68

Dietz KJ, Vogel MO, 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

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in Arabidopsis. Trends Plant Sci 15:573–581

Duplan V, Rivas S (2014) E3 ubiquitin-ligases and their target proteins during the regulation of plant innate im.munity. Front Plant Sci 5:42

Eddy SR (1996) Hidden Markov Models. Curr Opin Struct Biol 6:361-365

Eddy SR (2009) A New Generation of Homology Search Tools Based on Probabilistic Inference. Genome Inform 23:205-211

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends Plant Sci 5 199–206

Everett L, Vo A, Hannenhalli S (2009) PTM-Switchboard-a database of posttranslational modifications of transcription factors, the mediating enzymes and target genes. Nucleic Acids Res 37:D66-71

Farooq M, Wahid A, Kobayashi N, Fujita D, Basra SMA (2009) Plant drought stress: effects, mechanisms and management. Agron Sustain Dev 29:185-212

Filtz TM, Vogel WK, Leid M (2014) Regulation of transcription factor activity by interconnected post-translational modifications. Trends Pharm Sci 35(2):76-85

Finn RD, Clements J, Eddy SR (2011) HMMER Web Server: Interactive Sequence Similarity Searching. Nucleic Acids Research 39:W29-37

Freiman RN, Tjian R (2003) Regulating the regulators: lysine modifications make their mark. Cell 112(1):11-7

Fujita M, Fujita Y, Maruyama K, et al (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. Plant J 39:863–876

Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, Hiratsu K, Ohme-takagi M, Shinozaki K, Yamaguchi-Shinozaki K (2005) AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. Plant Cell 17:3470-3488

Fujita Y, Kakashima K, Yoshida T, Katagiri T, Kidokoro S, Kanamori N, Umezawa T, Fujita M, Maruyama K, Ishiyama K et al. (2009) Three SnRK2 protein kinases are the mains positive regulators of abscisic acid signaling in response to water stress in Arabidopsis. Plant Cell Physiol 50: 2123-2132

Fujita Y, Fujita M, Shinozaki K, Yamaguchi-Shinozaki K (2011) ABA-mediated transcriptional regulation in response to osmotic stress in plants. J Plant Res 124:509–525

Fujita Y, Yoshida T, Yamaguchi-Shinozaki K (2013) Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants. Physiol Plantarum 147:15–27

Funahashi A, Tanimura N, Morohashi M, Kitano H (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. BIOSILICO:159-162

Funnell AP, Crossley M (2012) Homo- and heterodimerization in transcriptional regulation. Ad Exp Med Biol 747:105-21

Gao SQ, Chen M, Xu ZS, Zhao CP, Li L, Xu HJ, Tang YM, Zhao X, Ma YZ (2011) The soybean GmbZIP1 transcription factor enhances multiple abiotic stress tolerances in transgenic plants. Plant Mol Biol 75:537–553

Georgii E, Salojarvi J, Brosche M, Kangasjarvi J, Kaski S (2012) Targeted retrieval of gene expression measurements using regulatory models. Bioinformatics 28:2349–2356

Gill G (2005) Something about SUMO inhibits transcription. Curr Opin Genet Dev 15(5):536-41

Gordan R, Murphy K, McCord RP, Zhu C, Vedenko A, Bulyk ML (2011) Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. Genome Biol 12(R125)

Grove CA, and Walhout AJ (2008) Transcription factor functionality and transcription regulatory networks. Mol Biosyst 4:309–314

Guan Q, Wen C, Zeng H, Zhu J (2013) A KH domain-containing putative RNA-binding protein is critical for heat stress-responsive gene regulation and thermotolerance in Arabidopsis. Mol Plant 6: 386–395

Guo A, He K, Liu D, Bai S, Gu X, Wei L, Luo J (2005) DATF: a database of Arabidopsis transcription factors. Bioinformatics 21:2568–2569

Hao YJ, Wei W, Song QX, Chen HW, Zhang YQ, Wang F et al. (2011) **Soybean NAC transcription factors promote abiotic stress tolerance and lateral root formation in transgenic plants**. *Plant J* **68(2)**:302-13

Harrison SC (1991) A structural taxonomy of DNA-binding domains. Nature 353:715–719

Harrison MA. In: Phytohormones and Abiotic Stress Tolerance in Plants. Khan NA, Nazar R, Iqbal N, Anjum NA, editor. Springer, New York; 2012. Cross-Talk Between Phytohormone Signaling Pathways Under Both Optimal and Stressful Environmental Conditions; pp. 49–73

Hauser M-T, Aufsatz W, Jonak C, Luschnig C (2011) Transgenerational epigenetic inheritance in plants. Biochim Biophys Acta 1809:459–468

Hay RT (2005) SUMO: a history of modification. Mol Cell 18(1):1-12

Hegedus D, Yu M, Baldwin D, Gruber M, Sharpe A, Parkin I, et al. (2003) Molecular characterization of *Brassica napus* NAC domain transcriptional activators induced in response to biotic and abiotic stress. Plant Mol Biol 53:383–397

Hirayama T, Shinozaki K. (2010) Research on plant abiotic stress responses in the post-genome era: past, present and future. Plant J 61: 1041–1052

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res 27:297–300

Hobo T, Kowyama Y, Hattori T (1999) A bZIP factor, TRAB1, interacts with VP1 and mediates abscisic acid-induced transcription. Proc Natl Acad Sci USA 96:15348–1535

Hu H, Dai M, Yao J, Xiao B, Li X, Zhang Q, Xiong L. 2006. Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. Proc Natl Acad Sci USA 103:12987–12992

Hu J, Wang D, Li J, Jing G, Ning K, Xu J (2014) Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae Nannochloropsis. Sci Rep 26(4):5454

Hu R., Qi G., Kong Y., Kong D., Gao Q., Zhou G. (2010) Comprehensive analysis of NAC domain transcription factor gene family in Populus trichocarpa. BMC Plant Biol. 10, 145.

Hu Y, Chen L, Wang H, Zhang L, Wang F, Yu D (2013) Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance. Plant J 74: 730–745

Huang X-S, Liu J-H, Chen X-J (2010) Overexpression of *PtrABF* gene, a bZIP transcription factor isolated from *Poncirus trifoliata*, enhances dehydration and drought tolerance in tobacco via scavenging ROS and modulating expression of stress-responsive genes. BMC Plant Biol 10:230

Hucka M, Finney A, Sauro H., Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19(4):524-31

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L et al. (2009) InterPro: the integrative protein signature database. Nucleic Acids Res 37:D211-D215

Hurst HC (1994) Transcription factors. 1: bZIP proteins. Protein Profile 1:123-168

Jackson SP (1992) Regulating transcription factor activity by phosphorylation. Trends Cell Biol 2(4):104-8.

Jakoby M, Weisshaar B, Droge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F (2002) bZIP transcription factors in Arabidopsis. Trends Plant Sci. 7:106–111

Jaspers P, Blomster T, Brosche M, Salojarvi J, Ahlfors R, Vainonen JP, Reddy RA, Immunk R, Angenent G, Turck F et al. (2009) Unequally redundant RCD1 and SRO1 mediate stress and developmental responses and Interact with transcription factors. Plant J 60:268-279

Jaspers P, Kangasjarvi J (2010) Reactive oxygen species in abiotic stress signaling. Physiol Plant 138:405-413

Jenuwein T, Allis CD (2001) Translating the histone code. Science 293:1074–1080

Jensen MK, Kjaersgaard T, Nielsen MM, Galberg P, Petersen K, O'Shea C et al. (2010) The*Arabidopsis thaliana* NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signaling. Biochem J 426:183–196 10

Jensen MK, Lindemose S, de Masi F, Reimer JJ, Nielsen M, Perera V, Workman CT, Turck F, Grant MR, Mundy J et al. (2013) ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene NCED3 in Arabidopsis thaliana. FEBS Open Bio 3:321–327

Jeong JS, Kim YS, Baek KH, Jung H, Ha SH, Do Choi Y, Kim M, Reuzeau C, Kim JK (2010) Root-specific expression of OsNAC10 improves drought tolerance and grain yield in rice under field drought conditions. Plant Physiol 153:185–197

Jeong JS, Kim YS, Redillas MC, Jang G, Jung H, Bang SW, Choi YD, Ha SH, Reuzeau C, Kim JK, et al. (2013) OsNAC5 over expression enlarges root diameter in rice plants leading to enhanced drought tolerance and increased grain yield in the field. Plant Biotechnol J 11:101–114

Ji XY, Liu GF, Liu YJ, Zheng L, Nie XG, Wang YC (2013) The bZIP protein from*Tamarix hispida*, ThbZIP1, is ACGT elements binding factor that enhances abiotic stress signaling in transgenic Arabidopsis. BMC Plant Biol. 13:15

Jin Z, Xu W, Liu A (2014) Genomic surveys and expression analysis of bZIP gene family in castor bean (*Ricinus communis* L.). Planta 239(2):299-312

Jin J, Zhang H, Kong L, Gao G, Luo J (2013) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res 42, D1182–1187

Jing Y, Lin R (2015) The VQ Motif-Containing Protein Family of Plant-Specific Transcriptional Regulators. Plant Physiology 169:371–378

Jiang Y, Deyholos MK (2009) Functional characterization of *Arabidopsis* NaCl-inducible *WRKY25* and *WRKY33* transcription factors in abiotic stresses. Plant MolecularBiology 69: 91–105

Johnson LS, Eddy SR, Portugaly E (2010) .Hidden Markov Model Speed Heuristic and Iterative HMM Search Procedure. BMC Bioinformatics 11:431

Jones JD, Dangl JL (2006) The plant immune system. Nature. 444(7117):323-9

Kakar K, Wandrey M, Czechowski T, Gaertner T, Scheible WR, Stitt M, Torres-Jerez I, Xiao Y, Redman JC, Wu HC, Cheung F, Town CD, Udvardi MK (2008) A community resource for high-throughput quantitative RT-PCR analysis of transcription factor gene expression in *Medicago truncatula*. Plant Methods 4:18

Kang J, Choi H, Im M, Kim SY (2002) Arabidopsis basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell 14**:** 343–357

Khidekel N, Hsieh-Wilson LC (2004) A 'molecular switchboard'--covalent modifications to proteins and their impact on transcription. Org Biomol Chem 2(1):1-7

Khoury GA, Baliban, R C, Floudas, CA (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep 1:90

Kim DY, Kwon SI, Choi C, Lee H, Ahn I, Park SR, Bae SC, Lee SC, Hwang DJ (2013) Expression analysis of rice VQ genes in response to biotic and abiotic stresses. Gene 529: 208–214

Kim JB, Kang JY, Kim SY *(*2004*)* Over-expression of a transcription factor regulating ABA-responsive gene expression confers multiple stress tolerance.*Plant Biotechnol J* 2*:* 459–466

Kim JS, Mizoi J, Yoshida T, Fujita Y, Nakajima J, Ohori T, et al. (2011) An ABRE promoter sequence is involved in osmotic stress-responsive expression of the *DREB2A* gene, which encodes a transcription factor regulating drought-inducible genes in Arabidopsis. Plant Cell Physiol. 52, 2136–2146

Kim N, Moon SJ, Min MK, Choi EH, Kim JA, Koh EY, Yoon I, Byun MO, Yoo SD, Kim BG, et al. (2015) Functional characterization and reconstitution of ABA signaling components using transient gene expression in rice protoplasts. Front Plant Sci 6:614

Kitano H, Funahashi A, Matsuoka Y, Oda K (2005) Using process diagrams for the graphical representation of biological networks. Nat Biotechnol 23(8):961-966

Kitchen J, Saunders RE, Warwicker J (2008) Charge environments around phosphorylation sites in proteins. BMC Struct Biol 8:19

Komeili A, O'Shea EK (1999) Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. Science 284(5416):977-80

Kosugi S, Ohashi Y (2002) E2F sites that can interact with E2F proteins cloned from rice are required for meristematic tissue-specific expression of rice and tobacco proliferating cell nuclear antigen promoters. Plant J. 29(1):45-59

Kouzarides T (2000) Acetylation: a regulatory modification to rival phosphorylation? EMBO J 19(6):1176-9

Kovacs D, Kalmar E, Torok Z, Tompa P (2008) Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. Plant Physiol 147:381-390

Krannich CT, Maletzki L, Kurowsky C, Horn R (2015) Network Candidate Genes in Breeding for Drought Tolerant Crops. Int J Mol Sci 16:16378-400

Krasensky J, Jonak C (2012) Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. J Exp Bot. 63: 1593–1608

Kummerfeld SK, Teichmann SA (2005) DBD: a transcription factor prediction database. Nucleic Acids Res 34:D74-D81

Lai Z, Li Y, Wang F, Cheng Y, Fan B, Yu JQ, Chen Z (2011) Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. Plant Cell 23: 3824–3841

Lata C, Yadav A, Prasad M (2011) Role of Plant Transcription Factors in Abiotic Stress Tolerance. In: Abiotic Stress Response in Plants. Physiological, Biochemical and Genetic Perspectives. Ed. ISBN: 978-953-307-672-0, InTech, pp 269-296

Lee SC, Choi HW, Hwang IS, Choi DS, Hwang BK (2006) Functional roles of the pepper pathogen-induced bZIP transcription factor, CAbZIPl, in enhanced resistance to pathogen infection and environmental stresses. Planta 224:1209-1225

Li S, Zhou X, Chen L, Huang W, Yu D (2010) *Functional characterization of Arabidopsis thaliana WRKY39 in heat stress*. *Mol. Cells 29:475–483*

Li SJ, Fu QT, Chen LG, Huang WD, Yu DQ (2011)*Arabidopsis thaliana* WRKY25, WRKY26, and WRKY33 coordinate induction of plant thermotolerance. *Planta* 233(6):1237–1252

Li XY, Liu X, Yao Y, Li YH, Liu S, He CY, Li JM, Lin YY, Li L (2013) Overexpression of *Arachis hypogaea* AREB1 gene enhances drought tolerance by modulating ROS scavenging and maintaining endogenous ABA content. Int J Mol Sci 14:12827–12842

Li Y, Jing Y, Li J, Xu G, Lin R (2014) Arabidopsis VQ MOTIFCONTAINING PROTEIN29 represses seedling deetiolation by interacting with PHYTOCHROME-INTERACTING FACTOR1. Plant Physiol 164: 2068–2080

Lindemose S, O'Shea C, Jensen MK, Skriver K (2013) Structure, Function and Networks of Transcription Factors Involved in Abiotic Stress Responses. Int J Mol Sci 14:5842-5878

Liu J, Chen N, Chen F, Cai B, Santo SD, Tornielli GB, Pezzotti M, Cheng Z-M (2014a) Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*). *BMC Genomics* **15**(281)

Liu C, Mao B, Ou S, Wang W, Liu L, Wu Y, Chu C, Wang X (2014b) OsbZIP71, a bZIP transcription factor, confers salinity and drought tolerance in rice. Plant Mol Biol 84:19–36

Liu QL, Zhong M, Li S, Pan YZ, Jiang BB, Jia Y, Zhang HG (2013) Overexpression of a chrysanthemum transcription factor gene, DgWRKY3, in tobacco enhances tolerance to salt stress. Plant Physiol Biochem 69:27-33

Liu X, Chu Z (2015) Genome-wide evolutionary characterization and analysis of bZIP transcription factors and their expression profiles in response to multiple abiotic stresses in Brachypodium distachyon. BMC Genomics 16:1-15

Lu G, Gao C, Zheng X, Han B (2009) Identification of OsbZIP72 as a positive regulator of ABA response and drought tolerance in rice. Planta 229: 605–615

Lu M, Sun Q-P, Zhang D, Wang T-Y, Pan J (2015) Identification of 7 stress-related NAC transcription factor members in maize (*Zea mays* L.) and characterization of the expression pattern of these genes. Biochem Biophys Res Commun 462(2), 144-150

Luo X, Bai X, Sun X, Zhu D, Liu B, Ji W, Cai H, Cao L, Wu J, Hu M, Liu X, Tang L, Zhu Y, et al. (2013) Expression of wild soybean WRKY20 in *Arabidopsis* enhances drought tolerance and regulates ABA signalling. J Exp Bot 64:2155–2169

Lyzenga WJ, Stone SL (2012) Abiotic stress tolerance mediated by protein ubiquitination. J Exp Bot 63(2):599-616

Martínez-Balbás MA, Bauer UM, Nielsen SJ, Brehm A, Kouzarides T (2002) Regulation of E2F1 activity by acetylation. EMBO J 19(4):662-71

Mao G, Meng X, Liu Y, Zheng Z, Chen Z, Zhang S (2011) Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in *Arabidopsis*. *The Plant Cell* 23(4):1639–1653

Mazzucotelli E, Belloni S, Marone D, De Leonardis A, Guerra D, Di Fonzo N, Cattivelli L, Mastrangelo A (2006) The e3 ubiquitin ligase gene family in plants: regulation by degradation. Curr Genomics 7(8):509-22

Minguez P, Letunic I, Parca L, Bork P (2013) PTMcode: a database of known and predicted functional associations between post-translational modifications in proteins. Nucleic Acids Res 41 (Database issue):D306-11

Mittler R, Kim Y, Song L, et al (2006) Gain- and loss-of-function mutations in Zat10 enhance the tolerance of plants to abiotic stress. FEBS Lett 580:6537–6542

Miura K, Lee J, Jin JB, Yoo CY, Miura T, Hasegawa PM (2009) Sumoylation of ABI5 by the Arabidopsis SUMO E3 ligase SIZ1 negatively regulates abscisic acid signaling. Proc Natl Acad Sci USA 106(13):5418-23

Miura K, Tada Y (2014) Regulation of water; salinity and cold stress responses by salicylic acid. Front Plant Sci 5:e4

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF Family transcription factors in plant abiotic stress responses. Biochim. Biophys. Acta 1819:86–96

Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS (2010) LegumeTFDB: an integrative database of *Glycine max, Lotus japonicus* and *Medicago truncatula* transcription factors. Bioinformatics 26:290–291

Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki KShinozaki K, Tran LSP (2013) TreeTFDB: An Integrative Database of the Transcription Factors from Six Economically Important Tree Crops for Functional Predictions and Comparative and Functional Genomics. DNA Res 20:151–162

Molinier J, Ries G, Zipfel C, Hohn B (2006). Transgeneration memory of stress in plants. Nature 442: 1046–1049

Moon S-J, Han S-Y, Kim D-Y, Yoon IS, Shin D, Byun M-O, Kwo H-B, Kim B-G (2015) Ectopic expression of a hot pepper bZIP-like transcription factor in potato enhances drought tolerance without decreasing tuber yield. Plant Mol Biol 1-11 pp

Mukhopadhyay D, Riezman H (2007) Proteasome-independent functions of ubiquitin in endocytosis and signaling. Science 315(5809):201-5

Musungu B, Bhatnagar D, Brown RL, Fakhoury AM, Geisler M (2015) A *predicted protein interactome* identifies conserved global networks and disease resistance subnetworks in maize. Frontiers in Genetics 6:201

Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009). Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. Plant physiology 149(1):88-95

Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) NAC transcription factors in plant abiotic stress responses. Biochim Biophys Acta 1819: 97–103

Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K (2014) The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. Front Plant Sci 5:170

Naika M, Shameer K, Mathew OK, Gowda R, Sowdhamini R (2013) STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in Arabidopsis and rice. Plant Cell Physiol. 54(2):e8

Nicolas P, Lecourieux D, Kappel C, Cluzet S, Cramer G, Delrot S, Lecourieux F (2014) **The basic leucine zipper transcription factor ABSCISIC ACID RESPONSE ELEMENT-BINDING FACTOR2 is an important transcriptional regulator of abscisic acid-dependent grape berry ripening processes.** *Plant physiol* **164**(1)**:**365-383

Niu, CF, Wei, W, Zhou, QY, Tian, AG, Hao, YJ, Zhang, WK, Ma, B, Lin, Q, Zhang, ZB, Zhang, JS, and Chen, SY (2012) Wheat WRKY genes TaWRKY2 and TaWRKY19 regulate abiotic stress tolerance in transgenic Arabidopsis plants. Plant Cell Environ 35: 1156-1170

Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, Kikuchi S (2010) Genome-Wide analysis of NAC transcription factor family in rice. Gene 465:30-44

Nuruzzaman M, Sharoni AM, Kikuchi S (2013) Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. Front Microbiol 4:248

Oh SJ, Song SI, Kim YS, Jang HJ, Kim SY, Kim M, et al. (2005) Arabidopsis CBF3/DREB1A and ABF3 in transgenic rice increased tolerance to abiotic stress without stunting growth. Plant Physiol 138: 341–351

Oktem HA, Eyidogan F, Selcuk F, Oz MT, Teixeira da Silva JA, Yucel M (2008) Revealing response of plants to biotic and abiotic stresses with microarray technology. Genes Genomes Genomics 2:1–35

Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, Matsubara K, Osato N, Kawai J, Carninci P, Hayashizaki Y, Suzuki K, Kojima K, Takahara Y, Yamamoto K, Kikuchi S (2003) Comprehensive analysis of NAC family genes in Oryza sativa and Arabidopsis thaliana. DNA Res 10(6):239-47

Orellana S, Yañez M, Espinoza A, Verdugo I, González E, Ruiz-Lara S, Casaretto JA (2010) The transcription factor SlAREB1 confers drought, salt stress tolerance and regulates biotic and abiotic stress-related genes in tomato. Plant, Cell and Environment 33:2191–2208

Osakabe Y, Osakabe K, Shinozaki K and Tran L-SP (2014) Response of plants to water stress. Front Plant Sci 5:86

Pan LJ, Jiang L (2014). Identification and expression of the WRKY transcription factors of *Carica papaya* in response to abiotic and biotic stresses. Mol Biol Rep 41:1215-1225

Peck SC (2003) Early phosphorylation events in biotic stress. Curr Opin Plant Biol 6(4):334-8

Pérez-Rodríguez P, Riaño-Pachón DM, Corrêa LGG, Rensing SA, Kersten B, Mueller-Roeber B (2010) PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic Acids Res 38:D822–827

Pickart CM, Fushman D (2004 Polyubiquitin chains: polymeric protein signals. Curr Opin Chem Biol. 8(6):610-6

Pinheiro C, Chaves MM (2011) Photosynthesis and drought: can we make metabolic connections from available data? J Exp Bot 62: 869–882

Pinheiro GL, Marques CS, Costa MD, Reis PA, Alves MS, Carvalho CM, Fietto LG, Fontes EP (2009) Complete inventory of soybean NAC transcription factors: sequence conservation and expression analysis uncover their distinct roles in stress response. Gene 444:10–23

Polevoda B, Sherman F (2002) The diversity of acetylated proteins. Genome Biol 3(5)

Puranik S, Sahu PP, Srivastava PS, Prasad M (2012) NAC proteins: Regulation and role in stress tolerance. Trends Plant Sci 17:369–381

Qiu YP, Yu DQ (2009) Over-expression of the stress-induced OsWRKY45 enhances disease resistance and drought tolerance in *Arabidopsis*. Environ Exp Bot 65:35–47

Rabara RC, Tripathi P, Rushton PJ (2014) The Potential of Transcription Factor-Based Genetic Engineering in Improving Crop Tolerance to Drought. OMICS 18:601-614

Rao VS, Srinivas K, Sujini GN, Sunand Kumar GN, (2014) Protein-Protein Interaction Detection: Methods and Analysis. Int J Proteomics 147648:1-12

Rasmussen S, Barah P, Suarez-Rodriguez MC, Bressendorff S, Friis P, Costantino P, Bones AM, Nielsen HB, Mundy J (2013) Transcriptome responses to combinations of stresses on Arabidopsis. Plant Physiol 161:1783–1794

Rea S, Eisenhaber F, O'Carroll D, et al (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406:593–599

Ren Q, Gorovsky MA (2001) Histone H2A.Z acetylation modulates an essential charge patch. Mol Cell 7(6):1329-35

Ren X, Chen Z, Liu Y, Zhang H, Zhang M, Liu Q, Hong X, Zhu J-K, Gong Z (2010) ABO3, a WRKY transcription factor, mediates plant responses to abscisic acid and drought tolerance in Arabidopsis. Plant J 63:417-429

Richards CL, Bossdorf O, Verhoeven KJF (2010) Understanding natural epigenetic variation. New Phytol 187:562–564

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105–2110

Rinerson CI, Rabara RC, Tripathi P, Shen QJ, Rushton PJ (2015) The evolution of WRKY transcription factors. BMC Plant Biol 15:1-18

Rushton DL, Tripathi P, Rabara RC, Lin J, Ringler P, Boken AK, Langum TJ, Smidt L, Boomsma DD, Emme NJ, et al. (2012) WRKY transcription factors: Key components in abscisic acid signalling. Plant Biotechnol J 10:2–11

Rushton PJ, Somssich IE, Ringler P, and Shen QJ (2010) WRKY transcription factors. Trends Plant Sci 15:247–258

Russo EA (1997) Epigenetic Mechanisms of Gene Regulation. Cold Spring Harb Lab Press, pp.0-3

Sakamoto H, Maruyama K, Sakuma Y, Meshi T, Iwabuchi M, et al. (2004) *Arabidopsis*Cys2/His2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. Plant Physiol 136: 2734-2746

Sanghera GS, Wani SH, Hussain W, Singh NB (2011) Engineering cold stress tolerance in crop plants. Curr. Genomics 12:30–43

Seo E, Choi DC (2015) Functional studies of transcription factors involved in plant defenses in the genomics era. Brief Funct Genomics:1-8

Shameer K, Ambika S, Varghese SM, Karaba N, Udayakumar M, Sowdhamini R (2009) STIFDB–Arabidopsis Stress-responsive Transcription Factor DataBase. Int J Plant Genomics 583429

Shen HS, Liu CT, Zhang Y, Meng XP, Zhou X, Chu CC, Wang XP (2012) OsWRKY30 is activated by MAP kinases to confer drought tolerance in rice. Plant Mol Biol 80:241–253

Shu Y, Zhang J, Ao Y, Song L, Guo C (2015) Analysis of the *Thinopyrum elongatum* transcriptome under water déficit stress. Int J Genomics 2015:265791

Siddiqui MH, Al-Khaishany MY, Al-Qutami MA, Al-Whaibi MH, Grover A, Ali HM, Al-Wahibi MS, Bukhari NA (2015) Response of different genotypes of faba bean plant to drought stress. Int J Mol Sci 16:10214-27

Schulman H (2013) Intrecellular signaling. In *Fundamental neuroscience*. 4th ed. Edited by Squire LR, Berg DK, Bloom FE, du Lac S, Ghosh A, Spitzer NC, 189-210, Chapter 9. New York: Academic Press

Schneider R, Bannister AJ, Myers FA, et al (2004) Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol 6:73–77

Spange S, Wagner T, Heinzel T, Krämer OH (2009) Acetylation of non-histone proteins modulates cellular signalling at multiple levels. Int J Biochem Cell Biol 41(1):185-98

Stam M, Viterbo A, Mol JN, Kooter JM (1998) Position-dependent methylation and transcriptional silencing of transgenes in inverted T-DNA repeats: implications for posttranscriptional silencing of homologous host genes in plants. Mol Cell Biol 18:6165–6177

Skrabanek L, Saini HK, Bader GD, Enright AJ (2008) Computational Prediction of Protein–Protein Interactions. Mol Biotechnol 38: 1–17

Smith GR, Sternberg MJ (2002) Prediction of protein-protein interactions by docking methods. Curr Opin Struct Biol 12: 28–35

Song SY, Chen Y, Chen J, Dai XY, Zhang WH (2011) Physiological mechanisms underlying OsNAC5-dependent tolerance of rice plants to abiotic stress. Planta 5:331–345

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006). BioGRID: A general repository for interaction datasets. Nucleic Acids Res 34(Database issue): D535–D539

Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403:41–4

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerater H, Li D, Meyer T, Muller R, et al. (2008).The Arabidopsis Information Resource (TAIR): Gene structure and function annotation. Nucleic Acids Res 36(Database issue):D1009–D1014

Szklarczyk, D; Franceschini, A; Kuhn, M; Simonovic, M; Roth, A; Minguez, P; Doerks, T; Stark, M; Muller, J; et al. (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids 39: D561-D568

Tang Y, Liu M, Gao S, Zhang Z, Zhao X, Zhao C, Zhang F, Chen X (2012). Molecular characterization of novel TaNAC genes in wheat and overexpression of TaNAC2a confers drought tolerance in tobacco. Physiol Plant 144:210–224

Tak H, Mhatre M (2013) Cloning and molecular characterization of a putative bZIP transcription factor VvbZIP23 from *Vitis vinifera*. Protoplasma 50(1)**:**333-345

Takuhara Y, Kobayashi M, Suzuki S (2011) Low-temperature-induced transcription factors in grapevine enhance cold tolerance in transgenic Arabidopsis plants. J Plant Physiol 168: 967–975

Takasaki H, Maruyama K, Kidokoro S, Ito Y, Fujita Y, Shinozaki K. et al. (2010) The abiotic stress-responsive NAC-type transcription factor OsNAC5 regulates stress-inducible genes and stress tolerance in rice. Mol Genet Genomics 5:173–183

Tao Z, Kou Y, Liu H, Li X, Xiao J, Wang S (2011) OsWRKY45 alleles play different roles in abscisic acid signalling and salt stress tolerance but similar roles in drought and cold tolerance in rice. J Exp Bot 62:4863-4874

Tootle TL, Rebay I (2005) Post-translational modifications influence transcription factor activity: A view from the ETS superfamily. BioEssays 27(3):285-98

Tran LS, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (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

Tran LS, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K (2007) Plant gene networks in osmotic stress response: from genes to regulatory networks. Methods in Enzymology 428: 109–128

Tripathi P, Rabara RC, Rushton PJ (2014) A systems biology perspective on the role of WRKY transcription factors in drought responses in plants. Planta 239:255–266

Tsutsui T, Kato W, Asada Y, Sako K, Sato T, Sonoda Y, Kidokoro S, Yamaguchi-Shinozaki K, Tamaoki M, Arakawa K, et al. (2009) DEAR1, a transcriptional repressor of DREB protein that mediates plant defense and freezing stress responses in Arabidopsis. J Plant Res 122:633–643

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K (2000) Arabidopsis basic leucine zíper transcription factos involved in na abscisic acid-dependent signal

trnasduction pathway under Drought and high-salinity conditions. Proc Natl Acad Sci USA 97:11632-11637

Verger A, Perdomo J, Crossley M (2003) Modification with SUMO. A role in transcriptional regulation. EMBO Rep (2):137-42

Verkest A, Byzova M, Martens C, et al (2015) Selection for improved energy use efficiency and drought tolerance in canola results in distinct transcriptome and epigenome changes. Plant Physiol 168:1185-1188

Vicente-Serrano SM, Gouveia C, Camarero JJ, Beguería S, Trigo R, López-Moreno JI, Azorín-Molina C, Pasho E, Lorenzo-Lacruz J, Revuelto J, Morán-Tejedo E, Sanchez-Lorenzo A, et al. (2013) Response of vegetation to drought timescales across global land biomes, P Natl Acad Sci 110:52–57

Vainonen JP, Jaspers P, Wrzaczek M, Lamminmaki A, Reddy RA, Vaahtera L, Brosche M, Kangasjarvi J (2012) RCD1-DREB2A interaction in leaf senescence and stress responses in Arabidopsis thaliana. Biochem J 442:573–581

Vieler A, Wu G, Tsai CH, Bullard B, Cornish AJ, Harvey C, Reca IB, Thornburg C, Achawanantakun R, Buehl CJ, et al. (2012) Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. PLoS Genet 8(11):e1003064

Wang JZ, Zhou JX, Zhang BL, Vanitha J, Ramachandran S, Jiang SY (2011) Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on sorghum. J Integr Plant Biol 53:212–231

Wang C, Deng P, Chen L, Wang X, Ma H, Hu W, Yao N, Feng Y, Chai R, Yang G, He G, et al. (2013) A wheat WRKY transcription fator TaWRKY10 confers tolerance to multiple abiotic stresses in transgenic tobacco. PLoS One 8:e65120

Wang Q, Guan Y, Wu Y, Chen H, Chen F, Chu C (2008) Overexpression of a rice OsDREB1F gene increases salt, drought, and low temperature tolerance in both *Arabidopsis* and rice. Plant Mol Biol 67:589–602

Wang L, Zhu W, Fang L, Sun X, Su L, Liang Z, Wang N, Londo JP, Li S, Xin H (2014) Genome-wide identification of WRKY family genes and their response to cold stress in *Vitis vinifera*. BMC Plant Biol 14:103

Wang M, Vannozzi A, Wang G, Zhong Y, Corso M, Cavallini E, Cheng ZM (2015) A comprehensive survey of the grapevine VQ gene family and its transcriptional correlation with WRKY proteins. Front Plant Sci 6:417

Wang Z., Libault M., Joshi T., Valliyodan B., Nguyen H., Xu D., Stacey G., Cheng J (2010) SoyDB: a knowledge database of soybean transcription factors. BMC Plant Biol 10, 14

Wang Z, Casas-Mollano JA, Xu J, et al (2015) Osmotic stress induces phosphorylation of histone H3 at threonine 3 in pericentromeric regions of *Arabidopsis thaliana*. Proc Natl Acad Sci 112:8487–8492

Watt F, Molloy PL (1998) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the Adenovirus major late promoter. Genes Dev 2:1136–1143

Wei K, Chen J, Wang Y, Chen Y, Chen S, Lin Y, Pan S, Zhong X, Xie D (2012) Genome-wide analysis of bZIP-encoding genes in maize. DNA Res (19):463–476

Whitmarsh AJ, Davis RJ (2000) Regulation of transcription factor function by phosphorylation. Cell Mol Life Sci 57(8-9):1172-83

Wu KL, Guo ZJ, Wang HH, Li J (2005) The WRKY family of transcription factors in rice and Arabidopsis and their origins. DNA Research 12(1):9-26

Wu X, Shiroto Y, Kishitani S, Ito Y, Toriyama K (2009) Enhanced heat and drought tolerance in transgenic rice seedlings overexpressing OsWRKY11 under the control of HSP101 promoter. Plant Cell Reports 28:21-30

Xiang Y, Tang N, Du H, Ye H, Xiong L (2008) Characterization of OsbZIP23 as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. Plant Physiol 148:1938–1952

Xiong L., Schumaker K.S., Zhu JK (2002) Cell signaling during cold; drought; and salt stress. Plant Cell 14:165–183

Xiong Y, Liu T, Tian C, Sun S, Li J, Chen M (2005) Transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots. Plant Mol Biol 59:191–203

Xu X, Chen C, Fan B, Chen Z (2006) Physical and functional interaction between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors. Plant Cell 18:1310-1326

Xu ZY, Kim SY, Hyeon Do Y, Kim DH, Dong T, Park Y et al. (2013) The *Arabidopsis*NAC transcription factor ANAC096 cooperates with bZIP-type transcription factors in dehydration and osmotic stress responses. *Plant Cell* 25 4708–4724

Yan H, Jia H, Chen X, Hao L, An H, Guo X (2014) The cotton WRKY transcription factor GhWRKY17 functions in drought and salt stress in transgenic *Nicotiana benthamiana* through ABA signaling and the modulation of reactive oxygen species production. Plant Cell Physiol 55(12):2060-76

Yang XJ, Seto E (2008) Lysine acetylation: codified crosstalk with other posttranslational modifications. Mol Cell 31(4):449-61

Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of cis-acting regulatory elements in osmotic-and cold-stress-responsive promoters. Trends Plant Sci 10:88–94

Yamaguchi-Shinozaki K, Shinozaki K. (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu Rev Plant Biol 57:781–803

Yokotani N, Ichikawa T, Kondou Y, Matsui M, Hirochika H, Iwabuchi M et al. (2009) Tolerance to various environmental stresses conferred by the salt-responsive rice gene ONAC063 in transgenic Arabidopsis. Planta 229:1065–107

Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2010) AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. Plant Journal 61:672–685

Yu Y, Xu W, Wang J, Wang L, Yao W, Yang Y, Xu Y, Ma F, Du Y, Wang Y (2013) The Chinese wild grapevine (Vitis pseudoreticulata) E3 ubiquitin ligase Erysiphe necator-induced RING finger protein 1 (EIRP1) activates plant defense responses by inducing proteolysis of the VpWRKY11 transcription factor. New Phyto (3):834-46

Zhang JY, Cruz DCM, Torres-Jerez I, Kang Y, Allen SN, Huhman DV, Tang Y, Murray J, Sumner LW, Udvardi MK (2014) Global reprogramming of transcription and metabolism in *Medicago truncatula* during progressive drought and after re-watering. Plant Cell Environ 37:2553–2576

Zhang X, Wang L, Meng H, Wen H, Fan Y, Zhao J (2012) Maize ABP9 enhances tolerance to multiple stresses in transgenic Arabidopsis by modulating ABA signaling and cellular levels of reactive oxygen species. Plant Mol Biol 75:365–378

Zhang Y, Gao P, Yuan JS (2010) Plant protein-protein interaction network and interactome**.** Curr Genomics 11**:**40-46

Zheng Y, Schumaker KS, Guo Y (2012) Sumoylation of transcription factor MYB30 by the small ubiquitin-like modifier E3 ligase SIZ1 mediates abscisic acid response in Arabidopsis thaliana. Proc Natl Acad Sci U S A 109(31):12822-7

Zheng L, Liu G, Meng X, Liu Y, Ji X, Li Y, et al (2013) A WRKY gene from *Tamarix hispida*, ThWRKY4, mediates abiotic stress responses by modulating reactive oxygen species and expression of stress-responsive genes *Plant Mol Biol* 82:303–32010

Zhou J, Wang X, Jiao Y, Qin Y, Liu X, He K, Chen C, Ma L, Wang J, Xiong L, et al. (2007) Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf, and panicle. Plant Mol Biol 63 591–608

Zhou Q-Y, Tian A-G, Zou H-F, Xie ZM, Lei G, Huang J, Wang CM, Wang HW, Zhang JS, Chen SY (2008) Soybean WRKY-type transcription factor genes,*GmWRKY13*, *GmWRKY21*, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic*Arabidopsis* plants *Plant Biotech J* 6(5):486–503

Zhu QH, Guo AY, Gao G, Zhong YF, Xu M, Huang M, Luo J (2007) DPTF: a database of poplar transcription factors Bioinformatics 23, 1307–1308

Zhu JK (2002) Salt and drought stress signal transduction in plants Annu Rev Plant Biol 53: 247–273

Zilberman D, Gehring M, Tran RK, et al (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription Nat Genet 39:61–69.

(Artigo para submissão na BMC Bioinformatics)

Transcription factor diversity in cowpea [*Vigna unguiculata*]: insight into water deficit responses

Transcription factor diversity in cowpea [*Vigna unguiculata*]: insight into water deficit responses

Bezerra-Neto, J.P.[1a], Pandolfi, V.[1], Barbosa-Silva, A.[2]; Kido, E.A.[1]; Nepomuceno, A.L. [3]; Brondani, R.P.V.[4]; Benko-Iseppon, A.M.[1*]

[1] Department of Genetics, Centre of Biological Sciences, Federal University of Pernambuco, Recife-PE, 50670-420, Brazil.

[2] Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Belvaux, Luxembourg.

[3] Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa da Soja, Londrina, PR, Brazil.

[4] Embrapa Arroz e Feijão, Rodovia GO-462, Km 12, Zona Rural, 75375, Santo Antonio de Goiás, GO, Brazil.

[a] e-mail: pacifico.joao@gmail.com

*Correspondence: ana.benko.iseppon@cnpq.pq.br

ABSTRACT

Due to their sessile condition, plants are often exposed to a broad range of both biotic and abiotic stresses. Therefore, they have developed intricate mechanisms to detect precise environmental changes, allowing optimal responses to adverse conditions. Transcription factors (TFs) are proteins that act together with other transcriptional regulators inducing or suppressing genes that may improve plant tolerance. These TFs often belong to large gene families, which are in some cases unique to plants, like DREB (Dehydration-Responsive Element Binding Factor), ERF (Ethylene Responsive Factor), NAC and WRKY. The present work aimed to perform an extensive computational identification of TFs represented on RNA-Seq data of two different cowpea cultivars (susceptible and resistant to drought) assembled on Trinity tool, using a python script implemented on the iTAK tool. The expression pattern profile under water deficit at different times from roots was conducted, as well as a differential expression analysis using an R/Bioconductor software package (EdgeR) for statistical analysis of replicated data counting. A total of 64 TF-families was identified, with differential expression in distinct time frames and cultivars. The results point to the fact that these genes probably have basal expression under normal conditions and are differentially up-regulated or repressed during stress conditions considering tolerant and susceptible cultivars. The identified cowpea TFs regard interesting candidates for improvement of legume resistance to abiotic stress. Evaluations including their identification, diversity, and expression may help the identification of potential candidates for biotechnological purposes.

Keywords: Computational biology; Fabaceae; Differential expression; Transcriptome assembly.

BACKGROUND

Current climate models indicate that global temperatures will increase 3-5 °C in the next 50-100 years, drastically affecting agricultural systems (IPCC, 2014), mainly by increasing of drought periods and intense heat (Mittler; Blumwald, 2010; IPCC, 2014). These changes, combined with increasing pressure on food productivity due to population growth, require crop varieties with tolerance and survivability under the main abiotic conditions, like drought, heat, and salinity (Takeda; Matsuoka, 2008; Newton et al., 2011).

Cowpea is one of the most important food and forage legumes in the semi-arid tropics that include parts of Asia, Africa, Southern Europe, Southern United States, and Central and South America (Timko; Singh, 2008). It represents a promising hope to combat food shortage due to its high nutritional value and capacity to grow under a broad range of climatic conditions such as semi-arid to subhumid conditions and is well adapted to moderate heat and drought stresses (Ishiyaku et al., 2005; Hall, 2012). In the north- and northeastern Brazilian regions, cowpea is the second most cultivated crop, representing 80% of the total grain production for human consumption (Freire-Filho et al., 2005). The ability of cowpea to survive under environmental pressure and its small genome (~620 Mb) make it an ideal crop model to study the molecular mechanisms of drought tolerance (Agbicodo et al., 2009).

Plant stress tolerance involves changes in the whole-plant, tissue, cellular, physiological and molecular levels. The combination of intrinsic changes ensures the capacity of a plant to sustain itself under unfavorable environmental conditions (Lata et al., 2011). Abiotic stresses, such as drought, high salinity, and low temperature, induce the expression of a vast number of genes. The induction of these genes under stress is regulated by complex transcriptional networks that include the perception, signal transduction, gene expression and ultimately metabolic changes in the plant, thus providing stress tolerance (Mittler; Blumwald, 2010; Lata et al., 2011; Todaka et al., 2015).

The activated genes in response to abiotic stresses and their products provide stress tolerance by the production of vital metabolic proteins and also in regulating downstream genes. Extensive transcriptome analyzes have shown that these gene products can largely be classified into two groups (Kavar et al., 2007; Benko-Iseppon et al., 2011). The first group that encodes for proteins that protect the cells from water-deficit effects, including genes that regulate the accumulation of compatible solutes (e.g. osmoprotectants, like proline, betaine, sugars, etc.); passive and active transport systems across membranes (water channel proteins and membrane transporters); and protection and stabilization of cell structures from damage by ROS (the detoxification enzymes such as catalase, superoxide dismutase, ascorbate peroxidase, etc.); fatty acid metabolism enzymes, proteinase inhibitors, ferritin and lipid-transfer proteins; and other proteins for the protection of macromolecules (LEA protein, osmotins and

chaperones). The second group of genes stimulated by abiotic stresses includes regulatory proteins that further regulate the stress signal transduction and alter gene expression and hence possibly function in stress response. They comprise several TFs, emphasizing the role of various transcriptional regulatory mechanisms in the stress signal transduction pathways. Examples include protein kinases (MAP kinase, CDP-kinase, receptor protein kinase, etc.); protein phosphatases and proteinases implicated in the regulation of stress signaling and gene expression (Seki et al., 2003; Shinozaki and Yamaguchi-Shinozaki, 2007). A comparison between rice (*Oryza sativa*) and *Arabidopsis thaliana* by Maruyama et al. (2014) demonstrated that different metabolites were accumulated under abiotic stress conditions.

TFs mediate cellular responses to the recognition of specific *cis*-regulatory DNA sequences at the promoters to turn on/off the transcription of their target genes, recruiting the RNA polymerase complex for the initiation of transcription (He et al., 2010; Franco-Zorrilla et al., 2014). The TFs can be classified into diverse families based on the sequence patterns and structural features of the DNA-binding Domain (DBD), the region in a TF that interacts directly with the DNA (Riechmann et al., 2000; He et al., 2010). In plants, TFs play many important roles in developmental processes, like organ development, and plant responses to environmental changes, since plants – as sessile organisms – have to face with a broad range of such stresses (Ramirez; Basu, 2009). Signaling cascades governing developmental and stress switches converge at the gene expression level. Transcriptional regulation may play important roles in plants, given a large number of TF-coding genes in plant genomes, ranging from 6% to 10%, depending on the organism database (Riechmann et al., 2000).

Plant genomes assign approximately 7% of their coding sequence to TFs that act together with other transcriptional regulators, including chromatin remodeling/modifying proteins, to employ or obstruct RNA polymerases to the DNA template, promoting the complexity of transcriptional regulation (Udvardi et al., 2007). The TFs interact with *cis*-elements in the promoter regions of several stress-related genes and thus up-regulate the expression of many downstream genes resulting in imparting abiotic stress tolerance (Agarwal and Jha, 2010). In *A. thaliana* genome about 1500 TFs were described which are considered to be involved in stress-responsive gene expression (Riechmann et al., 2000). Transcriptome data in numerous other plants also point to the fact that plants have several pathways that independently respond to environmental stresses, suggesting that stress tolerance or susceptibility is controlled at the transcriptional level by extremely intricate gene regulatory networks (Somerville; Somerville, 1999; Lata et al., 2011).

Transcriptomic approaches suggested that several strategies are developed by plants for survival under environmental stress conditions, with many genes up or down regulated during drought, heating or other abiotic conditions (Nakashima et al., 2014). The transcriptional regulation plays a critical role in the drought tolerance of cowpea (Diouf, 2011; Sadhukhan et al., 2014). Transcription profiling using computational approaches has therefore become the fundamental step in studies on the regulation of

gene expression, becoming a significant tool for the characterization of stress-responsive genes. Computational identification and annotation of TFs at genome scale are the first steps toward understanding the mechanism of gene expression and regulation (He et al., 2010). In the present study, we report an extensive computational identification and expression evaluation of TFs identified on transcriptome data of two different cowpea cultivars (susceptible and resistant to drought) submitted to water deficit, using computational approach.

RESULTS AND DISCUSSION

Identification of TFs in Vigna unguiculata De novo assembly: putative function and comparative analysis of TF-family profiles

The transcriptome of cowpea was obtained from sequences of RNA isolated from two cowpea cultivars, sensitive (Santo Inácio) and tolerant (Pingo de Ouro) to drought stress, as described in materials and methods section, using short reads on an Illumina HiSeq™ 2000 platform. The chosen tissue samples for RNA isolation and sequencing represented the main plant organs (leaves and roots) related with plant adaptation to environmental stresses. The *V. unguiculata* transcriptome dataset generated from this study is a new source for the identification of novel transcripts associated to abiotic stress (water deficit) responses, allowing the plant survivor under adverse conditions.

Approximately 453 million paired-end sequence reads, each 100 bp in length, were generated from RNA sequencing. Low-quality reads were filtered out before assembly. *De novo* transcriptome assembly was performed with Trinity (Grabherr et al., 2011), resulting in a total of 185.097 non-redundant transcripts from 117.1036 predicted loci. This number likely includes the alternatively spliced variants and non-coding transcripts. In fact, some transcripts generated, present no hits when aligned against GenBank, soybean and common bean genomes, with an e-value cut-off of $1e^{-05}$. The sequence dataset generated is available at NordEST databank, a private repository of cowpea consortium.

The identification of TFs in assembled transcripts from cowpea was performed via script python included on tool iTAK (http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi). The standalone version of iTAK is capable of large-scale data analyzes, comparing the sequences with Pfam models by HMM models. The presence or absence of the defining features of TFs, like DNA-binding domains and also auxiliary domains, are usually employed as the major

criteria for identification of TFs and moreover, for classification of the TFs into individual TF families. These features are used since the first attempt for the genome-wide analysis of TFs in *A. thaliana* (Riechman et al., 2000), and in several databases that are dedicated to the identification and annotation of plant TFs such as PlnTFDB (Jin et al., 2014).

By applying our pipeline, 4.822 transcripts were predicted as TF genes and classified into 64 families (Table 1), corresponding to 2.6% of all cowpea transcripts. Each predicted TF was then assigned into a specific TF-family based on its DNA-binding domain and the PlantTFDB criteria. The seven largest TF categories on cowpea are respectively bHLH, MYB-related and MYB, NAC, ERF, WRKY, Orphans and bZIP that together account 43.84 % for all TFs predicted (Figure 1).

To reinforce the TF identification, we annotated our predicted TFs against soybean (*Glycine max*) and common bean (*Phaseolus vulgaris*) genomes, via local alignments (BLAST) to check the existence of orthologous. The BLAST search used an e-value cut-off of $1e^{-05}$. Out of 4.822 transcripts, 4.788 (99.29%) resulted in hits with the soybean proteome, supporting that these are protein-coding transcripts. After alignment against *P. vulgaris* 4.668 (96.8%) transcripts presented matched with common bean genome. The high number of orthologs in *G. max* and *P. vulgaris* support their prediction in cowpea. A complete list of BLASTx hits is provided in the Additional file I.

The most abundant TF families found in cowpea include the bZIP, ERF, NAC and MYB-related and WRKY groups of TFs (Table 1 and Figure 1). TFs from these families are well-known regarding their activity in early responsive mechanisms after environmental signals perception, controlling the expression of stress-induced genes that are important for plants to be tolerant to abiotic stresses. Overexpression of NAC and bZIP families in rice significantly improved the tolerance to drought and high-salinity stresses (Jeong et al., 2010; Huang et al., 2012; Tang et al., 2012). Additionally, less prominent TFs are included here, allowing the achievement of a comprehensive study related to the diversity and action of the main TFs groups.

Table 1. Top hits among transcription factors identified by the iTAK tool on cowpea RNAseq assembly, associated with Pfam number or name, and the number of transcripts on each family.

Family Name	Best Match	Score	E-value	Domain	Pfam ID	Other Hits
bHLH	TR40688|c1_g2_i4	116.5	2,60E-36	bHLH	PF06507.8	375
MYB-related	TR26059|c0_g1_i5	76.3	1,00E-23	MYB_superfamily	PF04433.12	354
NAC	TR63852|c3_g1_i2	185.1	3,70E-57	NAC	PF02365.10	307
ERF	TR48671|c0_g1_i1	68.2	3,70E-21	AP2-ERF/ERF	PF00847.15	283
WRKY	TR16293|c0_g2_i1	107.2	1,90E-33	WRKY	PF03106.10	266
Orphans	TR49244|c0_g1_i2	102.5	5,90E-32	Orphans	G2-like	264
bZIP	TR20547|c1_g1_i5	71.6	2,70E-22	bZIP	PF00170.16	258
HB	TR39286|c0_g1_i20	92.6	6,70E-29	HB	PF02791.12	231
C3H	TR63114|c0_g1_i2	68.2	2,70E-21	C3H	PF02201.13	179
MYB	TR26460|c0_g1_i1	67.1	7,60E-21	MYB_superfamily	PF00249.26	151
FAR1	TR31004|c0_g1_i1	100.2	5,10E-31	FAR1	PF03101.10	136
B3	TR21506|c2_g5_i1	102.4	5,80E-32	B3_superfamily/B3	PF02362.16	134
G2-like	TR20554|c2_g3_i2	118.7	4,90E-37	GARP/G2-like	G2-like	125
GRAS	TR21956|c0_g2_i2	460.7	2,10E-140	GRAS	PF03514.9	124
HSF	TR50119|c2_g4_i1	117.2	2,30E-36	HSF	PF00447.12	117
C2C2-Dof	TR26905|c0_g2_i2	128.3	5,40E-40	C2C2/C2C2-Dof	PF02701.10	100
C2H2	TR38643|c0_g3_i1	124.7	1,40E-38	C2H2	PF02373.17	99
Tify	TR8114|c0_g2_i5	457.7	1,70E-139	Tify	PF03514.9	97
SBP	TR25253|c0_g1_i1	136.7	1,70E-42	SBP	PF03110.9	85
C2C2-GATA	TR8114|c0_g2_i6	457.7	1,70E-139	C2C2/C2C2-GATA	PF03514.9	77
ARF	TR86546|c1_g2_i6	123.0	2,30E-38	B3_superfamily/ARF	PF06507.8	77
Trihelix	TR24178|c0_g1_i1	95.8	8,40E-30	Trihelix	Trihelix	76
LOB	TR37432|c0_g2_i1	145.6	3,30E-45	LOB	PF03195.9	67
TUB	TR20454|c1_g2_i2	385.5	4,80E-118	TUB	PF01167.13	56
MIKC	TR38720|c0_g1_i1	104.1	2,00E-32	MADS/MIKC	PF01486.12	53
BES1	TR55873|c0_g1_i2	220.7	7,50E-68	BES1	PF05687.8	50
TCP	TR26873|c0_g1_i1	145.7	8,00E-45	TCP	PF03634.8	44
PLATZ	TR26108|c1_g1_i6	136.3	2,10E-42	PLATZ	PF04640.9	39
AP2	TR76205|c0_g1_i12	53.4	1,50E-16	AP2-ERF/AP2	PF00847.15	38

Table 1, continued

Family Name	Best Match	Score	E-value	Domain	Pfam ID	Other Hits
M-type	TR56477|c0_g1_i1	103.5	1,70E-32	MADS/M-type	PF00319.13	37
NF-YA	TR58689|c0_g1_i1	105.1	1,50E-32	CCAAT/NF-YA	PF02045.10	35
RWP-RK	TR40543|c0_g1_i7	91.9	1,20E-28	RWP-RK	PF02042.10	32
ARR-B	TR16309|c0_g1_i1	105.1	8,90E-33	GARP/ARR-B	G2-like	27
E2F-DP	TR8547|c0_g2_i2	88.1	1,50E-27	E2F-DP	PF02319.15	27
NF-YB	TR66739|c0_g1_i1	299.2	4,00E-92	CCAAT/NF-YB	CCAAT-Dr1	24
zf-HD	TR21739|c0_g1_i2	107.1	2,40E-33	zf-HD	PF04770.7	24
OFP	TR14940|c0_g1_i1	94.2	2,10E-29	OFP	PF04844.8	23
C2C2-YABBY	TR19022|c0_g2_i2	242.7	1,70E-74	C2C2/C2C2-YABBY	PF04690.8	22
C2C2-LSD	TR41676|c0_g1_i1	48.3	3,80E-15	C2C2/C2C2-LSD	PF06943.7	22
BBR-BPC	TR55848|c1_g1_i8	402.5	1,30E-122	BBR-BPC	PF06217.7	21
CAMTA	TR47417|c1_g1_i2	181.3	2,60E-56	CAMTA	PF03859.11	21
NF-YC	TR84113|c0_g2_i1	125.9	3,90E-39	CCAAT/NF-YC	NF-YC	21
SRS	TR57679|c0_g2_i1	220.8	6,90E-68	SRS	PF05142.7	16
Alfin-like	TR26957|c0_g1_i2	110.3	1,90E-34	Alfin-like	Alfin-like	16
GeBP	TR65979|c0_g1_i1	121.2	1,20E-37	GeBP	PF04504.9	15
EIL	TR82738|c1_g1_i1	519.1	4,60E-158	EIL	PF04873.8	14
DBB	TR20513|c1_g2_i10	31.2	1,10E-09	DBB	PF00643.19	14
CPP	TR65781|c0_g2_i1	54.4	6,10E-17	CPP	PF03638.10	13
DBP	TR61038|c0_g1_i8	233.5	1,80E-71	DBP	PF00481.16	12
C2C2-CO-like	TR21567|c0_g2_i1	69.5	9,40E-22	C2C2/C2C2-CO-like	PF06203.9	11
CSD	TR78228|c1_g2_i1	96.3	4,80E-30	CSD	PF00313.17	10
GRF	TR15541|c1_g1_i1	82.5	8,60E-26	GRF	PF08879.5	10
VOZ	TR19732|c0_g2_i1	423.4	9,10E-130	VOZ	VOZ	5
S1Fa-like	TR76181|c0_g2_i1	132.8	2,00E-41	S1Fa-like	PF04689.8	5
Whirly	TR14349|c0_g1_i1	202.1	1,70E-62	Whirly	PF08536.6	4
BSD	TR14264|c0_g1_i1	39.6	2,30E-12	BSD	PF03909.12	4
ULT	TR21141|c0_g1_i1	180.3	4,40E-56	ULT	ULT	3
HRT	TR22756|c0_g3_i1	56.7	9,20E-18	HRT	HRT	3
NF-X1	TR15486|c1_g1_i2	26.3	3,90E-08	NF-X1	PF01422.12	3
STAT	TR23212|c1_g1_i3	192.2	1,40E-59	STAT	STAT	1
RAV	TR53515|c3_g2_i2	97.5	2,10E-30	AP2-ERF/RAV	PF02362.16	1
SAP	TR25360|c0_g1_i1	387.0	5,50E-118	SAP	STER_AP	0
zn-clus	TR47030|c0_g1_i1	36.6	2,10E-11	zn-clus	PF00172.13	0
NOZZLE	TR39421|c0_g1_i1	29.6	3,80E-09	NOZZLE	NOZZLE_Angio	0

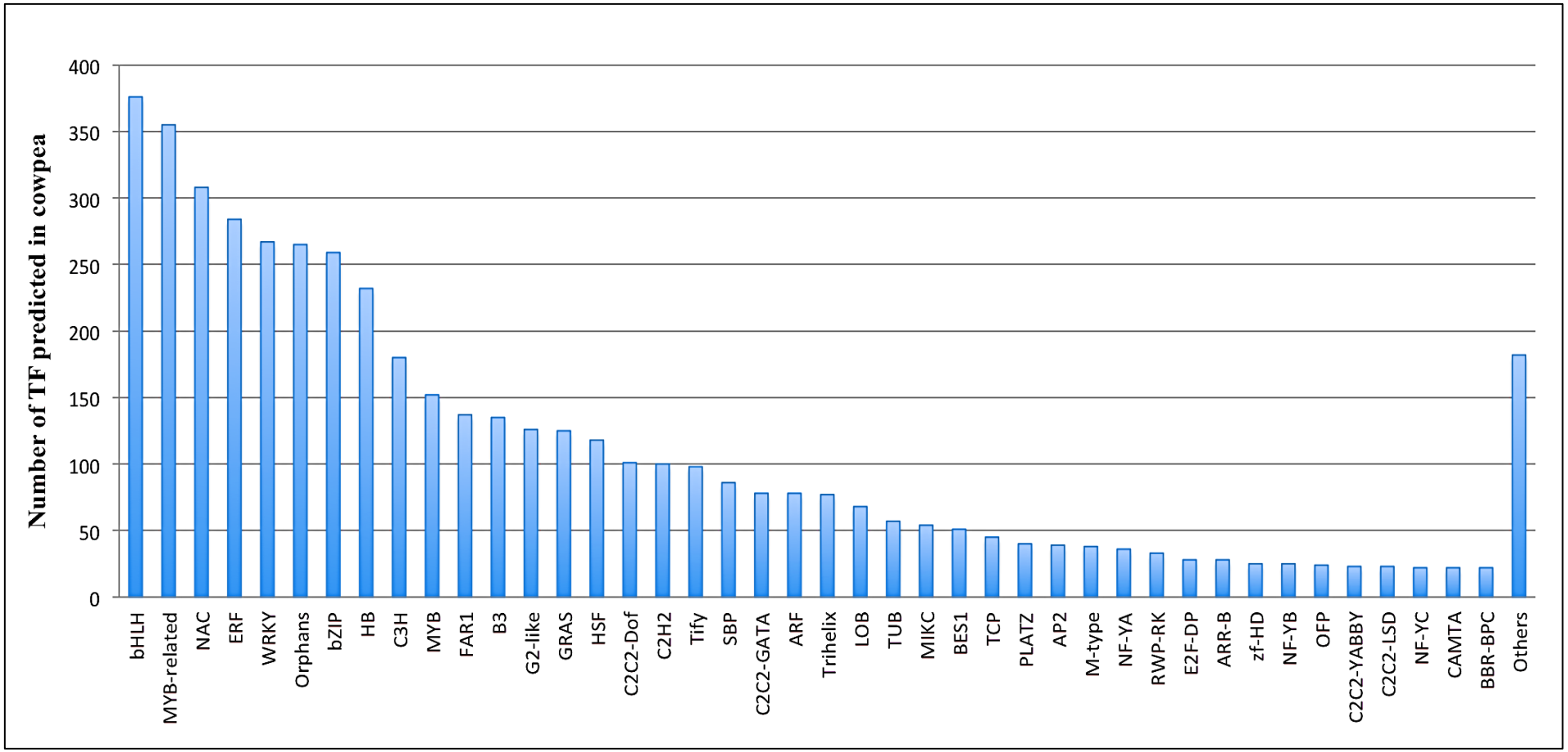


Figure 1. Distribution of *V. unguiculata* transcripts in different transcription factor families. A bar graph indicating the number of TFs predicted in cowpea distributed in various TF families. If the number of transcripts encoding for a particular TF family was lower than 20, those families were included in the 'Others' category.

The basic helix-loop-helix (bHLH) proteins constitute one of the largest TF families. They are found in all three eukaryotic kingdoms and are involved in many regulatory processes. Members of this family share the bHLH signature domain, mainly consisting of hydrophobic residues linked by a variable loop (the "helix-loop-helix" region). The involvement of bHLH members in plant developmental processes, light perception, and phytohormone signaling pathways has also been reported (Pires and Dolan, 2010; Seo *et al.*, 2011). In cowpea, 376 bHLH transcripts were identified, comprising the largest TF family found, representing 7.8% of all identified TFs. For instance, 155 genes encoding the bHLH domain were identified in common bean genome (Kavas et al., 2015) and a total of 225, 480 and 206 bHLH-encoding genes from *Arabidopsis*, soybean, and potato have been reported in plant TF databases, respectively (Jin et al., 2014).

The leucine zipper (bZIP) TFs present 259 transcripts, representing 5.37% of all TFs identified, composing one of the largest, well-known and most conserved gene families, playing key roles in diverse biological processes in plants (Yang et al., 2009; Liu and Chu, 2015). In cucumber (*Cucumis sativus*), 64 bZIP candidates were down-regulated after PEG (Polyethylene glycol) treatment (Baloglu et al., 2014). The bZIP domain shows about 60-80 amino acids (aa), with two distinct functional regions: a highly conserved basic region N-x7-R/K-x9 and a less conserved leucine zipper coiled-coil motif. Plant bZIP TFs are important in developmental processes (seed development, flower maturation, etc.) and multiple stress responses (pathogen defense, drought, cold, and salinity stresses) (Urano et al., 2010). Members of bZIP have been identified or predicted in most plant species, like *A. thaliana*, *Sorghum bicolor* and *Vitis vinifera* (Wang et al., 2011; Jin et al., 2014; Liu et al., 2014).

The APETALA2/Ethylene Responsive Factor (AP2/ERF) superfamily is one of the largest TF groups in plants and includes, at least, one APETALA2 (AP2) domain (Zhou et al., 2010). It is characterized by a conserved DNA-binding sequence of 60–70 aa, with subdivision into families based on the number of AP2 domains present in the protein sequence. The AP2 family is characterized by a double-tandem-repeat AP2 domain while the ERF family contains a single AP2 domain (Licausi et al., 2010; Hu et al., 2011). In *A. thaliana* this group can be divided into two major subfamilies (Sakuma et al., 2002), the ERF subfamily (group B) and the C-repeat binding factor/drought-responsive element binding factor (CBF/DREB) subfamily (group A). The highly conserved region from 14th to 19th aa distinguishes the DREB from the ERF group. DREBs are characterized by a valine residue at position 14 and a glutamic acid at position 19, while ERFs display an alanine and aspartic acid residue in position 14 and 19, respectively (Hu et al., 2011).

On cowpea transcriptome, 284 transcript sequences were assigned as ERFs, representing 5.88% of all TF sequences found in cowpea. Our findings are similar to those from other plant species, like *G. max*, where 120 genes were recorded (Zhang et al., 2008), as also in *Arabidopsis* and *Triticum aestivum*, with 122 and 117 putative AP2/ERF genes identified via in silico analyses (Nakano et al., 2006; Gil-Humanes et al., 2009). This high represented TF family regulates diverse processes such as vegetative and reproductive development, cell proliferation, abiotic and biotic stress responses, in addition to plant hormone responses (Sharoni et al., 2011; Kim et al., 2012).

The NAC family exhibits this name that regards an acronym derived from three genes that were initially discovered to contain the NAC domain, the gene NAM (no apical meristem), ATAF1 and 2, and CUC2 (cup-shaped cotyledon) (Souer et al., 1996; Aida et al., 1997; Nuruzzaman et al., 2015). The *NAC* genes constitute one of the largest families of plant-specific TFs and are present in a broad range of species. Cowpea presents 308 transcripts, representing 6.38% of the total amount of TFs identified, being most abundant than in *Arabidopsis* (117), rice (151) and 152 each in soybean and tobacco (Rushton et al., 2008; Le et al., 2011; Nuruzzaman et al., 2015). NAC proteins commonly possess a conserved 'NAC domain' at the N-terminus that consists of approximately 150–160 aa and is divided into five subdomains (A to E) (Shang et al., 2013). The NAC TFs are members of complex signaling pathways during plant stress responses. Numerous members play roles in transcriptional reprogramming associated with plant immune responses, and also regulating the transcriptional reprogramming related to plant abiotic responses. The tight regulation of *NAC* genes during plant stress responses contribute to the establishment of complex signaling, make them potential candidates for imparting stress tolerance (Nuruzzaman et al., 2013; Shang et al., 2013). In rice, *OsNAC6* gene is involved in both response and tolerance to biotic stress (Nakashima et al., 2007).

MYB comprises one of the largest TF families in plants, exhibiting a conserved DNA-binding domain homologous to the animal c-Myb. This domain consists of 1–4 imperfect aa sequence repeats (R0, R1, R2, and R3) of approximately 52 aa, each encoding three **α**-helices. According to the number of imperfect repeats, plant MYB proteins can be classified into three major subfamilies: the MYB-related group, which usually but not always contain a single MYB repeat, the R2R3-type group (two domains) and R1R2R3-MYB group (three domains) (Yanhui et al., 2006; Wang et al., 2015). Compared with R2R3-type MYB genes, few studies of MYB-related genes in abiotic stress response have been reported in plants (Xiong et al., 2014). We identified 355 MYB-related and 152 MYB transcripts, representing 7.36% and 3.15% of all

TFs identified in the cowpea transcriptome. Over the past decade, most of MYB genes have been associated with abiotic stress response. More than 190 MYB genes were identified in *Arabidopsis* and more than 180 members in rice (Yanhui et al., 2006). In potato, the single MYB domain TF *StMYB1R-1* has been shown to be involved in drought tolerance through activation of drought-related genes (Shin et al., 2011). The diverse functions of these genes in plant-specific processes also include primary metabolism, secondary metabolism, cell fate and identity and also organ development (Wang et al., 2015).

WRKY TFs have been identified in many different plant species, and their number in each species is highly variable. In our data 267 sequences were identified, representing 5.53% of the total amount of cowpea transcripts. In plants, WRKY comprises one of the largest TF families. They contain at least one DNA-binding domain named WRKY, which is composed of 60 aa residues, with a highly conserved heptapeptide WRKYGQK and following a non-canonical zinc finger motif (C-X_{4-5}-C-X_{22-23}-H-X_1-H or C-X_7-C-X_{23}-H-X_1-C) at its C-terminal (Rinerson et al., 2015). For instance, there are at least 74 WRKY TFs in *Arabidopsis* (Eulgem and Somssich, 2007) and 109 in rice (Ross et al., 2007). They have also been identified outside the plant kingdom, in a unicellular protist (*Giardia lamblia*), a non-photosynthetic eukaryote, indicating an ancient origin of WRKY proteins (Zhang and Wang, 2005; Rinerson et al., 2015).

The here identified TFs that were not readily aligned with other families, but have been documented and studied were placed together in the 'Orphans' class until their assignment in a given family. Here we found 265 (5.49%) Orphan TFs that comprise eight different groups of motifs described on Pfam platform. Some TFs do not belong to any of the families, thus called orphans. Some transcripts depicted vegetative specific expression and encoded for Orphan TFs promoting initiation/transition to flowering (Yilmaz et al., 2009). Orphan TFs often promote the initiation/transition to flowering, and in *Arabidopsis* its members were expressed in leaf structures, flowers, embryos, seeds, and shoots, playing a role in floral meristem determinacy and flower development (Purugganan et al., 1995; Sharma et al., 2013).

Expression patterns of TFs in roots under water deficit stress of two contrasting cowpea cultivars

Our initial objective was to perform an analysis of alterations in mRNA abundance in response to water deficit in two contrasting cowpea cultivars. Based on the FPKM (Fragments Per Kilobase Million) method, we obtained expression values for all predicted TFs in cowpea (Additional file II). The FPKM values have been achieved for each library/treatment,

uncovering the range of expression levels for cowpea TF transcripts, depending upon treatment type or cultivar. Most TFs were expressed in all libraries, with some induced expression associated to late experimental time (150 min. after water deficit). This observation confirms the fact that some TFs are activated in response to environmental changes after perception of stress signal (Farrant and Ruelland, 2015), as observed in several transcripts included in the cluster I of the heat-map (Figure 2B). The expression profile in response to external stimuli involves complex changes in defined temporal programs. In our data, the perception of an external stimulus leads to the activation of primary response TFs (see clusters I and III in Figure 2A), which may be involved in defining the proper signalization to abiotic stress tolerance after activation of other genes that act directly in plant survivor under adverse conditions (Lindemose et al., 2013; Yadav et al., 2013). Variable expression patterns were observed as well, including TFs that were expressed at initial time (without stress) and are down-regulated in a later time of stress (see cluster IV in Figure 2A), and other with constitutive expression, without drastic variation between libraries, like observed in cluster II.

Observing the top 200 high FPKM values on our data, 30 TFs families were represented (Additional file II). From these, 81 transcripts were more expressed in roots from tolerant cultivar as compared with sensitive ones at time zero (without water deficit stress), while at time 1 (75 minutes after stress) this number increased to 123 more expressed transcripts on the tolerant library, indicating that early response may be the main component of plant tolerance, with the fast activation of structural gene transcription allowing efficient root adaptation to new environmental adverse conditions. Of these 123 transcripts, 44 presented high expression level in all tolerant libraries, being more expressed in 'Pingo de Ouro' when compared with the respective treatment on the sensitive variety 'Santo Inácio'. These results reinforce the hypothesis that early expression of TFs is important to plant tolerance since 25 predicted TFs were more expressed in late tolerant tissues after stress, whereas the high expression at 75 minutes seems to represent the key strategy for the maintenance of water deficit tolerance. The top highly expressed candidates encoded for ERF, bZIP, WRKY, C3H, HSF, NAC and bHLH, most of them known to be involved in regulating plant responses to various stresses (Singh et al., 2002; Rushton et al., 2010; Mizoi et al., 2012; Nakashima et al., 2012).

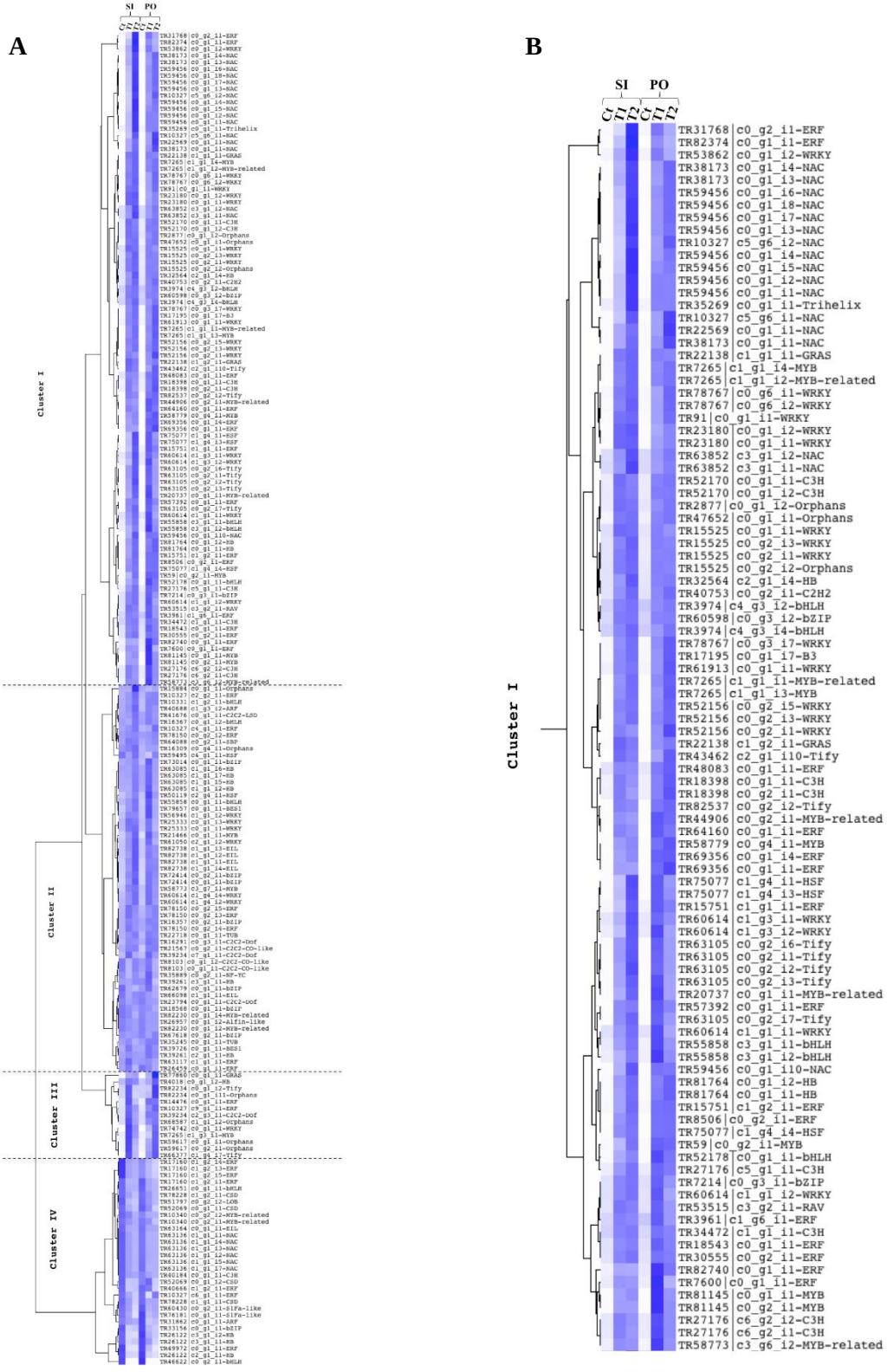


Figure 2. Gene expression profile analysis of the 200 most expressed cowpea transcription factors, based on FPKM values from the RNAseq in water deficit and control libraries. (A) Expression pattern of tolerant and sensitive genotypes and its respective experimental conditions (negative control, 75 min. and 150 min. of stress). (B) Cluster I zoomed image. SI: Santo Inácio; PO: Pingo de Ouro; Ct: negative control; T1: after 75 min. water deficit (WD); T2: after 150 min. WD Dark-blue means high expression level, light-blue means low level, and white indicates no expression. Dashed lines limit the different clusters.

Differential expression profiles of TFs from two contrasting genotypes of *V. unguiculata*

TF coding transcripts that were significantly up- or down-regulated in response to the experimental treatments were identified as differentially expressed genes (DEGs) by uniform processing of the data sets, using ≥2x-fold change (positive or negative) in abundance as criteria for differential expression in our RNAseq and HTSuperSAGE data (Additional file IV). The overall number of DEGs in each comparison between treatments regarding both cowpea cultivars is shown in Figure 3. In all cases, the number of induced DEGs exceeded the number of repressed DEGs. The highest number of TFs differentially expressed in each cultivar regarded 686 up-regulated transcripts on tolerant cultivar versus negative control and 591 on sensitive cultivar after 150 minutes of water deficit, against 302 and 431 down-regulated transcripts, respectively. The number of DEGs increased with the severity of stress in both cowpea cultivars. However, the number of reduced DEGs was greater when time after stress imposition was extended just in the sensitive cultivar. In comparison, the tolerant down-regulated transcripts presented lower induction values in late stress times when compared with early stress treatments. These results indicate that water deficit stress induced a programmed transcriptomic response in cowpea, with the most dramatic reconfiguration in plants, as described for other species. This marked adjustment of the plant transcriptome most likely reflects complex strategies of acclimatization and adaptation.

To analyze TF differential expression in cowpea cultivars, comparisons between stressed plants against their respective control in both varieties were investigated. The top 200 most differentially regulated transcripts (Up or Down) were selected to build the heat-map (Figure 4) based on RNA-Seq and SuperSAGE fold-change data (Additional file IV). A total of 1492 (30.94%) TFs were found to be differentially expressed under the analyzed the stress condition. Member of 56 TF families showed differential expression, with ERF family followed by NAC, bHLH, and WRKY been the most abundant. A significant number of TFs included in the orphan family (82 members) also showed a response to water deficit stress conditions. At least 30 TF families were differentially expressed under stress conditions with more than ten transcripts. The differential expression of several members of these TF families under stress conditions has previously also been reported (Jain et al., 2009; Ray et al., 2011; Yadav et al., 2013; Nguyen et al., 2015).

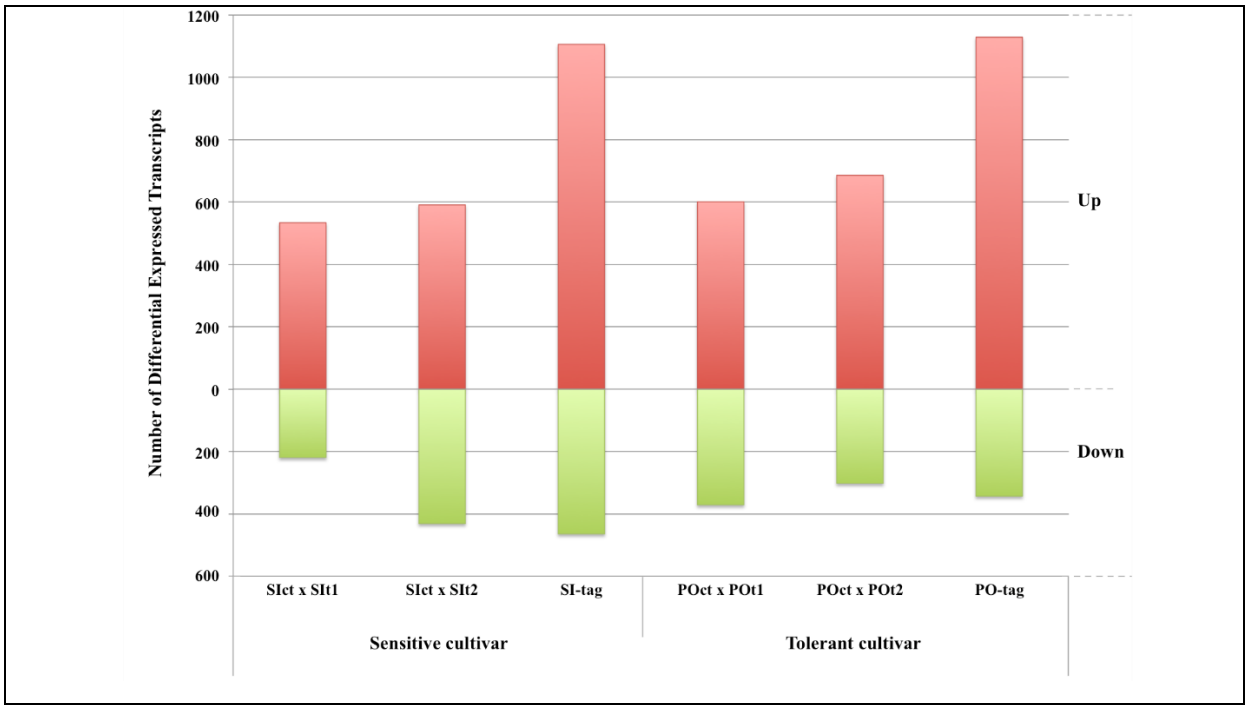


Figure 3. Number of induced and repressed genes of comparisons after water deficit treatments in different libraries of two cowpea cultivars (sensitive and tolerant) uncovered by RNAseq sequencing. Bars indicate the minimum and maximum number of DEGs recognized in each experimental condition and data (RNAseq and HTSuperSAGE methods). Induced DEGs are plotted as red bars, and repressed DEGs are plotted as green bars. SIct x SIt1 - roots sensitive control treatment *versus* roots under water deficit 75 min.; SIct x SIt2 - roots sensitive control treatment *versus* roots under water deficit 150 min.; SI-tag – roots sensible stressed cultivar *versus* control treatment based on SuperSAGE tags; POct x POt1 - roots tolerant control treatment *versus* roots under water deficit 75 min.; POct x POt2 - roots tolerant control treatment *versus* roots under water deficit 150 min.; PO-tag – roots sensitive stressed bulk *versus* control treatment based on SuperSAGE tags.

Plant adaptation to water deficit is dependent on molecular networks for stress perception, signal transduction, expression of a subset of genes and production of metabolites that protect and maintain the structure of cellular components (Umezawa et al., 2006; Hasanuzzaman et al., 2013). Transcriptional reprogramming is a central component of the response to water deficit and drought stress. The most differentially expressed TFs visualized in heat map facilitated classification of 21 TF families probably related with the early response to water deficit on roots, with ERF, NAC, WRKY, bHLH, MYB and MYB-related, representing together 68.5% of the differential expressed TFs, mostly regarding up-regulated transcripts under stress in root tissues (Figure 4). The heat-map revealed that most TFs differentially expressed in roots are up-regulated on the early stress time (75 min.) and also on the late stress treatment (150 min.) when compared with the control. This observation reinforces the presumption that plants need to regulate de TF expression levels at the beginning of stress perception, aiming the faster expression of protective genes in the bottom of signaling stress responsive cascade.

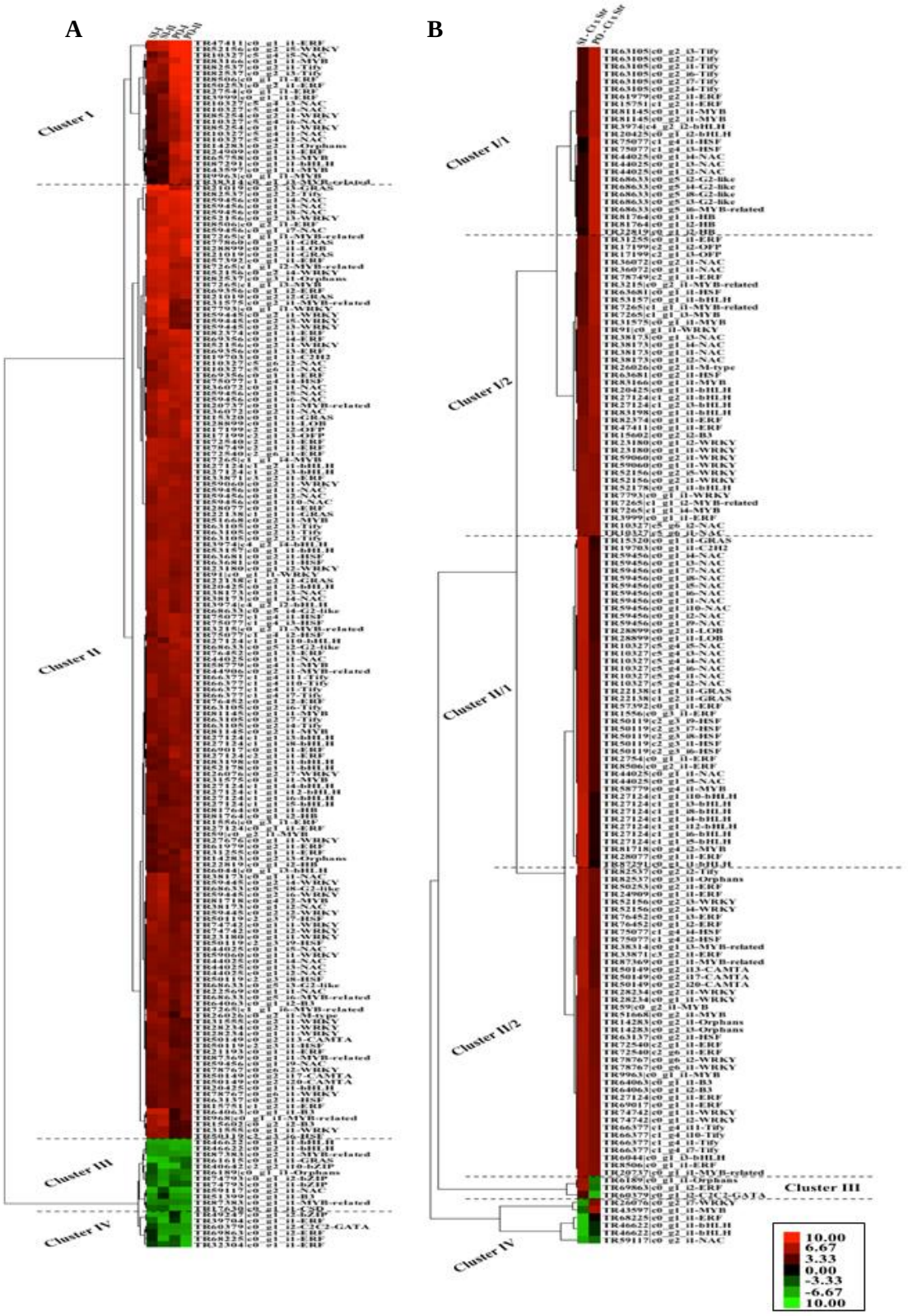


Figure 4. Hierarchical clusterization of differential expressed cowpea TFs recognized in each experimental library from RNAseq and HTSuperSAGE methods. (A) Heat-map of the 200 most differentially expressed transcript levels in roots under water deficit, based on the fold-change value of control versus stressed RNAseq libraries. SI-I: Santo Inácio (control vs. stressed 75 min.); SI-II: Santo Inácio (control vs. stressed 150 min.); PO-I: Pingo de Ouro (control vs. stressed 75 min.); PO-II: Pingo de Ouro (control vs. stressed 75 min.). (B) Differential expression of SuperSAGE tags associated with the 200 cowpea most expressed TFs based on the fold-change value of control versus stressed libraries. SI-Ct x Str: Santo Inácio (control vs. stressed bulk); PO-Ct x Str: Pingo de Ouro (control vs. stressed bulk). Red: up-regulated tags; green: down-regulated tags; black: constitutive expression. Dashed lines limit the different clusters.

The TFs related with plant response to water deficit are members of the same TF families in both cultivars. Comparing both cultivars it was evident that the tolerant one presented an increased expression of some TF families in the early time (75 min.) when compared with the sensitive cultivar. After 75 minutes of water deficit stress in root tissues, ERF, NAC, WRKY, bHLH, MYB and MYB-related were the main groups of up-regulated genes (124 transcripts) in the tolerant cultivar, while other TF families presented higher expression levels in the tolerant cultivar. A significant difference of time specificities between tolerant and sensitive cultivars was observed in cluster I (Figure 4A) for ERF, WRKY, NAC, MYB, Tify, Orphan and bHLH families, with drought-tolerant genotype presenting up-regulation since the first time of the experiment.

Comparing the two cultivars in the heat-map (Figure 4A, Clusters II, III, and IV), under dehydration treatments, it was possible to observe that the most differentially expressed TFs were regulated in the same way in both cultivars (tolerant and sensitive) under water deficit. Despite this observation, some transcripts presented different fold-change levels, indicating that both cultivars share signaling components in stress perception and response, with few TF members acting relatively independent in one cultivar or other. Regarding the NAC, ERF and MYB TF families, the up-regulation of their members are higher than the down-regulation, with 30, 28 and 24 members induced and one, four and two repressed, respectively. In turn, similarities in the expression patterns of different TF families may indicate functional correlation (Chen et al., 2013).

The tags aligned via BLASTn against the identified cowpea TFs returned 2.235 tags associated with TFs, from which 1.106 SuperSAGE tags were associated with the sensitive genotype and 1.129 with the tolerant one (Figure 3). For the majority of these anchored tags the differential gene expression was reinforced by RNA-Seq data, as indicated in the heat-map (Figure 4B and Additional file IV). A total of 159 tags (present in both genotypes) were classified as the most differentially expressed TFs, being used to build the heat-map. Regarding this differential gene expression response, the sensitive genotype (‘Santo Inácio’) showed more induced unitags than its tolerant counterpart (‘Pingo de Ouro’) as shown in the additional file IV. Comparison of the *in silico* expression based on fold-change values for both cultivars (stressed *vs*. control) can be classified into four main clusters, with clusters I and II presenting two subclusters each (Figure 4A).

Cluster I (Figure 4B) comprises tags available in both accessions and up-regulated preferentially in the tolerant cultivar in comparison with its control, especially regarding subcluster 1, including members of Tify, ERF, MYB, bHLH, HSF, NAC and other TFs. This was the case of the putative ERF (TR15751|c1_g2_i1; CpD93780) candidate that showed one of the highest expression modulation (SI - Ct x Str = 4.39 and PO - Ct x Str = 13.08). Our data present 32 candidates of ERF group differentially expressed (28 up-regulated and 4 down-regulated), with tolerant cultivars presenting the most high values. In addition to the regulation of developmental processes (Yin et al. 2010), hormonal signal

transduction and metabolism, the ERF family has been shown to perform diverse functions in imparting tolerance to a large number of biotic and abiotic stresses (Yamaguchi-Shinozaki and Shinozaki 2006; Sharoni et al., 2011; Rehman and Mahmood, 2015). Overexpression of an AP2/ERF representative improved tolerance to cold, drought and heat in *Arabidopsis* (Kang et al. 2011), and its expression under a constitutive promoter may reduce the plant growth under drought stress, but increases its capacity to preserve available water (Mawlong et al., 2015). The overexpression of this TF family in cowpea probably improves the tolerance to abiotic stress in agreement with the data obtained in RNA-seq in recent analysis of stress response in potato (*Solanum tuberosum*), where 210 AP2/ERF genes were upregulated under stress conditions (Bouaziz et al., 2013; Charfeddine et al., 2015).

In turn, cluster II includes tags mostly up-regulated in the sensitive cultivar, with the subcluster 1 comprising the most representative transcripts of this condition. This cluster comprises members of NAC, C2H2, ERF, HSF, MYB, WRKY, Orphans and others TF families, most of them presenting significant roles in abiotic stress signaling and gene expression modulation. The differences between clusters I and II, reinforce the fact that water deficit tolerance (and other stress types) in plants regards a complex process governed by multiple pathways and genes, with different TFs being up- or down-regulated in model systems and crop plants to improve stress tolerance (Lindemose et al., 2013; Pruthvi et al., 2014). Our data revealed a pattern between up- and downregulated gene sets, with genes belonging to different TF families, such as ERF, WRKY, MYB, and NAC, exhibiting transcriptional changes by dehydration in both genotypes, of which more drought-inducible genes were found than repressible genes (Figure 4A and B). Considering SuperSAGE data, 23 putative ERF genes were up-regulated in both drought sensitive and tolerant genotypes (stress/control comparison) while only one down-regulated ERF member was found in each (tolerant/sensible) root library. Other examples are among the NAC, MYB, and WRKY members, presenting 29, 18 and 16 induced genes on tolerant and sensitive cultivars, and just one member repressed of each family on sensitive cultivar, and one member of ERF and NAC repressed on tolerant cultivar, whereas no repressed members of MYB and WRKY were found (Figure 4B and Additional file IV).

Regarding the TF-encoding genes, many TF members, such as ERF, bZIP, MYB, WRKY and NAC, showed differential expression in both genotypes under water deficit. Moreover, the heat-map analysis also indicated that most of inducible TFs, such as ERF genes, exhibited higher expression levels in tolerant cultivar than sensitive, especially regarding the first time after stress. Members of these TF families and many others – that have not been studied before – may play important roles in plant responses to water deficit and other stresses by controlling transcription of downstream genes through their specific *cis*-acting elements located in the promoters of target genes (Yamaguchi-Shinozaki and Shinozaki, 2006; Jogaiah et al., 2013). In this regard, molecular tailoring of TF-encoding genes has

provided a promising approach for improvement of tolerance in a number of crops to various types of environmental stresses, including drought (Mickelbart et al., 2015; Todaka et al., 2015).

Also, four transcription factors classified like Orphan TF family were differentially expressed, with three members up-regulated and one member down-regulated in both cultivars. The Orphan group regards transcription factors that were not readily aligned with a particular family but have been documented and studied, been placed together in this new class (Davuluri et al., 2003). Our differentially expressed orphan genes present a CCT motif in its structure. The CCT (CONSTANS, CO-like, and TOC1) domain is a highly conserved basic module of ~43 amino acids, which is found near the C-terminus of plant proteins often involved in light signal transduction. The CCT domain is found in association with other domains, like regulatory domains, and has been shown to be involved in nuclear localization and probably also plays a role in protein-protein interaction (Priya and Jain, 2013).

Together, the present results suggest that induction of TFs responsive to water stress in roots of drought-tolerant cowpea cultivar (as compared with the drought-sensitive one) at early stages of stress exposure (75 min. and 150 min.) might contribute to its higher drought tolerability; and that the interaction/co-expression between different TFs are the key-role of diversity in controlling the gene expression. The diverse sets of plant stress-responsive genes are regulated by particular groups of TFs and the different combination of these factors that control the expression of different genes and favor plant survivor under unfavorable conditions (Hussain et al., 2011; Nakashima et al., 2014; Marco et al., 2015).

In summary, our comparative analysis of cowpea RNAseq data from two contrasting genotypes, tolerant and sensitive to drought, allowed us to identify putative known TFs families and their members that might contribute to the improved our knowledge about these genes, a first step for understanding plant signalization and environmental adaptation processes. Our study also provides a useful genetic resource to be applied to specific research to carry out further in-depth gene characterization and functional analyses about families here identified, also contributing to understanding the mechanisms of expression regulation triggered by drought/water deficit, comprising the main pathways of perception, signalization and adaptation responses in cowpea, which ultimately leads to development of cultivars with improved tolerance to drought.

MATERIAL AND METHODS

RNA preparation from plant materials

For root dehydration treatment, two cowpea (*V. unguiculata*) genotypes ['Pingo de Ouro' (drought-tolerant) and 'Santo Inácio' (drought-sensitive)] were used, based on the previous evaluation of Bastos *et al.* (2011). Plants were grown from seeds in a greenhouse at Embrapa-Soybean station (Londrina, Brazil) using an aerated hydroponic system in 30 L plastic containers with pH 6.6-balanced nutrient solution, as described by Kulcheski *et al*. (2010). Briefly, seeds were pre-germinated on moist filter paper in the dark at 25°C ± 1°C and in 65% ± 5% relative humidity. Plantlets were then placed in polystyrene supports, so the roots of the seedlings were completely immersed in the nutrient solution. Each seedling tray was maintained in a greenhouse at 25°C ± 2°C and in 60% ± 5% relative humidity under natural daylight (photosynthetic photon flux density (PPFD) = 1.5×10^3 µmoles m^{-2} s^{-1}, equivalent to 8.93×10^4 lux) for 12 h/day. After 15 days, seedlings with the first trifoliate leaf fully developed (V2 developmental stage) were submitted to different root dehydration periods, when the nutrient solution was removed from each plastic container where the roots were kept in the tray in the dark without nutrient solution or water for 0 minutes (negative control), 75 (T1) or 150 minutes (T2). At the end of each period, the roots of the seedlings were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction. The experimental design was a factorial (cultivars × duration of water deficit) with three replicates. Each replicate was composed of five plantlets sampled in bulk. To avoid the influence of volatile compounds, each treatment was carried out in isolated spaces presenting the same growing conditions.

RNA isolation, Sequencing, *De novo* assembly and Annotation

The samples of dried roots from hydroponic experiments were processed for RNA extraction with Trizol® Reagent (Invitrogen). To remove any DNA contamination, samples were treated with RNAse-free DNAseI (BioLabs). RNA concentration and purity were determined before and after DNAse I treatment using a NanoDropTM spectrophotometer ND-1000 (Thermo Scientific), and RNA integrity was verified by electrophoresis in a 1% agarose gel.

The cDNA library was sequenced using the Illumina HiSeq™ 2000 platform (Liu et al., 2011). After cleaning the raw reads and discarding low-quality reads, we ran Trinity (Grabherr et al., 2011) to assemble the clean reads into transcripts like described in Haas et al. (2013). First, reads were used to construct a K-mer dictionary, and each K-mer was used as an initial contig. Inchworm then examines each unique k-mer in decreasing order of abundance and generates transcript contigs using a greedy

extension based on (k-1)-mer overlaps. Next, Chrysalis cluster related Inchworm contigs into components, using the raw reads to group transcripts based on shared read support and paired reads links, when available. This procedure clusters together regions that have likely originated from alternatively spliced transcripts or closely related gene families. Chrysalis then encodes the structural complexity of clustered Inchworm contigs by building a de Bruijn graph for each cluster and partitions the reads amongst the clusters. Finally, the Butterfly tool processes the individual graphs in parallel and reconstructs transcript sequences in a manner that reflects the original cDNA molecules. The cowpea transcripts were annotated against the NR protein database in NCBI, soybean and common bean genomes by running locally the BLAST tool with a significance threshold of $e \leq 10^{-5}$.

Prediction of cowpea transcription factors

Cowpea TFs were predicted using an iTAK tool (standalone version; http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi) and alignment against the NCBI GenBank (www.ncbi.nlm.nih.gov/genbank). The iTAK is an unpublished analysis tool that has adopted the PFAM domain rules described in PlnTFDB to predict TFs genes (Jin et al., 2013). The standalone version of ITAK may be capable of large-scale data analyzes using command line optimized and deployed on a parallel computing cluster. The program compared all transcripts obtained in our assembly with the Pfam database (Punta et al., 2012) using HMMER (Eddy, 2011). Sequences that present significant similarity with any Pfam domains were kept for further analysis and assigning in different TF families according to rules described in Pérez-Rodríguez et al. (2010).

RNA-Seq Expression Analysis of cowpea transcription factors

Transcript quantification for RNA-Seq reads was performed with RSEM based on mapping the RNA-Seq reads of each experimental library [sensitive stressed (T1 and T2) and control, tolerant stressed (T1 and T2) and control] against the assembled transcriptome (Li, 2011). Given the Trinity-assembled transcripts and the RNA-Seq reads used to generate the transcriptome, RSEM aligned the reads to the *de novo* transcriptome, and then computed transcript abundance, estimating the number of RNAseq fragments corresponding to each transcript, including normalized expression values as FPKM. Hierarchical clustering analysis was carried out for the top 200 more expressed predicted TFs from tolerant (‘Pingo de Ouro’) sensitive (‘Santo Inácio’) genotypes. The FPKM counts for each transcript were clustered with the software Cluster 3.0, and the results were visualized using JAVA Treeview (Saldanha, 2004).

Analysis of differentially expressed cowpea transcription factors

To estimate differential gene expression between our libraries, we used the edgeR tool (Robinson et al., 2010), implemented in Bioconductor package (Huber et al., 2015), requiring R software for statistical computing. The differentially expressed transcripts in tolerant and sensitive cultivars ($log_2FC > 2$; FDR <5%) were identified for each experimental libraries based on comparisons between the experimental times after water deficit stress [75 (T1) or 150 minutes (T2)] against negative control (0 minutes), using the number of fragments mapping on each TF in each library in both cowpea cultivars. As the same way of expression analysis, the hierarchical clustering analysis was carried out for the top 200 more differentially expressed TF-transcripts predicted (Up- or Down-regulated) from tolerant ('Pingo de Ouro') sensitive ('Santo Inácio') genotypes. Fold-change values for each transcript in the different comparisons were clustered with the software Cluster 3.0, and the results were visualized using JAVA Treeview.

Cowpea transcription factors association with SuperSAGE tags

Identified TFs were used to screen four libraries in water deficit SuperSAGE experiment generated by the NordEST consortium. Water deficit libraries were generated using roots of two contrasting cowpea cultivars, 'Pingo de Ouro' (water deficit tolerant) and 'Santo Inácio' (water deficit sensitive). Both cultivars were submitted to water deficit in the dark for up to 150 min. Root tissues were collected at six time points (25, 50, 75, 100, 125 and 150 min post-dehydration), and all RNA extracted were bulked together to form a stressed library, in comparison with non-stressed control library. Libraries were constructed essentially as described by Matsumura et al. (2008) and were subsequently sequenced via SOLEXA platform (Illumina, Inc.). All statistical analyzes were performed like described in Ferreira-Neto et al. (2013), with values reflecting expression data (Up- or Down-regulation) associated with fold change values (FC).

The cowpea TFs were anchored against SuperTAGs from water deficit HT-SuperSAGE experiment. A local BLASTn was conducted between differentially expressed SuperTAGs (23745 tags), and TFs predicted from cowpea assembly. Only sequences that aligned with 25-26 bp transcripts without mismatches in the first bases and gaps were considered. To generate heat-maps regarding different comparisons, just unitags associated with the top 200 differentially TFs were hierarchically clustered on Cluster 3.0 software using default parameters and FC values as input data.

REFERENCES

Agarwal PK, Jha B: **Transcription factors in plants and ABA dependent and independent abiotic stress signaling**. *Biologia Plantarum* 2010,54:201-212.

Agbicodo E, Fatokun C, Muranaka S, Visser R: **Breeding drought tolerant cowpea: constraints, accomplishments, and future prospects**. *Euphytica* 2009,167:353–370.

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M: **Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant**. *Plant Cell* 1997,9:841–857.

Baloglu MC, Eldem V, Hajyzadeh M, Unver T: **Genome-wide analysis of the bZIP transcription factors in cucumber**. *PLoS One* 2014,9:e96014.

Bastos EA, Nascimento SP, Silva EM, Freire-Filho FR, Gomide RL: **Identification of cowpea genotypes for drought tolerance**. *Revista Ciência Agronômica* 2011, 42:100-107.

Benko-Iseppon AM, Soares-Cavalcanti NM, Belarmino LC, Bezerra-Neto JP, Amorim LLB, Ferreira-Neto JRC, Pandolfi V, Azevedo HMA, Silva RLO, Santos MG, Alves MVS, Kido EA: **Prospecção de genes de resistência à seca e à salinidade em plantas nativas e cultivadas**. *Revista Brasileira de Geografia Física* 2011, 6:1112-1134.

Bouaziz D, Pirrello J, Charfeddine M, Hammami A, Jbir R, Dhieb A, Bouzayen M, Gargouri-Bouzid R: **Overexpression of StDREB1 transcription factor increases tolerance to salt in transgenic potato plants**. Molecular Biotechnology 2013, 54:803–817.

Charfeddine M, Charfeddine S, Hammami A, Bouzid RG: **Genome-Wide Analysis and Expression Profiling of the ERF Transcription Factor Family in Potato (Solanum tuberosum L.).** *Mol Biotechnol.* 2015, 57(4):348-58.

Chen LM, Zhou XA, Li WB, Chang W, Zhou R, Wang C, Sha AH, Shan ZH, Zhang CJ, Qiu de Z, Yang ZL, Chen SL: **Genome-wide transcriptional analysis of two soybean genotypes under dehydration and rehydration conditions**. *BMC Genomics* 2013, 14(1):1.

Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, KurtzM, Grotewold E: **AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors**. *BMC Bioinformatics* 2003, 4:25.

Diouf D: **Recent advances in cowpea [*Vigna unguiculata* (L .) Walp.] " omics " research for genetic improvement**. *African Journal of Biotechnology* 2011, 10(15):2803-2810.

Eddy SR: **Accelerated Profile HMM Searches**. *PLoS Comput Biol* 2011, 7: e1002195.

Eulgem T, Somssich IE: **Networks of WRKY transcription factors in defense signaling.** *Curr Opin Plant Biol* 2007, 10:366–371.

Farrant JM, Ruelland E: **Plant signalling mechanisms in response to the environment**. *Environmental and Experimental Botany* 2015, 114:1–3.

Ferreira-Neto JRC, Pandolfi V, Guimaraes FCM, Benko-Iseppon AM, Romero C, Silva RLO, Rodrigues FA, Abdelnoor RV, Nepomuceno AL, Kido EA: **Early Transcriptional Response of Soybean Contrasting Accessions to Root Dehydration**. *Plos One* 2013, 8:e83466.

Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R: **DNA-binding specificities of plant transcription factors and their potential to define target genes**. *Proceedings of the National Academy of Sciences of the United States of America* 2014, 111(6):2367–72.

Freire-Filho FR, Rocha MM, Ribeiro VQ, Lopes ACA: **Adaptabilidade e estabilidade produtiva de feijão-caupi**. *Ciência Rural* 2005, 35:24-30.

Gil-Humanes J, Piston F, Martin A, Barro F: **Comparative genomic analysis and expression of the APETALA2-like genes from barley, wheat, and barley-wheat amphiploids**. *BMC Plant Biology* 2009, 9:66.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-seq data without a reference genome**. *Nat Biotechnol* 2011, 29(7):644-52.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A: **De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.** *Nat Protoc.* 2013, 8(8):1494-512. doi: 10.1038/nprot.2013.084. Epub 2013 Jul 11.

Hall AE: **Phenotyping cowpeas for adaptation to drought**. *Front Physiol* 2012, 3:155.

Hasanuzzaman M, Nahar K, Alam M, Roychowdhury R: **Physiological , Biochemical , and Molecular Mechanisms of Heat Stress Tolerance in Plants**. *Int. J. Mol. Sci.* 2013, 14(5):9643-9684.

He K, Guo AY, Gao G, Zhu QH, Liu XC, Zhang H, Chen X, Gu X, Luo J: **Computational identification of plant transcription factors and the construction of the PlantTFDB database**. *Methods Mol Biol*. 2010, 674:351-368.

Hu LF, Liu SQ: **Genome-wide identification and phylogenetic analysis of the ERF gene family in cucumbers**. *Genetics and Molecular Biology* 2011, 34:624–633.

Huang GT, Ma SL, Bai LP, Zhang L, Ma H, Jia P, Liu J, Zhong M, Guo ZF: **Signal transduction during cold, salt, and drought stresses in plants**. *Mol Biol Rep* 2012, 39: 969–987.

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M: **Orchestrating high-throughput genomic analysis with Bioconductor**. *Nat Meth*. 2015, 12(2):115-121.

Hussain SS, Kayani MA, Amjad M. **Transcription factors as tools to engineer enhanced drought stress tolerance in plants**. *Biotechnol Prog*. 2011, 27(2):297–306.

IPCC, 2014: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Ishiyaku MF, Singh BB, Craufurd PQ: **Inheritance of time to flowering in cowpea [*Vigna unguiculata* (L.) Walp.].** *Euphytica* 2005, 142:291–300.

Jain M, Khurana JP: **Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice.** *FEBS J.* 2009, 276:3148-3162.

Jeong JS, Kim YS, Baek KH, Jung H, Ha SH, Do Choi Y, Kim M, Reuzeau C, Kim JK: **Root-specific expression of *OsNAC10* improves drought tolerance and grain yield in rice under field drought conditions**. *Plant Physiol* 2010, 153:185–197.

Jin J, Zhang H, Kong L, Gao G, Luo J: **PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors.** *Nucleic Acids Res* 2014, 42:1182–1187.

Jogaiah S, Ramsandra-Govind S, Tran LS: **System biology-based approaches towards understanding drought tolerance in food crops**. *Crit. Rev. Biotechnol.* 2013, 33:23–39.

Kang HG, Kim J, Kim B, Jeong H, Choi SH, Kim EK, Lee HY, Lim PO: **Overexpression of *FTL1*/*DDF1*, an AP2 transcription factor, enhances tolerance to cold, drought, and heat stresses in *Arabidopsis thaliana*.** Plant Sci 2011, 180:634–641.

Kavar T, Maras M, Kidric M, Sustar-Vozlic J, Meglic V: **Identification of genes involved in the response of leaves of Phaseolus vulgaris to drought stress**. *Molecular Breeding* 2007, 21: 159–172.

Kavas M, Baloğlu MC, Atabay ES, Ziplar UT, Daşgan HY, Ünver T: **Genome-wide characterization and expression analysis of common bean bHLH transcription factors in response to excess salt concentration**. *Molecular Genetics and Genomics* 2015,1–15.

Kim YH, Jeong JC, Park S, Lee HS, Kwak SS: **Molecular characterization of two ethylene response factor genes in sweetpotato that respond to stress and activate the expression of defense genes in tobacco leaves**. *Journal of Plant Physiology* 2012, 169:1112–1120.

Kulcheski FR, Marcelino-Guimaraes FC, Nepomuceno AL, Abdelnoor RV, Margis R. **The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean**. *Anal Biochem*. 2010, 406:185–192.

Lata C, Yadav A, Prasad M: **Role of Plant Transcription Factors in Abiotic Stress Tolerance, Abiotic Stress Response in Plants** - *Physiological, Biochemical and Genetic Perspectives*, Prof. Arun Shanker (Ed.), ISBN: 978-953-307-672-0, InTech, DOI: 10.5772/23172.

Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K, Shinozaki K, Tran LS: **Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress.** *DNA Res.* 2011, 18:263–276.

Li B, Dewey C: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome**. *BMC bioinformatics* 2011, 12:323.

Licausi F, Giorgi FM, Zenoni S, Osti F, Pezzotti M, Perata P**: Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*.** *BMC Genomics* 2010, 11:719.

Lindemose S, O'Shea C, Jensen MK, Skriver K: **Structure, Function and Networks of Transcription Factors Involved in Abiotic Stress Responses.** *Int J Mol Sci* 2013, 14:5842-5878.

Liu J, Chen N, Chen F, Cai B, Dal Santo S, Tornielli GB, Pezzotti M, Cheng ZM: **Genome-wide analysis and expression profile of bZIP transcription factor gene family in grapevine (*Vitis vinifera*)**. *BMC Genomics* 2014,15:281–99.

Liu L, Hu N, Wang B, Chena M, Wanga J, Tiana Z, Hea Y, Lina, D: **"A brief utilization report on the Illumina HiSeq 2000 sequencer"**. *Mycology* 2011, 2(3):169–191.

Liu X, Chu Z: **Genome-wide evolutionary characterization and analysis of bZIP transcription factors and their expression profiles in response to multiple abiotic stresses in Brachypodium distachyon.** *BMC Genomics* 2015, 16:1-15.

Marco F, Bitrián M, Carrasco P, Rajam MV, Alcázar R, Tiburcio AF: **Genetic Engineering Strategies for Abiotic Stress Tolerance in Plants**. In B. Bahadur et al., eds. *Plant Biology and Biotechnology SE - 29*. Springer India, 2015:579–609.

Maruyama K, Urano K, Yoshiwara K, Morishita Y, Sakurai N, Suzuki H, Kojima M, Sakakibara H, Shibata D, Saito K, Shinozaki K, Yamaguchi-Shinozaki K: **Integrated analysis of the effects of cold and dehydration on rice metabolites, phytohormones, and gene transcripts**. *Plant Physiol.* 2014, 164:1759–1771.

Matsumura H, Kruger DH, Kahl G, Terauchi R: **SuperSAGE: A modern platform for genome-wide quantitative transcript profiling.** *Curr Pharm Biotechnol* 2008, 9:368–374.

Mawlong I, Ali K, Rai RSRD, Tyagi A: **Functional validation of a drought-responsive AP2 / ERF family transcription factor-encoding gene from rice in Arabidopsis**. *Molecular Breeding* 2015, 35(8), 1–14.

Mickelbart MV, Hasegawa PM, Bailey-Serres J: **Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability.** *Nature Publishing Group*, 2015, 16(4), 237–251.

Mittler R, Blumwald E: **Genetic engineering for modern agriculture: challenges and perspectives**. *Annual Review of Plant Biology* 2010, 61:443–462.

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K: **AP2/ERF Family transcription factors in plant abiotic stress responses.** *Biochim. Biophys. Acta* 2012, 1819:86–96.

Nakano T, Suzuki K, Fujimura T, Shinshi H: **Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice.** *Plant Physiology* 2006, 140:411–432.

Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K: **NAC transcription factors in plant abiotic stress responses.** *Biochim Biophys Acta* 2012, 1819:97–103.

Nakashima K, Tran LS, Van Nguyen D, Fujita M, Maruyama K, Todaka D, Ito Y, Hayashi N, Shinozaki K, Yamaguchi-Shinozaki K. **Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice.** *Plant J.* 2007, 51:617–630.

Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. **The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat**. *Front Plant Sci*. 2014, 5:170.

Newton AC, Johnson SN, Gregory PJ: **Implications of climate change for diseases, crop yields and food security.** *Euphytica* 2011, 179:3–18.

Nguyen HT, Stanton DE, Schmitz N, Farquhar GD, Ball MC: **Growth responses of the mangrove Avicennia marina to salinity: development and function of shoot hydraulic systems require saline conditions**. *Ann Bot.* 2015, 115(3):397-407.

Nuruzzaman M, Sharoni AM, Satoh K, Karim MR, Harikrishna JA, Shimizu T, Sasaya T, Omura T, Haque MA, Hasan SMZ, Ahmad A, Kikuchi S: **NAC transcription factor family genes are differentially expressed in rice during infections with Rice dwarf virus, Rice black-streaked dwarf virus, Rice grassy stunt virus, Rice ragged stunt virus, and Rice transitory yellowing vírus.** *Front Plant Sci*. 2015, 6: 676.

Nuruzzaman, M, Sharoni AM, Kikuchi S: **Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants.** *Front. Microbiol*. 2013, 4:248.

Pérez-Rodríguez P, Riaño-Pachón DM, Corrêa LGG, Rensing SA, Kersten B, Mueller-Roeber B: **PlnTFDB: updated content and new features of the plant transcription factor database.** *Nucleic Acids Res* 2010, 38:D822–827.

Pires N, Dolan L: **Origin and diversification of basic-helix-loop-helix proteins in plants**. *Mol Biol Evol*. 2010;27:862–874.

Priya P, Jain M: **RiceSRTFDB: A database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis**. *Database (Oxford)* 2013, 9;2013:bat027.

Pruthvi V, Narasimhan R, Nataraja KN: **Simultaneous Expression of Abiotic Stress Responsive Transcription Factors, Salinity and Drought Tolerance in Peanut (*Arachis hypogaea* L.)**. *PLoS One* 2014, 4;9(12):e111152.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD: **The Pfam protein families database.** *Nucleic Acids Res.* 2012, 40(Database issue):D290-301.

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF: **Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family.** *Genetics* 1995, 140(1)**:**345-356.

Ramirez SR, Basu C: **Comparative analyses of plant transcription factor databases.** *Curr Genomics*. 2009;10:10-17.

Ray S, Dansana PK, Giri J, Deveshwar P, Arora R, Agarwal P, Khurana JP, Kapoor S, Tyagi AK: **Modulation of transcription factor and metabolic pathway genes in response to water-deficit stress in rice.** *Funct. Integr. Genomics* 2011, 11:157-178.

Rehman S, Mahmood T: **Functional role of DREB and ERF transcription factors : regulating stress-responsive network in plants**. *Acta Physiologiae Plantarum* 2015, 37(9), 1–14.

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G: **Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes**. *Science* 2000, 290:2105–2110.

Rinerson CI, Rabara RC, Tripathi P, Shen QJ, Rushton PJ: **The evolution of WRKY transcription factors.** *BMC Plant Biol* 2015, 15:1-18.

Robinson MD, McCarthy DJ, Smyth GK: **"edgeR: a Bioconductor package for differential expression analysis of digital gene expression data".** *Bioinformatics* 2010, 26(1):139-40.

Ross CA, Liu Y, Shen QJ: **The WRKY gene family in rice (*Oryza sativa*).** *J Integr Plant Biol* 2007, 49:827–842.

Rushton PJ, Bokowiec MT, Han S, Zhang H, Brannock JF, Chen X, Laudeman TW, Timko MP: **Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae.** *Plant Physiol.* 2008, 147:280–295.

Rushton PJ, Somssich IE, Ringler P, Shen QJ: **WRKY transcription factors**. *Trends Plant Sci* 2010, 15:247–258.

Sadhukhan A, Kobayashi Y, Kobayashi Y, Tokizawa M, Yamamoto YY, Iuchi S, Koyama H, Panda SK, Sahoo L: **VuDREB2A, a novel DREB2-type transcription factor in the drought-tolerant legume cowpea, mediates DRE-dependent expression of stress-responsive genes and confers enhanced drought resistance in transgenic Arabidopsis**. *Planta*. 2014, 240(3):645-64.

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K: **DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression.** *Biochemical and Biophysical Research Communications* 2002, 290:998–1009.

Saldanha AJ: **Java Treeview - extensible visualization of microarray data**. *Bioinformatics* 2004, 20:3246-3248.

Seki M, Kamei A, Satou M, Sakurai T, Fujita M, Oono Y, Yamaguch-Shinozaki K, Shinozaki K: **Transcriptome analysis in abiotic stress conditions in higher plants.** *Topics Curr Genet* 2003, 4:271–295.

Seo J-S, Joo J, Kim M-J, Kim Y-K, Nahm BH, Song SI, Cheong J-J, Lee JS, Kim J-K, Choi YD: **OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice.** *Plant Journal* 2011, 65:907–921.

Shang H, Li W, Zou C, Yuan Y: **Analyses of the NAC Transcription Factor Gene Family in Gossypium raimondii Ulbr.: Chromosomal Location , Structure , Phylogeny , and Expression Patterns**. *J Integr Plant Biol*. 2013, 55(7):663-676.

Sharma R, De Vleesschauwer D, Sharma MK, Ronald PC: **Recent advances in dissecting stress-regulatory crosstalk in rice.** *Mol. Plant* 2013, 6:250–260.

Sharoni AM, Nuruzzaman M, Satoh K, Shimizu T, Kondoh H, Sasaya T, Choi IR, Omura T, Kikuchi S: **Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice.** *Plant and Cell Physiology* 2011, 52:344–360.

Shin D, Moon SJ, Han S, Kim BG, Park SR, Lee SK, Yoon HJ, Lee HE, Kwon HB, Baek D, Yi BY, Byun MO: **Expression of *StMYB1R-1,* a novel potato single MYB-like domain transcription factor, increases drought tolerance.** *Plant Physiol* 2011, 155: 421–432.

Shinozaki K, Yamaguchi-Shinozaki K: **Gene networks involved in drought stress tolerance and response.** *Journal of Experimental Botany* 2007, 58:221-227.

Singh AK, Choudhury BU, Bouman BAM: **"Water-wise rice production," in Proceedings of the international workshop on water-wise rice production**, eds B. A. M. Bouman, H. Hengsdijk, B. Hardy, P. S. Bindraban, T. P. Tuong, and J. K. Ladha (Los Baños, Philippines*: International Rice Research Institute*). 2002, 237–248.

Somerville C, Somerville S: **Plant Functional Genomics**. *Science* 1999, 28(5426):380-383.

Souer E, van Houwelingen A, Kloos D, Mol J, Koes R: **The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries.** *Cell* 1996, 85(2):159-70.

Takeda S, Matsuoka M: **Genetic approaches to crop improvement: responding to environmental and population changes.** *Nature Reviews Genetics* 2008, 9:444–457.

Tang N, Zhang H, Li X, Xiao J, Xiong L: **Constitutive activation of transcription factor OsbZIP46 improves drought tolerance in rice**. *Plant Physiol* 2012, 158:1755–1768.

Timko M, Singh BB: **Cowpea, a Multifunctional Legume**. In P. Moore & R. Ming (Eds.), *Genomics of Tropical Crop Plants SE - 10* (Vol. 1, pp. 227–258). *Springer New York* 2008.

Todaka D, Shinozaki K, Yamaguchi-Shinozaki K: **Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants.** *Frontiers in Plant Science*, 205, 6:1–20.

Udvardi MK, Kakar K, Wandrey M, Montanri O, Murray J, Andraiankaja A, Zhang J-Y, Benedito V, Hofer JMI, Cheng F, Town CD: **Legume transcription factors: global regulators of plant development and response to the environment**. *Plant Physiology* 2007, 144:538-549.

Umezawa T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K: **Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future**. *Curr Opin Biotechnol* 2006, 17:113–122.

Urano K, Kurihara Y, Seki M, Shinozaki K: **'Omics' analyses of regulatory networks in plant abiotic stress responses.** *Curr Opin Plant Biol.* 2010, 13:132–138.

Wang JZ, Zhou JX, Zhang BL, Vanitha J, Ramachandran S, Jiang SY: **Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on sorghum.** *J Integr Plant Biol*. 2011, 53:212–31.

Wang Z, Tang J, Hu R, Wu P, Hou X, Song X, Xiong A: **Genome-wide analysis of the R2R3-MYB transcription factor genes in Chinese cabbage (Brassica rapa ssp . pekinensis) reveals their stress and hormone responsive patterns.** *BMC Genomics* 2015, 16:17.

Xiong H, Li J, Liu P, Duan J, Zhao Y, Guo X, Li Y, Zhang H, Ali J, Li Z: **Overexpression of OsMYB48-1, a Novel MYB-Related Transcription Factor, Enhances Drought and Salinity Tolerance in Rice**. *PLoS One* 2014, 9(3):1–13.

Yadav NR, Taunk J, Rani A, Aneja B, Yadav RC**: Role of Transcription Factors in Abiotic Stress Tolerance in Crop Plants**. In *Climate Change and Plant Abiotic Stress Tolerance*. Wiley-VCH Verlag GmbH & Co. KGaA, 2013:605–640.

Yamaguchi-Shinozaki K, Shinozaki K: **Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses**. *Annu. Rev. Plant Biol.* 2006, 57:781–803.

Yang O, Popova OV, Suthoff U, Luking I, Dietz KJ, Golldack D: **The Arabidopsis basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance.** *Gene* 2009, 436:45–55.

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q: **The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family.** *Plant Mol Biol* 2006, 60:107–124.

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q: **The MYB transcription factor superfamily of *Arabidopsis:* expression analysis and phylogenetic comparison with the rice MYB family.** *Plant Molecular Biology* 2006, 60:107–124.

Yilmaz A, Nishiyama MY, Fuentes BG, Souza GM, Janies D, Gray J, Grotewold E: **GRASSIUS : A Platform for Comparative Regulatory Genomics across the Grasses**. *Plant Physiol.* 2009 149(1):171–180.

Yin XR, Allan AC, Chen KS, Ferguson IB: **Kiwifruit EIL and ERF genes involved in regulating fruit ripening.** *Plant Physiol* 2010, 153:1280–1292.

Zhang G, Chen M, Chen X, Xu Z, Guan S, Li LC, Li A, Guo J, Mao L, Ma Y: **Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (*Glycine max* L.).** *Journal of Experimental Botany* 2008, 59:4095–4107.

Zhang YJ, Wang LJ: **The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants**. *BMC Evol Biol* 2005:1–28.

Zhou ML, Ma JT, Pang JF, Zhang ZL, Tang YX, Wu YM: **Regulation of plant stress response by dehydration responsive element binding (DREB) transcription factors.** *African Journal of Biotechnology* 2010, 9:9255–9269.

(Artigo a ser submetido ao periódico *Genetics and Molecular Biology*)

Mining and comparative analysis of reactive oxygen gene network in cowpea and soybean transcriptomes

Mining and comparative analysis of reactive oxygen gene prevalence and diversity in cowpea and soybean transcriptomes

João Pacifico Bezerra-Neto[1a], Valesca Pandolfi[1], Bruna Piereck[1], Marx Oliveira de Lima[1], Ana Maria Benko-Iseppon[1*]

[1]Department of Genetics, Centre of Biological Sciences, Federal University of Pernambuco, Recife PE, 50670-420, Brazil.

[a] e-mail: pacifico.joao@ @gmail.com

*Correspondence: ana.benko.iseppon@cnpq.pq.br

Abstract

It is well known that plants exposed to different types of stress exhibit an increased expression of ROS (Reactive Oxygen Species) genes. As a consequence, plants may activate signaling pathways and cell responses including the production of antioxidant proteins, aiming to minimize damage and prevent the toxic action of ROS in plant cells. Our searches returned 1.273 cowpea transcripts and 451 soybean transcripts distributed in 15 categories of ROS genes that play important roles in oxidative stress. The number of orthologs in cowpea and soybean reflects the high complexity of detoxification system in angiosperms, which relies on antioxidant molecules and enzymes, found in all cellular compartments in higher plants. Generation of ROS occurs under a diverse range of conditions and plant tissues, with almost the same number of reads, except for seed library that presented 4.5x more reads than all other together, with 82.12% of reads identified. The high number of ROS gene members in the soybean chromosomes showed the preferential distribution in subterminal regions or next to them, with no gene-rich regions found in pericentromeric regions and moderately distrusted in intercalary regions. The ROS genes found in the present work comprised a group of antioxidant enzymes that work in concert to sustain a steady-state intracellular level of ROS promoting plant growth, development, cell cycle, hormone signaling, and reinforces responses to abiotic and biotic environmental stressors, similar to other plant species. Plant evolution lead to a tight regulation of ROS equilibrium attained through a complex gene network that operates in all subcellular compartments, being a reduction of the cell toxic levels of ROS crucial and thus conferring increased tolerance to oxidative stress.

Key-words: *oxidative stress, bioinformatics, ROS genes, Vigna unguiculata, Glycine max.*

Introduction

Cowpea [*Vigna unguiculata* (L.) Walp.] is a widely grown legume crop of semi-arid regions, especially in developing countries of South America, Africa and Asia (Ehlers and Hall, 1997). It represents a promising hope to combat food shortage due to its high nutritional value and capacity to grow under adverse conditions (Benko-Iseppon, 2001; Pimentel et al., 2002; Wang et al., 2003; Ishiyaku et al., 2005). In the north- and northeastern Brazilian regions, cowpea is the second most cultivated crop, representing 80% of the total grain production for human consumption (Freire-Filho et al., 2005). Environmental stresses, such as drought, salinity and pathogen attack, limit seed production on high-yielding sensitive cowpea genotypes, inducing phenotypic, physiological, biochemical and molecular changes in plants and also leading to significant yield shortage (Wang et al., 2003; Newton et al., 2011).

Soybean (*Glycine max* [L.] Merr.) is an important crop plant in the commodities market because of its seed protein and oil content, with a production of over 91 million tons in forecasting crop of 2014/2015 (CONAB, 2014; USDA, 2014). With the completion of soybean genome sequencing, the soybean genome has been well characterized (Schmutz et al., 2010). Currently, the soybean genome is being investigated by several groups in different countries, including studies of structural and functional genomics, transcriptomics and proteomics. Running projects also aim to study the structure and function of genes and proteins, as well as to generate an expression profile of soybean under biotic and abiotic stresses, such as rust, nematodes, drought, and nitrogen fixation. The generated data have been deposited in public databases (GenBank) and also in repositories of restricted access, as the GENOSOJA consortium (Abdelnoor et al., 2009; Benko-Iseppon et al., 2012). Multiple genes have been annotated using functional genomics methods and their functions tentatively identified.

It is well known that plant exposition to environmental stresses such as drought, salinity, high temperature (Larkindale and Knight, 2002), metal toxicity, UV-B radiation, and biotic stress conditions (such as herbivory and pathogen attack), lead to enhanced generation of ROS due to disruption of cellular homeostasis (Gill and Tuteja, 2010; Peer et al., 2013; Voothuluru and Sharp, 2013). ROS molecules are recognized for playing a dual role as both deleterious and beneficial effects, depending on their concentration in plants. At high concentrations, ROS causes damage to biomolecules, whereas at low/moderate concentration it acts as the second messenger in intracellular cascades, as signaling molecules for regulating and coordinating vital processes, including growth, cell cycle, programmed cell death, abiotic stress responses, pathogen defense, signaling and development (Tamás et al., 2012; Gill et al., 2013).

During stress conditions, the enhanced production of ROS under environmental stresses can pose a threat to cells by causing peroxidation of lipids, oxidation of proteins, damage to nucleic acids, enzyme inhibition, activation of programmed cell death (PCD) pathway and ultimately leading to cell death (Mittler, 2002; Meriga et al., 2004; Srivastava and Dubey, 2011). Major ROS molecules in cells include free radicals such as superoxide anion ($O2^{\bullet -}$), hydroxyl radical (•OH), as well as non-radical molecules like hydrogen peroxide (H_2O_2), singlet oxygen ($^1O_2$), among others. In plants, ROS are always formed by the inevitable leakage of electrons onto O_2 from the electron transport activities of chloroplasts and other cell compartments like mitochondria, peroxisomes, endoplasmic reticulum, plasma membrane, cell wall and apoplast, aspects that have been extensively reviewed (Gill and Tuteja, 2010; Sharma et al., 2012) Hydrogen peroxide is one of the major ROS compounds produced in and outside the cells during abiotic and biotic stress conditions (Voothuluru and Sharp, 2013).

Under steady state conditions, the ROS molecules are scavenged by various antioxidative defense mechanisms. This network is highly dynamic and redundant, and encodes ROS-scavenging and ROS-producing proteins, with different regulation and protective roles, besides modulation of signaling that control growth and development. Because of their multifunctional role, it is necessary for the cells to control the level of ROS tightly to avoid any oxidative injury and not to eliminate them completely. Scavenging or detoxification of ROS in excess is achieved by an efficient antioxidative system comprising non-enzymatic as well as enzymatic antioxidants (Foyer and Noctor, 2005).

The ROS-scavenging pathways from different cellular compartments are coordinated, and under normal conditions, potentially toxic oxygen metabolites are generated at a low level, and there is an appropriate balance between production and quenching of ROS (Pang and Wang, 2008). The balance may be perturbed by a number of adverse environmental factors cited previously, giving rise to rapid increases in intracellular ROS levels, which can induce oxidative damage. To avoid the oxidative damage, higher plants raise the level of endogenous antioxidant defense (Sharma et al., 2010).

The enzymatic components of the antioxidative defense system comprise several antioxidant enzymes such as superoxide dismutase (SOD), catalase (CAT), guaiacol peroxidase (GPX), enzymes of ascorbate- glutathione (AsA-GSH) cycle ascorbate peroxidase (APX), monodehydroascorbate reductase (MDHAR), dehydroascorbate reductase (DHAR), and glutathione reductase (GR) (Miller et al., 2010; Krishnamurthy and Rathinasabapathi, 2013). These enzymes operate in different subcellular compartments and respond in concert when cells are exposed to oxidative stress. Non-enzymatic components of the antioxidative defense system

include the major cellular redox buffers ascorbate (AsA) and glutathione (γ-glutamyl-cysteinyl-glycine, GSH) as well as tocopherol, carotenoids, and phenolic compounds. They interact with numerous cellular components and, in addition, play crucial roles in defense and as enzyme cofactors, these antioxidants influence plant growth and development by modulating processes from mitosis and cell elongation to senescence and cell death (Pinto and Gara, 2004; Gill et al., 2013; Krishnamurthy and Rathinasabapathi, 2013).

Osmotic stress and salinity enhance the production of ROS and cause ROS-associated injury (Serrato et al., 2004; Zhu et al., 2007; Giraud et al., 2008), whereas ROS-scavenging mechanisms have an important role in protecting plants against environmental stresses, being found in almost all cellular compartments, highlighting their importance for detoxification and cellular survival (Rizhsky et al., 2004; Koussevitzky et al., 2008; Ding et al., 2012). The rapid increase in ROS production, referred to as 'the oxidative burst', was shown to be essential for many of these processes, encoding enzymatic products like NADPH oxidases, that is one of the main ROS signal transductors in cells (Mittler et al., 2004; Torres and Dangl, 2005). Additionally, it has been shown that ROS influences the expression of a number of genes and also associated signal transduction pathways, indicating that cells have evolved strategies to use ROS as biological stimuli and as signals to activate and control various genetic stress-response programs (Dalton et al., 1999; Turpaev, 2002; Bhattacharjee, 2012).

Elucidating the distribution, structure and expression profile under different kinds of stress (biotic or abiotic), including important ROS signaling elements, focusing on its pathways, could, therefore, provide a relevant strategy to understand and possibly enhance the tolerance of important crops, like cowpea, to these environmental stress conditions. The present work aimed to identify and characterize key genes related to oxidative stress responses in cowpea transcriptome using bioinformatic tools and their differential expression levels related to environmental stresses, based on RNAseq and EST data developed by our group, also determining the structure and expression profile of the major ROS-related genes. The diversity and abundance of the identified candidates will also be comparatively evaluated using the available soybean genome and transcriptome, whereas their genomic distribution will be inferred on the basis of soybean available pseudochromosomes.

Materials and Methods

Annotation routine

The selection of seed sequences was based on a literature search in the PubMed database using the keywords "Reactive Oxygen Species" AND "Plant Stress". To identify the genes related to ROS elimination in cowpea, 125 proteins, distributed in 15 groups, that play important roles in oxidative stress response known from *Arabidopsis thaliana* were selected (Mittler et al., 2004) as seed sequences from GenBank at NCBI (National Center for Biotechnology Information). For the identification of these gene analogs in soybean and cowpea transcriptomes, tBLASTn alignments were carried out against two platforms: GENOSOJA (http://bioinfo03.ibi.unicamp.br/soja/) (Benko-Iseppon et al., 2012; Nascimento et al., 2012) and the *V. unguiculata* database (NordEST - http://bioinfo03.ibi.unicamp.br/vigna/; Benko-Iseppon et al., 2010) using the tBLASTn algorithm (Altschul et al., 1990) with a cut-off value of $1e^{-5}$. The results were annotated for further evaluations and comparisons with target organisms (cowpea and soybean).

Obtained transcripts from two databanks were annotated and analyzed for score, e-values, sequence size and presence of conserved domains, as shown in Table 1. For this purpose, all clusters were translated in six frames using a Perl script (distributed by BioPerl project) and screened for conserved motifs with the aid of rps-BLAST using CD-search tool (Altschul et al., 1990). Only orthologs presenting the fully characteristic of each ROS group selected (domain and motifs) were considered for subsequent analysis. The best matches for putative orthologs in the studied species were submitted to a BLASTx alignment in NCBI GenBank to confirm their putative function.

In silico expression assay based on Cowpea and Soybean data

A preliminary analysis of the prevalence regarding putative ROS genes in cowpea and soybean libraries was verified by direct correlation of the EST (soybean) and RNAseq (cowpea) read frequencies regarding each putative orthologs in different libraries. For practical purposes we combined some libraries that comprised various stages of the same tissue/organ in soybean, resulting in a total of seven libraries (C: cotyledons; F: flowers; H: hypocotyls; L: leaves; R: roots; S: seeds and UK: unknown). The prevalence of ROS candidates in cowpea RNAseq data was verified by direct correlation of transcript frequencies of each treatment time of NordEST

experiments. ROS gene candidates were used to screen six libraries generated by the NordEST consortium. Drought-treated libraries were generated using roots of two contrasting cowpea cultivars, 'Pingo de Ouro' (drought tolerant) and 'Santo Inácio' (drought sensitive). Both cultivars were submitted to dehydration up to 150 min. Root tissues of both accessions were collected at three time points [0 (negative control), 75 and 150 min after dehydration], and all RNA samples extracted were used to form a control (time 0) and each stressed library (times 75 and 150 min). To generate an overall picture of selected ROS genes expression patterns in both species, a hierarchical clustering approach (Eisen et al., 1998) was applied using normalized data and a graphic representation constructed with CLUSTER program, and visualization generated with TreeView program (Page, 1996). In these graphics, light blue means no or low expression and dark blue indicates all degrees of expression.

Studying cowpea ROS genes distribution among soybean genome

Best matches for five selected cowpea-ROS genes were aligned against the *G. max* pseudogenome aiming to anchor these sequences in virtual chromosomes through a BLAST tool implemented on soybean browser (Phytozome - http://www.phytozome.net/). This step aimed to infer on the distribution, relative position, and abundance of ROS genes in legumes. BLAST algorithm parameters (score, e-value, and percentage of identity) were adjusted to allow the anchoring of soybean sequences position along the soybean virtual chromosomes. Afterward, the identified anchoring positions were submitted to the Circos program (Krzywinski et al., 2009) and edited to generate a picture of higher resolution. This approach allowed the generation of a graph based on a circular organization of the soybean chromosomes (n = 20).

Results

Description and distribution of ROS genes in cowpea and soybean

The ROS gene families were classified into 15 main groups (Table 1) that comprised all known ROS-scavenging genes and NADPH oxidases, acting in concert to sustain an intracellular steady-state level of ROS, allowing plant growth, development, normal cell cycle, hormone signaling, and reinforced response to abiotic and biotic environmental stressors

(Mittler et al., 2004; Foyer and Noctor, 2005; Van Breusegem and Dat 2006). The tBLASTn alignment against the cowpea and soybean transcriptome using the 125 known ROS gene probes returned 1,273 non-redundant sequences from the EST and RNAseq data deposited in the NordEST and 451 non-redundant sequences (contigs and singlets) in the GENOSOJA database. Among them in cowpea, 667 were obtained from 'Santo Inácio' cultivar (drought sensitive accession), 366 from 'Pingo de Ouro' cultivar (drought tolerant accession) and 240 related to cowpea EST data (contigs and singlets of abiotic and biotic stresses). A graphical representation regarding the number of identified candidates per gene family in cowpea and soybean transcriptomes is shown in Figure 1, including data from all available cowpea ESTs (≈446.000) from different accessions and tissues), RNAseq data from Pingo de Ouro (drought tolerant) and Santo Inácio (drought sensitive) as compared with soybean EST database (≈~1,460,000) also from different tissues and accessions).

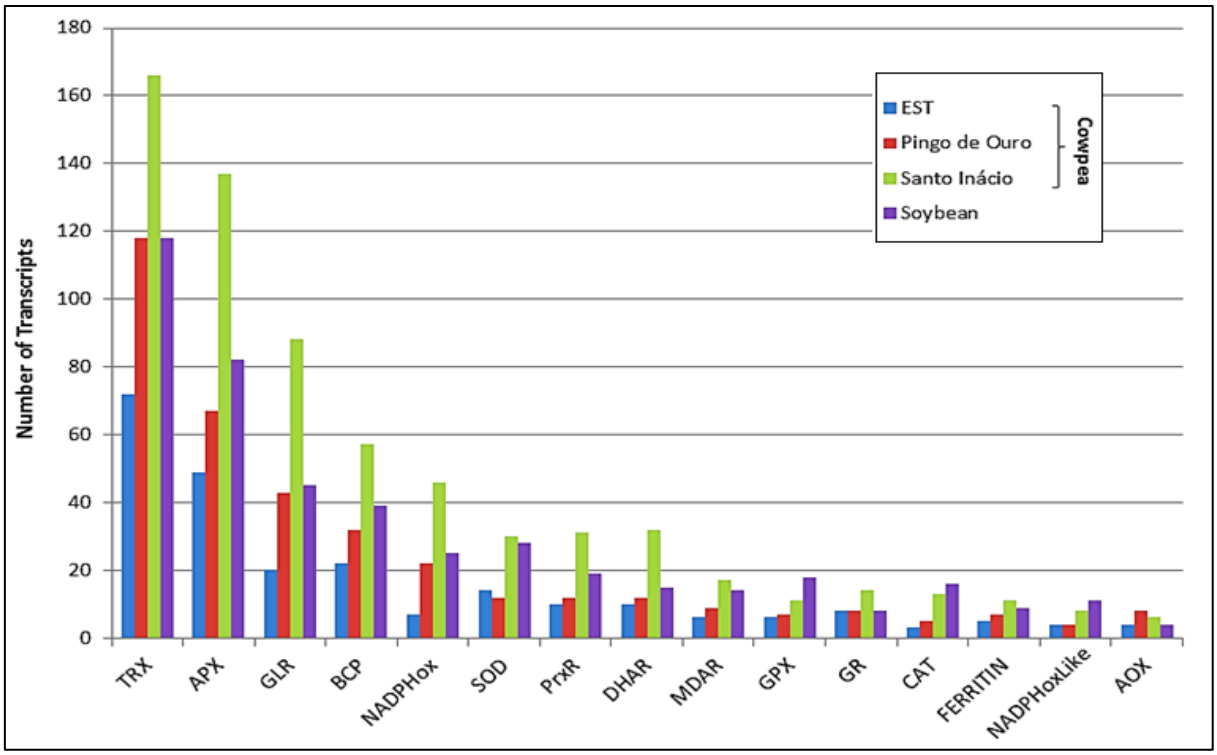


Figure 1. Number of ROS (Reactive Oxygen Species) gene family representatives, including 15 selected categories in cowpea and soybean transcriptome libraries evaluated, based on EST and RNAseq data.

Table1. Cowpea and soybean contigs matching results for each procured ROS family, showing the best match for each gene family in both species, e-value, score, size in nucleotide (nt) and amino-acid (aa), presence and kind of conserved domains and number other matches in cowpea (EST and RNAseq). Abbreviations: PO: Pingo de Ouro; SI: Santo Inácio; [c]: complete; [i]: incomplete.

	ROS Gene Family	Features of Transcripts						#Other Matches		
		Best Match	e-value	Score	ORF Size		Conserved Domain	EST	PO	SI
					nt	aa				
Cowpea	Alternative Oxidase (AOX)	SI40716.1	1e-137	488.8	963	320	Alternative Oxidase [c]	4	8	5
	Ascorbate Peroxidase (APX)	PO12393.1	2e-160	565.1	1308	435	Ascorbate peroxidase [c]	49	66	137
	Blue copper protein	PO6428.1	4,3e-53	206.5	459	152	Divalent Ion Tolerance Protein [c]	22	31	57
	Catalase (Cat)	Contig1831	0	927.2	1479	492	Catalase_Like Superfamily [c]	2	5	13
	Dehydroascorbate Reductase (DHAR)	Contig299	9,6e-94	342.4	789	262	GST_C_DHAR [c]	9	12	32
	Ferritin	Contig14061	4,1e-89	327.0	786	261	Eukaryotic ferritin [c]	4	7	11
	Glutaredoxin (GLR)	PO15669.2	4e-160	564.7	1869	622	Glutaredoxin (GRX) family [c]	20	42	88
	Glutathione Peroxidase (GPX)	Contig10886	1,7e-86	318.2	819	272	Glutathione (GSH) peroxidase family [c]	5	7	11
	Glutathione Reductase (GR)	Contig4373	0	810.1	1635	544	Pyr_redox_dim [c]	7	8	14

	Monodehydroascorbate Reductase (MDAR)	Contig6090	0	295.0	810	269	Pyr_redox [c]	5	9	17
	NADPH oxidase (NADPHox)	PO13480.1	0	347.4	1479	492	Respiratory burst NADPH oxidase [c]	7	21	46
	NADPH oxidase-like (NADPHoxLike)	PO13041.1	0	683.3	2028	675	NOX_Duox_like_FAD_NADP [c]	4	3	8
	Peroxiredoxin (PrxR)	Contig7018	5e-105	380.2	783	260	Peroxiredoxin (PRX) family [c]	9	12	31
	Superoxide Dismutase (SOD)	Contig12178	6e-102	370.2	933	310	Sod_Fe_C/Sod_Fe_N [c]	13	12	30
	Thioredoxins (Trx)	Contig11319	0	678.3	1407	468	Thioredoxin-dependent reductase [c]	71	118	166
Soybean	Alternative Oxidase (AOX)	Contig8218	2e-140	494.6	966	321	Alternative Oxidase [c]	3	-	-
	Ascorbate Peroxidase (APX)	Contig17102	7e-141	496.1	1152	383	Ascorbate peroxidase [c]	81	-	-
	Blue copper protein	Contig8226	5,7e-52	199.1	459	152	Divalent Ion Tolerance Protein [c]	38	-	-
	Catalase (Cat)	Contig14835	0	936.4	1479	492	Catalase_Like Superfamily [c]	15	-	-
	Dehydroascorbate Reductase (DHAR)	Contig24859	6,9e-95	342.0	642	213	GST_C_DHAR [c]	14	-	-
	Ferritin	Contig19687	2,5e-89	323.9	780	259	Eukaryotic ferritin [c]	8	-	-
	Glutaredoxin (GLR)	Contig19134	1,2e-95	317.0	564	187	Glutaredoxin (GRX) family [i]	44	-	-

Glutathione Peroxidase (GPX)	Contig737	1,6e-87	317.8	705	234	Glutathione (GSH) peroxidase family [c]	17	-	-
Glutathione Reductase (GR)	Contig13210	0	813.9	1506	501	Pyr_redox_dim [c]	7	-	-
Monodehydroascorbate Reductase (MDAR)	Contig26245	1e-118	422.5	735	244	Pyr_redox [c]	13	-	-
NADPH oxidase (NADPHox)	Contig23237	5e-157	550.8	1005	334	Respiratory burst NADPH oxidase [c]	24	-	-
NADPH oxidase-like (NADPHoxLike)	Contig18739	5e-139	490.7	1121	373	NOX_Duox_like_FAD_NADP [c]	10	-	-
Peroxiredoxin (PrxR)	Contig27234	4e-108	386.7	777	258	Peroxiredoxin (PRX) family [c]	18	-	-
Superoxide Dismutase (SOD)	Contig27392	4,7e-96	346.7	735	244	Sod_Fe_C/Sod_Fe_N [c]	27	-	-
Thioredoxins (Trx)	Contig18445	0	680.2	1413	470	Thioredoxin-dependent reductase [c]	117	-	-

A comparison of the distribution of non-redundant sequences in the two species revealed that the TRX (Thioredoxin) family was the most frequent one in all data (RNAseq and EST) while the AOX (Alternative Oxidase) class was the least represented category in the studied organisms (Figure 1). The distribution of ROS genes in both species were the same, with TRX, APX (Ascorbate Peroxidase), GLR (Glutaredoxin), BCP (Blue copper protein) and NADPHox (NADPH oxidase) been the five families most represented (Figure 2). Due to this abundance, all five were selected for anchoring in soybean genome and construction of a synteny map.

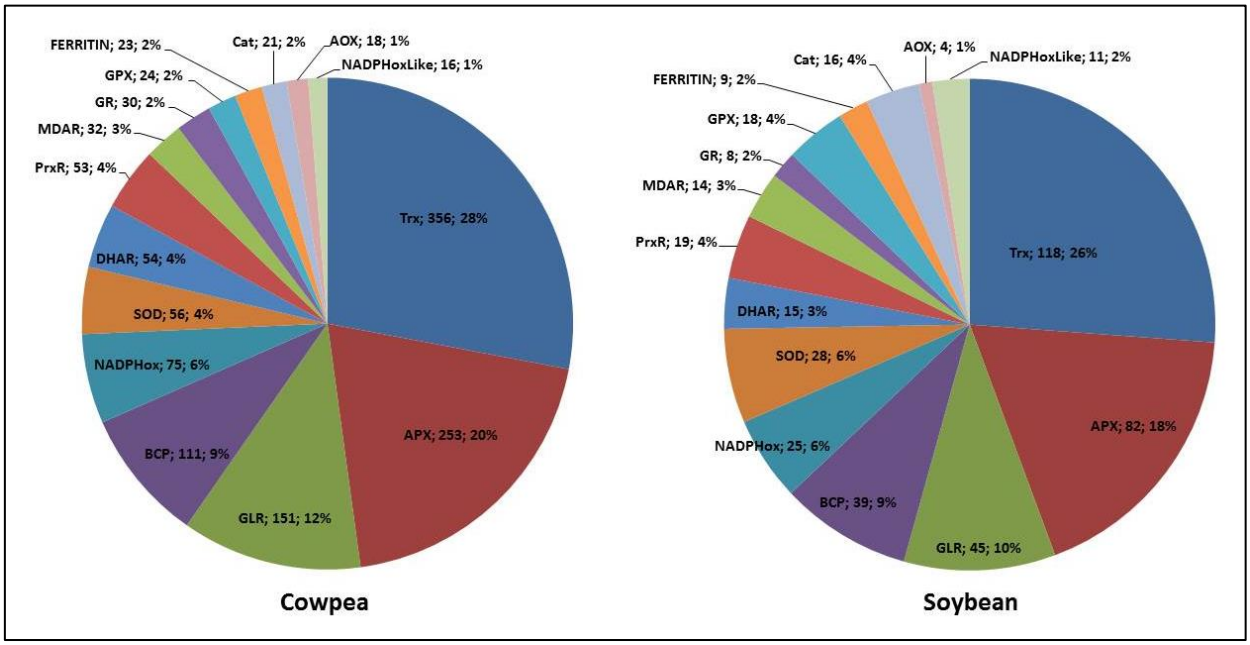


Figure 2. Number of candidate transcripts for 15 ROS gene families in cowpea and soybean transcriptome, based on proportional abundance in both species.

Expression analysis of ROS genes in cowpea and soybean transcriptome

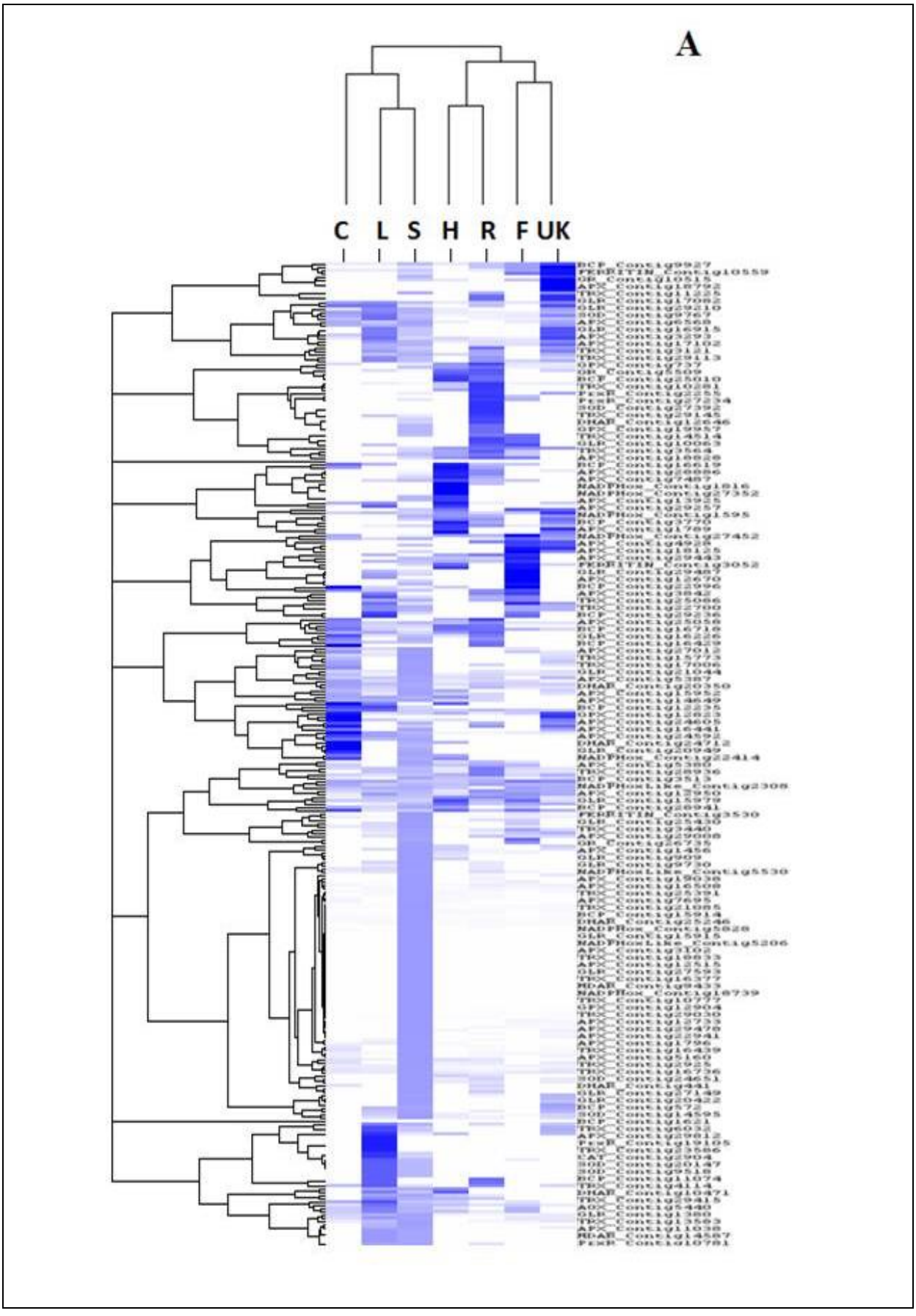
The analysis of soybean expression pattern comprised by 19.235 reads, whereas an *in silico* expression assay was carried out considering transcripts from all 15 ROS families. Subsequently, the best matches in each family were used to construct a summarized expression profile that allowed inferences on their prevalence, that after data normalizing allowed the generation of a heat-map including tissues and conditions represented in the different libraries. Graphic illustrations of these comparisons are available in Figure 3. Libraries analyzed presented almost the same number of reads, except for seed library that presented 4.5x more reads that all other together, with 82,12% of reads identified. Expression in tissues from other

libraries presented low read percentage, with 4.53% regarding roots (R), 3.67% for leaves (L), and 3.2% unknown (mix of tissues). The remaining tissues presented reads ranging from 2.6% to 1.6% of the total analyzed.

The number of times that a particular transcript was identified in the libraries does not represent the entire transcript population, but only an estimate of their abundance. Soybean ROS transcripts were distributed in seven libraries, being members of APX, MDAR, SOD, TRX, BCP and CAT, the most expressed of all libraries (absolute data). A total of 153 transcripts were found less than 10 times in the libraries (~50%), indicating that some isoforms of all families were poorly represented.

Plants generally are tolerant to oxidative stress due to ROS gene isoform activity, where members of a given family often respond differentially to various stress stimuli. Constructed expression profiles by hierarchical clustering based on transcript abundance (Figure 3) revealed that ROS representatives exhibited a constitutive expression pattern in many plant organs, including seed, leave and root libraries, with expression differences observed in tissues and organs related to many isoforms analyzed, pointing to a site-specific expression for each transcript isoform.

To obtain a panel of cowpea ROS candidates associated with drought response in sensitive (Santo Inácio) and tolerant (Pingo de Ouro) genotypes a hierarchical clustering analysis was performed using data from RNAseq data for each accession regarding five more well-annotated transcript candidates (best e-values). Considering the upper dendrogram and the hydric stress experiment (Figure 4) greater proximity was observed to the stressed data groups of tolerant cultivar, but not found in sensitive variety. ROS genes present similar levels of expression in both, generally indicating a similar response of both cultivars when subjected to dehydration stress. Despite this, ROS gene recruited were different in times of treatment and cultivars, with scavengers and genes related to cell signaling and growth expressed in early time preferentially in resistant cultivar. Moreover, control treatment in both pattern expressions presenting constitutive expression to most of ROS genes, point to the importance of them to normal cell growth and signaling.


A
C
L
S
H
R
F
UK

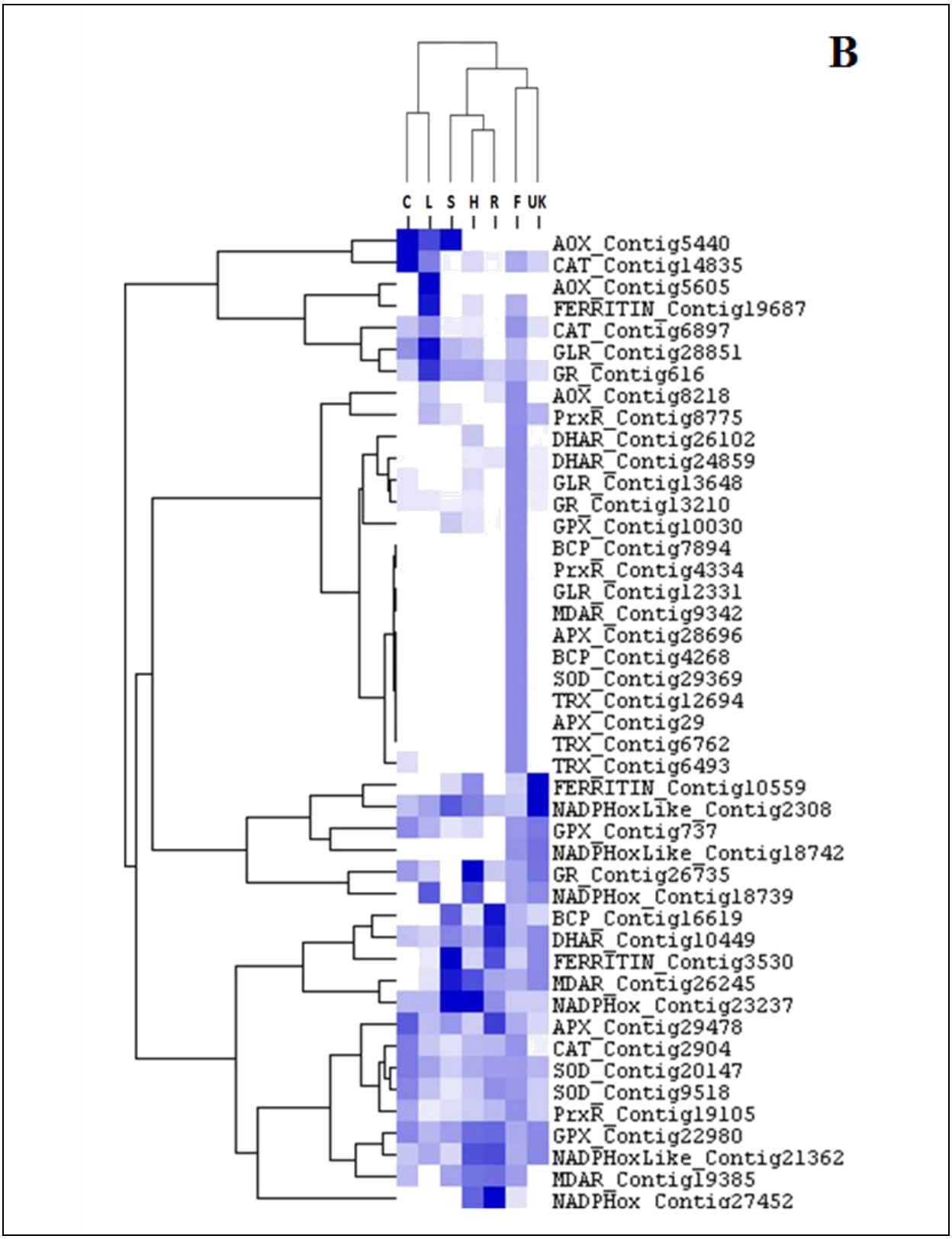


Figure 3. (A) ROS transcripts expression pattern of *Glycine max* EST data. (B) Expression pattern of best annotated RPS genes in soybean. C: Cotyledons; L: Leaves; S: Seed; H: Hypocotyls; R: Root; UK: unknown. Red: high expression and Black: absence of expression.

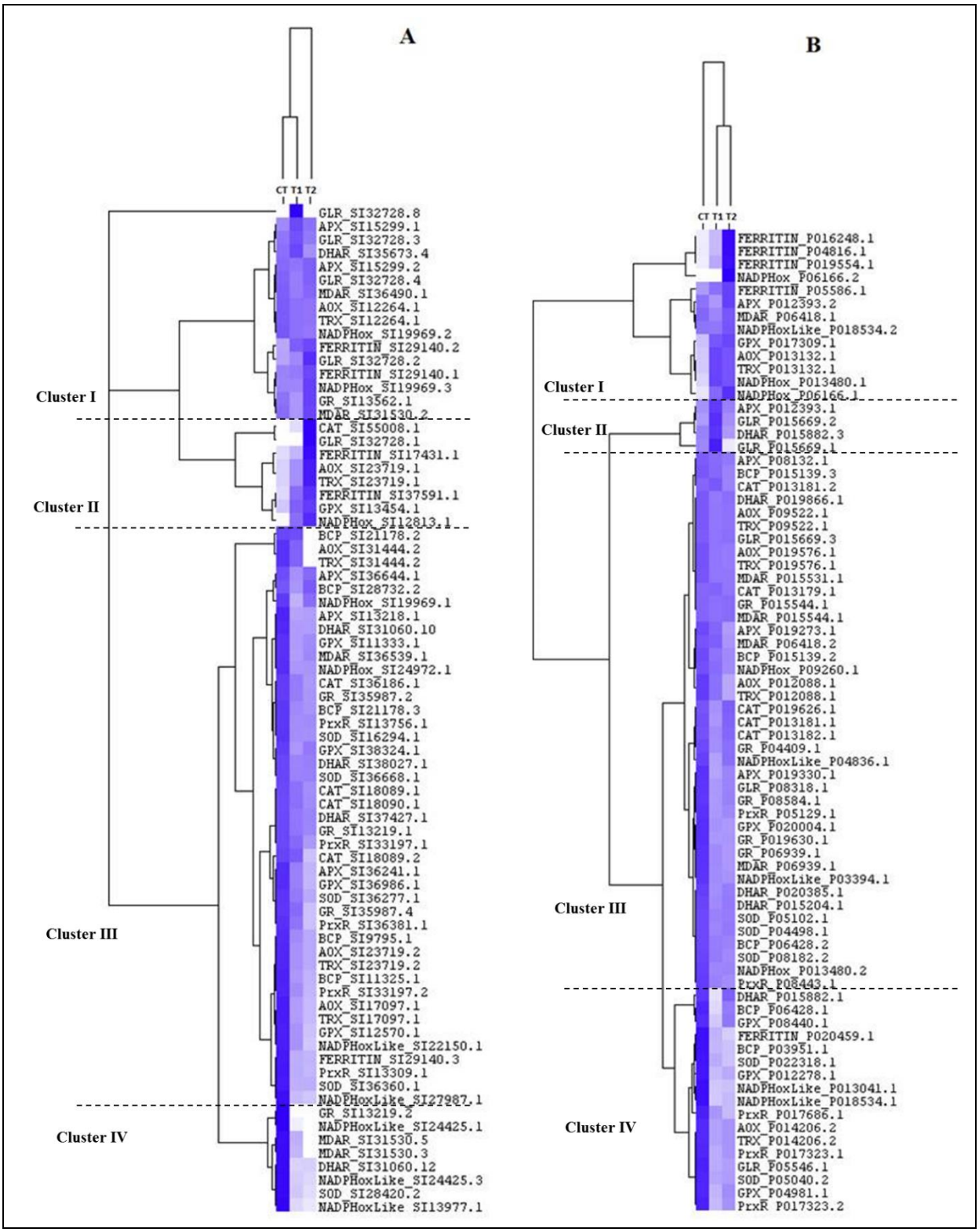


Figure 4. (A) ROS transcripts expression pattern of *Vigna unguiculata* (Santo Inácio – sensitive accession) RNAseq data. (B) ROS transcripts expression pattern of cowpea (Pingo de Ouro – tolerant accession) RNAseq data. CT: control treatment; T1: roots collected at 75 min. after stress; T2: roots collected at 150 min. after stress. Red: high expression and Black: absence of expression.

Considering both heat-maps to dry experiment, it was possible to identify two groups of putatively co-expressed ROS genes. Both groups comprised the same kind of ROS genes, with some candidates (e.g. Ferritin) high expressed in late times of stress in the tolerant cultivar. The analysis further revealed that the tolerant accession submitted to stress responded earlier than the sensitive, with ROS gene activation at 75 min after stress, as compared with the sensitive that presented a similar response just at 150 min after stress. It is possible that an efficient response to oxidative stress was associated with an early activation of gene machinery and signaling by ROS genes.

Circos mapping of ROS gene in soybean genome

The search for the five most significant ROS gene groups against soybean genome that matched with soybean transcripts recovered 779 genomic sequences, each one representing a specific locus, suggesting an important role of nuclear ROS genes in vital processes during normal development and abiotic stresses in plants, probably inducing specific transcripts (isoforms) for environmental conditions. The anchoring of these transcripts revealed that four of five chosen groups are present in all 20 soybean chromosomes and that just BCP family was absent considering chromosomes 2, 12, 13, 14 and 16 (Figure 5).

Most sequences presented syntenic regions among non-homologous chromosomes, often forming gene clusters mainly on both arms of chromosomes 7, 12, 15 and 16, with a large number of copies per chromosome. Some chromosomes presented a lower number of copies, as in the case of chromosomes 5, 7, 16 and 19 with less than 30 loci each, while chromosomes 1, 3 and 13 presented more than 50 loci identified. The high number of ROS gene members in the soybean chromosomes shows the preferential distribution in subterminal regions or next to them, with no gene-rich regions found in pericentromeric regions and moderately distrusted in intercalary regions.

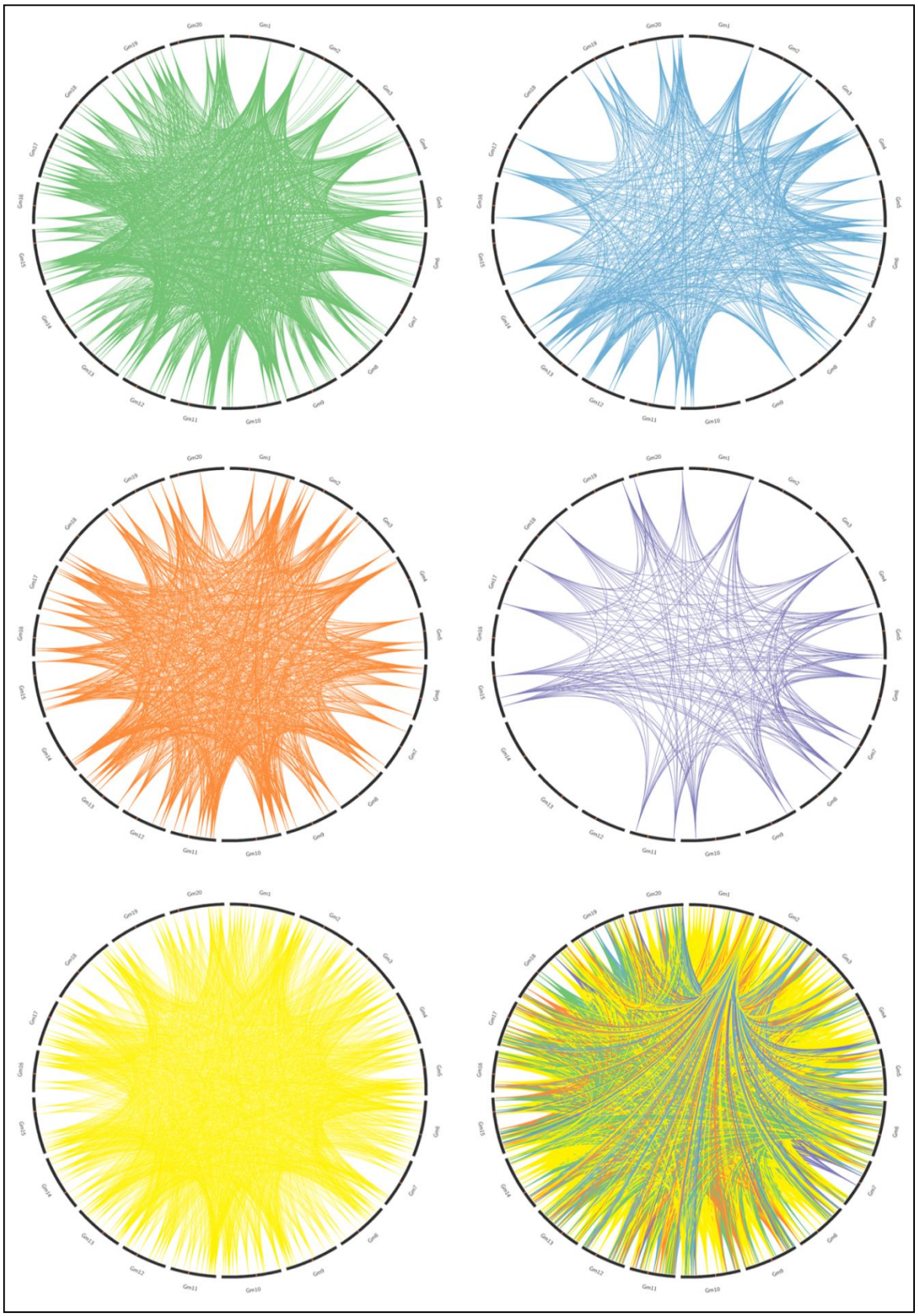


Figure 5. Distribution of five selected ROS categories in soybean (Green: TRX; Blue: APX; Orange: GLR; Violet: BCP; Yellow: NADPHox) using Circos program based on 779 gene loci. Centromere positions in each soybean pseudochromosomes are marked in red. Last Circos map includes a superposition of all five categories.

DISCUSSION

In recent years, ROS have been implicated in the control and regulation of biological functions, such as growth, cell cycle, programmed cell death, hormone signaling, biotic and abiotic stress responses and also developmental stages. Emerging evidence indicates that production of ROS and activation of redox-dependent signaling cascades are involved in the regulation of the antioxidant genes, which in turn affect the intracellular level of ROS and may provide a feedback control of the ROS-dependent biological processes. Here we found 1.273 different transcripts in cowpea and 451 transcripts in soybean grouped into 15 categories previously described in *A. thaliana*. Under normal and abnormal physiological conditions this antioxidant machinery is sufficient to maintain equilibrium between production and scavenging of ROS, commonly known as redox homeostasis (Sharma et al., 2012; Krishnamurthy and Rathinasabapathi, 2013).

The high number of members found in our data is related to a high complexity of detoxification system, which comprises antioxidant molecules and enzymes, related to the fact that plant evolution has necessitated a tight regulation of ROS equilibrium attained through a complex gene pathway that operates in all subcellular compartments. Known as "scavengers" or "recycling" several regulatory antioxidant enzymes prevent the toxic action of ROS in plant cells, acting in electron mobilization from the electron transport chain to use them in oxygen reduction to water (Mittler et al., 2004).

By the present approach members of the enzymatic and non-enzymatic ROS gene families have been detected. Major ROS-scavenging enzymes included the superoxide dismutase (SOD) that dismutates O_2^- to H_2O_2 (56 cowpea and 28 soybean members) followed by the coordinated action of a set of five enzymes namely catalase (CAT; 21 cowpea and 16 soybean members), ascorbate peroxidase (APX; 253 cowpea and 82 soybean members), glutathione peroxidase (GPX; 24 cowpea and 18 soybean members) and peroxiredoxins (PrxR; 53 cowpea and 19 soybean members) that reduce H_2O_2. All ROS-detoxifying enzymes known to date are encoded by nuclear genes that are processed properly to function in various subcellular compartments (Table 1). The ROS genes found comprise a group of antioxidant enzymes that work in concert to sustain an intracellular steady-state level of ROS promoting plant growth, development, cell cycle, and hormone signaling, reinforcing responses to abiotic and biotic environmental stressors, like similar data in other plant species (Mittler et al., 2004; Francoz et al., 2014).

One member that represents non-enzymatic group is glutathione, which acts as a reducing substrate for glutathione peroxidase and glutathione reductase (Sacandalios, 2005, Navrot et al., 2006) enzymes identified in cowpea as having eight in soybean and 22 transcripts in cowpea distributed in EST and RNAseq data, respectively.

Elimination of H_2O_2 is undertaken by a set of antioxidant enzymes, encoded by nuclear genes (Table 1), including CAT, APX, GPX, and Prx. Among them, catalases are unique in decomposing H_2O_2 without additional reductant, thus providing the cell with an energy efficient mechanism. CATs are indispensable, being responsible for the gross removal of intracellular H_2O_2 generated in peroxisomes during photorespiration, here we found just 16 and 21 for soybean and cowpea representatives, respectively, what points to the fact that these ROS genes are high conserved in number and specificity/function in plant evolution (Mhamdi et al., 2010). They are responsible for the gross removal of H_2O_2 generated in peroxisomes of photosynthetic plant tissues, predominantly localized in peroxisomes, whereas their presence in plant mitochondria is still unclear.

Related to CAT, we found transcripts of the SOD family that comprises a multigene family of nuclear-encoded enzymes present in every plant subcellular compartment (including chloroplasts, mitochondria, peroxisomes, glyoxysomes, cytosol and apoplast). SODs are the front line of defense as they rapidly dismutate O^{-}_2 to H_2O_2. The number of SOD genes varies among plant species, here were found 54 in cowpea data and 28 in soybean. According to their metal cofactor at their active site, plant SODs are classified into three groups: the copper-zinc SOD (Cu/ZnSOD), the manganese SOD (MnSOD), and the iron SOD (FeSOD). The Cu/ZnSODs present 23 candidates in cowpea and 12 in soybean, probably present in the cytosol, apoplast, peroxisomes and chloroplasts like described in other plants. FeSODs show 17 and nine members in cowpea and soybean, localized in the chloroplasts in other species, and MnSODs, usually found in mitochondria and are resistant to H_2O_2 inhibition, here were found 14 and seven members of cowpea and soybean respectively. Noteworthy, that both Cu/ZnSOD and FeSOD are inhibited by H_2O_2, thus they coexist with robust H_2O_2-scavenging systems such as enzymes of the ascorbate-glutathione cycle. The intracellular balance of SODs and the different H_2O_2-scavenging enzymes is apparently crucial in determining the steady-state level of O_2^{-} and H_2O_2 (Alscher et al., 2002).

On the other hand, APXs (the second most abundant ROS group found in this work) catalyze the reduction of H_2O_2 with concomitant consumption of ascorbate as the reducing

agent, while reduced glutathione can be used in some instances. Under normal conditions, the cellular pool of ascorbate is kept in the reduced state by a set of enzymes, namely mono-dehydroascorbate reductase (MDAR) and dehydro ascorbate reductase (DHAR) capable of using NADPH to regenerate oxidized ascorbate (Mittler et al., 2004).

Expression of ROS genes influence

Identification of all changes in gene expression regulated by oxidative stress is of considerable importance for developing stress tolerant plants. However, a global analysis of the effect of ROS on the transcriptome of any plant species has not yet been completely described. In higher plants more than 100 genes can be induced by ROS and this number is growing with the application of transcriptomic analysis of oxidative stress-regulated gene expression. Our analysis uncovered an expressive number of isoforms in cowpea and soybean, indicating a significant number of ROS representatives, representing probably a large pool of alternative transcripts, associated with different stresses, tissues and stages of development (Figures 3 and 4) (Varnov´a et al., 2002; Bhattacharjee, 2012).

ROS scavenging enzymes comprise a high number of enzymes, involved in maintaining redox homeostasis either by scavenging directly particular ROS and ROS-byproducts or by replenishing antioxidants. In this respect, these enzymes could be also considered antioxidants. Under stress conditions plants enhancing production of ROS as observed here (Figure 4), with increased expression of a cluster of ROS genes at 75 and 150 min after drought stress. It is important to note that most ROS genes are also expressed at control libraries in both cultivars (sensitive and tolerant). This fact indicate that ROS genes function in maintaining the plant homeostasis, also highlighting their role in signaling and growth pathways, adjusting the level of secondary messengers like hydrogen peroxide (H2O2). This behavior was observed in all 15 groups, including AOX, that presented the smaller number of isoforms, comprising a small multigenic family, but playing significant role under adverse environmental conditions in two ways: (i) modulating plant growth and development, and (ii) protecting cells from oxidative stress (Polidoros et al., 2009; Aluru et al., 2009). In both cultivars, this ROS gene family presented high expression levels, with higher expression in drought-tolerant cowpea cultivar.

In concordance with other researchers, our data suggests too that ROS play a fundamental role in plants as signal transduction molecules involved in mediating responses to environmental stresses, programmed cell death and different developmental stimuli (Torres and

Dangl, 2005). Modulation of ROS levels may also be linked to cell and tissue growth, as shown in Figure 3. When compared the different tissues, it is noteworthy that ROS genes normally present a constitutive expression, with members of the 15 groups here considered expressed in all tissues. It is important to note that many isoforms presented a preference for some tissues or developmental stages, with isoforms expressing just in initial times of development, like cluster IV that present high expression level in seeds, hypocotyl and roots, and cluster I, that was induced just in leaves. The rapid increase in ROS production referred to as 'oxidative burst' was shown to be essential for many processes in plant tissues, associated with growth, development, and death. In addition to genes likely to be directly involved in ROS production and scavenging, many other genes may be stimulated through ROS signaling cascades that ultimately influence gene expression. Serine/threonine protein kinases, for example, play a major role in ROS sensing and the activation of MAPKs (Rentel et al., 2004).

The constitutive expression of 15 groups of isoforms, cited previously, reflect the beneficial signal transduction molecules needed to control the steady-state level of ROS in cells during normal metabolism, as well as in response to different stress situations (Dietz, 2008; Miller et al., 2010).

Molecular mechanisms associated with significant stress tolerance, leading to the expression of genes as an early stress response in cowpea are largely unknown. However, for plants in general it became gradually evident that the gene expression associated with such responses is highly sensitive to the redox state of the cell (Foyer et al., 1997; Bhattacharjee, 2012). ROS genes are essential to control growth and morphogenesis of plant cells. In cowpea, the imposition of water deficit increased the level of ROS (Figure 4), with earlier induction of ROS genes in the tolerant variety (T1; 75 min in clusters I and II, Fig. 4B), whereas the sensitive cultivar showed a similar behavior just in T2 (150 min in cluster II, Fig. 4A). It was not possible to find a gene or isoform exclusive to one of the contrasting cultivars. This observation indicates that the early induction of some groups of genes is relevant for plant signaling and tolerance, as observed in other plants where the speed of expression activation of given genes was associated to resistance or tolerance to a given stress (Bhattacharjee, 2005; 2012).

It was expected that the expression of most ROS genes would increase in both cultivars and also that other genes would be induced, but most of them were already expressed at the control library, indicating that ROS generation may occur under diverse conditions. It appears that ROS accumulation in specific tissues and appropriate quantities is of benefit to plants and

can mediate cross-tolerance towards other stresses, involved in plant defense response, but not exclusive to them (Desikan et al., 1999; Bhattacharjee, 2012). The steady-state level of ROS in a cell is largely determined by the efficiency of the antioxidative systems. Antioxidative defenses play a pivotal role in preventing stress-induced injuries and toxicities. Various antioxidants and quenchers like glutathione, ascorbate together with antioxidant enzymes are increased in plants under stressful conditions and correlate significantly with enhanced tolerance like the observed in cluster I of figure 4 (Bhattacharjee, 2012).

The induction of the cellular antioxidant machinery is necessary for protection against ROS. Overexpression of ROS scavenging enzymes like isoforms of SOD (Mn-SOD, Cu/Zn-SOD, Fe-SOD), CAT, APX, GR, DHAR and others, probably contribute to abiotic stress tolerance in our cowpea data, due to efficient ROS scavenging capacity in early times, and prevent the occurrence of other damages occasioned by a secondary stress. Therefore, plants with the ability to scavenge and/or control the level of cellular ROS may be useful in future to withstand harsh environmental conditions (Gill and Tuteja, 2010).

Circos chromosome mapping

The completion of soybean genome sequencing provides an opportunity to improve our understanding of the distribution, evolution mechanisms, and other characteristics of many gene families, including ROS-related genes in this crop. The observed preference and a high number of ROS gene clusters corroborate previous suggestions of soybean octoploid nature. The redundancies observed probably reflect past duplication events, increasing the number of ROS genes in soybean genome (Liu et al., 2009; Schmutz et al., 2010). Gene duplications are considered to be a significant primary driving force towards the evolution of genomes and genetic systems (Moore and Purugganan, 2003). Duplicated genes provide the raw material for the generation of new allelic variants and genes, which, in turn, facilitate the generation of new functions. Segmental duplications of multiple genes through polyploidy followed by chromosome rearrangements (Yu et al., 2005) occur most frequently in plants because most angiosperms regard diploidized polyploids and retain numerous duplicated chromosomal blocks within their genomes (Cannon et al., 2004; Zhu et al., 2014).

The present work approached the distribution of five ROS gene groups in subterminal regions, creating clusters of genes uncovering a similar distribution and organization as those observed for aquaporin (Belarmino et al., 2012) and osmoprotectants genes in soybean (Kido

et al., 2013), both gene categories associated with drought and also salinity stress. Besides redundancies among chromosomes, the five gene families were prevalent in the terminal and subterminal gene clusters, out of centromere region, with NADHox family (group with most of the locus identified) also found along the chromosome length. Similar clustering was described for another previous approach, where 59 soybean defense genes (two super-families: R, resistance and PR, pathogen-related genes) were anchored in the virtual chromosomes of the legume *Medicago truncatula*, revealing 1,253 sites, most of them clustered in subterminal or terminal positions (Wanderley-Nogueira et al., 2012), indicating that such a distribution may occur in regard to different gene families and plant groups, associated probably a "genic islands" in euchromatin regions . The predicted number of coding genes in soybean is higher than that of Arabidopsis and grape, possibly due to the genome duplication events in soybean's history, being this analysis the first on legume plants, revealing a high number and specific distribution of these five families in soybean (Sterck et al., 2007; Cannon and Shoemaker, 2012).

Based on the integrated data from recently duplicated genomic segments, homeologous blocks of duplicated segments were found in all 20 chromosomes, but not for all five ROS families (Figure 5). Multiple blocks of ROS genes were observed in most chromosome clusters, composed of five family members or a combination of families. The observed clustering and prevalence in some chromosomes, especially those combining different QTLs (Quantitative Trait Loci) useful for mapping approaches and marker assisted selection (Lestari et al., 2013). Therefore, the integration of soybean genomics with relative phenotypic trait resources should facilitate the identification of homeologous chromosomal rearrangements and new duplicate gene copies.

CONCLUDING REMARKS

The identified candidates provide an improved picture of genes involved in osmotic stress response in cowpea and soybean, and may be helpful for the discovery of useful genes for breeding and biotechnological purposes, especially if they are associated to SNPs (Single Nucleotide Polymorphisms). The here analyzed transcripts reveal a significant diversity and abundance of isoforms, with a higher number of candidates in cowpea than in soybean. In cowpea, some of them were expressed before stress imposition whereas other their expression increased after water deficit. Regarding the genome distribution of evaluated ROS categories

in soybean, some categories are distributed genome-wide (as NADPHox members) while other are preferably concentrated in a given chromosome arm (as APX candidates) and, in turn, some are less prevalent, being absent in some chromosomes (case of BCP), possibly also representing different roles, diversity, and expression levels.

ACKNOWLEDGMENTS

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

REFERENCES CITED

Abdelnoor, R.V. et al. GENOSOJA – A Brazilian Soybean Genome Consortium. In: Plant & Animal Genome XVII, San Diego. Plant & Animal Genome Conference XVII Abstracts (www.intl-pag.org/17/abstracts/P01_ PAGXVII_034.html), 2009.

Alscher RG, Erturk N, Heath LS (2002) Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. Journal of Experimental Biology 531(1):1331–1341.

Altschul SF, Gish W, Miller W, Myers E (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403–410.

Aluru MR, Zola J, Foudree A, Rodermel SR (2009) Chloroplast photooxidation induced transcriptome reprogramming in *Arabidopsis immutans* white leaf sectors. Plant Physiology 150: 904–923.

Belarmino LC, Oliveira ARS, Brasileiro-Vidal AC, Bortoleti KCA, Bezerra-Neto JP, Abdelnoor RV, Benko-Iseppon AM (2012) Mining plant genome browsers as a means for efficient connection of physical, genetic and cytogenetic mapping: An example using soybean. Genetics and Molecular Biology 35(1):335-47.

Benko-Iseppon AM (2001) Estudos moleculares no caupi e em espécies relacionadas: Avanços e perspectivas. Embrapa Documentos 56:327-332.

Benko-Iseppon AM, Kido EA, Pandolfi V et al., (2010) Brazilian cowpea transcriptome project: over 20 million expressed sequence tags to analyze and breed salinity and virus resistance. Annals of III Congresso Brasileiro de Biotecnologia. Fortaleza, Brazil, 97-98.

Benko-Iseppon AM, Nepomuceno AL, Abdelnoor RV (2012) GENOSOJA - The Brazilian Soybean Genome Consortium: High Throughput Omics and Beyond. Genetics and Molecular Biology 35(1suppl):i–iv.

Bhattacharjee S (2005) "Reactive oxygen species and oxidative burst: roles in stress, senescence and signal transduction in plants". Current Science 89(7):1113–1121.

Bhattacharjee S (2012) The Language of Reactive Oxygen Species Signaling in Plants. Journal of Botany ID:985298,1–22.

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004)The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biology 14(1):10-21.

Cannon SB, Shoemaker RC (2012) Evolutionary and comparative analysis of the soybean genome. Breed Science 61:437–444.

CONAB. Companhia Nacional de Abastecimento. Disponível em <http://www.conab.gov.br>. Acesso em agosto/2014.

Dalton TP, Shertzer HG, Puga A (1999) Regulation of gene expression by reactive oxygen. Annual Review of Pharmacology and Toxicology 39:67e101.

Desikan R, Clarke A, Hancock JT, Neill SJ (1999) H_2O_2 activates a MAP kinase-like enzyme in Arabidopsis thaliana suspension cultures. Journal of Experimental Botany 50(341):1863–1866.

Dietz KJ (2008) Redox signal integration: from stimulus to networks and genes. Physiologia Plantarum 133:459–468.

Ding S, Lei M, Lu Q, Zhang A, Yin Y, Wen X, Zhang L, Lu C (2012) Enhanced sensitivity and characterization of photosystem II in transgenic tobacco plants with decreased chloroplast glutathione reductase under chilling stress. Biochimica et Biophysica Acta Bioenergetics 1817(11):1979-91.

Ehlers JD, Hall AE (1997) Cowpea (*Vigna unguiculata* (L.) Walp). Field Crops Research 53:187-204.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences USA 95:14863–14868.

Foyer CH, Lopez-Delgado H, Dat JF, Scott IM (1997) Hydrogen peroxide and glutathione-associated mechanisms of acclimatory stress tolerance and signaling. Physiologia Plantarum 100:241–254.

Foyer CH, Noctor G (2005) Redox homeostasis and antioxidant signaling: a metabolic interface between stress perception and physiological responses. Plant Cell 17:1866e1875.

Francoz E, Ranocha P, Nguyen-Kim H, Jamet E, Burlat V, Dunand C (2014) Roles of cell wall peroxidases in plant development. Phytochemistry 2014 pii:S0031-9422(14)00307-0. [In press]

Freire-Filho FR, Rocha MM, Ribeiro VQ et al., (2005) Adaptabilidade e estabilidade produtiva de feijão-caupi. Ciência Rural 35:24-30.

Gill SS, Anjum NA, Hasanuzzaman M, Gill R, Trivedi DK, Ahmad I, Pereira E, Tuteja N (2013) Glutathione and glutathione reductase: a boon in disguise for plant abiotic stress defense operations. Plant Physiology and Biochemistry 70:204–212.

Gill SS, Tuteja N (2010) Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. Plant Physiology and Biochemistry 48:909-30.

Giraud E, Ho LH, Clifton R. et al., (2008) The absence of ALTERNATIVE OXIDASE 1a in Arabidopsis results in acute sensitivity to combined light and drought stress. Plant Physiology 147:595–610.

Ishiyaku MF, Singh BB, Craufurd PQ (2005) Inheritance of time to flowering in cowpea [*Vigna unguiculata* (L.)Walp.]. Euphytica 142:291–300.

Kido EA, Ferreira-Neto JR, Silva RL, Belarmino LC, Bezerra-Neto JP, Soares-Cavalcanti NM, Pandolfi V, Silva MD, Nepomuceno AL, Benko-Iseppon AM (2013) Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress. BMC Bioinformatics. 14 Suppl 1:S7.

Koussevitzky S, Suzuki N, Huntington S, Armijo L, Sha W, Cortes D, Shulaev V, Mittler R (2008) Ascorbate peroxidase 1 plays a key role in the response of *Arabidopsis thaliana* to stress combination. Journal of Biological Chemistry 283:34197– 34203.

Krishnamurthy A, Rathinasabapathi B (2013) Novel interplay between auxin and reactive oxygen species signaling Oxidative stress tolerance in plants. Plant Signaling & Behavior 8(10):1–5.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Research 19:1639-1645.

Larkindale J, Knight MR (2002) Protection against heat stress-induced oxidative damage in Arabidopsis involves calcium, abscisic acid, ethylene, and salicylic acid. Plant Physiology 128:682-95.

Lestari P, Van K, Lee J, Kang YJ, Lee SH (2013) Gene divergence of homeologous regions associated with a major seed protein content QTL in soybean. Frontiers in Plant Science 4:1-8.

Liu Q, Wang H, Zhang Z, Wu J, Feng Y, Zhu Z (2009) Divergence in function and expression of the NOD26-like intrinsic proteins in plants. BMC Genomics 10:313.

MerigaB, Reddy BK, Rao KR, Reddy LA, Kishor PBK (2004) Aluminium-induced production of oxygen radicals, lipid peroxidation and DNA damage in seedlings of rice (*Oryza sativa*). Journal of Plant Physiology 161(1):63–68.

Mhamdi A, Queval G, Chaouch S, Vanderauwera S, Van Breusegem F, Noctor G (2010) Catalase function in plants: a focus on Arabidopsis mutants as stress-mimic models. Journal of Experimental Botany 61(15):4197-4220.

Miller G, Suzuki N, Ciftci-Yilmaz S, Mittler R (2010) Reactive oxygen species homeostasis and signalling during drought and salinity stresses. Plant Cell Environmental 33:453–67.

Mittler R (2002) Oxidative stress, antioxidants and stress tolerance. Trends in Plant Science 7(9):405–410.

Mittler R, Vanderauwera S, Gollery M, Van Breusegem F (2004) Reactive oxygen gene network of plants. Trends in Plant Science 9:490–498.

Moore RC, Purugganan MD (2003).The early stages of duplicate gene evolution. Proceedings of the National Academy of Sciences 14(26):15682–15687.

Nascimento LC, Costa GGL, Binneck E, Pereira GAG, Carazzolle MF (2012) A web-based bioinformatics interface applied to Genosoja Project: Databases and pipelines. Genetics and Molecular Biology 35(suppl1):203–211.

Navrot N, Collin V, Gualberto J, Gelhaye E, Hirasawa M, Rey P, Knaff DB, Issakidis E, Jacquot J-P, Rouhier N (2006) Plant glutathione peroxidases are functional peroxiredoxins distributed in several subcellular compartments and regulated during biotic and abiotic stresses. Plant Physiology 142:1364–1379.

NCBI, National Center for Biotechnology Information [http://www.ncbi.nlm.nih.gov/].

Newton AC, Johnson SN, Gregory PJ (2011) Implications of climate change for diseases, crop yields and food security. Euphytica 179:3–18.

Page RD (1996) Treeview program, ver. 1.6.1. Computer Applications in the Biosciences 12:357–358.

Pang CH, Wang BS (2008) Oxidative stress and salt tolerance in plants. In: Progress in Botany, Luttge U, Beyschlag W, Murata J, Eds. 231–245, Springer, Berlin, Germany.

Peer WA, Cheng Y, Murphy AS (2013) Evidence of oxidative attenuation of auxin signalling. Journal of Experimental Botany 64:2629-39.

Pimentel C, Abboud ACS, Roy-Macauly H et al. (2002) Leaf protoplasmic tolerance to drought, in two cowpea genotypes cultivated at the field. Revista Universidade Rural, Série Ciências da Vida 22(1):07–14.

Pinto MC, Gara L (2004) Changes in the ascorbate metabolism of apoplastic and symplastic spaces are associated with cell differentiation. Journal of Experimental Botany 55(408):2559–2569.

Polidoros AN, Mylona PV, Arnholdt-Schmitt B (2009) AOX gene structure, transcript variation and expression in plants. Physiology Plantarum 137(4):342-53.

PubMed Database [http://www.ncbi.nlm.nih.gov/pubmed].

Rentel MC, Lecourieux D, Ouaked F, Usher SL, Petersen L, Okamoto H, Knight H, Peck SC, Grierson CS, Hirt H, Knight MR (2004) OXI1 kinase is necessary for oxidative burst-mediated signalling in Arabidopsis. Nature 427(6977):858-861.

Rizhsky L, Liang H, Shuman J, Shulaev V, Davletova S, Mittler R (2004) When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. Plant Physiology 134:1683–1696.

Scandalios JG (2005) Oxidative stress: Molecular perception and transduction of signals triggering antioxidant gene defenses. Brazilian Journal of Medical and Biological Research 38:995–1014.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Hyten DL, Song Q, Thelen JJ, Cheng J et al. (2010) Genome sequence of the paleopolyploid soybean. Nature 463:178–183.

Serrato AJ, Perez-Ruiz JM, Spinola MC, Cejudo FJ (2004) A novel NADPH thioredoxin reductase, localized in the chloroplast, which deficiency causes hypersensitivity to abiotic stress in *Arabidopsis thaliana*. Journal of Biological Chemistry 279:43821–43827.

Sharma P, Dubey RS, Pessarakli M (2012) Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. Journal of Botany ID: 217037.

Sharma P, Jha AB, Dubey RS (2010) Oxidative stress and antioxidative defense systemin plants growing under abiotic Stresses. In: Handbook of Plant and Crop Stress, Pessarakli M, Eds. 89–138, CRC Press, Taylor and Francis Publishing Company, Fla, USA, 3rd edition.

Srivastava S, Dubey RS (2011) Manganese-excess induces oxidative stress, lowers the pool of antioxidants and elevates activities of key antioxidative enzymes in rice seedlings. Plant Growth Regulation 1–16.

Sterck L, Rombauts S, Vandepoele K, Rouze P, VandePeer Y (2007) How many genes are there in plants (and way are they there)? Current Opinion in Plant Biology 10:199–203.

Tamás L, Bočová B, Huttová J, Liptáková L, Mistrík I, Valentovičová K, et al., (2012) Impact of the auxin signaling inhibitor p-chlorophenoxyisobutyric acid on short-term Cd-induced hydrogen peroxide production and growth response in barley root tip. J Plant Physiol 169:1375-81.

Torres MA, Dangl JL (2005) Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. Current Opinion in Plant Biology 8:397–403.

Turpaev KT (2002) Reactive oxygen species and regulation of gene expression. Biochemistry 67(3):281-92.

USDA, United States Department of Agriculture. World agricultural supply and demand estimates (2013). [http://www.usda.gov/oce/commodity/wasde/latest.pdf]

Van Breusegem F, Dat JF (2006) Reactive oxygen species in plant cell death. Plant Physiology 141:384–390.

Varnov´a E, Van Breusegem F, Dat J, Belles-Bolx E, Inz´e D (2002) The role of reactive oxygen species in signal transduction, Oxford University Press, New York, NY, USA, 2002.

Voothuluru P, Sharp RE (2013) Apoplastic hydrogen per- oxide in the growth zone of the maize primary root under water stress. I. Increased levels are specific to the apical region of growth maintenance. Journal of Experimental Botany 64:1223-1233.

Wanderley-Nogueira AC, Soares-Cavalcanti NM, Belarmino LC, Bezerra-Neto JP, Kido EA, Pandolfi V, Abdelnoor R, Bineck E, Carazzole MF, Benko-Iseppon AM (2012) An overall evaluation of the Resistance (R) and Pathogenesis Related (PR) superfamilies in soybean, as compared with Medicago and Arabidopsis. Genetics and Molecular Biology 35(1):260–271.

Wang W, Vinocur B, Altman A (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218:01–14.

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Mi P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li S, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Wang J, Deng Y, Ran L, Shi X (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biology 14(2):e38.

Zhu J, Fu X, Koo YD et al. (2007) An enhancer mutant of Arabidopsis salt overly sensitive 3 mediates both ion homeostasis and the oxidative stress response. Molecular Cell Biology 27:5214–5224.

Zhu Y, Wu N, Song W, Yin G, Qin Y, Yan Y, Hu Y (2014) Soybean (*Glycine max*) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. BMC Plant Biology 14:1-19.

CONCLUSÕES

Nossos estudos nos permitiram concluir que:

- A tecnologia de sequenciamento de RNA (RNAseq) e a montagem '*de novo*'de sequências mostrou-se como uma estratégia robusta e eficiente para análises de transcriptomas, em especial no que tange ao déficit hídrico em feijão-caupi.

- A construção de bibliotecas utilizando variedades contrastantes de feijão-caupi submetidas a déficit hídrico, permitiu a identificação de famílias gênicas intimamente relacionadas com a resposta vegetal frente ao estresse, avaliando-se a modulação de sua expressão em diferentes backgrounds genéticos e condições de construção das bibliotecas.

- Membros de diferentes famílias de fatores de transcrição identificados (FTs) em feijão-caupi por meio de métodos computacionais contribuem para diferenças na tolerância vegetal frente a estresses, considerando os acessos contrastantes analisados. No geral FTs foram induzidos nos tempos iniciais de estresse, com funções nas principais vias de sinalização e adaptação ao ambiente, havendo uma indução mais precoce no acesso tolerante comparativamente ao susceptível.

- Transcritos associados aos genes de resposta a ROS são expressos em ambos os acessos, tanto na ausência como na presença de estresse (déficit hídrico), provavelmente desempenhando importante papel na manutenção da fisiologia das plantas, atuando na manutenção da homeostase; além de ajustar os níveis de mensageiros secundários que atuam na sinalização de respostas a estresses.

- Mecanismos moleculares associados com a percepção e sinalização frente a estresses ambientais compreendem uma rede complexa, com expressão gênica modulada por fatores de transcrição que se ligam em regiões promotoras específicas do DNA, podendo

ocorrer *crosstalk* (resposta cruzada) em vias de sinalização para diferentes tipos de estresse abiótico.

- A validação da expressão de candidatos selecionados por PCR quantitativa em tempo real (RT-qPCR) deverá auxiliar na identificação de candidatos com maior potencial para uso no melhoramento de fejão-caupi e outras leguminosas, com ênfase para a categoria de fatores de transcrição, especialmente no caso dos candidatos com indicação de indução pelos dois métodos experimentais usados (RNAseq e SuperSAGE).

(Artigos de colaboração publicados em Periódicos)

Apêndice A

Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress

BMC Bioinformatics 2013, 14(Suppl 1):S7. **Qualis B1**

Apêndice B

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

Genet Mol Biol. 2012 Jun, 35(1 (suppl)):315-21. **Qualis B2**

Apêndice C

An overall evaluation of the Resistance (R) and Pathogenesis-Related (PR) superfamilies in soybean, as compared with Medicago and Arabidopsis.

Genet Mol Biol. 2012 Jun, 35(1 (suppl)):260-71. **Qualis B2**

Apêndice D

Mining plant genome browsers as a means for efficient connection of physical, genetic and cytogenetic mapping: An example using soybean.

Genet Mol Biol. 2012 Jun, 35(1 (suppl)):335-47. **Qualis B2**

Apêndice E

Insight on Pathogen Defense Mechanisms in the Sugarcane Transcriptome. In: Sundar A.R., Viswanathan R. (Org.).

Global Science Books. 2012, 6:134-148. **Capítulo de Livro**

Apêndice A

Kido et al. BMC Bioinformatics 2013, 14(Suppl 1):S7
http://www.biomedcentral.com/1471-2105/14/S1/S7

BMC Bioinformatics

RESEARCH Open Access

Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress

Ederson A Kido[1], José RC Ferreira Neto[1], Roberta LO Silva[1], Luis C Belarmino[1], João P Bezerra Neto[1], Nina M Soares-Cavalcanti[1], Valesca Pandolfi[1], Manassés D Silva[1], Alexandre L Nepomuceno[2], Ana M Benko-Iseppon[1*]

From Seventh International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB 2010)
Palermo, Italy. 16-18 September 2010

Abstract

Background: Despite the importance of osmoprotectants, no previous *in silico* evaluation of high throughput data is available for higher plants. The present approach aimed at the identification and annotation of osmoprotectant-related sequences applied to short transcripts from a soybean HT-SuperSAGE (High Throughput Super Serial Analysis of Gene Expression; 26-bp tags) database, and also its comparison with other transcriptomic and genomic data available from different sources.

Methods: A curated set of osmoprotectants related sequences was generated using text mining and selected seed sequences for identification of the respective transcripts and proteins in higher plants. To test the efficiency of the seed sequences, these were aligned against four HT-SuperSAGE contrasting libraries generated by our group using soybean tolerant and sensible plants against water deficit, considering only differentially expressed transcripts ($p \leq 0.05$). Identified transcripts from soybean and their respective tags were aligned and anchored against the soybean virtual genome.

Results: The workflow applied resulted in a set including 1,996 seed sequences that allowed the identification of 36 differentially expressed genes related to the biosynthesis of osmoprotectants [Proline (*P5CS*: 4, *P5CR*: 2), Trehalose (*TPS1*: 9, *TPPB*: 1), Glycine betaine (*BADH*: 4) and *Myo*-inositol (*MIPS*: 7, *INPS1*: 8)], also mapped *in silico* in the soybean genome (25 loci). Another approach considered matches using Arabidopsis full length sequences as seed sequences, and allowed the identification of 124 osmoprotectant-related sequences, matching ~10.500 tags anchored in the soybean virtual chromosomes. Osmoprotectant-related genes appeared clustered in all soybean chromosomes, with higher density in some subterminal regions and synteny among some chromosome pairs.

Conclusions: Soybean presents all searched osmoprotectant categories with some important members differentially expressed among the comparisons considered (drought tolerant or sensible *vs.* control; tolerant *vs.* sensible), allowing the identification of interesting candidates for biotechnological inferences. The identified tags aligned to corresponding genes that matched 19 soybean chromosomes. Osmoprotectant-related genes are not regularly distributed in the soybean genome, but clustered in some regions near the chromosome terminals, with some redundant clusters in different chromosomes indicating their involvement in previous duplication and rearrangements events. The seed sequences, transcripts and map represent the first transversal evaluation for osmoprotectant-related genes and may be easily applied to other plants of interest.

* Correspondence: ana.iseppon@gmail.com
[1]Departament of Genetics/Biological Sciences Center, Federal University of Pernambuco, Recife, Pernambuco, CEP 50.670-420, Brazil
Full list of author information is available at the end of the article

Apêndice B

Genetics and Molecular Biology, 35, 1 (suppl), 315-321 (2012)

www.sbg.org.br

Research Article

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

Nina M. Soares-Cavalcanti[1], Luis C. Belarmino[1], Ederson A. Kido[1], Ana C. Wanderley-Nogueira[1], João P. Bezerra-Neto[1], Rafaela Cavalcanti-Lira[1], Valesca Pandolfi[1], Alexandre L. Nepomuceno[2], Ricardo V. Abdelnoor[2], Leandro C. Nascimento[3] and Ana M. Benko-Iseppon[1]

[1]*Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Recife, PE, Brazil.*
[2]*Embrapa Soja, Londrina, PR, Brazil.*
[3]*Laboratório 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 Genosoja 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. Morais Rego 1235, 50.670-420 Recife, PE, Brazil. E-mail: ana.iseppon@gmail.com.

Apêndice C

Genetics and Molecular Biology, 35, 1 (suppl), 260-271 (2012)

www.sbg.org.br

Research Article

An overall evaluation of the Resistance (*R*) and Pathogenesis-Related (*PR*) superfamilies in soybean, as compared with *Medicago* and *Arabidopsis*

Ana C. Wanderley-Nogueira[1], Luis C. Belarmino[1], Nina da M. Soares-Cavalcanti[1], João P. Bezerra-Neto[1], Ederson A. Kido[1], Valesca Pandolfi[1], Ricardo V. Abdelnoor[2], Eliseu Binneck[2], Marcelo F. Carazzole[3,4] and Ana M. Benko-Iseppon[1]

[1]*Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Recife, PE, Brazil.*
[2]*Embrapa Soja, Londrina, PR, Brazil.*
[3]*Laboratório de Genômica e Expressão, Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.*
[4]*Centro 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*.

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 a of 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. Morais Rego 1235, 50.670-420 Recife, PE, Brazil. E-mail: ana.iseppon@gmail.com.

Apêndice D

Genetics and Molecular Biology, 35, 1 (suppl), 335-347 (2012)

www.sbg.org.br

Research Article

Mining plant genome browsers as a means for efficient connection of physical, genetic and cytogenetic mapping: An example using soybean

Luis C. Belarmino[1], Ana R. da S. Oliveira[1,2], Ana C. Brasileiro-Vidal[1,2], Kyria C. de A. Bortoleti[1,3], João Pacífico Bezerra-Neto[1], Ricardo V. Abdelnoor[4] and Ana M. Benko-Iseppon[1]

[1]*Laboratório de Genética e Biotecnologia Vegetal, Departamento de Genética, Universidade Federal de Pernambuco, Recife, PE, Brazil.*
[2]*Programa de Pós Graduação em Melhoramento Genético de Plantas, Universidade Federal Rural de Pernambuco, Recife, PE, Brazil.*
[3]*Campus Ciências Agrárias, Universidade Federal do Vale do São Francisco, Petrolina, PE, Brazil.*
[4]*Embrapa Soja, Londrina, PR, Brazil.*

Abstract

Physical maps are important tools to uncover general chromosome structure as well as to compare different plant lineages and species, helping to elucidate genome structure, evolution and possibilities regarding synteny and colinearity. The increasing production of sequence data has opened an opportunity to link information from mapping studies to the underlying sequences. Genome browsers are invaluable platforms that provide access to these sequences, including tools for genome analysis, allowing the integration of multivariate information, and thus aiding to explain the emergence of complex genomes. The present work presents a tutorial regarding the use of genome browsers to develop targeted physical mapping, providing also a general overview and examples about the possibilities regarding the use of Fluorescent *In Situ* Hybridization (FISH) using bacterial artificial chromosomes (BAC), simple sequence repeats (SSR) and rDNA probes, highlighting the potential of such studies for map integration and comparative genetics. As a case study, the available genome of soybean was accessed to show how the physical and *in silico* distribution of such sequences may be compared at different levels. Such evaluations may also be complemented by the identification of sequences beyond the detection level of cytological methods, here using members of the aquaporin gene family as an example. The proposed approach highlights the complementation power of the combination of molecular cytogenetics and computational approaches for the anchoring of coding or repetitive sequences in plant genomes using available genome browsers, helping in the determination of sequence location, arrangement and number of repeats, and also filling gaps found in computational pseudochromosome assemblies.

Key words: gene families, FISH, BAC, SSR, aquaporin, bioinformatics.

Introduction

Scientific advances in the field of genomics have been promising for crop improvement in quality, productivity and resistance against pathogens, meeting the demands for food, fiber and biofuels. Such an interest has led to the production of large quantities of biological data from diverse sources. The continuous increase in the amount of available data on genomes and gene expression studies requires efficient storage, organization and data analysis. So the next logical step is to develop various graphical user interfaces or genome browsers, which provide logical access to data flows that otherwise would be unintelligible (Sen *et al.*, 2010). According to the Entrez Genome Project, in 2009 more than 150 projects related to the Viridiplantae genomes were initiated, including several species of agronomic, industrial and biotechnological interest, emphasizing the importance of bioinformatics platforms for the promotion of comparative genomics of model plants so as to enable us to understand the biological properties of each species, as well as accelerating gene discovery and functional analysis.

In this scenario, several genome browsers were developed, especially dedicated to generate information on cultivated and model plants. Gramene, for example, is a free online tool for genome comparison, providing a total of 15 genomes, including those of *Oryza sativa* (cv. japonica and cv. indica), *Arabidopsis thaliana*, *A. lyrata*, *Brachypodium distachyon*, *Populus trichocarpa*, *Sorghum bicolor* and *Vitis vinifera* (Youens-Clark *et al.*, 2010).

Send correspondence to Ana M. Benko-Iseppon. Laboratório de Genética e Biotecnologia Vegetal, Departamento de Genética, Universidade Federal de Pernambuco, Av. Prof. Morais Rego 1235, 50.670-420 Recife, PE, Brazil. E-mail: ana.iseppon@gmail.com.

Apêndice E

Functional Plant Science and Biotechnology ©2012 Global Science Books

Insight on Pathogen Defense Mechanisms in the Sugarcane Transcriptome

Ana Carolina Wanderley-Nogueira[1] • Ederson Akio Kido[1] • Nina da Mota Soares-Cavalcanti[1] • Luis Carlos Belarmino[1] • João Pacífico Bezerra-Neto[1] • William Lee Burnquist[2] • Sabrina Moutinho Chabregas[2] • José Ivo Baldani[3] • Ana Maria Benko-Iseppon[1*]

[1] Genetics Department, Center of Biological Sciences, Universidade Federal de Pernambuco, Av. Prof. Morais 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: * **ana.benko.iseppon@pq.cnpq.br**

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 pathogenesis-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, 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 their distribution, as compared with rice. For this purpose a collection of 282,818 expressed sequences tags have been evaluated 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 also 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 1,460 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.

Keywords: bioinformatics, crosstalk among biotic and abiotic stresses, *PR*-genes, *R*-genes
Abbreviations: ABA, abscisic acid; **AS**, salicylic acid; **avr**, avirulence; **BLAST**, basic local alignment search 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 expression

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 nematodes are the main disease agents, resulting in serious losses to agriculture and also native plants, reducing the productivity, nutritional value and overall quality of the produced biomass.

During the plant-pathogen co-evolution, plants develop a complex network of synergic mechanisms to defend against pathogen attack (Pinzón *et al.* 2009). Considering this complex response, one of the most important steps includes detection of the possible invaders by the plant, a step where Resistance (*R*) genes play a crucial role (Mofet 2009). This sensing involves the recognition of a pathogen gene product called avirulence (avr) factor by a matching *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) including local cell death to impair spreading of the pathogen (Jones and Dangl 2006). Besides this localized response, the HR activates a signal cascade which is able to establish resistance against a spectrum of different pathogens (Wang *et al.* 2005), corroborating the observations made at the beginning of the last century that plants, as animals, may be immunized against the attack a of given pathogen after infection by another pathogen (Chester 1933).

In the past decade many aspects of the systemic acquired resistance (SAR) have been elucidated (Ingle *et al.* 2006). The SAR pathway is also common in many incompatible plant-pathogen interactions (Park *et al.* 2010). As soon as the pathogen is detected, the plant induces a set of complex signal molecules that may activate defense proteins (Humphry 2010) or that may have direct antimicrobial effect, as it is the case of the pathogenesis-related (*PR*) genes (Durrant and Dong 2004) or alternatively the production of secondary metabolites that impair pathogen movement or growth within the plant tissues (Sparla *et al.* 2004; Benko-Iseppon *et al.* 2010).

Categories of *R* genes

Altogether *R* genes have been recently classified into five different groups or classes, defined by their conserved domains (CD; Bent 1996; Hammond-Kosak and Jones 1997; Ellis and Jones 2000).

The first class is represented by the *HM1* gene of maize that codes for a reductase able to inactivate toxins produced by the fungus *Helminthosporium carbonum* (Joahal and

Received: 12 April, 2011. Accepted: 15 November, 2011.

Original Research Paper

ANEXO I. Súmula Curricular associada à Tese

João Pacifico Bezerra Neto (CV: http://lattes.cnpq.br/7394341362881575)

Biólogo, Mestre em Genética pela Universidade Federal de Pernambuco, atualmente é Doutorando do Programa de Pós-Graduação em Ciências Biológicas da mesma instituição, estando vinculado ao Laboratório de Genética e Biotecnologia Vegetal. Tem experiência na área de Genética, com ênfase em Bioinformática e Biologia Molecular, atuando principalmente na pesquisa relacionada a prospecção de genes de tolerância a estresses bióticos e abióticos em culturas de importância econômica.

1. FORMAÇÃO COMPLEMENTAR

Phylogeny, Biogeography and Spatial Modeling – 65hs
Universidade Federal de Pernambuco, UFPE, Brasil.

Curso de curta duração em Bioinformática estrutural e análises do proteoma – 60hs
Centro Brasileiro-Argentino de Biotecnologia – CBAB e Universidade Federal de Minas Gerais, UFMG, Belo Horizonte, Brasil.

Systems Biology Applied to Biotic and Abiotic Interactions – 60hs
Universidade Federal de Pernambuco, UFPE, Brasil.

2. TRABALHOS REALIZADOS

Artigos completos publicados em periódicos

Kido, E.A.; Ferreira Neto, J.R.C.; Silva, R.L.; Belarmino, L.C.; **Bezerra Neto, J.P.;** Soares-Cavalcanti, N.M.; Pandolfi, V.; Silva, M.D.; Nepomuceno, A.L.; Benko-Iseppon, A.M. Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress. **BMC Bioinformatics**, v.14, p.S7, 2013.

Wanderley-Nogueira, A.C.; Belarmino, L.C.; Soares-Cavalcanti, N.M.; **Bezerra-Neto, J.P.;** Kido, E.A.; Pandolfi, V.; Abdelnoor, R.V.; Binneck, E.; Carazzole, M.F.; Benko-Iseppon, A.M. An overall evaluation of the resistance (R) and pathogenesis-related (PR) superfamilies in soybean, as compared with Medicago and Arabidopsis. **Genetics and Molecular Biology**, v.35, p.260 - 271, 2012.

Soares-Cavalcanti, N.M.; Belarmino, L.C.; Kido, E.A.; Wanderley-Nogueira, A.C.; **Bezerra-Neto, J.P.;** Cavalcanti-Lira, R.; Pandolfi, V.; Nepomuceno, A.L.; Abdelnoor, R.V.;

Nascimento, L.C.; Benko-Iseppon, A.M. In silico identification of known osmotic stress responsive genes from Arabidopsis in soybean and Medicago. **Genetics and Molecular Biology**, v.35, p.315 - 321, 2012.

Belarmino, L.C.; Oliveira, A.R.S.; Brasileiro-Vida, A.C.; Bortoleti, K.C.A.; **Bezerra-Neto, J.P.;** Abdelnoor, R.V.; Benko-Iseppon, A.M. Mining plant genome browsers as a means for efficient connection of physical, genetic and cytogenetic mapping: an example using soybean. **Genetics and Molecular Biology**, v.35, p.335 - 347, 2012.

Benko-Iseppon, A.M.; Soares-Cavalcanti, N.M.; Belarmino, L.C.; **Bezerra-Neto, J.P.;** 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, E.A. Prospecção de genes de resistência à seca e à salinidade em plantas nativas e cultivadas. Revista Brasileira de Geografia Física. v.6, p.1112, 2011.

Artigos aceitos para publicação

Wanderley-Nogueira, A.C.; **Bezerra-Neto, J.P.;** Kido, E.A.; Araújo, F.T.; Amorim, L.L.B.; Crovella, S.; Benko-Iseppon, A.M. Plant Elite Squad: first defense line and resistance genes – identification, diversity and functional roles. **Current Protein and Peptide Science**, 2016.

Silva-Lima, S.C.B.; Benko-Iseppon, A.M.; **Bezerra-Neto, J.P.**; Amorim, L.L.B.; Ferreira Neto, J.R.C.; Crovella, S.; Pandolfi, V. Plants Defense-related Cyclic Peptides: Diversity, Structure and Applications. **Current Protein and Peptide Science**, 2016.

Amorim, L.L.B; Santos, R.F.; **Bezerra-Neto, J.P.**; Guida-Santos, M.; Crovella, S.; Benko-Iseppon, A.M. Transcription factors involved in plant resistance to pathogens. **Current Protein and Peptide Science**, accepted Feb/2015.

3. CAPÍTULO DE LIVRO

Benko-Iseppon, A.M.; Wanderley-Nogueira, A.C.; Vasconcelos, S.S.; Amorim, L.L.B.; Bezerra Neto, J.P.; Marx Lima, M.O.; Azevedo, H.M.A.; Pandolfi, V. (2015). Mendel e suas exceções à luz das ômicas e da biologia de sistemas. In: Francisco J. L. Aragão & José Roberto Moreira. Mendel: 150 anos depois. **Embrapa Edições (Brasília, DF)**. Aceito.

Wanderley-Nogueira, A.C.; Kido, E.A.; Mota, N.S.C.; Bellarmino, L.C.; Bezerra-Neto, J.P.; Burnquist, W.; Chabregas, S.; Baldani, J.I.; Benko-Iseppon, A.M.. Insight on Pathogen Defense Mechanisms in the Sugarcane Transcriptome. In: Sundar A.R., Viswanathan R. (Org.). Sugarcane Pathology. 2ed. Isleworth: **Global Science Books**, 2012, v. 6, p. 134-148.

4. RESUMOS PUBLICADOS E APRESENTADOS EM CONGRESSOS (2015-2012)

1. Araújo, F.T., Matos, M.K.S., Oliveira, A.R.S., Silva-Lima, S.C.B., **Bezerra-Neto, J.P.**, Amorim, L.L.B., Benko-Iseppon, A.M. Caracterização e ancoragem de gene de resistência do tipo NBS-LRR em soja [*Glycine max* (L.) Merr, Fabaceae]. In: XX ENGENE Encontro de Genética do Nordeste, 2014, Campina
Grande - PB. **Anais do XX ENGENE - Encontro de Genética do Nordeste**. Campina Grande: Editora da Universidade Estadual da Paraíba, 2014, v.10.

2. Silva-Lima, S.C.B., **Bezerra-Neto, J.P.**, Benko-Iseppon, A. M., Pandolfi, V. Desenho de motivos, busca e caracterização in silico de genes codificadores de ciclotídeos em espécies vegetais. In: XX ENGENE Encontro de Genética do Nordeste, 2014, Campina Grande - PB. **Anais do XX ENGENE - Encontro de Genética do Nordeste**. Campina Grande: Editora da Universidade Estadual da Paraíba, 2014, v.10.

3. Araújo, F.T., Matos, M.K.S., **Bezerra-Neto, J.P.**, Amorim, L.L.B., Benko-Iseppon, A. M. Desenvolvimento de marcadores do tipo RGA (Resistance Gene Analog) em feijão-caupi. In: XX ENGENE Encontro de Genética do Nordeste, 2014, Campina Grande - PB. **Anais do XX ENGENE - ENCONTRO DE GENÉTICA DO NORDESTE**. Campina Grande: Editora da Universidade Estadual da Paraíba, 2014, v.10.

4. Matos, M.K.S., **Bezerra-Neto, J.P.**, Araújo, F.T., Amorim, L.L.B., Benko-Iseppon, A. M. Diversidade estrutural in silico de candidatos a fatores de transcrição da família MYB no transcriptoma do feijão-caupi. In: XX ENGENE Encontro de Genética do Nordeste, 2014, Campina Grande - PB. **Anais do XX ENGENE - Encontro de Genética do Nordeste**. Campina Grande: Editora da Universidade Estadual da Paraíba, 2014, v.10.

5. **Bezerra-Neto, J.P.**, Pandolfi, V., Benko-Iseppon, A. M. Ethylene Responsive (AP/ERF) transcription factors: abundance of stress response mediators in cowpea (*Vigna unguiculata*). In: XX ENGENE Encontro de Genética do Nordeste, 2014, Campina Grande - PB. **Anais do XX ENGENE - Encontro de Genética do Nordeste**. Campina Grande: Editora da Universidade Estadual da Paraíba, 2014, v.10.

6. Amorim, L.L.B., Kido, E.A., Ferreira Neto, J.R.C, **Bezerra-Neto, J.P.**, Benko-Iseppon, A.M. Expressão de fatores de transcrição em feijão-caupi sob estresse salino In: XX ENGENE Encontro de Genética do Nordeste, 2014, Campina Grande - PB. **Anais do XX ENGENE - Encontro de Genética do Nordeste**. Campina Grande: Editora da Universidade Estadual da Paraíba, 2014, v.10.

7. Amorim, L.L.B., Ferreira Neto, J.R.C., Kido, E.A., **Bezerra-Neto, J.P.**, Silva, M.D., Benko-Iseppon, A.M. Fatores de Transcrição: Diversidade e expressão diferencial em acessos contrastantes de feijão-caupi submetidos a estresse hídrico In: XX ENGENE Encontro de Genética do Nordeste, 2014, Campina Grande - PB. **Anais do XX ENGENE - Encontro de Genética do Nordeste**. Campina Grande: Editora da Universidade Estadual da Paraíba, 2014, v.10.

8. Benko-Iseppon, A.M., Pandolfi, V., Amorim, L.L.B., Santana, K.C.B., Belarmino, L.C., **Bezerra-Neto, J.P.**, Brasileiro-Vidal, A.C., Araújo, S.S., Rivas-Costa, R., Santos, M.G., Decuadro, S.B.R., Costa, A.F., Monte, S.J.H.M., Kido, E.A., Crovella, S. Genetic diversity and prospection of bioactive molecules from plants of the Brazilian northeastern region. In: German Latin American Conference on Knowledge and Technology Transfer & Biotechnology and Life Science DeLaTec, 2014, Potsdam. **Annals of the German Latin American Conference on Knowledge and Technology Transfer & Biotechnology and Life Science DeLaTec**. Potsdam: Potsdam University, 2014, v.1. p.9 – 9.

9. Piereck, B., **Bezerra-Neto, J.P.**, Brasileiro-Vidal, A.C., Benko-Iseppon, A.M. Identificação de domínios transposase (TNP) da superfamília Mutator e sequências relacionas em Glycine max.. In: XX ENGENE Encontro de Genética do Nordeste, 2014, Campina Grande - PB. **Anais do XX ENGENE - Encontro de Genética do Nordeste**. Campina Grande: Editora da Universidade Estadual da Paraíba, 2014, v.10.

10. Benko-Iseppon, A.M., Araújo, F.T., **Bezerra-Neto, J.P.**, Amorim, L.L.B., Wanderley-Nogueira, A.C. Identificação e ancoragem de genes de resistência TIR-NBS-LRR no genoma do feijão-caupi. In: XI Congresso Latinoamericano de Botánica/LXV Congresso Nacional de Botânica, 2014, Salvador. **Anais do XI Congresso Latinoamericano de Botánica/LXV Congresso Nacional de Botânica**, 2014.

11. Araújo, F.T., **Bezerra-Neto, J.P.**, Matos, M.K.S., Amorim, L.L.B., Wanderley-Nogueira, A.C., Benko-Iseppon, A. M. Identificação e ancoragem de genes de resistência TIR-NBS-LRR no genoma do feijão-caupi. In: VII SICBIO Simpósio de Ciências Biológicas, 2014, Recife - PE. **Anais do VII SICBIO Simpósio de Ciências Biológicas**, 2014.

12. Matos, M.K.S., **Bezerra-Neto, J.P.**, Amorim, L.L.B., Araújo, F.T., Benko-Iseppon, A.M. Prospecção de Genes MYB no Transcriptoma do Feijão-Caupi e Ancoragem em Pseudocromossomos de *Phaseolus vulgaris* L.. In: XX ENGENE - ENCONTRO DE GENÉTICA DO NORDESTE, 2014, Campina Grande - PB. **Anais do XX ENGENE - ENCONTRO DE GENÉTICA DO NORDESTE**. Campina Grande: Editora da Universidade Estadual da Paraíba, 2014, v.10.

13. Matos, M.K.S., **Bezerra-Neto, J.P.**, Amorim, L.L.B., Araújo, F.T., Benko-Iseppon, A. M. Structural domain analysis and 'in silico' expression of r2r3-myb subfamily in Cowpea (*Vigna unguiculata* (L.) Walp.) Transcriptome In: 60º CONGRESSO BRASILEIRO DE GENÉTICA, 2014, Guarujá-SP. **Anais do 60º Congresso Brasileiro de Genética**, 2014.

14. Vasconcelos, S., **Bezerra-Neto, J.P.**, Huettel, B., Benko-Iseppon, A.M. Characterization of microsatellites and gene content from genome shotgun sequences of Philodendron solimoesense (Araceae), a thermogenic arum lily from the Amazon basin. In: IV Simpósio Brasileiro de Genética Molecular de Plantas, 2013, Bento Gonçalves. **Resumos do IV Simpósio Brasileiro de Genética Molecular de Plantas**, 2013.

15. Benko-Iseppon, A.M., **Bezerra-Neto, J.P.**, Pandolfi, V., Wanderley-Nogueira, AC., Belarmino, L.C., Ferreira Neto, J.R.C., Abdelnoor, R.V., Marcelino-Guimaraes, F.C., Nepomuceno, A.L, Nascimento, L. C., C, Marcelo F., Pereira, G.A.G., Kido, E.A. Drought and salinity stress response in legume crops which genes really matter? In: IV Simpósio Brasileiro de Biologia Molecular de Plantas, 2013, Bento Gonçalves. **Anais do IV Simpósio Brasileiro**

de Biologia Molecular de Plantas, 2013, v.1, p.5 – 5.

16. **Bezerra-Neto, J.P.**, Belarmino, L.C., Pandolfi, V., Ferreira Neto, J.R.C., Marcelino-Guimaraes, F.C., Romero, C., Kido, E.A., Nepomuceno, A.L., Benko-Iseppon, A.M. Soybean Aquaporins: Differential Expression, Genome Distribution and Structure In: IX World Soybean Research Conference, 2013, Durban. **Annals of the IX World Soybean Research Conference**, 2013.

17. **Bezerra-Neto, J.P.**, Amorim, L.L.B., Pandolfi, V., Benko-Iseppon, A.M. Transcription Factor abundance of cowpea bean (*Vigna unguiculata*) an important insight in abiotic stress responses In: X-meeting: International Conference of th AB3C + Brazilian Symposium of Bioinformatics, 2013, Recife - PE. **Annals of the X-meeting 2013**, 2013.

18. Lima, M.O., Lima, J.P.P., Silva, S.C.B., Santos, R.F., Amorim, L.L.B., **Bezerra-Neto, J.P.**, Pandolfi, V., Benko-Iseppon, A.M. Characterization and macrosynteny analysis of the SOS pathway in castor bean In: 8th International Conference of the Brazilian Association for Bioinformatics and Computational Biology, 2012, Campinas. **Anais do 8th International Conference of the Brazilian Association for Bioinformatics and Computational Biology**, 2012.

19. Piereck, B., Amorim, L.L.B., **Bezerra-Neto, J.P.**, Belarmino, L.C., Benko-Iseppon, A.M. Identificação de domínios conservados de transposons em sequências de feijão-caupi (*Vigna unguiculata*) In: XIX Encontro de Genética do Nordeste, 2012, Petrolina. **Anais do XIX Encontro de Genética do Nordeste**, 2012.

20. Amorim, L.L.B., Lima, M.O., Piereck, B., **Bezerra-Neto, J.P.**, Pandolfi, V., Benko-Iseppon, A.M. Key members for flavonoid biosynthesis in cowpea *Vigna unguiculata* (L.) Walp.]: abundance, diversity and in silico expression In: 58º Congresso Brasileiro de Genética, 2012, Foz do Iguaçu. **Resumos do o 58º Congresso Brasileiro de Genética**. , 2012.

21. **Bezerra-Neto, J.P.**, Amorim, L.L.B., Pandolfi, V., Santos, R. F., Lima, J.P.P., Silva, S. C. B., Belarmino, L.C., Kido, E.A., Benko-Iseppon, A. M. An overall evaluation of the expressed reactive oxygen species (ROS) genes in cowpea. In: 7th International Conference Of The Brazilian Association For Bioinformatics And Computational Biology (AB3C) And 3rd International Conference Of The Iberoamerican Society For Bioinformatics (SolBio), 2011, Florianópolis, SC. **Abstract book**, 2011.

22. Piereck, B., Amorim, L.L.B., Belarmino, L.C., **Bezerra-Neto, J.P.**, Onofre, A.V.C, Brasileiro-Vidal, A. C., Benko-Iseppon, A. M. Identification Od Mutator-Like Transposable Elements From Cowpea Expressed Sequences In: 7th International Conference Of The Brazilian Association For Bioinformatics And Computational Biology (AB3C) And 3rd International Conference Of The Iberoamerican Society For Bioinformatics (SolBio), 2011, Florianópolis, SC. **Abstract book**, 2011.

23. Lira, R.C., **Bezerra-Neto, J.P.**, 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, 2011, Ilhéus, BA. **Resumos do III simpósio Brasileiro de Genética Molecular de Plantas**, 2011.

24. Amorim, L.L.B., Santana, K.C.B., Gazzaneo, L.R.S., Pandolfi, V., Belarmino, L.C., **Bezerra-Neto, J.P.**, Crovella, S., Kido, E.A., Benko-Iseppon, A.M. Sequence analysis and homology modeling of the first defensin from *Etlingera elatior* In: 7th International Conference Of The Brazilian Association For Bioinformatics And Computational Biology (AB3C) And 3rd International Conference Of The Iberoamerican Society For Bioinformatics (SolBio), 2011, Florianópolis, SC. **Abstract book**, 2011.

25. Araujo, G.I.R, **Bezerra-Neto, J.P.**, Belarmino, L.C., Benko-Iseppon, A.M. Thionin Encoding Genes in Soybean: Identification and Expression Profiling in Response to Biotic and Abiotic Factors In: III Simpósio Brasileiro de Genética Molecular de Plantas, 2011, Ilhéus, BA. **Resumos do III simpósio Brasileiro de Genética Molecular de Plantas**, 2011.

26. **Bezerra-Neto, J.P.**, Soares-Cavalcanti, N.M., Benko-Iseppon, A.M. Unraveling the cowpea (*Vigna unguiculata* (L.) Walp.) response to osmotic stress via aquaporin genes In: III Simpósio Brasileiro de Genética Molecular de Plantas, 2011, Ilhéus, BA. **Resumos do III simpósio Brasileiro de Genética Molecular de Plantas**. , 2011.

27. Silva, A.M., **Bezerra-Neto, J.P.**, Belarmino, L.C., Amorim, L.L.B., Pandolfi, V., Benko-Iseppon, A.M. Catalase I in the Transcriptome of Soybean (*Glycine max* (l.) Merr.): redox state maintainers in seeds and leaves? In: 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy, 2010, Recife. **Abstracts of 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy**, 2010, p.22 – 22.

28. **Bezerra-Neto, J.P.**, Belarmino, L.C., Cavalcanti, N.S.M., Benko-Iseppon, A.M. In silico Characterization of Major Intrinsic Proteins in the Soybean Transcriptome (*Glycine max* (L.) Merr.) In: 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy, 2010, Recife. **Abstracts of 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy**, 2010, p.8 – 8.

29. Lira, R.C., Cavalcanti, N.S.M., **Bezerra-Neto, J.P.**, Benko-Iseppon, A.M. In Silico Characterization of the SOS (Salt Overly Sensitive) Pathway in the Expressed Genome of the Sugar Cane In: 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy, 2010, Recife. **Abstracts of 1st Brazilian-German Meeting of Plant Systems Biology and Bioenergy**, 2010, p.13 – 13.

Trabalhos publicados em anais de eventos (resumo expandido)

1. Lima, M.O., **Bezerra-Neto, J.P.**, Lima, J.P.P., Santos, R.F., Soares-Rodrigues, S.S., Belarmino, L.C., Wanderley-Nogueira, A.C., Kido, E.A., Benko-Iseppon, A.M. Caracterização e expressão in silico de esnaquinas no feijão-caupi (*Vigna unguiculata*). In: III CONAC, Congresso Nacional do Feijão-Caupi, 2013, Recife - PE. **Anais do III CONAC, Congresso Nacional do Feijão-Caupi**, 2013, v.1, p.1 – 5.

2. Piereck, B., Amorim, L.L.B., Belarmino, L.C., **Bezerra-Neto, J.P.**, Wanderley-Nogueira, A.C., Pandolfi, V., Brasileiro-Vidal, A.C., Benko-Iseppon, A.M. Identificação e caracterização de domínios TNP1 e TNP2 do elemento transponível CACTA em transcritos de feijão-caupi. In: III CONAC, Congresso Nacional do Feijão-Caupi, 2013, Recife - PE. **Anais do III CONAC, Congresso Nacional do Feijão-Caupi**, 2013, v.1, p.1 – 5.

3. **Bezerra-Neto, J.P.**, Pandolfi, V., Kido, E.A., Monte, S.J.H.M., Benko-Iseppon, A.M.

Modelagem comparativa e análise estrutural de uma aquaporina de feijão-caupi (*Vigna unguiculata* (L.) Walp.). In: III CONAC, Congresso Nacional do Feijão-Caupi, 2013, Recife - PE. **Anais do III CONAC, Congresso Nacional do Feijão-Caupi**, 2013, v.1, p.1 – 5.

4. Lima, J.P.P., **Bezerra-Neto, J.P.**, Lima, M.O., Pandolfi, V., Wanderley-Nogueira, A.C., Benko-Iseppon, A.M. Via SOS: análise da expressão in silico no transcriptoma do feijão-caupi. In: III CONAC, Congresso Nacional do Feijão-Caupi, 2013, Recife - PE. **Anais do III CONAC, Congresso Nacional do Feijão-Caupi**, 2013, v.1, p.5 – 5.

5. Lima, J.P.P., **Bezerra-Neto, J.P.**, Lima, M.O., Santos, R.F., Pandolfi, V., Benko-Iseppon, A.M. Identificação e análise estrutural in silico de genes associados à via SOS (Salt-Overly-Sensitive) no transcriptoma do feijão-caupi In: Simpósio de Inovação em Ciências Biológicas, 2012, Recife. **Anais do V SICBIO Simpósio de Inovação em Ciências Biológicas**, 2012.

6. ORIENTAÇÕES E SUPERVISÕES CONCLUIDAS

Sheyla Carla Barbosa da Silva Lima. Fitorremediação: uma alternativa verde. (2014)
Especialização em Perícia e Auditoria Ambiental - Faculdade Frassinetti do Recife.

Ivanberto Pacifico de Souza. Caracterização e Modelagem Estrutural in silico de Lectinas Vegetais em Espécies de Interesse Econômico. (2013)
Graduação em Biomedicina - Universidade Federal de Pernambuco.