PETRA BARROS DOS SANTOS

ANÁLISE *IN SILICO* DE OSMOPROTETORES NO GENOMA EXPRESSO DA CANA-DE-AÇÚCAR, EUCALIPTO E FEIJÃO-CAUPI

RECIFE
2009

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE GENÉTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA
DISSERTAÇÃO DE MESTRADO

ANÁLISE *IN SILICO* DE OSMOPROTETORES NO GENOMA EXPRESSO DA CANA-DE-AÇÚCAR, EUCALIPTO E FEIJÃO-CAUPI

PETRA BARROS DOS SANTOS

Dissertação apresentada ao Programa de Pós-graduação em Genética da Universidade Federal de Pernambuco como requisito para obtenção do grau de Mestre em Genética pela UFPE

Orientador: Dra. Ana Maria Benko Iseppon

Co-orientador: Dr. Tercílio Calsa Junior

RECIFE
2009

Santos, Petra Barros dos

Análise in silico de osmoprotetores no genoma expresso da cana-de-açúcar, eucalipto e feijão-caupi/ Petra Barros dos Santos – Recife: O Autor, 2009

142 folhas: il., fig., tab.

Dissertação (mestrado) – Universidade Federal de Pernambuco. CCB. Genética, 2009.

Inclui bibliografia e anexos

1. Melhoramento genético. 2. Osmoprotetores. 3. Estresse abiótico. 4. Bioinformática I Título.

631.52 CDU (2.ed.) UFPE
631.52 CDD (22.ed.) CCB – 2009- 137

UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM
GENÉTICA E BIOLOGIA MOLECULAR

Parecer da comissão examinadora da dissertação de Mestrado de:
Petra Barros dos Santos

Análise *in silico* de Osmoprotetores no Genoma Expresso da Cana-de-açúcar, Eucalipto e Feijão-caupi

A comissão examinadora considera o presente trabalho:

APROVADO

Portanto, cumpridas todas as exigências regimentais, Petra Barros dos Santos faz jus ao grau de Mestre em Genética e Biologia Molecular pela UFPE.
Recife, 2/07/09

1º Examinador: Drª. Kátia Silva Guimarães (UFPE-CIN)

2º Examinador: Dr. Mauro Guida (UFPE)

3º Examinador: Dr. Valdir Queiroz Balbino (UFPE)

Orientador: Drª. Ana Maria Benko-Iseppon (UFPE)

Co-orientador: Dr. Tercílio Calsa Júnior (UFPE)

“O verdadeiro homem mede a sua força, quando se defronta com o obstáculo.”

Antoine de Saint-Exupéry”

Dedico:

Aos meus pais, Ana Maria Barros dos Santos e Dacier de Barros e Silva. *A minha irmã,* Bruna Dacier Barros dos Santos.

Agradecimentos

Agradeço primeiramente a Deus, pois sem a fé que nele deposito nada seria possível.

Agradeço aos meus pais "ANA MARIA BARROS DOS SANTOS e DACIER BARROS E SILVA", pessoas a quem mais amo neste mundo, pois me proporcionaram o melhor que puderam para a construção da minha vida pessoal e acadêmica.

Agradeço a minha irmã BRUNA DACIER BARROS que esteve sempre comigo nas horas de dificuldade.

A THIAGO ALBERTO DE LIMA MORAIS pelo seu amor, carinho, paciência e estímulo para finalizar este trabalho.

À GABRIELA SOUTO, amiga verdadeira que ganhei no laboratório, pela ajuda, pelas risadas e milhares de confidências trocadas.

Aos meus amigos MARIANA DURANT, FABÍOLA MARTINS, PETRA BRISSANTT, MOACYR BARRETO, MARCELA RANDAU e MIRELLA SOARES, pela sinceridade e vontade de ajudar no que foi preciso.

Agradeço à NINA MOTA CAVALCANTI pela dedicação, pela ajuda, pelos ensinamentos enquanto juntas trabalhamos. Também sou grata a ANA CAROLINA WANDERLEY, pela ajuda no laboratório.

À professora ANA MARIA BENKO ISEPPON, minha orientadora, responsável pela concretização deste projeto.

Ao professor e co-orientador TERCÍLIO CALSA JUNIOR, pela atenção e conselhos nas horas de dúvidas.

À Professora MÔNICA CARVALHO e ao professor TERCÍLIO pela oportunidade de lecionar e aprender no estágio de iniciação à docência.

A todos os funcionários do Departamento de Genética e também às técnicas do LGBV, CLAUDETE e VANESSA, pelo auxílio e apoio dispensados.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por viabilizar a realização deste trabalho com a concessão de bolsa de estudo.

Enfim, agradeço a todos que de uma forma ou de outra contribuíram para minha realização acreditando sempre, dando-me força e torcendo para que todos os meus projetos continuem a ser realizados.

MUITO OBRIGADA!

SUMÁRIO

LISTA DE ABREVIATURAS

AtNHX	Trocador Na^+/H^+ de *Arabidopsis thaliana*
BADH	Betaína Aldeído Desidrogenase
BLAST	Ferramenta Básica de Alinhamento Local de Sequências
BRACELPA	Associação Brasileira de Celulose e Papel
CD	Domínio Conservado
CDH	Colina Desidrogenase
cDNA	DNA Complementar ao mRNA
CGKB	Base de Conhecimentos Genômicos de Caupi
CMO	Colina Mono-oxigenase
COPERSUCAR	Cooperativa Central dos Produtores de Açúcar e Álcool do Estado de São Paulo
COX	Colina Oxigenase
CSC	Complexo Cisteína Sintase
DMT	Dimetilglicina Metiltransferase
EST	Etiquetas de Sequências Expressas
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FORESTs	Projeto Consórcio de Sequenciamento do Genoma do Eucalyptus
GB	Glicina-Betaína
Glu	Ácido Glutâmico
GSMT	Sarcosina Glicina Metiltransferase
GenBank	Banco de Genes do NCBI
GSA	Ácido Glutâmico γ-semialdeído
GSS	Seqüências Ricas em Genes
HDF	Fibra de Alta Densidade
IAC	Instituto Agronômico de Campinas
INPS1	Mio-Inositol fosfato sintase
LEA	Abundante na Embriogênese Tardia
MAPK	Cascata de Proteínas Quinases
MDF	Fibra de Média Densidade
MEGA	Análises Genéticas e Evolução Molecular

Mbp	Mega Pares de Bases
NCBI	Centro Nacional de Biotecnologia e Informação – EUA
NJ	Agrupamento por Vizinhança
NordEST	Rede Nordeste de Biotecnologia
OASTL	O-Acetil Serina (Tiol) Liase
ONSA	Organização para Sequenciamento e Análise de Nucleotídeos
ORF	Quadro Aberto de Leitura
Orn	Ornitina
PALP	Piridoxal-5-Fosfato
PIB	Produto Interno Bruto
PMGCA	Programas de Melhoramento Genético da Cana-de-açúcar
P5C	Δ^1-Pirrolina-5-Carboxilato
P5CR	Δ^{1-}Pirrolina-5-Carboxilato Redutase
P5CS	Δ^{1-}Pirrolina-5-Carboxilato Sintetase
ROS	Espécies Reativas de Oxigênio
SAT	Serina Acetil-Transferase
SAGE	Análise Serial da Expressão Gênica
SOS	Altamente Sensível ao Sal
SUCEST	Projeto EST da Cana-de-açúcar
SWISS-PROT	Banco de Dados de Sequências de Proteínas Curado
TPP	Trealose 6-fosfato Fosfatase
TPS	Trealose 6-fosfato Sintase
UPGMA	Método não Polarizado de Agrupamentos aos Pares com Médias Aritméticas

LISTA DE FIGURAS

CAPÍTULO 1

CAPÍTULO 2

CAPÍTULO 3.

Supplementary Data. Dendrograms generated after maximum parsimony analysis illustrating relationship revealed by conserved domains similarity in (I) OASTL-like and (II) SAT-like (III) TPP-like and (IV) TPS-like sequences and putative cowpea and sugarcane orthologs. 132

LISTA DE TABELAS

CAPÍTULO 1

CAPÍTULO 2

CAPÍTULO 3.

RESUMO

Estresses abióticos, como seca e salinidade, limitam severamente o crescimento, desenvolvimento e produtividade das plantas. Uma das estratégias que muitos vegetais utilizam, a fim de minimizar os danos causados por esses estresses aos processos metabólicos e fisiológicos, é a síntese de osmólitos compatíveis. Essas substâncias são trocadas pelo excesso de sais inorgânicos na célula, permitindo o ajustamento tanto da concentração interna de sais quanto do volume celular. Este trabalho objetivou identificar *in silico* candidatos aos genes que codificam osmoprotetores (trealose, glicina-betaína, mio-inositol, prolina e cisteína) nos transcriptomas de *Eucalyptus* spp., *Saccharum* spp. e *Vigna* spp. As análises revelaram a presença destes genes em todos os transcriptomas estudados; ao todo foram observados 51 ortólogos em eucalipto, 56 em cana e 73 em feijão-caupi. De um modo geral, com relação ao padrão de expressão foi percebido o envolvimento destes genes tanto no estresse biótico, quanto no abiótico, sendo mais abundantes os transcritos identificados nas bibliotecas de caule de mudas submetidas a déficit hídrico e plântulas cultivadas no escuro, no caso do eucalipto e em tecidos de calos (CL) submetidos a um ciclo de luz/escuro e estresse de temperatura no caso da cana-de-açúcar. Adicionalmente, no caupi, a partir da contagem direta dos transcritos, foi identificado que a maioria dos transcritos integrava a biblioteca em raiz de plantas sensíveis à salinidade após 2 horas de exposição ao estresse. Nos alinhamentos múltiplos gerados para a comparação das sequências de cada um desses genes entre diferentes organismos (desde bactérias até animais), percebe-se que a síntese de osmólitos compatíveis é uma característica que foi bastante conservada no curso da evolução. Nos dendrogramas gerados a partir dos dados supracitados, correspondendo ao esperado, observou-se a separação das espécies de acordo com a sua classe. Além disso, no que tange os vegetais, mono e dicotiledôneas foram agrupadas em clados distintos. Considerando que os organismos estudados apresentam grandes diferenças no metabolismo, fisiologia, hábito e ciclo de vida, os resultados sugerem que as mutações acumuladas, principalmente nas diferentes classes, são resultado da adaptação às pressões seletivas ambientais e estão, assim, relacionadas com a funcionalidade destes genes nos diferentes organismos. Assim, os resultados apontam para a conservação das principais vias de síntese de osmólitos compatíveis tanto em eucalipto, quanto em cana e feijão-caupi. Ademais, este estudo pode servir como modelo para a identificação de genes osmoprotetores em espécies relacionadas e apresenta grande potencial de aplicação ao melhoramento genético, com o desenvolvimento de culturas economicamente importantes mais adaptadas aos solos salinos e secos.

Palavras-chave: Estresse abiótico, bioinformática, ESTs, mineração de dados.

ABSTRACT

Abiotic stresses such as drought and salinity, severely limiting the growth, development and productivity of plants. One of the strategies that many plants use in order to minimize the damage caused by these stresses to the physiological and metabolic processes is the synthesis of compatible osmolytes. These substances are exchanged by the excess of inorganic salts in the cell, allowing the adjustment both of the internal concentration of salts and the cell volume. This study aimed to identify in silico candidates for genes encoding osmoprotectants (trehalose, glycine-betaine, myo-inositol, proline and cysteine) in the *Eucalyptus* spp., *Saccharum* spp. and *Vigna* spp transcriptomes. The analysis revealed the presence of these genes in all studied transcriptome; totally were observed 51 ortologs in eucalyptus, 56 in sugarcane and 73 in cowpea. In general, regarding the expression pattern was seen the involvement of these genes in both biotic and abiotic stress, being the most abundant transcripts identified in libraries from stems of seedlings subjected to water deficit and seedlings grown in the dark, in the case of eucalyptus and tissues of calli (CL) submitted to light/dark and temperature stress in the case of sugarcane. Additionally, in cowpea, from the direct counting of the transcripts, was identified that the majority of transcripts contained in the library root of plants sensitive to salinity after two hours exposure to salt stress. In the generated multiple alignments for the comparison of sequences of each of these genes between different organisms (from bacteria to animals), we find that the synthesis of compatible osmolytes is a feature that was kept in the course of evolution. In the generated dendrograms from the data above, corresponding to the expected, there was the separation of species according to their class. Furthermore, with regard to plants, monocots and dicots were grouped in distinct clades. Whereas the studied organisms show large differences in metabolism, physiology, habits and life cycle, results suggest that the mutations accumulated, mainly in the different classes, are the result of adaptation to environmental selective pressures and are thus related to the functionality of these genes in different organisms. Thus, the results point to the conservation of the main routes of synthesis of compatible osmolytes both in *Eucalyptus*, as in sugarcane and cowpea. Moreover, this study can serve as a model for identifying osmoprotectants genes in related species and has great potential for application to breeding, with the development of economically important crops better adapted to saline and dry soils.

Key-words: Abiotic stress, bioinformatics, ESTs, data-mining.

INTRODUÇÃO

A produtividade e a qualidade de muitas culturas de importância econômica estão diretamente relacionadas às condições edafo-climáticas de cada região, sendo necessário um maior domínio dessas condições e o conhecimento das necessidades hídricas das plantas nos seus diferentes estágios de desenvolvimento. As zonas áridas e semi-áridas são caracterizadas, na maioria das vezes, pela limitação no suprimento de água, tanto em termos de quantidade quanto de qualidade. Tratam-se de regiões onde as taxas de evapo-transpiração são elevadas resultando na reposição de altas concentrações de sais na superfície do solo, o que consequentemente favorece a perda de água pela cultura.

Os processos fisiológicos e metabólicos utilizados pelas plantas em resposta a estresses abióticos têm sido bastante investigados. Os mais conhecidos estão relacionados à desestabilização das membranas celulares e aos níveis de proteínas. O alto teor de sais no solo, principalmente dos íons Na^+ e Cl^-, pode inibir da absorção dos íons K^+ e Ca^{++} pelas raízes, provocando sintomas de deficiências, como redução no crescimento, afetando também o teor de aminas e proteínas. Outra consequência da salinidade é o aumento na concentração de radicais livres (ROS - *Reactive oxygen species*; Espécies oxigênio-reativas), os quais reduzem a capacidade fotossintética das plantas, levando consequentemente, à redução do seu crescimento.

Com tantos efeitos causados pela salinidade/seca e sua associação a outros fatores, espera-se que a adaptação das plantas a estes estresses envolva um grande número de modificações, tais como, excreção de sais, compartimentalização iônica, adaptação osmótica, controle da expressão de alguns genes, síntese de osmólitos compatíveis, entre outros.

A produção de osmólitos compatíveis (metabólitos osmoprotetores) é conhecida como um dos mecanismos que os organismos, inclusive as plantas, possuem para driblar os efeitos nocivos, causados por alguns tipos de estresses, em especial causados pela salinidade e a seca. Essas substâncias são trocadas pelo excesso de sais inorgânicos na célula, permitindo que tanto a concentração interna de sais quanto o volume celular sejam ajustados. A vantagem desse acúmulo é que essas substâncias não perturbam a funcionalidade das macromoléculas celulares, como ocorre com os sais inorgânicos. Esses

compostos osmoprotetores incluem açúcares como a trealose, polióis, aminoácidos como a prolina e a cisteína, entre outros.

A cana-de-açúcar, o eucalipto e o feijão-caupi, além da significativa importância sócio-econômica para o sistema agroindustrial brasileiro, têm adquirido considerável destaque no âmbito científico e tecnológico com os resultados do sequenciamento de ESTs (Projetos SUCEST; FORESTs e Rede NordEST, entre outros) e da categorização de milhares de genes cujas funções biológicas vêm sendo foco de estudos. Essa crescente geração e disponibilização de sequências genômicas e protéicas oriundas de espécies de interesse agronômico irão auxiliar na identificação de genes envolvidos nas respostas às condições estressantes. Assim, além de estudos de genômica funcional, a análise proteômica, apresenta-se como um fator importante para a obtenção de variedades e híbridos com características comerciais mais favoráveis.

O presente trabalho teve como objetivo identificar, caracterizar e analisar estruturalmente, com o auxílio de ferramentas computacionais, sequências candidatas a genes envolvidos na síntese de osmoprotetores nos transcriptomas de *Vigna unguiculata, Saccharum* spp. e *Eucalyptus* spp., comparando-as entre si, bem como àquelas depositadas em bancos de dados públicos e descritas na literatura, inferindo sobre sua estrutura, expressão e função.

Capítulo 1

Revisão da Literatura

---

1. REVISÃO DA LITERATURA

1.1. Feijão-caupi

O feijão-caupi [*Vigna unguiculata* (L.) Walp.], também conhecido como feijão macassar, fradinho, ou de corda, constitui uma das principais leguminosas cultivadas no Brasil e no mundo. O centro de origem e diversidade dessa espécie ainda é bastante controverso. Muitos afirmam ser a África, embora outras regiões tenham sido sugeridas, como a Índia, Etiópia e a América do Sul. Contudo, através de análises moleculares Simon et al. (2007) sugerem que o gênero e a espécie tenham se originado no sudoeste da Ásia, enquanto o sul da África representaria a região de diversificação da espécie.

Especula-se que o caupi foi introduzido na América Latina no século XVI, pelos colonizadores espanhóis e portugueses, primeiramente nas colônias espanholas e em seguida no Brasil, provavelmente pelo estado da Bahia, sendo a partir deste levado para outras áreas do país (Freire Filho, 1988). Por outro lado, com base em análises moleculares Simon et al. (2007) postulam a ocorrência de mais de um evento de introdução do feijão-caupi no Brasil por colonizadores e escravos de diferentes regiões da África durante os séculos XVI e XVII.

O feijão-caupi é cultivado predominantemente nas regiões Norte e Nordeste, onde é usado principalmente para fins alimentares (subsistência ou comercial), representando 95 a 100% do total das áreas ocupadas nessas regiões. O sucesso de seu plantio se deve à sua alta capacidade de adaptação às condições do semi-árido, região marcada por períodos prolongados de estiagem, alta salinidade nos solos e temperaturas elevadas (Silveira et al., 2001).

Nas áreas urbanas, não-metropolitanas, do Nordeste o caupi contribui com 41% do feijão consumido, tratando-se de alimento básico para a população, sendo fundamental como fonte nutritiva para as camadas carentes. Este grão é considerado uma fonte barata no suprimento de proteína (média de 23-25%), apresentando todos os aminoácidos essenciais, carboidratos (média de 62%), vitaminas, minerais, além de possuir quantidade significativa de fibras dietéticas, baixa quantidade de gordura (teor de óleo em torno de 2%) e não conter colesterol. Desta forma, não apenas as propriedades nutricionais do

feijão-caupi são relativamente superiores às do feijão comum (*Phaseolus*), havendo a vantagem adicional do baixo custo de sua produção, fazendo com que esta cultura desempenhe grande importância econômica e social (Embrapa Meio-Norte, 2008).

1.1.1. Taxonomia, Importância Econômica e Melhoramento

O feijão-caupi é uma dicotiledônea (Magnoliopsida) que pertence à ordem *Fabales*, família *Fabaceae* (leguminosas), subfamília *Faboideae,* tribo *Phaseoleae,* subtribo *Phaseolinea* e gênero *Vigna* (Freire Filho et al., 1999). Possui um dos menores genomas do grupo das leguminosas (450-500 Mb), apresentando o nível diplóide com 2n=22 cromossomos (Benko-Iseppon, 2001).

Devido seu alto valor nutritivo, o feijão-caupi é cultivado principalmente para a produção de grãos, secos ou verdes, *in natura*, em conserva ou desidratado. Além disso, também pode ser empregado na forragem verde, feno, ensilagem, farinha para alimentação animal, como adubo verde e na proteção do solo (Embrapa Meio-Norte, 2008).

O Brasil é o terceiro produtor mundial de feijão-caupi com 11,5 milhões de hectares plantados e rendimento médio de 317 $kg.ha^{-1}$ (Freire Filho et al., 2005; Bezerra et al., 2008), sendo o Ceará o maior produtor nacional, com estimativa em cerca de 20% da produção brasileira. Com relação aos aspectos socioeconômicos, a cultura do caupi é responsável pela geração de quase 1.500.000 empregos/ano no Brasil, com o valor de produção estimado em US$ 250.000.000/ano (IBGE, 2008). Esses dados são de extrema importância, por refletir a participação da cultura na geração de emprego, de renda e da produção de alimentos do país, o que a credencia para receber maior atenção por parte dos órgãos de apoio à pesquisa.

Nesse contexto, nos últimos anos cresceram as perspectivas da utilização da irrigação para aumento da produtividade, ampliação do período de colheita e melhoria da qualidade através de programas de melhoramento no país, visando ganhos na qualidade nutricional dos grãos e na resistência a pragas e doenças. Os primeiros trabalhos que visavam o melhoramento do feijão de corda foram iniciados na década de 1960 e tinham como objetivo o aumento da produtividade (Paiva et al., 1970).

O melhoramento do caupi visa principalmente: (I) Desenvolver cultivares com alta qualidade de grãos (aspecto visual, culinário e alimentar), alto potencial produtivo e bem adaptado aos sistemas de cultivo de sequeiro e/ou irrigado; (II) Desenvolver cultivares com resistência múltipla a vírus e a outras doenças fúngicas e bacterianas; (III) Identificar fontes de resistência a insetos, pragas e vetores, bem como ampliar os cultivares com essas características; (IV) Desenvolver cultivares com porte mais compacto, mais ereto, que possibilitem a colheita mecanizada; entre outros (Freire Filho et al., 2005).

Segundo Freire Filho et al. (2005) é necessário que se iniciem estudos mais avançados de genética do caupi e que os programas de melhoramento convencional busquem suporte na biologia molecular a fim de que se tornem mais ágeis e eficientes, para que esta cultura possa competir em qualidade, oferta e custo de produção com outros feijões.

1.1.2. Projetos NordEST, HARVEST e CGKB

Projetos iniciados em 2001 visando o sequenciamento de várias culturas, incluindo o feijão-caupi, o café, a laranja, a soja e o arroz, deram origem a um banco de dados internacional, o Banco de Dados HARVEST. Integrante deste, o HARVEST *Cowpea* compreende ESTs geradas a partir de 17 bibliotecas de cDNA (DNA complementar ao mRNA) oriundos de vários tecidos. Essas ESTs (*expressed sequence tag,* ou região de sequência expressa) foram sequenciadas pelo Instituto Genoma, do Departamento de Energia dos Estados Unidos (*U.S. Department of Energy Joint Genome Institute*; 141.050 ESTs), Instituto J. Craig Venter (*J. Craig Venter Institute;* 41.505 ESTs) e Universidade de Califórnia, Campus de Riverside (*University of California, Riverside*; 488 ESTs), estando desde 2005 disponíveis para acesso pela comunidade científica mundial (HARVEST, 2008).

Em 2004, foi iniciado o projeto "Genômica funcional, estrutural e comparativa de feijão-caupi (*V. unguiculata*)", integrante da Rede Nordeste de Biotecnologia (NordEST). O projeto abrange a organização de uma rede de laboratórios da Região Nordeste visando efetuar análises de genômica funcional e estrutural no feijão-caupi, com intuito de identificar novos genes candidatos potencialmente úteis para fins de melhoramento desta

cultura. Um dos focos do projeto relaciona-se à obtenção de bibliotecas genômicas contrastantes (sob estresse/controle; resistente/susceptível) a condições de alta salinidade (associada ou não à seca), com a geração de 10.000 sequências expressas para esta finalidade (Tabela 1).

O projeto NordEST pretende obter 78.000 sequências de ESTs e 24.000 sequências de SAGE (*tags*; *Serial analysis of gene expression*, ou análise serial da expressão gênica) relacionadas à expressão diferencial de genes de resistência a fatores bióticos e abióticos da maior importância para o melhoramento do feijão-caupi com técnicas de biotecnologia (Benko-Iseppon et al., 2005). Além disso, o projeto tem gerado mais de 600.000 *tags* através da utilização da metodologia SuperSAGE.

Tabela 1: - Descrição das bibliotecas geradas pelo projeto NordEST, incluindo o código da biblioteca, bem como quantidade total de ESTs dentro de uma categoria de bibliotecas, descrição dos tecidos de extração das ESTs.

Código da biblioteca	Nº Total de ESTs	Descrição	Tecido
CT00	288	Controle	Raiz
BM90	1624	Genótipo BR14-Mulato	Folha
IM90	464	Genótipo IT85F coletado 90 minutos após infecção com vírus do mosaico	Folha
SS00	1433	Genótipo sensível à salinidade sem estresse salino	Raiz
SS02	2204	Genótipo sensível à salinidade após duas horas de estresse*	Raiz
SS08	3647	Genótipo sensível à salinidade após oito horas de estresse*	Raiz
ST00	2500	Genótipo tolerante à salinidade sem estresse salino	Raiz
ST02	3646	Genótipo tolerante à salinidade após duas horas de estresse*	Raiz
ST08	3142	Genótipo tolerante à salinidade após oito horas de estresse*	Raiz

* Genótipos submetidos a estresse salino foram cultivados com 200mM NaCl

O CGKB (*Cowpea Genomics Knowledge Base*; Base de Conhecimentos Genômicos de Caupi) trata-se de um banco de dados de sequências genômicas baseado em informações derivadas de 298.848 sequências ricas em genes (GSS - *cowpea genespace sequences*) isoladas através do método de filtragem de DNA metilado. O banco contém principalmente sequências nucleares (263.425), além de organelas e de transposons/retrotransposons (http://cowpeagenomics.med.virginia.edu/CGKB/; Chen et al., 2007).

1.2. Cana-de-açúcar

A cana-de-açúcar é uma das principais culturas agrícolas do mundo, sendo cultivada em mais de 120 países (Matsuoka et al., 1999). É cultivada predominantemente em áreas subtropicais entre 15° e 30° de latitude, podendo se estender até 35° de latitude tanto ao norte quanto ao sul, sendo produzida comercialmente em mais de 70 países. Os maiores produtores são: Brasil, Cuba, Índia, México, China, Filipinas, Austrália, África do Sul, Estados Unidos da América, República Dominicana e Formosa (Barbosa, 2005).

O centro de origem da cana ainda é muito discutido. Alguns pesquisadores consideram a região leste da Indonésia e a Nova Guiné (Daniels e Roach, 1987), tendo em vista a grande diversidade genética ali encontrada para o gênero *Saccharum.* Outros afirmam ser precisamente nas regiões de Assam e Bengala, localizadas na Índia.

A colonização da América do Sul permitiu extraordinária expansão das áreas de cultura da cana. As primeiras mudas chegaram ao Brasil em 1502, e, já em 1550, numerosos engenhos nas regiões Centro-Oeste, Sudeste e, principalmente, Nordeste, espalhados pelo litoral, produziam açúcar de qualidade equivalente ao produzido pela Índia (Barbosa, 2005).

Embora grande produtor desde o Brasil Colônia, o país expandiu muito a cultura da cana a partir da década de 1970, com o advento do Pro-Álcool; programa do governo criado em 1975 que substituiu parte do consumo de gasolina por etanol (obtido a partir da cana-de-açúcar), sendo pioneiro no uso em larga escala deste álcool como combustível automotivo (Biodiesel, 2008).

Os dez principais produtores mundiais são responsáveis por 74% da produção de açúcar. Dentre estes, o Brasil é o maior produtor, sendo seus produtos largamente utilizados na produção de açúcar, álcool combustível, ocupando um papel de liderança, tanto na produção como na exportação. Esse potencial competitivo do Brasil se deve não só à produtividade e rendimento industrial, como também às pesquisas sobre novas variedades e utilização de novas técnicas agrícolas e do solo, condições que garantem ao país os menores custos de produção do mundo (Amaral et al., 2003).

1.2.1. Taxonomia, Importância Econômica e Melhoramento

A cana-de-açúcar é classificada taxonomicamente como membro da divisão *Embriophyta,* incluída na subdivisão *Angiospermae*, classe *Monocotyledoneae* (Liliopsida), família *Poaceae* (gramíneas juntamente com os gêneros *Zea* e *Sorghum*), tribo *Andropogoneae* e gênero *Saccharum* (*Angiosperm Phylogeny Website*, 2008). O gênero *Saccharum* compreende seis espécies: *S. officinarum, S. robustum, S. spontaneum, S. sinense, S. barberi*, e *S. edule*. Dentre estas, *S. robustum* e *S. spontaneum* são reconhecidas como espécies selvagens (Guimarães et al., 1999).

Os genomas das espécies do gênero *Saccharum* são altamente complexos, com números cromossômicos variando de 2n=36 a 2n=170 (Takahashi et al., 2005). Ingelbrecht et al. (1999) sugerem que o genoma básico da cana-de-açúcar seja composto por 10 cromossomos (x=10), como a maioria das gramíneas; porém, há especulações de que o número básico possa ser x=8 e x=12. Além disso, muitas espécies apresentam variações no número cromossômico não só entre as células de uma mesma planta, como também entre células de um mesmo tecido, fenômeno chamado de mosaicismo (Heinz et al., 1969).

A cana-de-açúcar é uma planta perene e alógama (com fecundação cruzada), cujas cultivares são decorrentes de complexos híbridos interespecíficos envolvendo *S. officinarum,* que contribui com elevado teor de açúcar e *S. spontaneum*, responsável por outras características desejáveis como a rusticidade, de suma importância para o cultivo (Berding e Roach 1987). Para os produtores de cana, é particularmente importante o acesso à diversidade existente nas diferentes espécies, pois grandes avanços no melhoramento da cana foram atingidos por hibridizações interespecíficas (Takahashi et al., 2005).

Além do açúcar e do etanol são diversos os subprodutos que podem ser extraídos a partir da cana, como o álcool anidro e hidratado, o vinhoto, o bagaço que pode ser hidrolisado para alimentação animal, o melaço que também pode ser usado na indústria química, farmacêutica e cosmética, na produção de leveduras, mel, ácido cítrico, etc. (Scandiffio, 2005).

A estimativa da produção nacional para a safra 2008/09 de cana-de-açúcar destinada à indústria sucroalcooleira é de 473,16 milhões de toneladas, das quais 46,92% são para a fabricação de açúcar e 53,08% são para a produção de álcool. Quando comparada à safra 2006/07, verifica-se um crescimento de 11,1% (47,2 milhões de

toneladas). Do total da safra 2008/09, 6,0% é produzido no Norte/Nordeste, devido principalmente às boas condições climáticas, e 11,9% na região Centro-Sul, em função da fertilidade do solo e de ajustes para maior produtividade nas condições ambientais locais (UNICA, 2008).

A área ocupada, atualmente, com essa cultura é de 6,92 milhões de hectares, sendo 5,7 milhões de ha na região Centro-Sul e 1,23 milhões de ha na região Norte/Nordeste. Estima-se que para a safra 2008 houve uma produtividade média de 79.034 kg/ha, valor superior em 2,60% à safra 2006/07 (CONAB, 2008). A agroindústria do açúcar e do álcool tem gerado para o Brasil, em produto final, dez bilhões de dólares por ano, mais de um milhão de empregos diretos e indiretos, bem como o sequestro de 20% das emissões de carbono que o setor de combustíveis fósseis emite no país (Rodrigues, 2004).

Atualmente esta cultura consiste em variedades híbridas obtidas por um criterioso trabalho de seleção, melhoramento genético a partir de cruzamentos entre espécies conhecidas, obtendo características desejáveis conforme regiões e situações específicas. A escolha da variedade é considerada um fator de produção e desenvolvimento tecnológico de maior importância em uma usina sucroalcooleira (Matsuoka, 1999). Grande parte das pesquisas com a cana-de-açúcar tem dedicado esforços à procura e ao desenvolvimento de novas variedades com características agronômicas de produtividade, rusticidade, algumas características industriais, como alto teor de sacarose, teor médio de fibras essenciais e resistência a pragas e doenças.

A habilidade de sobreviver em condições de estresse requer múltiplas adaptações das plantas. Tais adaptações evolutivas envolvem um grupo complexo de parâmetros, que ainda não foram totalmente esclarecidos, o que reforça a importância de estudos direcionados às respostas fisiológicas, metabólicas e moleculares das plantas aos diferentes tipos de estresse. Neste sentido, com o avanço das técnicas bioquímicas e moleculares nas últimas décadas estes estudos estão sendo intensificados, fornecendo maiores informações sobre a fisiologia das cultivares utilizadas e auxiliando programas de melhoramento genético de diversas espécies agrícolas.

Com relação aos Programas de Melhoramento Genético da cana-de-açúcar (PMGCA), com o objetivo de atender à demanda varietal do país, existem atualmente quatro programas: RIDESA – Rede Inter-universitária de Desenvolvimento do Setor Sucroalcooleiro; o Instituto Agronômico de Campinas (IAC); a Cooperativa Central dos Produtores de Açúcar e Álcool do Estado de São Paulo (COPERSUCAR) e o da empresa CanaVialis, que juntamente com a Alellyx foram criadas recentemente pelo esforço de

pesquisadores da área de biologia molecular e recursos de investimento do Grupo Votorantin (*venture capital*), no Estado de São Paulo (Tavares de Melo, 2005).

A CanaVialis é a maior empresa privada de melhoramento de cana-de-açúcar do mundo. Por meio de suas pesquisas de melhoramento, estão sendo desenvolvidas, e consequentemente patenteadas, variedades de cana com características genéticas superiores, que devem proporcionar rendimentos significativos sobre as variedades hoje em dia disponíveis. A Alellyx é uma empresa de genômica aplicada que se dedica ao desenvolvimento de pesquisas com biotecnologia, principalmente para a cana-de-açúcar (JBonline, 2008).

1.2.2. Projeto SUCEST

Devido à grande importância da cana-de-açúcar na economia mundial, a FAPESP, pela rede ONSA (*Organization for Nucleotide Analysis and Sequencing*; Organização para Sequenciamento e Análise de Nucleotídeos), financiou o programa SUCEST (*Sugarcane Expressed Sequence Tag Project*) de sequenciamento de genes expressos da cultura. O programa SUCEST gerou sequências ESTs com alto valor científico agregado, empregando 37 bibliotecas, originadas de diversos tecidos, estágios de desenvolvimento e condições, tais como: meristemas apicais vegetativos e florais, folhas em diversas fases de desenvolvimento, vários tipos radiculares, sementes, caules, células de calos cultivados, além de tecidos infectados com bactérias simbióticas (Tabela 2) (Vettore et al., 2001).

O SUCEST disponibiliza atualmente 291.689 sequências de EST, dispostas em 43.141 *clusters* (Fusaro et al., 2001). Dados gerados no programa, como análises da expressão diferencial de genes em diferentes situações fisiológicas, permitirão desvendar os mecanismos moleculares do desenvolvimento e da fisiologia da cana-de-açúcar, bem como, encontrar funções gênicas significativas responsáveis pelas características de superioridade varietal desejadas pelos melhoristas da cultura (Molinari, 2006).

Tabela 2 - Descrição sucinta das bibliotecas geradas pelo projeto SUCEST, incluindo o código da biblioteca e sigla, bem como quantidade total de ESTs dentro de uma categoria de bibliotecas, descrição dos tecidos e situações de extração de ESTs.

Código da biblioteca	Nº Total de ESTs	Descrição
AD1	18137	Infecção de tecidos de plantas cultivadas *in vitro* por *Glauconacetobacter diazotroficans*
AM1, AM2	28128	Meristema apical
CL3, CL4, CL6	11872	Calos tratados por 12h à 4ºC e 37ºC no escuro e no claro
FL1, FL3, FL4, FL5, FL8	83899	Flor em diferentes estágios de desenvolvimento
HR1	12000	Infecção de tecidos de plantas cultivadas *in vitro* por *Herbaspirilum diazotroficans*
LB1, LB2	18047	Ramo lateral de plantas adultas
LR1, LR2	18141	Primórdio foliar
LV1	6432	Crescimento foliar *in vitro*
NR1, NR2	768	Todas os tecidos Normalizados
RT1, RT2, RT3	31487	Ápice radicular e 0,3 cm a partir do ápice radicular em plantas maduras
RZ1, RZ2, RZ3	24096	Tansição raiz-caule de plantas jovens
SB1	16318	Colmo
SD1, SD2	21406	Desenvolvimento de sementes
ST1, ST3	20762	Primeiro ou quarto internó do caule de plantas jovens

Fonte: Banco de Dados do SUCEST, https://sucest.lad.dcc.unicamp.br/en/

1.3. Eucalipto

O eucalipto (*Eucalyptus* spp.) possui como centro de origem a Austrália e regiões próximas, como o Timor, Indonésia, Papua Nova Guiné e sul das Filipinas, abrangendo uma faixa compreendida entre as latitudes 9º N e 44º S (Eldridge et al., 1994). Trazido para o Brasil, da Austrália, em meados do século XIX, pelo cientista paulista Edmundo Navarro, o eucalipto adaptou-se bem ao clima do país. Contratado pela antiga Companhia Paulista de Estradas de Ferro, após seis anos de estudos com várias espécies vegetais brasileiras e estrangeiras, Navarro plantou 38 milhões de pés de eucalipto, correspondendo a 144 variedades. O cientista concluiu que o eucalipto trazido da Austrália, era a árvore

mais apropriada para ser cultivada em larga escala para a fabricação de dormentes de linhas de trem (Martini, 2004).

Hoje em dia, espécies do gênero *Eucalyptus* são utilizadas em larga escala no estabelecimento de florestas industriais no Brasil. O clima tropical e subtropical, na maioria do território brasileiro permite um crescimento ininterrupto e, consequentemente, um rápido acúmulo de biomassa (SBS, 2008). A idade de rotação de plantios de *Eucalyptus* no Brasil está entre cinco e sete anos, enquanto que, em países de clima temperado, situa-se ao redor de 12 anos (Ho et al., 1998). Além disso, segundo o IBGE (2008) o país tem a maior área plantada de eucalipto do mundo, mais de três milhões de hectares e alcança o maior índice médio de produtividade ($40m^3$ha/ano).

1.3.1. Taxonomia, Importância Econômica e Melhoramento

O gênero *Eucalyptus*, pertencente à classe *Magnoliopsida* (Dicotiledôneas), subclasse *Rosidae*, ordem *Myrtales,* família *Myrtaceae*, compreendendo cerca de 700 espécies, além de um grande número de variedades e alguns híbridos. Dentre as espécies de *Eucalyptus*, podem-se destacar: *E. grandis, E. citriodora, E. viminalis, E. dunnii E. globulus, E. saligna, E. camaldulensis* e *E. urophylla* (Franco, 1999).

A classificação do gênero *Eucalyptus* é relativamente complexa, pois envolve um grande número de espécies (400 a 700), dependendo do sistema de classificação utilizado. De acordo com Pryor e Johnson (1971), o gênero *Eucalyptus* está subdividido em oito subgêneros: *Blakella; Corymbia*; *Eudesmia*; *Gaubaea*; *Idiogenes*; *Monocalyptus*; *Telocalyptus* e *Symphyomyrtus,* cada um com nove, 35, 15, duas, uma, 100, quatro e 300 espécies, respectivamente. O panorama mais recente sobre filogenia de *Eucalyptus*, que consiste basicamente de análises de caracteres morfológicos e moleculares, relata dois grupos: o primeiro compreendendo o gênero *Angophora* e *Corymbia* e o segundo incluindo todos os demais subgêneros (Hill e Johnson, 1995; Steane et al., 1999).

Com relação ao cariótipo do gênero, nos trabalhos realizados por Eldridge et al. (1994) a análise cariotípica de 108 espécies de *Eucalyptus* indicou que 95 espécies apresentaram o número cromossômico 2n=22, presente na maioria dos gêneros da família *Myrtaceae*. As demais 13 espécies apresentaram 2n=24. Estimativas do conteúdo de DNA

nuclear, para algumas espécies estudadas por Grattapaglia e Bradshaw (1994), variaram de 370-700 Mpb. Ademais, alguns híbridos apresentaram uma quantidade de DNA intermediária entre as espécies parentais, porém sem nenhuma evidência de poliploidia.

A diversidade de usos e as características do eucalipto – como rápido crescimento, ótima capacidade de adaptação, resistência a pragas e características físico-mecânicas da madeira – justificam o fato da espécie ser uma das árvores mais plantadas no mundo. Hoje em dia, tudo no eucalipto é aproveitado: como a celulose (para impressão de cadernos, revistas, papel higiênico, viscose, tencel (fibra de celulose feita a partir da polpa de madeira), acetato para filmes, ésteres para tintas, cápsulas para medicamentos; Óleos essenciais (para produção de fármacos, produtos de higiene e limpeza), produtos apícolas (como o mel, própolis e geléia real); Madeira serrada (para a construção civil, postes e mourões, laminados, HDF (*High density fiberboard*; Fibra de alta densidade), MDF (*Medium density fiberboard*, Fibra de média densidade), carvão, lenha e confecção de móveis (Bracelpa, 2008).

Com relação ao Brasil, as principais espécies de eucalipto cultivadas são *E. grandis* e *E. urophylla* (representando mais de 15% do total plantado), seguidas por *E. saligna* (menos de 5%), e híbridos, que constituem a maior parte das plantações, com mais de 50% (Bracelpa, 2008).

As indústrias brasileiras que usam o eucalipto como matéria-prima para a produção de papel, celulose e dos demais derivados da madeira, representam 4% do PIB (Produto Interno Bruto) e 8% das exportações. Segundo a Bracelpa (Associação Brasileira de Celulose e Papel), a produção de papel superou os nove milhões de toneladas em 2007, e neste ano houve um aumento de 3,25%. Ainda em relação à produção de papel, de janeiro até agosto de 2008 o setor já havia ultrapassado os seis milhões de toneladas.

Devido à grande importância que a cultura exerce sobre a agroindústria brasileira, o interesse no melhoramento deste gênero tem aumentado nas últimas décadas. No melhoramento genético de árvores para plantio comercial, o principal interesse é a obtenção de árvores o mais retilíneas possível, com mínima quantidade de galhos laterais e copa, e com caules de maior volume. Além destas, diversas outras características são também contempladas nos processos de seleção, como o teor de lignina; a densidade da madeira e seu poder calorífico; a presença de tiloses; o teor de cinzas; o teor de extrativos; resistência a pragas, entre outros (Bracelpa, 2008). Analisando o genoma expresso do eucalipto, Barbosa-da-Silva et al. (2005) contribuíram para a identificação de 210 genes envolvidos na resposta específica à invasão de patógenos (Genes *R*). Trabalhos como estes

podem ser utilizados em futuros estudos visando ao mapeamento destes genes, minorando perdas, especialmente devido ao ataque de fungos e bactérias em plantios recentes feitos durante a estação chuvosa.

A geração de híbridos inter e intra-específicos tem assumido destacada importância dentro dos programas de melhoramento genético do gênero *Eucalyptus.* A possibilidade de associação de características diferenciadas em espécies importantes, bem como a manifestação de heterose verificada nos cruzamentos entre vários pares de espécies têm levado os melhoristas a buscar na hibridação um meio mais rápido de promover o melhoramento de características florestais desejáveis (Assis, 2008).

1.3.2. Projeto FORESTs

Em 2001 a FAPESP deu início ao projeto FORESTs (*Eucalyptus Genome Sequencing Project Consortium*), que teve como objetivo o sequenciamento do genoma expresso do eucalipto. O projeto foi coordenado por docentes da ESALQ (Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo), e da UNESP (Universidade Estadual Paulista) (Jornal da Unicamp, 2001). O consórcio consistiu no sequenciamento e identificação de genes do eucalipto e análises de genômica funcional, ou seja, associação de genes às características da planta, tendo como objetivo melhorar a matéria-prima utilizada na produção de papel e celulose (Jornal Gazeta Mercantil, 2008).

Inicialmente, o projeto FORESTs teve como finalidade identificar 15.000 genes, através do sequenciamento de aproximadamente 100.000 ESTs. Essas sequências de ESTs foram preparadas a partir de bibliotecas de diferentes tecidos, incluindo plântulas, folhas, raízes, caule e madeira (Tabela 3). O resultado da primeira fase do projeto foi um banco de dados, concluído em Março de 2002, com 112.152 sequências de cDNA de espécies de *Eucalyptus.* A principal finalidade deste projeto é permitir o desenvolvimento de variedades melhoradas geneticamente com base nos dados da análise do transcriptoma desta cultura (Ripasa, 2002).

Tabela 3 - Descrição sucinta das bibliotecas geradas pelo projeto FORESTs, incluindo o código da biblioteca e sigla, bem como quantidade de ESTs geradas por biblioteca, seu total dentro de uma categoria de bibliotecas, descrição dos tecidos e situações de extração de ESTs .

Código da biblioteca	Sigla	ESTs (bib)	ESTs (Total)	Descrição
BK1	BK	1.052	1.052	Casca, alburno, cerne e medula de árvores de *E. grandis* com oito anos.
CL1	CL	9.998	12.533	Calos de *E. grandis* formados no escuro.
CL2		2.535		
FB1	FB	12.275	12.275	Botões, flores e frutos.
LV2	LV	7.352	11.693	Folhas de mudas.
LV3		4.341		
RT3	RT	13.252	20.129	Raízes de mudas de viveiro.
RT6		6.877		
SL1	SL	6.182	27.437	Plântulas de *Eucalyptus* sp. Cultivadas no escuro.
SL4		6.718		
SL5		7.165		
SL6		1.217		
SL7		4.120		
SL8		2.035		
ST2	ST	11.032	26.318	Caule de mudas submetidas a déficit hídrico
ST6		12.558		
ST7		2.728		
WD2	WD	10.224	10.224	Madeira de *E. grandis.*

Fonte: Banco de Dados do FORESTs, https://forests.esalq.usp.br
Obs.: As espécies de *Eucalyptus* utilizadas nas bibliotecas SL (Plântulas de *Eucalyptus* sp. Cultivadas no escuro) foram: SL1 - *E. grandis* (expostas a três horas de luz); SL4 - *E. globulus*; SL5 - *E. saligna*; SL6 - *E. urophylla*; SL7 - *E. grandis* e SL8 - *E. camaldulensis*.

1.4. Estresses Abióticos

O termo estresse deriva da palavra latina *stringere* e, do ponto de vista fisiológico, é definido como um desvio das condições ótimas para a sobrevivência e reprodução, o que leva a mudanças e respostas em todos os níveis funcionais do organismo. Qualquer fator que reduz a reprodução, o crescimento e perturba a estabilidade de um sistema, aumentando os gastos energéticos do organismo, é denominado como fator de estresse (Nogueira et al., 2005).

Estresses considerados negativos, como alta salinidade, congelamento e seca (estresses abióticos) ou ainda ataque de patógenos e injúria (estresses bióticos), podem gerar danos ao desenvolvimento vegetal e, numa situação extrema e irreversível, resultar na morte da planta. A capacidade para sobreviver ao estresse é dirigida por mecanismos que conferem resistência ou tolerância. Esses mecanismos podem ocorrer por meio da combinação de processos comportamentais, morfológicos, anatômicos, fisiológicos e

bioquímicos e dependem, primordialmente, de processos moleculares (Nogueira et al., 2005).

No que diz respeito às expressões tolerância e resistência de plantas a fatores de estresse, o emprego dos dois conceitos parece ser, muitas vezes, equivocado. Normalmente, em termos biológicos, o conceito de resistência está associado com a capacidade da planta em sobreviver e/ou crescer em meio a um determinado estresse, seja ele biótico ou abiótico, enquanto que "tolerância" está relacionada com evolução genética da mesma no sentido de conviver com a presença daquele agente estressor (Silveira et al., 2001).

Geralmente, o estresse hídrico ocorre quando a taxa de transpiração excede a taxa de absorção de água, o que pode ocorrer devido à baixa disponibilidade hídrica do solo, em solos salinos ou a temperaturas muito baixas e altas (Bray, 1997). A capacidade das plantas em tolerar a seca existe em função de diversas características anatômicas, morfológicas e fisiológicas, de caráter constitutivo ou indutivo, que interagem, permitindo a manutenção dos processos de crescimento e desenvolvimento. O acúmulo de solutos orgânicos de baixo peso molecular e íons inorgânicos, os quais estão diretamente relacionados com o ajuste osmótico das células, normalmente ocorre em plantas submetidas a diferentes fatores de estresse, incluindo a deficiência hídrica e salinidade (Martim, 2003).

Algumas espécies e variedades de plantas diferem em termos de susceptibilidade aos diferentes tipos de estresses. Através do desenvolvimento de algumas estratégias, as espécies podem tolerar com mais facilidade altos níveis de desidratação. Porém, condições ambientais adversas induzem, na maioria das vezes, respostas e adaptações similares nos diferentes vegetais. Algumas dessas respostas incluem alterações nas propriedades das membranas celulares; redução no crescimento vegetativo e reprodutivo; inibição da fotossíntese; senescência e a abscisão foliar; síntese de proteínas; acúmulo de osmoprotetores, bem como osmorregulação (Zhu, 2002).

1.4.1. Respostas aos Estresses Hídrico e Salino

As plantas enfrentam o problema de falta de água desde quando conquistaram o ambiente terrestre, há cerca de 450 milhões de anos. Para se proteger dos efeitos nocivos

causados pela escassez hídrica as plantas desenvolveram algumas estratégias para sua proteção, como o desenvolvimento de cutícula serosa e espessa, raízes mais profundas, acúmulo e reserva de água em tecidos carnosos, células especializadas da epiderme que formam poros para a troca gasosa ou ainda fechamento dos estômatos contra a transpiração. Uma vez que todas as estruturas subcelulares devem existir em um meio aquoso, um importante mecanismo de tolerância à desidratação deve consistir na capacidade das células em manter a integridade da membrana e prevenir a desnaturação das proteínas (Zeiger, 1990).

Muitas das áreas agricultáveis do Brasil se encontram com problemas de salinidade devido ao manejo inadequado das águas de irrigação aliado ao uso intensivo de fertilizantes. Nas regiões áridas e semi-áridas uma maior atenção deve ser despendida devido à insuficiência hídrica, chuvas mal distribuídas e à alta demanda evaporativa, o que dificulta a lixiviação dos sais localizados na camada arável do solo. No Brasil, considera-se que existem, mais de 9.000.000 ha com problemas de salinidade, sendo a maior parte dessa área localizada nos perímetros irrigados do Nordeste (Carneiro et al., 2002).

Altas concentrações de sais alteram a estrutura básica do solo, resultando na diminuição da porosidade e consequentemente na redução da aeração e da condutância da água (Mahajan e Tuteja, 2005). O limite de tolerância às condições de estresse salino depende da concentração salina em solução, do tempo de exposição, bem como do estágio de desenvolvimento da planta e pode variar entre genótipos de uma mesma espécie e de acordo com a fase de germinação de cada cultura. Segundo Dias et al. (2003) a germinação e o desenvolvimento inicial são as fases mais sensíveis aos efeitos da salinidade.

O principal dano causado pelo estresse salino nas plantas acontece pela quebra do equilíbrio iônico. A entrada de Na^+ altera o potencial da membrana e facilita a entrada de Cl^- através do gradiente químico. O acúmulo destes íons nas folhas, por exemplo, afetam processos fisiológicos, como a fotossíntese e a transpiração, por provocar alterações no metabolismo do carbono, deficiência nutricional, déficit hídrico e estresse oxidativo (Dooki et al., 2006; Wahid e Ghazanfar, 2006).

A toxicidade do sódio ao metabolismo celular tem efeitos deletérios no funcionamento de algumas enzimas. Além disso, em altas concentrações o elemento Na^+ causa desbalanço osmótico, desorganização da membrana plasmática, redução no crescimento, inibição da divisão e expansão celular e também redução da fotossíntese e da produção de espécies de oxigênio reativo (ROS). A expansão e a divisão celular são

afetadas diretamente pelo estresse salino por reduzir a atividade de proteínas quinases dependentes de ciclina, que participam do ciclo celular (Wang et al., 2003). Outro efeito causado pelo NaCl é a perturbação da integridade da membrana na raiz e a redução na captação de íons nitrato causando restrição no carregamento do mesmo ao xilema, o que provoca diminuição na absorção da água pela planta. Juntos esses efeitos podem diminuir o fluxo de NO_3^- das raízes para as folhas afetando o desenvolvimento vegetal (Abd El-Baki et al., 2000).

Com isso, a capacidade das plantas em se ajustar osmoticamente ao desbalanço iônico provocado pelo excesso de Na^+ e Cl^- nas células é muito importante para a manutenção do desenvolvimento e reprodução. Um dos mecanismos comumente citados para a maior tolerância à salinidade é a capacidade das plantas em acumular íons no vacúolo e/ou solutos orgânicos de baixo peso molecular no citoplasma, em um processo denominado de ajustamento osmótico, que pode permitir a manutenção da absorção de água e da turgescência celular (Hopkins, 1999). Outro componente de fundamental importância na conservação da turgescência, em resposta ao decréscimo do potencial hídrico, é a diminuição da condutância estomática, que é responsável pelo controle da abertura dos estômatos; o fechamento dos estômatos para proteção contra a perda de água, simultaneamente, restringe a difusão do CO_2 atmosférico, provocando queda da taxa fotossintética (Maruyama et al., 2005).

Essas mudanças provocam alterações na membrana celular e em vários de seus componentes, assim como na concentração celular de metabólitos. Essas alterações na conformação da membrana provocam mudanças em canais de transporte ativados por pressão, modificando a conformação de proteínas sensoriais críticas. Desta forma, há modificações do espaço entre a parede e a membrana celular, levando a alterações que ativam complexos enzimáticos, bem como ativando uma cascata de eventos moleculares que levam à expressão de várias famílias gênicas, como as que codificam enzimas de detoxificação, tais como a ascorbato peroxidase, superóxido dismutase e glutamina S-transferase (Ingram e Bartels, 1996; Shinozaki e Yamaguchi-Shinozaki, 1997).

Deste modo, em suma, os estresses abióticos iniciam nas plantas um complexo de respostas, que podem ser resumidas em três principais eventos: percepção dos sinais, mudanças em nível molecular e respostas morfofisiológicas (Meneses et al., 2006).

De maneira geral, as vias envolvidas nas respostas à seca e à salinidade são interconectadas, uma vez que ambos os estresses induzem danos similares nas células, como mudanças no potencial osmótico. Uma das mais bem documentadas respostas

fisiológicas/moleculares a esses estresses é a habilidade que algumas espécies possuem em ajustar osmoticamente suas células. Dentre as vias de sinalização, destaca-se a via SOS (*Salt Overly-Sensitive*; Altamente Sensível ao Sal) como mediadora da homeostase iônica e tolerância ao Na^+. Genes dessa via, junto com os genes *AtNHX* (Trocador Na^+/H^+ de *Arabidopsis thaliana*), codificam proteínas de antiporte responsáveis pela exclusão e compartimentalização em vacúolos do íon Na^+. Ademais, pode-se citar a cascata de proteínas quinases (MAPK), mediadora da homeostase osmótica, e a cascata de proteínas do tipo LEA (*Late Embryogenesis Abundant*; Abundante na Embriogênese Tardia), responsáveis pelas vias que promovem a desintoxicação ou minimização dos danos oxidativos (Zhu, 2001). Várias funções têm sido propostas para os grupos de proteínas LEA, incluindo a conservação de água, a estabilização de proteínas, a proteção de membranas, o sequestro de íons e a degradação de espécies reativas de oxigênio (Bray, 1993; Close, 1997; Zhu et al., 1997).

A desidratação e a salinidade requerem mudanças no fluxo da água para permitir que as células e os tecidos adaptem-se à situação estressante. As aquaporinas são componentes centrais nas relações hídricas das plantas. Essas moléculas participam do controle dinâmico da permeabilidade da água nas células e órgãos das plantas; tal controle ocorre durante o desenvolvimento e, também, em resposta aos estímulos externos como condições de estresse (Clarkson et al., 2000; Javot e Maurel, 2002).

De acordo com Meneses et al. (2006) a resistência a estresses abióticos é uma complexa via que ainda precisa ser elucidada completamente, uma vez que a expressão de diversos genes dependem da ação e interação fisiológica (transpiração reduzida, fechamento estomático, ajuste osmótico), morfológica (área reduzida da folha, murcha da folha) e bioquímica (aumento da atividade da nitrato redutase, aumento no armazenamento de hidratos de carbono, acúmulo de solutos como prolina, trealose, entre outros).

1.5. Osmoprotetores

Quase todos os organismos, desde bactérias até plantas, sintetizam osmoprotetores em resposta ao estresse osmótico. Essas substâncias incluem moléculas pequenas, eletricamente neutras e atóxicas, que são acumuladas em concentrações molares às células

(Yancey, 2001). Nos vegetais que acumulam naturalmente osmoprotetores, os níveis destes compostos são tipicamente 5-50 μmol g^{-1} do peso fresco, tornando-se mais elevados durante a exposição a algum agente estressor (Bohnert et al., 1995). É importante ressaltar que esses solutos osmoprotetores são altamente confinados ao citoplasma (incluindo organelas) sendo praticamente ausentes nos vacúolos, que geralmente ocupam cerca de 90% do volume celular (Matoh et al., 1987; Yancey, 2005).

Muitos autores discutem o acúmulo de osmoprotetores em vegetais. Acredita-se que o aumento da osmolaridade celular como resultado do acúmulo desses solutos não-tóxicos ("compatíveis") osmoticamente ativos, acompanhada do influxo, ou pelo menos na redução da saída de água nas células, proporciona uma turgescência necessária para expansão celular (Ingram e Bartels, 1996).

Algumas evidências mostram que em concentrações acima do normal (fato que ocorre em situações de estresse osmótico), tais substâncias estabilizam macromoléculas celulares, diminuindo as perdas, tanto no que tange à integridade das membranas celulares, quanto à atividade de algumas enzimas. Além de substituírem o excesso de sais inorgânicos, os osmoprotetores permitem que as células recobrem seu volume e suas concentrações internas de sal (Génard et al., 1991). Inicialmente, atribuiu-se aos solutos compatíveis o papel principal no ajuste osmótico, porém alguns estudos também os relacionam com outras possíveis funções (Serraj e Sinclair, 2002). Uma hipótese adicional é que, esses solutos compatíveis estão igualmente envolvidos na degradação de espécies reativas de oxigênio (Chen e Murata, 2002).

Quimicamente, os osmoprotetores estão agrupados em quatro famílias: (I) metilaminas, (II) polióis, (III) açúcares e (IV) aminoácidos ou derivados. Nem todos estes grupos ocorrem naturalmente em todas as plantas; a síntese de glicina-betaína, prolina e trealose, por exemplo, são alguns dos principais solutos orgânicos acumulados em plantas superiores sob estresse hídrico ou salino, facilitando a adaptação às condições adversas (Rhodes e Hanson, 1993; Brigotti et al., 2003).

1.5.1. Glicina-Betaína

A glicina-betaína (GB) é um composto quaternário do amônio, sendo encontrado em um grande número de organismos, incluindo bactérias, vegetais, mamíferos, etc. É um

dos solutos compatíveis mais extensivamente estudados, relacionados à tolerância ao estresse hídrico, salino, baixas temperaturas e estresse oxidativo (Jagendorf e Takabe, 2001; Shirasawa et al., 2006). O nível de betaína aumenta durante a exposição a condições adversas, pois ambas as enzimas envolvidas na sua biossíntese são estresse-induzidas (McCue e Hanson, 1992; Rathinasabapathi et al., 1997).

A via de oxidação da colina em plantas compreende também a via de síntese da betaína aldeído, porém envolve diferentes enzimas. Na maioria das plantas superiores, animais e também microorganismos que sintetizam GB a oxidação da colina em betaína aldeído é catalisada pela enzima colina monooxigenase, codificada pelo gene *CMO*. Já a oxidação da betaína aldeído em GB é catalisada pela enzima betaína aldeído desidrogenase codificada pelo gene *BADH* (Figura 1a). No entanto, nos microorganismos *Arthrobacter globiformis* e *A. pascens* a síntese de GB requer somente uma enzima: a colina oxidase (Figura 1b) (Rathinasabapathi et al., 1997).

Waditee et al. (2005) demonstraram que a cianobactéria *Aphanothece halophytica* sintetiza GB a partir da glicina em três passos, nos quais duas enzimas N-metiltranferases estão envolvidas. A enzima sarcosina glicina metiltransferase (ApGSMT) catalisa a reação de metilação da glicina em sarcosina e da sarcosina para dimetilglicina, respectivamente, enquanto a dimetilglicina metiltransferase (ApDMT) catalisa a metilação da dimetilglicina em betaína.

A GB parece ser um determinante crítico na tolerância ao estresse. Estudos *in vitro* demonstraram que ela é eficaz como um soluto compatível na estabilização de estruturas quaternárias de enzimas e proteínas complexas, pois possui a propriedade de interagir com macromoléculas tanto hidrofílicas quanto hidrofóbicas, enzimas e complexos protéicos. Esta molécula contribui para manter o potencial osmótico no citoplasma e uma quantidade de água apropriada, bem como manter a integridade das membranas, quando da exposição celular a concentrações elevadas de sais ou temperaturas extremas (Papageorgiou e Murata, 1995).

A glicina betaína não possui, exclusivamente, a função de osmoprotetor nas células, uma vez que é usada como precursora nas vias metabólicas para formar compostos ligados à síntese do hormônio etileno e da molécula piruvato, tratando-se de uma rica fonte de carbono, nitrogênio e energia (Munôz-Clares e Velasco-Garcia, 2004).

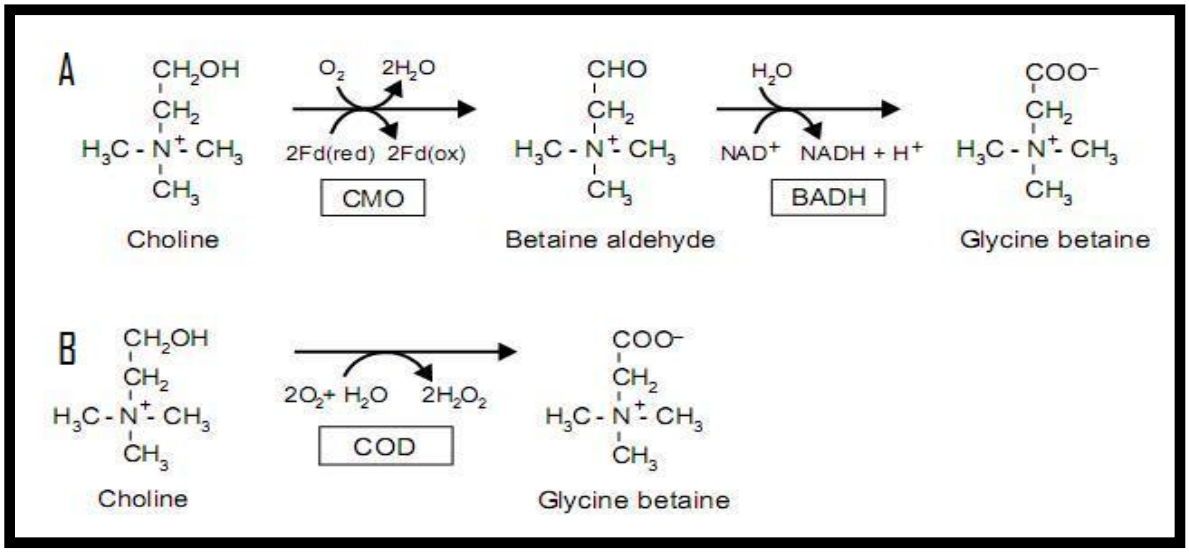


Figura1: Via de síntese de Glicina Betaína. **(A)**: Via utilizada pelas algas, plantas, animais e maioria dos microorganismos. **(B)**: Via utilizada pelas *Arthrobacter globiformis* e *A. pascens*. Fonte: Chen e Murata, 2002.

A GB é amplamente distribuída nas plantas superiores de um modo geral, porém algumas espécies como *Arabidopsis thaliana*, arroz (*Oryza sativa*) e tabaco (*Nicotiana tabacum*) são consideradas como não acumuladores desse osmólito (McNeil et al., 1999).

1.5.2. Prolina

A Prolina (L-prolina) é um dos aminoácidos presentes nas proteínas de todos os organismos vivos, sendo provavelmente um dos osmólitos compatíveis mais extensamente distribuídos entre os organismos. Sua síntese em decorrência do estresse osmótico ocorre em praticamente todo o reino vegetal (McCue e Hanson, 1990; Delauney et al., 1993). Uma característica marcante de distúrbios no metabolismo das proteínas decorre de mudanças nas proporções dos aminoácidos, que frequentemente levam a um aumento na concentração de prolina (Larcher, 2000).

Em plantas superiores, a síntese de prolina ocorre por duas vias: a do ácido glutâmico (Glu) e a da ornitina (Orn). Na primeira via, considera como a mais importante,

especialmente em relação ao estresse osmótico (Delauney e Verma, 1993), a prolina é sintetizada a partir do Glutamato através de dois intermediários, o ácido glutâmico γ-semialdeído (GSA) e Δ^1-pirrolina-5-carboxilato (P5C). Duas enzimas catalisam esta via, a Δ^1-pirrolina-5-carboxilato-sintase (P5CS) na primeira etapa e Δ^1-pirrolina-5-carboxilato-redutase (P5CR) na etapa final. Genes que codificam as enzimas P5CS e P5CR foram isolados de diversas plantas, sendo apontados como genes chave na síntese da prolina (Delauney e Verma, 1990; Verbruggen et al., 1993).

Hu et al. (1992) em análises de *northern blot* com plantas de *Vigna aconitifolia* submetidas a um tratamento com 200 mM de NaCl, mostraram que o gene *P5CS* foi superexpressado, particularmente, em tecidos da folha. Também Yoshiba et al. (1995), trabalhando com a planta modelo *A. thaliana*, demonstraram que conjuntamente estresses salino e hídrico induziram fortemente a expressão do gene *P5CS*.

Em plantas a prolina constitui menos de 5% dos aminoácidos totais livres em condições normais. Porém, sob várias formas de estresse, a concentração de prolina pode chegar a 80% do conjunto total de aminoácidos (Alia e Matysik, 2001).

Além do papel no ajuste osmótico, outras funções foram propostas para esse aminoácido, como na proteção e manutenção da integridade da membrana plasmática (Mansour, 1998), dissipação ou redução de energia (Verbruggen et al., 1996), eliminação de radicais hidroxil (Hong et al., 2000) e como fonte de carbono e nitrogênio (Ahmad e Hellebust, 1988; Peng et al., 1996) .

1.5.3. Cisteína

O enxofre desempenha um papel importante na regulação do crescimento e desenvolvimento nos vegetais. O sulfato inorgânico é reduzido, assimilando à cisteína a partir da qual uma grande variedade de compostos são sintetizados. A cisteína pode ser utilizada para a síntese de proteínas ou glutationa, funcionando como um doador de S (enxofre) para metionina e para a biossíntese de metabólitos secundários (Barroso et al., 1997; Hell et al., 2001).

Em plantas e microrganismos, a síntese da cisteína, a partir da serina, acontece em duas etapas: (I) Conversão da serina em O-acetil serina pela enzima SAT (serina acetil

transferase); (II) Troca do acetato do O-acetil serina por um sulfeto, advindo da redução de oito elétrons do sulfato, pela ação da enzima OASTL (O-acetil serina (tiol) liase) (Campanini et al., 2005). De acordo com Bonner et al. (2005), nesta etapa o OASTL usa o piridoxal-5-fosfato (PALP) como co-fator para catalisar a formação da cisteína, a partir da O-acetil serina e sulfeto. Essas duas enzimas (OASTL e SAT) estão localizadas nos três compartimentos celulares envolvidos na síntese de proteínas: citosol, cloroplastos e mitocôndrias, assumindo isoformas diferentes (Barroso et al., 1997).

Em experimentos conduzidos por Romero et al. (2001) com folhas de arabidopsis expostas a NaCl, foi visto que a isoforma citosólica de *OASTL* foi induzida e que o nível dessa enzima aumentou, continuando progressivamente por mais de cinco dias após exposição ao sal. O mesmo foi observado por Barroso et al. (1999), onde a exposição de plantas maduras ao NaCl induziu uma rápida acumulação de mRNA em toda a lâmina foliar, nas raízes e nos caules. Ruiz e Blumwald (2002) em experimentos com canola (*Brassica napus*) viram que os níveis das enzimas SAT e OASTL aumentaram significativamente em consequência da exposição a 150 mM de NaCl durante 15 dias. Os trabalhos sugeriram que a taxa elevada da biossíntese da cisteína, foi necessária e imprescindível para a proteção e/ou adaptação das plantas a níveis mais elevados de Na^+, caracterizando esta proteína como um osmoprotetor nesses vegetais.

1.5.4. Mio-Inositol

A isomerização, metilação e epimerização do mio-inositol e de outros estereoisômeros (*scyllo, chiro, muco,* e *neo*) leva à formação de compostos O-metil-inositólicos (*sequoyitol, bornesitol, quebrachitol, pinitol, ononitol*, etc.), e sua síntese aumenta em resposta a estresses abióticos, contribuindo para a regulação osmótica (Donald et al., 1998). Além disso, segundo Nelson et al. (1998), essas substâncias facilitam o sequestro de sódio e protegem o aparato fotossintético das plantas, quando submetidas a condições estressantes.

O mio-inositol é um álcool cíclico bastante importante para o crescimento e desenvolvimento das plantas. Ele participa de vários processos metabólicos, tais como síntese de ácido fítico, estoque e transporte de auxina nas células e produção de moléculas

ligadas à resposta ao estresse (Loewus e Loewus, 1983); desempenha também papel importante na germinação de sementes, transporte de açúcar, nutrição mineral, metabolismo de carboidratos, estrutura de membranas, formação de parede celular e homeostase de hormônios (Styer, 2000).

A partir da glicose-6-fosfato, é formado o mio-inositol-1-fosfato, através de uma reação interna de oxi-redução que envolve NAD^+ catalisada pela enzima INPS1 (mio-inositol-1-fosfato sintase) que é codificada pelo gene *INPS1*. Posteriormente, ocorre a desfosforilação em mio-inositol e ortofosfato (Majumder et al., 1997).

O mio-inositol é distribuído em abundância em todo o sistema biológico, podendo permanecer como açúcar livre, éster de fosfato e metil-derivados, podendo ainda ser precursor de importantes metabólitos (Loewus e Murthy, 2000).

1.5.5. Trealose

A trealose é um dissacarídeo constituído de duas moléculas de glicose. Este açúcar está presente em vários organismos, incluindo, bactérias, leveduras, fungos, insetos, invertebrados e plantas superiores e inferiores (Elbein, 1974). Em muitos desses organismos, pode atuar como sinalizador molecular para dirigir ou controlar determinadas vias metabólicas, podendo afetar o crescimento desses organismos. Em fungos, essa molécula, ocorre em esporos, corpos frutíferos e células vegetativas. Clegg e Filosa (1961) reportaram a presença de mais de 7% de trealose do peso seco em esporos de *Dictyostelium mucoroides*; porém quando esses esporos germinaram o conteúdo de trealose desapareceu rapidamente, sugerindo que esse açúcar serve como uma fonte de carbono e energia. Em micobactérias, foi demonstrado que a trealose é integrante de diversos glicolipídeos que são importantes componentes estruturais da parede celular (Elbein et al., 2003).

No reino animal, a trealose encontra-se amplamente distribuída, tendo sido primeiramente reportada em insetos, estando presente nos estágios de larva e pupa, como constituinte da hemolinfa. Durante o vôo, os níveis desse açúcar caem drasticamente, indicando sua função como fonte de carbono e energia para esses animais. Em alguns organismos que são capazes de sobreviver em condições de escassez de água (anidrobiose)

– como algumas plantas, fungos, esporos, nematóides, etc. – foi demonstrado que o nível de trealose pode subir drasticamente, indicando que esse açúcar é necessário para sobrevivência durante a ausência de água (Elbein et al., 2003). Em estudos com o nematóide *Aphelenchus avenae*, Madin e Crowe (1975) observaram que após desidratação lenta do animal, 20% do seu peso seco é convertido em trealose, indicando a estreita correlação do conteúdo desse açúcar com a sobrevivência desse organismo nestas condições.

Em relação ao seu papel osmoprotetor em vegetais, foi demonstrado que a trealose pode proteger proteínas e membranas celulares da inativação ou desnaturação causada por uma variedade de estresses, incluindo a seca, salinidade, frio e oxidação. Durante o estresse hídrico a trealose parece ser um dos osmoprotetores mais eficientes na manutenção dos lipídios em uma fase fluida, evitando, assim, separação de fases e a fusão das membranas. Este açúcar liga-se às membranas celulares, diminuindo sua temperatura de fusão, mantendo-as, assim, na fase líquido-cristalina. Além disso, age como estabilizador de enzimas e vesículas durante a desidratação, o que permite que a célula mantenha suas funções essenciais por períodos maiores (Crowe et al., 1987).

A biossíntese da trealose se dá através da formação da trealose-6-fosfato a partir da UDP-glicose e glicose-6-fosfato, onde essa reação é catalisada pela enzima trealose-6-fosfato sintase (TPS). A glicose-6-fosfato é então desfosforilada em trealose e ortofosfato pela enzima trealose-6-fosfato fosfatase (TPP) (Cabib e Leloir, 1958). Já a hidrólise da trealose pode ocorrer sob várias formas. Por exemplo, em *Euglena gracilis* e *Pichia fermentans*, a trealose é catalisada pela enzima trealose fosforilase (Schick et al., 1995), em *Escherichia coli* pela fosforilação e subsequente hidrólise pela enzima trealose-6-fosfato hidrolase (Rimmele e Boos, 1994), enquanto em plantas, fungos, bactérias e animais o processo ocorre através da enzima trealase (Horlacher et al., 1996).

1.6. Potencial Biotecnológico de Osmoprotetores

Atualmente, grande parte das pesquisas com plantas têm dedicado esforços à procura e ao desenvolvimento de novas variedades resistentes a estresses de origem biótica ou abiótica. O potencial de utilização da engenharia genética para o melhoramento vegetal

é extremamente amplo, até o presente, diversos genes foram introduzidos estavelmente em plantas, conferindo resistência a muitos tipos de estresses (Mansur e Margis-Pinheiro, 1996).

Para a compreensão da função de um gene é essencial conhecer e identificar sua expressão local e espacialmente. Ultimamente, várias estratégias foram desenvolvidas para permitir o estudo da função de vários genes simultaneamente, como a utilização da genética reversa (geração de mutações específicas em genes de interesse), *screening* após aplicação de agentes mutagênicos (geração de mutações randômicas e seleção de mutantes para identificação de fenótipos de interesse) e bioinformática (análise em larga escala de dados biológicos). As estratégias acima, associadas ao sequenciamento do genoma ou transcriptoma do organismo, têm possibilitado o estudo sistemático da expressão diferencial de genes de interesse (Leal-Bertioli et al., 2003).

Análises da expressão diferencial, a partir de bibliotecas construídas sob diferentes tecidos e condições (ex. salinidade X controle) ou utilizando mutantes (mutante X selvagem) apresentam-se como importantes ferramentas no isolamento de genes de resistência, possibilitando a identificação de genes–chave, revelando prováveis candidatos a fatores envolvidos nas respostas aos estresses abióticos e bióticos. O conjunto de dados da expressão diferencial de vários genes sob diferentes condições pode ser usado para reconstruir as vias reguladas por estes genes, permitindo predições sobre tecidos e vias onde os mesmos atuam, auxiliando na identificação de novos genes associados ao processo (Leal-Bertioli et al., 2003).

Uma das principais características dos osmoprotetores baseia-se no fato de que seus efeitos benéficos geralmente não são espécie-específicos, o que permite que tais genes sejam expressos em plantas transgênicas, a fim de conferir maior proteção contra os danos causados pela exposição a altas concentrações de sais no solo e/ou à escassez de água (Rontein et al., 2002). Consideráveis progressos têm sido alcançados nas áreas da engenharia e biossíntese de solutos compatíveis em uma variedade de espécies, incluindo algumas culturas agrícolas de importância econômica, resultando na produção de plantas transgênicas com maior tolerância aos estresses abióticos (McNneil et al., 1999). Esses benefícios foram observados em estudos *in vitro* e *in vivo* que permitiram averiguar o crescimento normal e a sobrevivência das células/plantas cultivadas sob tais condições (Rontein et al., 2002).

Estudos relacionados à via biossintética da prolina em plantas em geral objetivaram avaliar sua atividade durante o estresse osmótico. A enzima P5CR se encontra no ponto de

convergência de ambas as vias de síntese da prolina (via da Orn e Glu). Embora esta não seja limitante para a taxa de acumulação da prolina (Szoke et al., 1992), a P5CR é diferencialmente regulada durante o desenvolvimento, sendo induzida durante o estresse salino (Verbruggen et al., 1993). Já a atividade da enzima P5CS é apontada como o passo limitante para a acumulação de prolina em resposta ao estresse osmótico.

Evidências do envolvimento direto da prolina durante o estresse osmótico foram observadas em abordagens com transgênicos. Os genes *P5CS* e *P5CR* e seus correspondentes cDNAs foram clonados e caracterizados em *Vigna aconitifolia* e arabidopsis (Delauney e Verma, 1990; Hu et al., 1992; Verbruggen et al., 1993), conferindo maior tolerâncias ao estresse osmótico. Kishor et al. (1995), obtiveram plantas transgênicas de tabaco super-expressando o gene *P5CS*, as quais apresentaram maior resistência tanto ao déficit hídrico quanto ao estresse salino. Do mesmo modo, o gene *P5CS* foi introduzido no trigo utilizando a transferência mediada por *Agrobacterium*; os testes com as plantas transgênicas indicaram que o excesso de prolina resultava no aumento da tolerância ao estresse salino (Kishor et al., 2005). Assim, todas estas abordagens resultaram em níveis elevados de prolina, aumentando a tolerância ao estresse osmótico em plantas estressadas.

Genes que codificam as enzimas envolvidas na biossíntese GB têm sido frequentemente clonados. Alguns deles incluem os genes *CMO* e *BADH* de plantas superiores (Nuccio et al., 1998; Selvaraj, 2000); *CDH* (colina desidrogenase) e *BADH* de *Escherichia coli* (Landfald e Strom, 1986), *COX* (colina oxigenase) de *Arthrobacter pascens* (Rozwadowski et al., 1991) e Ap*GSMT* e Ap*DMT* de *A. halophytica* (Waditee et al., 2005). Plantas transgênicas de diversas espécies foram produzidas expressando todos esses genes, individualmente, com exceção dos dois últimos que foram co-expressos em plantas de arabidopsis. Estas plantas acumularam GB em vários níveis, exibindo maior tolerância a vários tipos de estresse, principalmente à salinidade (Sakamoto e Murata, 2000; 2001; 2002).

Quan et al. (2004) desenvolveram uma linhagem de milho DH4866 transformado com o gene *betA* de *E. coli* codificando a enzima colina desidrogenase (EC. 1.1.99.1). O milho transgênico acumulou níveis mais elevados da GB e as plantas apresentaram maior tolerância ao estresse hídrico quando comparadas com o tipo selvagem (plantas não-transgênicas), tanto durante a germinação das sementes quanto nas mudas jovens. O rendimento dos grãos das plantas transformadas também foi significativamente superior ao das plantas selvagens após exposição à condição de escassez de água. O aumento do

acúmulo da GB no milho transgênico proporcionou uma maior proteção da integridade da membrana celular e maior atividade das enzimas comparadas com plantas não transformadas em condições de estresse hídrico. Além dessas abordagens envolvendo a transformação genética de plantas, a aplicação exógena de GB também parece proporcionar melhores condições de crescimento e a sobrevivência de uma grande variedade de plantas submetidas a diferentes tipos de estresses (Hayashi et al., 1998; Alia et al., 1998; Chen et al., 2000).

Em estudos conduzidos por Romero et al. (1997) o gene *TPS* de levedura foi introduzido em plantas de *Nicotiana tabacum* submetidas a escassez hídrica por um período de 15 dias. O gene foi integrado ao DNA genômico e constitutivamente expresso no vegetal. Os pesquisadores quantificaram o açúcar acumulado e o resultado foi de 0,17 mg/g do peso fresco das folhas das plantas transgênicas, acúmulo que produziu várias alterações fenotípicas, incluindo o crescimento atrofiado, folhas em forma de lança e teor de sacarose reduzido. No que se refere ao estresse hídrico, plantas-controle de tabaco foram drasticamente afetadas, enquanto que os sintomas de murcha foram minimizados nas plantas transgênicas que expressaram o gene *TPS*. Os autores concluíram que a trealose acumulada nas plantas transgênicas conferiu uma maior tolerância ao estresse.

Garg et al. (2002) introduziram os genes *otsA* e *otsB* (genes para a biossíntese da trealose) de *E. coli* em arroz visando conferir maior tolerância ao estresse hídrico. Em comparação com o controle, as plantas transgênicas tiveram um melhor crescimento, diminuição do dano foto-oxidativo, bem como aumento da fotossíntese e um balanço nutricional favorável sob condições de estresse. Dependendo das condições de crescimento, as plantas transformadas acumularam três a 10 vezes mais trealose que as plantas-controle. Esse trabalho confirmou a tolerância a diversos tipos de estresse (salinidade, seca e baixas temperaturas) e também o aumento da produtividade em plantas de arroz super-expressando os genes da via de síntese da trealose sem os efeitos pleiotrópicos negativos observados em trabalhos anteriores conduzidos com outros vegetais.

De acordo com Sirko et al. (2004) muitas plantas transgênicas, submetidas a condições de estresse, super-expressando o gene *OASTL* são mais tolerantes a condições adversas que as plantas não transformadas, sugerindo que um aumento na síntese de cisteína é uma importante adaptação na resposta a diferentes estresses. Em vários estudos onde o gene *OASTL* foi introduzido em plantas observou-se um aumento também na tolerância a metais pesados. Por exemplo, Dominguez-Solis et al. (2001) observaram um

aumento do nível de cisteína em resposta ao estresse causado pelo enxofre, também existindo constatações de que níveis aumentados conferem proteção contra estresse oxidativo (Noji et al., 2001; Youssefian et al., 2001).

A partir das muitas abordagens disponíveis, observa-se que a identificação de genes e enzimas envolvidas na síntese de osmoprotetores atuantes nas vias de tolerância aos estresses abióticos compreende importante fonte de candidatos para uso em engenharia genética de plantas cultivadas em áreas acometidas por tais estresses (Rathinasabapathi, 2000).

1.7. Bioinformática

1.7.1. Retrospectiva e Atualidades

Um dos aspectos decisivos para o grande avanço da genética é o uso intensivo de técnicas computacionais. A Bioinformática surgiu na segunda metade da década de 90, como uma nova ciência envolvendo a união de diversas linhas de conhecimento, entre elas, a engenharia de softwares, matemática, estatística, ciência da computação e a biologia molecular. Com o advento dos sequenciadores automáticos, houve um crescimento exponencial na quantidade de sequências gênicas e protéicas tornando imprescindível a aplicação da informática para análise e administração dessa grande quantidade de dados (Prosdocimi et al., 2002). No entanto, as demandas evoluem diariamente, demonstrando que o trabalho dos bioinformatas é um desafio constante. Os benefícios que essa nova abordagem oferece para a sociedade, através de uma compreensão mais profunda dos genes e proteínas, doenças, drogas e tratamentos, irão incentivar cada vez mais um número maior de pesquisadores em ambos os domínios, da biologia e da informática (Lesk, 2002).

O Brasil vem crescendo quanto a pesquisas e estudos genéticos e também em relação à bioinformática. Boa parte deste reconhecimento se deve ao Laboratório de Bioinformática (LBI) do Instituto de Computação da Unicamp, responsável pela bioinformática dos primeiros projetos de sequenciamento genético do país. Outros projetos puderam ser acompanhados, como o Programa Genoma FAPESP, reunidos, desde maio de

1997 na Rede ONSA. Com esta rede, foram estabelecidas as bases para o sequenciamento de vários genomas e transcriptomas, entre eles o Genoma da bactéria *Xylella fastidiosa*, o Genoma do Câncer Humano, o Genoma da *Xanthomonas*, o Genoma Bovino e transcriptomas da cana-de-açúcar e do Eucalipto. Esses programas mostram as etapas dessa progressiva conquista científica e tecnológica da Genômica no Brasil, em que a bioinformática esteve presente como instrumento indispensável, de um lado e como conhecimento, em dinâmica de evolução do outro (Vogt, 2003).

1.7.2. Bancos de Dados e Ferramentas Computacionais

Em consequência da grande quantidade de informações de sequências de nucleotídeos e de aminoácidos que são produzidas atualmente, principalmente em projetos de genômica, transcriptômica e proteômica, o uso dos bancos de dados se tornou indispensável. Grupos de pesquisa do mundo inteiro precisavam de acesso rápido e fácil a esses dados, demandando a construção de bancos de dados públicos e privados que permitissem a interação entre grupos de mineração de dados, bem como o acesso e depósito contínuo de sequências (Morais, 2003).

Os bancos de dados envolvendo sequências de nucleotídeos, de aminoácidos ou estruturas de proteínas podem ser classificados em bancos primários e secundários. Os primários são formados por sequências de nucleotídeos, aminoácidos ou estruturas protéicas, depositadas diretamente, enquanto os bancos de dados secundários derivam dos primários, ou seja, foram construídos usando as informações depositadas nos bancos primários. Um exemplo é o SWISS-PROT (Banco de Dados de Sequências de Proteínas Curado), um banco de dados onde as informações sobre sequências de proteínas foram anotadas e associadas a informações sobre função, domínios funcionais, proteínas homólogas, etc. (Prosdocimi et al., 2002).

Um dos principais bancos de dados é o *GenBank* (Banco de Genes - banco de dados norte-americano de sequências de DNA e proteína), que está hospedado no Centro Nacional para Informação em Biotecnologia (NCBI; *National Center for Biotechnology Information* - http://www.ncbi.nlm.nih.gov/), contendo mais de 60 bilhões de pb, incluindo

sequências de mais de 170.000 organismos (NCBI, 2008). O NCBI oferece ainda informações sobre a taxonomia de todos os organismos depositados, genomas completos, mapas gênicos, informações sobre estruturas protéicas e o PubMed, uma ferramenta de busca bibliográfica, com milhares de artigos disponíveis para *download* (Benson et al., 2000).

Paralelamente à criação dos bancos de dados, houve a necessidade da criação de ferramentas computacionais que fossem capazes de analisar as sequências geradas pelos projetos de sequenciamento. Essas ferramentas aliadas aos bancos de dados permitem uma análise de predição *in silico* de sequências desconhecidos e novos genes, por meio de análises comparativas (Kent et al., 2002).

O BLAST (*Basic Local Alignment Search Tool*; Ferramenta de Busca por Alinhamento Local), foi desenvolvido para comparar sequências de genes e proteínas tanto de diferentes organismos como de um mesmo organismo, sendo constantemente utilizado na busca por sequências homólogas e/ou análogas (Altschul et al., 1990). Esta ferramenta possui várias abordagens e uma delas é o *Blast2Sequences*, que permite avaliar a identidade entre duas sequências com o objetivo de compará-las minuciosamente. Outra possibilidade é o BLASTx que traduz a sequência de DNA para proteína e a compara com banco de dados protéico.

Ferramentas computacionais são necessárias para a predição de genes, de estruturas, descoberta e reconhecimento de padrões, entre outros. Para permitir uma comparação de sequências genômicas existem *softwares* que podem comparar várias sequências ao mesmo tempo. O CLUSTAL-X (Thompson et al., 1997) é um programa baseado em um algoritmo para alinhamento múltiplo global de sequências biológicas informativas, o qual pode ser usado tanto para nucleotídeos quanto aminoácidos.

A análise comparativa de sequências tornou-se essencial não só pela possibilidade de avaliar as similaridades entre sequências de vários organismos, como por proporcionar a reconstrução da história evolutiva das espécies e das famílias multigênicas, estimando as taxas da evolução molecular e inferindo sobre a natureza e a extensão das forças seletivas que atuaram no curso da evolução dos genes e dos genomas (Kumar et al., 2004).

O MEGA (*Molecular Evolutionary Genetic Analysis,* Análises Genéticas Moleculares Evolutivas) é um dos programas que permitem inferir sobre caracteres evolutivamente informativos, podendo ser usado para análises com sequências nucleotídicas, protéicas e marcadores moleculares (Sudhir et al., 1993). Este programa permite, por exemplo, a análise da matriz de dados através de métodos distâncias

genéticas, como distância P (Nei, 1991), distância de Junkes-Cantor (Junkes-Cantor, 1969), entre outras. Além disso, o programa disponibiliza os algoritmos UPGMA (*Unweighted Pair Group Method with Arithmetic Means*; Método não Polarizado de Agrupamentos aos Pares com Médias Aritméticas; Sneath e Sokal, 1973), NJ (*Neighbor-Joining*; Agrupamento por Vizinhança; Saitou e Nei, 1987) e Máxima Parcimônia (Eck e Dayhoff, 1966; Fitch, 1971) para reconstruções fenéticas e filogenéticas com a geração de dendrogramas.

Finalmente, a análise do padrão de expressão de genes induzidos durante situação de estresse (ou não) transformou-se em uma ferramenta importante para investigar como e onde o organismo responde às condições adversas do ambiente. O CLUSTER (Eisen et al. 1998) e TREEVIEW (Page, 1996) fazem parte de um grupo integrado de programas para analisar e visualizar os resultados de experimentos de *microarray*. O programa CLUSTER inclui as abordagens: *Hierarchical Clustering* (*Clusterização* Hierárquica), *Self Organizing Maps* (Auto Organização de Mapas), *K-Means Clustering* (*Clusterização* de Médias K) e *Principal Component Analysis* (Análises de Componentes Principais). Através da normalização dos transcritos este programa permite a visualização dos padrões de expressão relativa dos genes estudados nos diferentes tecidos e condições em que seu isolamento ocorreu.

REFERÊNCIAS BIBLIOGRÁFICAS

Abd-El Baki GK, Siefritz F, Man HM, Weiner H, et al. (2000). Nitrate reductase in *Zea mays* L. under salinity. *Plant Cell Environ*. 23: 515–521.

Ahmad I e Hellebust JA (1988). The relationship between inorganic nitrogen metabolism and proline accumulation in osmoregulatory responses of two Euryhaline microalgae. *Plant Physiol*. 88(2): 348–354.

Alia, Chen THH e Murata N (1998). Transformation with a gene for choline oxidase enhances the cold tolerance of Arabidopsis during germination and early growth. *Plant Cell Environ*. 21:232-239.

Alia PM e Matysik J (2001). Effect of proline on the production of singlet oxygen. *Amino Acids*. 21(2):195-200.

Altschul SF, Gish W, Miller W, Myers E et al. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

Amaral TM, Neves MF e Moraes MAD (2003). Cadeias produtivas do açúcar do Estado de São Paulo e da França: Comparação dos sistemas produtivos, organização, estratégias e ambiente institucional. *Agric. SP*. 50(2):65-80.

Angiosperm Phylogeny Website. http://www.mobot.org/mobot/research/apweb/. Acessado 20 Outubro, 2008.

Assis TF (2008). Melhoramento genético do eucalipto. http://www.celsofoelkel.com.br/artigos/outros/Arquivo%2014.%20melhoramento%20eucalipto.doc. Acessado 11 Novembro, 2008.

Barbosa EA (2005). Avaliação fitotécnica de cinco variedades de cana-de-açúcar para o município de Salinas – MG. Tese de Mestrado, UESB, Bahia.

Barbosa da Silva A, Wanderley-Nogueira AC, Silva RRM, Belarmino LC et al. (2005). In silico survey of resistance (*R*) genes in *Eucalyptus* transcriptome. *Genet. Mol. Biol.* 28:562-574.

Barroso C, Romero LC, Vega JM e Gotor C (1997). Molecular characterization of the sulfur metabolism in plants. *Curr. Top. Phytochem*. 1:19–29.

Barroso C, Romero LC, Cejudo FJ, Vega JM, et al. (1999). Salt-specific regulation of the cytosolic O-acetylserine(thiol)lyase gene from *Arabidopsis thaliana* is dependent on abscisic acid. *Plant Mol. Biol*. 40:729–736.

Benko-Iseppon AM (2001). Estudos moleculares no caupi e em espécies relacionadas: Avanços e perspectivas. *Embrapa Docs.* 56:327-332.

Benko-Iseppon AM, Soares-Cavalcanti NM, Wanderley-Nogueira AC, Berlarmino LC, et al. (2005). Genes associados a estresses bióticos e abióticos em feijão-caupi [*Vigna*

unguiculata (L.) Walp.] e outras angiospermas. In: Estresses Ambientais: Danos e/ou benefícios à vida das plantas, (Lima-Araújo E, ed), . Sociedade Brasileira de Fisiologia Vegetal, Campinas, Brasil. 43-53.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. (2000). GenBank. *Nucleotide. Acid. Res.* 28:15-18.

Berding N e Roach BT (1987). Germplasm collection, maintenance, and use. In: Sugarcane improvement through breeding, (Heinz DJ, ed). Elsevier, Amsterdam, 143–210.

Bezerra AAC, Távora FJAF, Freire Filho FR e Ribeiro VQ (2008). Morfologia e produção de grãos em linhagens modernas de feijão-caupi submetidas a diferentes densidades populacionais. *Rev. Biol. Cienc. Terra.* 8(1):85-93.

Biodiesel. www.biodiesel.gov.br/. Acessado 20 Novembro, 2008

Bohnert HJ, Nelson DE e Jensen RG (1995). Adaptations to environmental stress. *Plant Cell*. 103:1035-1040.

Bonner ER, Cahoon RE, Knapke SM e Jez JM (2005). Molecular basis of cysteine biosynthesis in plants: Structural and functional analysis of O-acetylserine sulfhydrylase from *Arabidopsis thaliana*. *J. Biol. Chem*. 280: 38803–38813.

Bracelpa. Associação Brasileira de Celulose e Papel. www.bracelpa.org.br/. Acessado 30 Outubro, 2008

Bray EA (1993). Molecular responses to water deficit. *Plant Physiol.* 103:1035-1040.

Bray EA (1997). Plant responses to water deficit. *Trends. Plant Sci.* 2:48-54.

Brigotti M, Petronini PG, Carnicelli D, Alfieri RR, et al. (2003). Effect of osmolarity, ions and compatible osmolytes on cell-free protein synthesis. *Biochem. J*. 369:369–374.

Cabib E e Leloir LF (1958). The biosynthesis of trehalose phosphate. *J. Biol. Chem*. 231:259–275.

Campanini B, Speroni F, Salsi E, Cook PF, et al. (2005). Interaction of serine acetyl-transferase with O-acetylserine sulfhydrylase active site: Evidence from fluorescence spectroscopy. *Protein Sci*. 14:2115-2124.

Carneiro PT, Fernandes PD, Gheyi HR e Soares FAL (2002). Germination and initial growth of precocious dwarf cashew genotypes under saline conditions. *Rev. Bras. Eng. Agr*. 6(2):199-206.

Chen TH e Murata N (2002). Enhancement of tolerance of abiotic stress by metabolic engineering of betaines and other compatible solutes. *Curr. Opin. Plant Biol.* 5:250–257.

Chen WP, Li PH e Chen THH (2000). Glycinebetaine increases chilling tolerance and reduces chilling-induced lipid peroxidation in *Zea mays* L. *Plant Cell Environ*. 23:609-618.

Chen X, Laudeman TW, Rushton PJ, Spraggins TA, et al. (2007). Database CGKB: an annotation knowledge base for cowpea (*Vigna unguiculata* L.) methylation filtered genomic genespace sequences. *BMC. Bioinfo.* 8:129.

Clarkson DT, Carvajal M, Henzler T, Waterhouse RN, et al. (2000). Root hydraulic conductance: diurnal aquaporin expression and the effects of nutrient stress. *J. Exp. Bot.* 51:61-70.

Clegg JS e Filosa MF (1961). Trehalose in the cellular slime mold, *Dictyostelium mucoroides*. *Nature*. 192:1077-1078.

Close TJ (1997). Dehydrins: a commonalty in the response of plants to dehydration and low temperature. *Physiol. Plant* 100:291-296.

CONAB. Companhia Nacional de Abastecimento. http://www.conab.gov.br/. Acessado 9 Dezembro, 2008.

Crowe JH, Crowe LM, Carpenter JF e Wistrom AC (1987). Stabilization of dry phospholipid bilayers and proteins by sugars. *Biochem. J*. 242: 1–10.

Daniels J e Roach BT (1987). A review of the origin and improvement of sugarcane In: COPERSUCAR International Sugarcane Workshop COPERSUCAR, São Paulo.

Delauney AJ, Hu C-AA, Kishor PBK e Verma DPS (1993). Cloning of ornithine 6-aminotransferase cDNA from *Vigna aconitifolia* by trans-complementation in *Escherichia coli* and regulation of proline biosynthesis. *J. Biol. Chem.* 268:18673-18678.

Delauney AJ e Verma DPS (1990). A soybean gene encoding Δ'pyrroline-5-carboxylate reductase was isolated by functional complementation in *Escherichia coli* and is found to be osmoregulated. *Mol. Gen. Genet*. 221:299-305.

Delauney AJ e Verma DPS (1993). Proline biosynthesis and osmoregulation in plants. *Plant J*. 4:215-223.

Dias NS, Gheyi HR e Duarte SN (2003). Prevenção, manejo e recuperação dos solos afetados por sais. Série Didática, 13, ESALQ/USP, Piracicaba.

Dominguez-Solis JR, Gutierrez-Alcala G, Vega JM, Romero LC, et al. (2001). The cytosolic O-acetylserine(thiol)lyase gene is regulated by heavy metals and can function in cadmium tolerance. *J. Biol. Chem*. 276:9297–9302.

Donald EN, Gerald R e Hans JB (1998). Regulation of cell-specific inositol metabolism and transport in plant salinity tolerance. *Plant Cell*. 10:753–764.

Dooki AD, Mayer-Posner FJ, Askari H, Zaiee A-a, et al. (2006). Proteomic responses of rice young panicles to salinity. *Proteomics.* 6:6498-6507.

Eisen MB, Spellman PT, Brown PO e Botstein D (1998). Cluster analysis and display of genomic-wide expression pattern. *Proc. Natl. Acad. Sci. USA*. 95:14863-14868.

Elbein AD (1974). The metabolism of alpha,alpha-trehalose. *Adv. Carb. Chem. Biochem*. 30:227-256.

Elbein AD, Pan YT, Pastuszak I e Carroll D (2003). New insights on trehalose: a multifunctional molecule. *Glycobiology*. 13:17-27.

Eldridge K, Davidson J, Harwood C, Van Wyk G (1994). Eucalypt domestication and breeding. Oxford University Press, New York.

Eck RV e Dayhoff MO (1966). Atlas of protein sequence and structure. Silver Spring, Md: National Biomedical Research Foundation. 1-68.

Embrapa Meio-Norte. Empresa Brasileira de Pesquisa Agropecuária http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Feijao/FeijaoCaupi/index.htm Acessado 11 Maio, 2008

Fitch WM (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool*. 20:406-416.

Franco J (1999). *Eucalipto.* Enciclopédia Verbo Luso-Brasileira da Cultura, Edição Século XXI, v.11, Editorial Verbo, Braga.

Freire-Filho FR (1988). Cowpea taxonomy and introduction to Brazil. In: Cowpea research in Brazil IITA, (Watt EE e Araújo JPP eds), EMBRAPA, Brasília.

Freire-Filho FR, Ribeiro VQ, Barreto PD e Santos CAF (1999). Melhoramento genético do caupi (*Vigna unguiculata* (L.) Walp.) na região nordeste. In: Recursos genéticos e melhoramento de plantas para o Nordeste Brasileiro, (Queirós MA de, Goedert CO e Ramos SRR, eds.). EMBRAPA, CPATSA, Petrolina.

Freire-Filho FR, Rocha MM, Ribeiro VQ e Lopes ACA (2005). Adaptabilidade e estabilidade produtiva de feijão-caupi. *Ciên. Rural*. 35:24-30.

Fusaro A, Mangeon A, Junqueira RM, Magran R, et al. (2001). Classification, expression pattern and comparative analysis of sugarcane expressed sequences tags (ESTs) encoding glycine-rich proteins (GRPs). *Gen. Mol. Biol*. 24:263-273.

Garg AK, Ju-Kon K, Owens TG, Ranwala AP, et al. (2002). Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. *Proc. Natl. Acad. Sci. USA*. 99(25):15898-15903.

Génard H, Le-Saos J, Billard JP, Trémolières A, et al. (1991). Effect of salinity on lipid composition, glycine betaine content and photosynthetic activity in chloroplasts *of Suaeda maritima*. *Plant Physiol. Biochem*. 29:421-427.

Grattapaglia D e Bradshaw HD (1994). Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Canad. J. For. Res*. 24(5):1074-1078.

Guimarães CT, Honeycutt RJ, Sills GR e Sobral BWS (1999). Genetic maps of *Saccharum officinarum* L. and *Saccharum robustum* Brandes & Jew. ex grassl. *Gen. Mol. Biol*. 22(1):1415-4757.

Hayashi H, Alia, Sakamoto A, Nonaka H, et al. (1998). Enhanced germination under high-salt conditions of seeds of transgenic *Arabidopsis* with a bacterial gene (codA) for choline oxidase. *J. Plant. Res.* 111:357-362.

Heinz DJ, Mee GWP e Nickel LG (1969). Chromosome number of some *Saccharum* species, hybrids and their cell suspension cultures. *Am. J. Bot.* 56:450-454.

Hell R, Wirtz M, Berkowitz O e Droux M (2001). The cysteine synthase complex from plants. Mitochondrial serine acetyltransferase from *Arabidopsis thaliana* carries a bifunctional domain for catalysis and protein-protein interaction. *Eur. J. Biochem*. 268:686-683.

Hill KD e Johnson LAS (1995). Systematics studies in the *Eucalyptus*: a study of the bloodwoods, genus *Corymbia* (*Myrtaceae*). *Telopea*. 6:185-504.

Ho CK, Chang SH, Tsay JY, Tsay CJ, et al. (1998) *Agrobacterium tumefaciens* mediated transformation of *Eucalyptus camaldulensis* and production of transgenic plants. *Plant Cell Rep*. 17:675-680.

Hong Z, Lakkineni K, Zhang Z e Verma DPS (2000). Removal of feedback inhibition of Δ-pyrroline-5-carboxylate synthetase results in increased proline accumulation and protection of plants from osmotic stress. *Plant Physiol*. 122:1129-1136.

Hopkins GW (1999). Introduction to plant physiology, (John Wiley e Sons eds.), New York. 512.

Horlacher R, Uhland K, Klein W, Ehrmann M, et al. (1996). Characterization of a cytoplasmic trehalase of *Escherichia coli*. *J. Bacteriol*. 178:6250–6257.

Hu CA, Delauney AJ e Verma DP (1992). A bifunctional enzyme (delta^{1}-pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants. *Proc. Natl. Acad. Sci. USA.* 89:9354-9358.

IBGE. Instituto Brasileiro de Geografia e Estatística. Acessado 15 Outubro, 2008

Ingelbrecht IL, Irvine JE e Mirkov AC (1999). Posttranscriptional and silencing in transgenic sugarcane: Dissection of homology-dependent virus resistance in monocot that has a complex polyploid genome. *Plant Physiol.* 199:1187-1197.

Ingram J e Bartels D (1996). The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol.* 47:377-403.

Javot H e Maurel C (2002). The role of aquaporins in root water uptake. *Ann. Bot.* 90:301-313.

Jagendorf AT e Takabe T (2001). Inducers of glycinebetaine synthesis in barley. *Plant Physiol*. 127:1827-1835.

JBonline. http://jbonline.terra.com.br/extra/2008/11/04/e041119021.html. Acessado 15 Dezembro, 2008.

Jornal da Unicamp (2001). http://www.unicamp.br/unicamp/unicamp_hoje/out2001/unihoje_ju167pag09.html/. Acessado 15 Dezembro, 2008.

Jornal Gazeta Mercantil. http://www.gazetamercantil.com.br/Gazeta/home.aspx/. Acessado 25 Outubro, 2008.

Junkes TH e Cantor CR (1969). Evolution of protein molecules In: Mammalian protein metabolism. (Munro HN ed) Academic Press, New York.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, et al. (2002). The human genome browser at UCSC. *Genome Res.* 12:996-1006.

Kishor PBK, Hong Z, Miao GH, Hu CAA, et al. (1995). Overexpression of Δ'-pyrroline-5-carboxylate synthetase increases proline production and confers osmotolerance in transgenic plants. *Plant Physiol*. 108:1387-1394.

Kishor PBK, Sangam S, Amrutha RN, Laxmi PS, et al. (2005). Regulation of proline biosynthesis, degradation,uptake and transport in higher plants: Its implications in plant growth and abiotic stress tolerance. *Curr. Sci*. 88(3):424-438.

Kumar S, Tamura K e Nei M (2004). MEGA 3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Brief. Bioinf*. 5:150-163.

Landfald B e Strom AR (1986). Choline-glycine betaine pathway confers a high level of osmotic tolerance in *Escherichia coli*. *J. Bacteriol*. 165:849-855.

Larcher W (2000). Ecofisiologia Vegetal. In: A planta sob o estresse. (Prado CHBA, ed.) São Carlos, 341-448.

Leal-Bertioli SC, Messenberg P, Fávero GAP, Moretzsohn MC, et al. (2003). Amendoim Selvagem - Uma fonte de resistência a pragas. *Ver. Biotec. Cien. Desenv.* 31:116-119.

Lesk AM (2002). Introduction to bioinformatics. Oxford, New York.

Loewus FA e Loewus MW (1983). Myo-inositol: its biosynthesis and metabolism. *Plant Physiol*. 34:137-161.

Loewus FA e Murthy PPN (2000). Myo-Inositol metabolism in plants. *Plant Sci*. 150:1–19.

Madin KAC e Crowe JH (1975). Anhydrobiosis in nematodes: carbohydrate and lipid metabolism during dehydration. *J. Exp. Zool*. 193:335-342.

Mahajan S e Tuteja N (2005). Cold salinity and drought stresses: An overview. *Arch. Biochem. Biophys*. 444:139-158.

Majumder AL, Johnson MD e Henry SA (1997). 1L-myo-inositol-1-phosphate synthase. *Biochem. Biophys.* 1348:245-256.

Mansur E e Margis-Pinheiro M (1996). Plantas geneticamente modificadas: um caminho para o melhoramento vegetal. In: ABCTP Notícias.http://www.cnph.embrapa.br/laborato/ biocel/abctp23.htm. Acessado 15 Novembro, 2008.

Mansour MMF (1998). Protection of plasma membrane of onion epidermal cells by glycinebetaine and proline against NaCl stress. *Plant Physiol. Biochem*. 36:767-772.

Maruyama Y, Nakamura S, Marenco RA, Vieira G, et al. (2005). Photosynthetic traits of seedlings of several tree species in an Amazonian forest. *Tropics.* 14(3):211-219.

Martini AJ (2004). O plantador de Eucaliptos: A questão da preservação florestal no Brasil e o resgate documental do legado de Edmundo Navarro de Andrade. Tese de Mestrado, USP, São Paulo.

Martim SA (2003). Pulverização do cafeeiro com açúcar: potencial de uso em mudas submetidas à deficiência hídrica e na recuperação de plantas atingidas por Glyphosate. Tese de Mestrado, UFLA, Lavras.

Matoh T, Watanabe J e Takahashi E (1987). Sodium, potassium, chloride, and betaine concentrations in isolated vacuoles from salt-grown *Atriplex gmelini* leaves. *Plant Physiol.* 84:173–177.

Matsuoka S, Garcia AAF e Arizono H (1999). Melhoramento da cana-de-açúcar In: Melhoramento de espécies cultivadas. (Bórem A, ed), Viçosa, Minas Gerais.

McCue KF e Hanson AD (1990). Drought and salt tolerance: towards understanding and application. *Trends. Biotechnol*. 8:358–362.

McCue KF e Hanson AD (1992). Effects of soil salinity on the expression of betaine aldehyde dehydrogenase in leaves: investigation on hydraulic, ionic and biochemical signals. *Aust. J. Plant Physiol*. 19:555–564.

McNeil SD, Nuccio ML e Hanson AD (1999). Betaines and related osmoprotectants. Targets for metabolic engineering of stress resistance. *Plant. Physiol.* 120:945-949.

Meneses CHSG, Morais-Lima LHG, Lima MMA e Vidal MS (2006). Aspectos genéticos e moleculares de plantas submetidas ao déficit hídrico. *Rev. bras. ol. fibros*. 10:1039-1072.

Molinari HBC (2006). Expressão estresse-induzida do gene *P5CS* em plantas transgênicas de cana-de-açúcar submetidas ao déficit hídrico. Tese de Mestrado, UFPR, Curitiba.

Morais DAL (2003). Análise bioinformática de genes de resistência à patógenos no genoma expresso da cana-de-açúcar. Tese de Mestrado, UFPE, Recife.

Munôz-Clares RA e Velasco-Garcia R (2004). Genio y figura de la betaína aldehído deshidrogenasa. *Mensaje Bioquímico*. 28:203-223.

Nei M (1991). Relative efficiencies of different tree making methods for molecular data. In: Recent advances in phylogenetic studies of DNA sequences, (Miyamoto MM e Cracraft JL, eds) Oxford University Press, New York, 90-128.

Nelson DE, Rammesmayer G e Bohnert HJ (1998). Regulation of cell-specific inositol metabolism and transport in plant salinity tolerance. *Plant Cell.* 10:753-764.

Nogueira RMC, Araújo EL, Willadino LG e Cavalcante UMT (2005). Estresses ambientais: danos e benefícios em plantas. Imprensa Universitária, Recife.

Noji M, Saito M, Nakamura M, Aono M, et al. (2001). Cysteine synthase overexpression in tobacco confers tolerance to sulfur-containing environmental pollutants. *Plant Physiol.* 126:973-980.

Nuccio ML, Russell BL, Nolte KD, Rathinasabapathi B, et al. (1998). The endogenous choline supply limits glycine betaine synthesis in transgenic tobacco expressing choline monooxygenase. *Plant J*. 16:487-496.

Page RDM (1996). TreeView: An application to display phylogenetic trees on personal computers. *Comp. Applic. Biosci.* 12: 357–358.

Paiva JB, Carmo CM, Tavora FJA, Almeida FG, et al. (1970). Melhoramento, experimentação e fitossanidade com feijão (*Vigna simensis*), realizadas no estado do Ceará (1967/68). *Pesq. Agrop. Nord*. 2(2):99-113.

Papageorgiou GC e Murata N (1995). The unusually strong stabilizing effects of glycine betaine on the structure and function of the oxygen-evolving photosystem II complex. *Photosynth. Res*. 44:243-252.

Peng Z, Lu Q e Verma DP (1996). Reciprocal regulation of Δ^1-pyrroline-5-carboxylate synthetase and proline dehydrogenase genes control proline levels during and after osmotic stress in plants. *Mol. Gen. Genet*. 253:334-341.

Prosdocini F, Cerqueira GC, Binneck E, Silva AF, et al (2002). Bioinformática: manual do usuário. *Biotecnol. Cien. Desenvol*. 29:12-25.

Pryor LD e Johnson LA (1971). A classification of the Eucalyptus. Canberra, Australian National University, 192.

Quan R, Shang M, Zhang H, Zhao Y, et al. (2004). Improved chilling tolerance by transformation with *betA* gene for the enhancement of glycinebetaine synthesis in maize. *Plant Sci.* 166:141–149.

Rathinasabapathi B, Burnet M, Russell BL, Gage DA, et al. (1997). Choline monooxygenase, an unusual iron-sulfur enzyme catalyzing the first step of glycine betaine synthesis in plants: Prosthetic group characterization and cDNA cloning. *Proc. Natl. Acad. Sci. USA.* 94:3454–3458.

Rathinasabapathi B (2000). Metabolic engineering for stress tolerance: installing osmoprotectant synthesis pathways. *Ann. Bot*. 86:709-716.

Rhodes D e Hanson AD (1993). Quaternary ammonium and tertiary sulfonium compounds in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. 44:357-384.

Rimmele M e Boos W (1994). Trehalose-6-phosphate hydrolase of *Escherichia coli*. *J. Bacteriol.* 176:5654-5664.

RIPASA (2002). Ripasa S/A Celulose e Papel. <http://www.ripasa.com.br/index.cfm?cg=NEWS&noticia=946>. Acessado 23 August, 2008

Rodrigues R (2004). Século XXI, o novo tempo da agroenergia renovável. *Visão Agrícola*. 1(1):4-7.

Romero C, Belles JM, Vaya JL, Serrano R, et al. (1997). Expression of the yeast trehalose-6-phosphate synthase gene in transgenic tobacco plants: pleiotropic phenotypes include drought tolerance. *Planta.* 201(3):293-297.

Romero LC, Domínguez-Solís JR, Gutiérrez-Alcalá G e Gotor C (2001). Salt regulation of O-acetylserine(thiol)lyase in *Arabidopsis thaliana* and increased tolerance in yeast. *Plant Physiol. Biochem.* 39:643−647.

Rontein D, Basset G e Hanson AD (2002). Metabolic engineering of osmoprotectant accumulation in plants. *Metab. Eng*. 4(1):49-56.

Rozwadowski KL, Khachatourians GG e Selvaraj G (1991). Choline oxidase, a catabolic enzyme in *Arthrobacter pascens*, facilitates adaptation to osmotic stress in *Escherichia coli*. *J. Bacteriol*. 173:472-478.

Ruiz JM e Blumwald E (2002). Salinity-induced glutathione synthesis in *Brassica napus*. *Planta*. 214:965–969

Saitou N e Nei M (1987). The neighbor-joinning method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. 4:406-425.

Sakamoto A e Murata N (2000). Genetic engineering of glycinebetaine synthesis in plants: current status and implications for enhancement of stress tolerance. *J. Exp. Bot*. 51:81-88.

Sakamoto A e Murata N (2001). The use of bacterial choline oxidase, a glycinebetaine-synthesizing enzyme, to create stress-resistant transgenic plants. *Plant Physiol*. 125:180-188.

Sakamoto A e Murata N (2002). The role of glycine betaine in the protection of plants from stress: clues from transgenic plants. *Plant Cell Environ*, 25:163-171.

Sociedade Brasileira de Silvicultura (SBS), http://www.sbs.org.br. Acessado 11 Novembro, 2008.

Scandiffio IGS (2005). Análise prospectiva do álcool combustível no Brasil - Cenários 2004-2024. Tese de Doutorado. UNICAMP, São Paulo.

Schick I, Haltrich D e Kulbe KD (1995). Trehalose phosphorylase from *Pichia fermentans* and its role in the metabolism of trehalose. *Appl. Microbiol. Biotechnol*. 43:1088-1095.

Selvaraj G (2000). Genetic engineering of glycinebetaine production toward enhancing stress tolerance in plants: metabolic limitations. *Plant Physiol*. 122:747-756.

Serraj R e Sinclair TR (2002). Osmolyte accumulation: can it really help increase crop yield under drought conditions? *Plant Cell Environ*. 25:333-341.

Shinozaki K e Yamaguchi-Shinozaki K (1997). Gene expression and signal transduction in water-stress response. *Plant Physiol.* 115:327–334.

Shirasawa K, Takabe T, Takabe T e Kishitani S (2006). Accumulation of glycine-betaine in rice plants that overexpress choline monooxygenase from spinach and evaluation of their tolerance to abiotic stress. *Annals. Bot.* 98:565–571.

Silveira JAG, Melo ARB, Viégas RA e Oliveira JTA (2001). Salinity-induced effects on nitrogen assimilation related to growth in cowpea plants. *Environ. Exp. Bot*. 46:171-179.

Sneath PH e Sokal RR (1973). Numerical taxonomy**:** The principles and practice of numerical classification. San Francisco: W.H. Freeman, 573.

Simon MV, Benko-Iseppon AM, Resende LV, Winter P, et al. (2007). Genetic diversity and phylogenetic relationships in *Vigna Savi* germplasm revealed by DNA amplification fingerprinting. *Genome*. 50: 538-547.

Sirko A, Blaszczyk A e Liszewska F (2004). Overproduction of SAT and/or OASTL in transgenic plants: a survey of effects. *J. Exp. Bot*. 55(404):1-8.

Steane D, Mckinnon GE, Vaillancourt RE e Potts BM (1999). Its sequence data resolve higher level relationships among the *Eucalyptus*. *Mol. Phylogenet. Evol*. 12:215-223.

Styer JC (2000). Regulating inositol biosynthesis in plants: myo-Inositol phosphate synthase and myo-Inositol monophosphate. Tese de Mestrado, Faculty of Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Sudhir K, Koichiro T e Masatoshi N (1993). MEGA: Molecular evolutionary genetics analysis version 2.0. The Pennsylvania State University. University Park. Pennsylvania, USA.

Szoke A, Miao GH, Hong Z e Verma DPS (1992). Subcellular location of delta^{1}-pyrroline-5-carboxylate reductase in root/nodule and leaf of soybean. *Plant Physiol*. 99:1642-1649.

Takahashi S, Furukawa T, Asano T, Terajima Y, et al. (2005). Very close relationship of the chloroplast genomes among *Saccharum* species. *Theor. Appl. Genet*. 110:1523–1529.

Tavares de Melo LJO (2005). Análise agronômica e genética de genótipos de cana-de-açúcar nas regiões litoral sul e mata norte de Pernambuco. Tese de Mestrado, UFRPE, Recife.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin J, et al. (1997). The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic. Acids. Res.* 25:4876-4882.

UNICA. União da Indústria de Cana-de-açúcar. www.unica.com.br/. Acessado 10 Novembro, 2008

Verbruggen N, Hua XJ, May M e van Montagu M (1996). Environmental and developmental signals modulate proline homeostasis: evidence for a negative transcriptional regulator. *Proc. Natl. Acad. Sci. USA*. 93:8787–8791.

Verbruggen N, Villarroel R e Montagu MV (1993). Osmoregulation of a pyrroline-5-carboxylate reductase gene in *Arabidopsis thaliana. Plant Physiol.* 103:771-781.

Vettore AL, da-Silva FR, Kemper EL e Arruda P (2001). The libraries that made SUCEST. *Genet. Mol. Biol.* 24:1-7.

Vogt C (2003). Bioinformática, genes e inovação. http://www.comciencia.br/reportagens/bioinformatica/bio01.shtml. Acessado 14 Novembro, 2008.

Wahid A e Ghazanfar A (2006). Possible involvement of some secondary metabolities in salt tolerance of sugarcane. *J. Plant Physiol*. 163:723-730.

Wang W, Vinocur B e Altman A (2003). Plant responses to drougth salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta.* 218:1-14.

Waditee R, Bhuiyan MdNH, Rai V, Aoki K, et al. (2005). Genes for direct methylation of glycine provide high levels of glycinebetaine and abiotic-stress tolerance in *Synechococcus* and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA*. 102(5):1318–1323.

Yancey PH (2001). Water stress, osmolytes and proteins. *Amer. Zool.* 41:699–70.

Yancey PH (2005). Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J. Experim. Biol*. 208:2819-2830.

Yoshiba Y, Kiyosue T, Katagiri T, Ueda H, et al. (1995). Correlation between the induction of a gene for Δ^1 -pyrroline-5-carboxylate synthetase and the accumulation of proline in *Arabidopsis thaliana* under osmotic stress. *Plant J.* 7:751-760.

Youssefian S, Nakamura M, Orudgev E e Kondo N (2001). Increased cysteine biosynthesis capacity of transgenic tobacco overexpressing an *O*-acetylserine (thiol) lyase modifies plant responses to oxidative stress. *Plant. Physiol*. 126:1001–1011.

Zeiger E (1990). Light perception in guard cells. *Plant Cell Environ.* 13:739747.

Zhu J-K (2001). Cell signaling under salt water and cold stresses. *Curr. Opin. Plant Biol*. 4:401–406.

Zhu J-K (2002). Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53:247-273.

Zhu J-K, Hasegawa PM e Bressan RA (1997). Molecular aspects of osmotic stress in plants. *CRC. Crit. Rev. Plant Sci.* 16:253-277.

Fontes da Internet

Eucalyptus EST Database (FORESTs), https://forests.esalq.usp.br

Harvest Database (HARVEST), harvest.ucr.edu/

National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov

Sugarcane EST Genome Project (SUCEST), http://sucest.lad.dcc.unicamp.br

Capítulo 2

Artigo Científico

***In silico* Evaluation of Osmoprotectants genes in the *Eucalyptus* Transcriptome.**

Artigo aceito para publicação na revista *Lecture Notes in Computer Science*.

In Silico Evaluation of Osmoprotectants in *Eucalyptus* Transcriptome

Petra Barros dos Santos[1], Nina M. Soares-Cavalcanti[1], Gabriela S. Vieira-Mello[1], Ana C. Wanderley-Nogueira[1], Tercílio Calsa-Junior[1], Ana M. Benko-Iseppon[2]

[1] Universidade Federal de Pernambuco, Center for Biological Sciences (CCB), Genetics Department, Laboratory for Plant Genetics and Biotechnology, Av. Prof. Moraes Rêgo s/no. Recife, PE, Brazil. Phone: +55-81-2126-8520, Fax: +55-81-2126-8569

[2] Corresponding author: ana.benko.iseppon@pesquisador.cnpq.br

Abstract. The synthesis of osmoprotectants is one of the main mechanisms to dribble harmful effects caused by abiotic stresses. In the present work, osmoprotectant genes were selected and used in a search for orthologs in the FORESTs database using *in silico* procedures. The studied genes were here analyzed for the first time in *Eucalyptus*, including 51 identified orthologs from nine gene families (*P5CS = 3*, *P5CR = 4*, *TPS1 = 10*, *TPPB = 8*, *SAT = 6*, *OASTL = 13*, *CMO = 1*, *BADH = 3* and *INPS1 = 3*). In general, the most represented genes were those involved in the trehalose and cysteine synthesis, probably due to their vital importance in the maintenance of cellular homeostasis. Overall expression pattern revealed significant amounts of reads in most libraries. Dendrograms based on analysis of selected genes from other flowering plants revealed no clear separation between mono and dicots, since diversity regarding the evolution and diversity in gene signatures may be more related to fitness and adaptation, than evolutionary patterns.

1 Introduction

Despite of significant information regarding response to abiotic stresses in herbaceous plants, as rice and arabidopsis, scarce information is available for understanding such mechanisms in woody plants, principal components of arid and semi-arid regions of the word [1]. Physiological characters widely associated with drought resistance in eucalyptus include greater osmotic adjustment, tissue elasticity, higher leaf water content and higher ability to resist desiccation [2]. Drought and salt stresses are often interconnected and cause cellular damages, such as osmotic and oxidative stresses. To cope with these environmental influences, plants have developed arrays of physiological and biochemical strategies to adapt to the adverse conditions. One of the most common metabolic adaptations is the accumulation of certain organic solutes (known as osmoprotectants) that protect proteins and membranes against damage by high concentrations of inorganic ions [3].

Trehalose helps to stabilize enzymes, membrane proteins and lipids, acting also as protective of biological structures for several stress types. It is synthesized by a two-step process in which trehalose-6-phosphate synthase (TPS1) synthesizes T6P from UDP-glucose and glucose-6-phosphate, followed by dephosphorylation of trehalose by trehalose-6-phosphate phosphatase (TPPB) [4].

Myo-inositol is a cyclic alcohol formed from the glucose-6-phosphate by an internal reaction of oxy-reduction involving NAD^+ and INPS1 enzyme (myo-inositol-1-phosphate synthase), a reaction catalyzer. Methylated forms of myo-inositol accumulate in plants in response to salinity stress to act during osmotic adjustment [5].

Proline (Pro) is one of the most common compatible osmolytes in water-stressed plants. The accumulation of Pro in dehydrated plants is caused both by activation of the Pro biosynthesis and by inactivation of the Pro degradation. In plants, L-Pro is synthesized from L-glutamic acid (l-G1u) via Δ^1-pyrroline-S-carboxylate (P5C) by two enzymes, P5C synthetase (P5CS) and P5C reductase (P5CR) [6].

The final steps of cysteine biosynthesis are catalyzed by a bi-enzyme complex composed of serine acetyltransferase (SAT) and *O*-acetyl-serine(thiol)lyase (OASTL). Some previous studies revealed an increased synthesis of cysteine when the plants are exposed to high salinity levels. The authors suggested that the synthesis of this amino acid is essential for the protection and/or adaptation of the plants to the stress condition [7].

In higher plants the glycine-betaine (GB) is synthesized from two steps, with the first is catalyzed by the enzyme CMO (choline mono-oxygenase), which carries out the oxidation of choline in aldehyde betaine and the second by the enzyme BADH, which converts aldehyde betaine in GB. Under salt stress, GB helps to maintain the osmotic potential in cytosol, also protecting macromolecules, including proteins [8].

None of the above described enzymes have been described before in eucalyptus. The *Eucalyptus* Genome Sequencing Consortium (FOREST project) generated 112,857 expressed sequence tags (ESTs) from 19 libraries regarding specific tissues and stages or conditions using five *Eucalyptus* species [9]. The present work performed a data-mining based identification of osmoprotectant genes in the FORESTs database, using well known arabidopsis sequences as seed sequences, also evaluating their expression in FORESTs libraries, and comparing them with sequences from public databases and literature data.

2 Methods

Descriptions of the FORESTs cDNA libraries, sequence nomenclature and clustering methods are described in Vicentini et al. (2005) [9]. Complete A. thaliana full length cDNA sequences of osmoprotectants were blasted against the FORESTs database using the tBLASTn tool [10]. Only sequences with e-value $e^{-}05$ or less were annotated. *Eucalyptus* ESTs were translated with the aid of ORFfinder program and screened for similarity to entries in GenBank (NCBI) using BLASTx.

Moreover, Blink function, was used to obtain similar sequences containing the searched conserved domains (CDs) in other organisms. The RPS-BLAST tool was used to inquire the presence and integrity of the CDs. A multiple alignment using CLUSTALx [11] program was generated including only sequences with complete CDs, aiming to compare the eucalyptus' and other osmoprotectants. The MEGA software (Version 4.0) [12] allowed the construction of dendrograms using the Neighbor Joining method, gamma distance and bootstrap test (with 1000 replications) to infer the relationship among sequences. To evaluate the ESTs expression pattern in eucalyptus sequences, a normalized data matrix including annotated reads for each library was constructed and used for a hierarchical clustering analysis using CLUSTER_v.2.20 program [13], with data displayed in the TreeView_v.1.60 program [14].

3 Results

Osmoprotectants in Eucalyptus – The search for orthologs osmoprotectant genes in the FORESTs database revealed the presence of several clusters with high degree of similarity, with best e-value for each gene ranging from 0.0 to $3e^{-74}$ (Table 1). In g eneral, 90.6% sequences of eucalyptus presented higher similarity to dicots, while only 5.66% were more similar to monocots and in few cases (3.74%) with other taxa. This was the case of *TPS1*, where one of the 13 clusters identified showed similarity with a bryophyte (*Physcomitrella patens*). In the case of *INPS1*, one of the three orthologs clusters presented high similarity with the fungi *Aspergillus terreus* and *Giberella zeae*.

Moreover, the results in FORESTs revealed the presence of 13 and seven clusters similar to the *OASTL* and *SAT* genes respectively, which jointly take role in cysteine synthesis. From the 13 candidates to *OASTL*, five had the full conserved domain PALP, while in the other eight this domain was incomplete. Regarding the *SAT*, three clusters showed the full domains SATase_N and Lbeta H while these domains were absent in the other four. Regarding *TPS1* and *TPPB13*, eight candidates were identified. After reverse alignment most ortholog candidates of these genes showed similarity with *A. thaliana*. In conserved domain analyses for each gene, two sequences (e^{-74} and e^{-107}, respectively) were found to contain the complete procured domains (glico-transferase and TPPase) in the first case, and TPPase, in the second.

The search for similar sequences to *P5CS* gene revealed three clusters, with sequences ranging from 829 to 2666 nt, the last sequence with e-value 0.0. The BLASTx results showed that eucalyptus sequences presented high similarity with the same gene from kiwi fruit (*Actinidia chinensis*, gi:6225816), *Saccharum arundinaceum* (gi:157061835) and from ice-plant (*Mesembryanthemum crystallinum*, gi:6225818). Regarding the CD analysis, only one eucalyptus sequence had the desired domains AA_Kinase and ProA, although the latter was incomplete. In the search for orthologs of *P5CR*, two of the four clusters found presented a partial *P5CR* domain.

The search for ortholog candidates to *CMO* gene revealed one sequence with high similarity (e-value $e^{-}91$), but without the procured CD. In the search for *BADH*, 30 clusters were observed; however the analysis of these candidates in NCBI database, showed that three of them were in fact ortholog to the gene in question, one with the searched domain in its entirety, while the other clusters showed similarity to aldehyde dehydrogenase (ALDH) genes of the same family than BADH.

Library Specific Expression – 625 osmoprotectant transcripts were identified at the FOREST database and included in the evaluation. The exceptions were the candidates to *P5CR* and *CMO* genes, where the number of reads was not sufficient to build a representative analysis of the expression profile.

Table 1: *Clusters* of *Eucalyptus* (best matches) in the FORESTs database as compared with osmoprotectants known from *A. thaliana*. Abbreviations: nt, nucleotide; aa, amino acid; Fr, Frame; Gi, GenBank identifier.

Gene	Results from FORESTs					Results from BLASTx in NCBI				
	FORESTs code	e-value	Size nt	Fr	# Hits	Gi number	Organism	e-value	Size aa	Fr
P5CS	P5CS_RT3103C01.g	0.0	2666	+2	3	6225816	*Actinidia chinensis*	0.0	717	+2
P5CR	P5CR_CL1286B03.g	$2e^{-84}$	851	+2	4	130972	*Glycine max*	3e-84	274	+2
CMO	CMO_CL1347G03.g	e^{-91}	815	+1	1	33300598	*Oryza sativa*	1e-90	410	+1
BADH	BADH_RT3200D09.g	0.0	1837	+1	3*	40850676	*Gossypium hirsutum*	0.0	503	+1
SAT	SAT_ST2201B08.g	e^{-93}	1309	+2	7	23343585	*Nicotiana tabacum*	$3e^{-118}$	300	+2
OASTL	OASTL_ST2203B10.g	e^{-243}	1333	+2	13	148562451	*Glycine max*	$2e^{-183}$	372	+2
TPPB	TPPB_RT5001B05.g	$2e^{-74}$	1059	+2	8	51458330	*Nicotiana tabacum*	6^{*-101}	384	+1
TPS1	TPS1_RT3140G03.g	0.0	3089	+2	13	15224213	*Arabidopsis thaliana*	0.0	862	+2
INPS1	INPSI_LV2201G02.g	0.0	2106	+3	3	14548095	*Sesamum indicum*	0.0	510	+3

*From 30 clusters with high e-values only three sequences were, in fact, orthologs to the BADH.

Considering all identified osmoprotectants, most reads came from stems (ST) and seedlings (SL) libraries, the two largest FORESTs libraries (with 26,318 and 27,437 reads, respectively). However, they were abundant also in smaller libraries, nominally 100 reads in callus libraries (CL, 12.533 reads), and 91 in root tissues (RT, 20.129 reads) of eucalyptus seedlings from nurseries. 625 identified transcripts were used for a hierarchical clustering analysis, to allow the identification of expression distribution and intensity, as well as co-expressed sequences (co-regulated), pertaining gene categories (pink dendrogram, in vertical) or libraries (gray dendrogram, in horizontal) (Figure 1).

Expression was evidently higher in stem of plants submitted to water deficit (ST) regarding all gene families studied. Increased but less strong expression of osmoprotectants could be observed in seedlings grown in the dark (SL) and roots in the absence of stress (RT). Associated spatial co-expression was stronger among STRT, while SL had higher association with FB (buttons, flower and fruits) and CL (calli) libraries. Co-expression of genes regarded mainly BADH/-TPS1+OASTL and SAT/P5CS+TPPB (Figure 1)

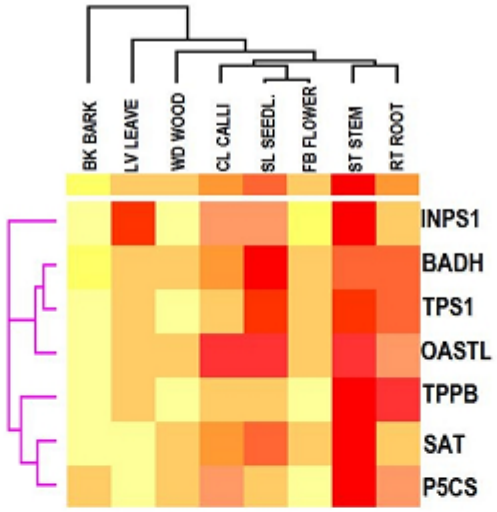


Fig. 1: Graphic representation of expression pattern analysis of seven osmoprotectants in the FORESTs database. Light yellow means no expression, and degrees from yellow to orange and red regard increased expression intensity. Abbreviations for libraries: BK (bark, duramen, pith and alburnum of trees of *E. grandis* with eight years); CL (calli grown in the dark), FB (buttons, flowers and fruits); LV (leaves of seedlings); RT (roots of seedlings from nursery), SL (seedlings grown in the dark), ST (stem seedlings subjected to water deficit) and WD (wood of *E. grandis*).

Phenetic analysis – Most dendrograms (Figure 2 and supplementary data, Figure 3), grouped genes from monocots and dicots together in the same clades, with exception of BADH (Fig. 2A), where members of the two angiosperm classes grouped into different branches. However, within dicots classic taxonomic grouping was not observed for BADH sequences. With respect to TPS1 (Fig. 2B), the dendrogram included primitive plant members (Lycopodiophyta and Ginkgophyta) together with Angiosperms. Additional dendrograms for SAT, OASTL and TPPB comprised only flowering plants (Figs. 3A, 3B and 3C), with exception of INPS1 (Fig. 3D), where a fungus sequence (genus *Giberella*) was included.

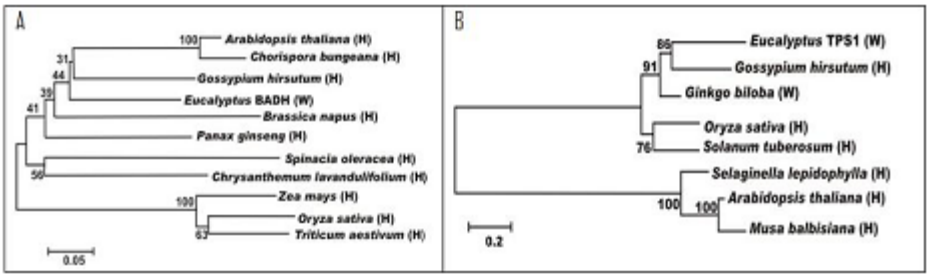


Fig. 2: Dendrograms generated after Neighbor Joining analysis illustrating relationship revealed by conserved domains similarity in (A) BADH-like and (B) TPS1-like sequences and putative *Eucalyptus* orthologs. The bar represents genetic distance, while numbers in the base of clades regard bootstrap values. The letters in parenthesis mean herbaceous (H) and woody (W) plants.

4 Discussion

Osmoprotectants in Eucalyptus – All procured gene classes were present in the FOREST database bearing structural conserved regions as compared to arabidopsis seed sequences, exhibiting also considerable divergence, sometimes even regarding the domain structure, enough to be indicative of long divergent evolution. Our findings confirmed observations [15] concerning to Resistance (R) genes analysis in eucalyptus, which revealed considerable discrepancies among genes of eucalyptus and those available in databases for angiosperms, especially considering the procured conserved domains. The authors suggested that the low number of sequences from woody plants available in public databases can be one of the factors that justify novel structural forms in eucalyptus.

The fact that OASTL gene family presents a larger number of representatives in eucalyptus may be due to the important role of this protein in the cysteine synthesis. The OASTL activity is higher than that of SAT [16], since SAT is generally found only in association with OASTL, forming the complex CSC, while OASTL may be also found in its free form in the cell.

The presence of restricted number of *TPS1* and *TPPB* representatives may be explained by previous evidences in arabidopsis [17] where some plants did not accumulate trehalose effectively when subjected to stresses, probably due to the high activity of the trehalase enzyme (found in all tissues of *A. thaliana* and also in *Glycine max*) that hydrolyses trehalose in two glucose molecules. With reference to *P5CR*, none of the candidates presented complete domains. Regardless that, the high similarity of the identified sequences (e-value 0.0 to *A. thaliana*) indicates the existence of this route in *Eucalyptus* transcriptome. Previous identification of this protein in many organisms, from prokaryotes to eukaryotes (bacteria, algae, higher plants, mammals, etc.) provide strong evidences that the processes involved in the proline formation are conserved [18]. Physiological experiments performed with *Eucalyptus tereticornis* showed that when subjected to salt stress, the leaves presented increased proline content [19]. Since the FORESTs project presents only a library subjected to conditions of water deficit (regarding the stem tissue of seedlings), it is possible that additional libraries from different tissues and treatments (as drought and salinity) would allow the observation of a greater number of representatives of these genes.

Considering glycine-betaine, it was expected to find at least one representative of the *CMO* with its complete CD. Despite of that, BLASTx search revealed that the identified CMO eucalyptus cluster showed high similarity with the same gene in rice, indicating that even in the absence of the CD the eucalyptus transcriptome included orthologs of this gene. Evidences in favor to this supposition are given by the effective accumulation of glycine-betaine in *Eucalyptus* and in many woody plants as *Pinus*, *Larix*, *Fraxinus*, [20].

Besides the already mentioned compatible osmolytes, the *INPS1* also participates in cellular protection against dehydration and salt accumulation. Three sequences with significant similarity were found in eucalyptus for this gene category. The first two aligned significantly (e-values 0.0 and e^{-95}) with *Sesamum indicum* after BLASTx. Together these results point to the presence of all im-

portant gene classes for the synthesis of compatible osmolytes in *Eucalyptus* transcriptome, considering the knowledge available to *A. thaliana* model plant.

In silico expression – The observed abundance of osmoprotectants in stems of seedlings subjected to water deficit library was expected, since stem tissues are responsible for the support of young plants, preventing its falling down when under drought or salinity [21]. The existence of a higher expression in two libraries subjected to other abiotic stresses (SL=seedlings grown in the dark and CL=calli grown in the dark) confirms that these genes are recruited under additional adverse conditions, other than those related to drought and salinity, in the present case, the absence of light. It is interesting to note that co-expression of osmoprotectants was evident in the above mentioned libraries, with closer co-expression of *BADH*, *TPS1* and *OASTL*, in contrast to *SAT*, *P5CS* and *TPPB*. Previous evaluations of PR (pathogenesis-related) proteins in sugarcane, [22] identified comparable standards, with authors arguing that co-expressed genes associated to metabolic pathways may be physically "clusterized" in the chromosomes, a fact confirmed in completely sequenced genomes [23]. It is worth noticing that besides higher expression in some libraries and tissues, almost all tissues presented some expression of compatible osmolytes, even considering that only one library was built under abiotic stress. This indicates a relative importance of some basal osmoprotection in plants, even when the environmental conditions are not stressful.

Dendrograms – It is noteworthy that none of the dendrograms followed a strict taxonomic grouping forming often mixed non taxonomic grouping. This should be surprising, considering that alignments included mainly conserved domains. As emphasized by Weston et al. [24], despite the advances made possible by "omics"-based technologies, we still struggle to accurately associate the genes, transcriptional cascades, and signaling networks with physiological performance and ecological fitness. This is especially true for osmoprotectants, with gene evolution and functional structure probably more related to fitness and adaptation, than to casual evolutionary patterns, revealing groupings that seem to associate different gene signatures to physiological performance. Results regarding genes related to abiotic stress responses have shown that dendrograms very often group sequences in functional arrangements, with common signatures probably reflecting environmental driven adaptation or selection [25]. This may be interpreted as the result of common ancestry, partially lost (or silenced/rarely expressed) genes in some taxa, but also to convergent mutation (or reverse mutation) associated to ecological fitness.

The separation of the BADH dendrogram (Fig. 2A) in two major groups (dicots and monocots) can be explained by the difference of RNA processing pattern of BADH homologs among both groups. Moreover, poaceous species as *Zea mays*, *Triticum aestivum* and *Oryza sativa* accumulate less glycinebetaine (GB) in comparison with other families. In dicots clade, spinach is a chenopod member that accumulates far more GB than do cereal crops in response to osmotic stresses [26]. In our dendrogram rice and wheat grouped together, maybe because the rice sequence used came from salt stressed plants fed with

betaine aldehyde for one week, leading to expression of some new candidates and producing GB levels close to that observed in wheat.

The presence of trehalose-6-phosphate synthase (Fig. 2B) in Ginkgophyta, Magnoliophyta and Algae indicates that TPS1 represents an arqueomorphic character. The monocots, as well as Rosidae, also formed a polyphyletic group while the branch composed by eucalyptus, *G. hirsutum* and *G. biloba* formed a paraphyletic group. TPS1 appear to be an essential gene in plants, present in primitive plants including *Selaginella*, a known desert plant from the spikemoss Lycopodiophyta group (known for its ability to survive almost complete desiccation) and also Ginkgo that is remarkably resistant to cold and drought, being regarded as one of the first terrestrial plants.

Considering the SAT dendrogram (Fig. 3A), monocot plants (rice and maize) grouped together, both sensible to abiotic factors. In the second clade all dicots grouped (again with exception of eucalyptus) with both Brassicaceae joining a distinct sub-clade with high bootstrap value (99). The here aligned gene from *A. thaliana* was an ortholog to a *Salmonella typhimurium* gene [27], also revealing its ancestral character. Furthermore, previously observations revealed that elevated glutathione, driven by constitutive activation of SAT gene, plays a role in the Ni tolerance of *T. goesingense* (Brassicaceae). A recent research using this organism [28] indicated that elevated glutathione levels, driven by constitutively enhanced SAT activity in the hyperaccumulator *T. goesingense*, plays an important role in the Ni, Co and Zn tolerance. The soybean SAT was positioned in an intermediary position between the two Brassicaceae and two *Nicotiana* species. Furthermore, it was shown [29] that a serine acetyltransferase from soybean (GmSerat2;1) is a substrate of CDPK (calcium dependent protein kinase), so it seems that SAT, at least in *Glycine max*, is strongly involved with tolerance to different kinds of stresses.

After analysis using OASTL sequence, the distribution of species in this dendrogram may be explained by the existence of different isoforms of this gene (OASTLA, cytosolic; OASTLB, plastid; and OASTLC, mitochondrial), a fact already observed in arabidopsis and spinach [30]. In fact previous analysis [31] showed that the here used sequences of *Z. mays*, *O. sativa*, *N. plumbaginifolia*, *Citrullus lanatus*, *Populus alba* X *P. tremula* and *Allium tuberosum*, are the cytosolic OASTL isoforms and therefore are grouped together in the dendrogram. It seems clear that the here analyzed eucalyptus sequence represents the cytosolic isoform, *Hevea brasiliensis* may represent another isoform, considering its behavior as an out-group in the dendrogram. Regarding the few complete TPPB proteins from data banks, a junction of several organisms from different taxonomic groups are mixed in the formed clades, with eucalyptus composing an outgroup together with tobacco (Fig. 3C) with high bootstrap value (100). The availability of a larger number of orthologs may bring useful evidences concerning diversity and also regarding complex regions in conserved/non conserved segments of this gene category. With exception of the eucalyptus positioning, the *INPS1* dendrogram (Fig. 3D) grouped most species according to their taxonomic affinities. Three legumes joined a clade, with *Phaseolus* and *Medicago* in

the same subclade together with soybean in a separate subclade, what may also reflect adaptive features. Beans and alfalfa are more tolerant to drought than to saline soils [32] and also share the ability to develop in soils with high levels of aluminum [33], while soybean appear to be more tolerant to salinity than to drought [34].

Arabidopsis and sesame also joined a clade being both characterized by their ability to grow in arid and semi-arid deserts [35]. Rice and maize are annual grasses that need high temperatures, both sensible to abiotic factors as day length for maize and availability of flooded environments in the case of rice [36]. As for other osmoprotectants, the eucalyptus position near the fungus *Giberella zeae* was unexpected, again confirming their slow evolving manner as woody most primitive plant group, as compared with all other herbaceous species. It is interesting to note that this fungus presents tolerance to heavy metals, a feature also present in *Neurospora crassa* [37].

5 Concluding Remarks

The present study allowed the identification of all procured osmoprotectants in eucalyptus (FOREST database). Nevertheless different aspects of structure, abundance, function and pattern of expression in all tissues could be observed and reported by the first time.

Osmoprotectant sequences were identified in low levels in most tissues, but were more abundant (and co-expressed) in libraries produced under abiotic stresses, providing important evidences for physiological studies, expression assays and germplasm evaluation by means of quantitative real-time PCR (TaqMan) assays.

Dendrograms generated using eucalyptus osmoprotectants, as compared to other organisms bearing procured conserved domains revealed similarities to plants adapted to different kinds of abiotic stresses and also to model plants, abundant in public databases. Mostly divergent patterns in eucalyptus sequences were also identified, probably due to the prevalence of herbaceous (most derived) plants in databases, as compared with the woody (primitive) habit and taxonomic position of eucalyptus, indicating urgent need of more information regarding basal Angiosperm and also other primitive plants (especially that adapted to dry environments) to clarify structure, functionality and evolutionary patterns regarding osmoprotectants.

Acknowledgments

The authors thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FACEPE (Fundação de Amparo à Pesquisa do Estado de Pernambuco) for the concession of fellowships. To FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and FORESTs coordination we thank for the access to the *Eucalyptus* EST data bank.

References

1. Lemcoff, J., Guarnaschelli, A., Garau, A., Bascialli, M., Ghersa, C.: Osmotic adjustment and its use as a selection criterion in eucalyptus seedlings. Canadian journal of forest research **24**(12) (1994) 2404–2408
2. Ladiges, P.: Variation in drought tolerance in eucalyptus viminalis labill. Australian Journal of Botany **22**(3) (1974) 489–500
3. Paston, G., Foyer, C.: Common components, networks, and pathways of cross-tolerance to stress. the central role of âĂŸâĂŸredoxâĂŹâĂŹ and abscisic acid-mediated controls. Plant Physiology **129** (Year) 460–468
4. Elbein, A., Pan, Y.T., Pastuszak, I., Carroll, D.: New insights on trehalose: a multifunctional molecule. Glycobiology **13** (2003) 17–27
5. Hegeman, C., Good, L., Grabau, E.: Expression of d-myo-inositol-3-phosphate synthase in soybean. implications for phytic acid biosynthesis. Plant Physiology **125** (2001) 1941–1948
6. Hu, C.A., Delauney, A., Verma, D.: A bifunctional enzyme (delta 1-pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants. Proceedings of the National Academy of Sciences of the United States of America **89**(19) (1992) 9354–9358
7. Sirko, A., Blaszczyk, A., Liszewska, F.: Overproduction of sat and/or oastl in transgenic plants: a survey of effects journal of experimental botany. Sulphur Metabolism in Plants Special Issue **55**(404) (2004) 1881–1888
8. Sakamoto, A., Murata, N.: The role of glycine-betaine in the protection of plants from stress: clues from transgenic plants. Plant Cell Environment **25**(2) (2002) 163–171
9. Vicentini, R., Sassaki, F., Gimenes, M., Maia, I., Menossi, M.: In silico evaluation of the eucalyptus transcriptome. Genetic and Molecular Biology **28**(3) (2005) 487–495
10. Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D.: Basic local alignment search tool. Journal of Molecular Biology (1990)
11. Thompson, J., Higgins, D., Gibson, T.: Clustalx, multiple sequence alignment program version 1.63b (1997)
12. Tamura, K., Dudley, J., Nei, M., Kumar, S.: Mega4: Molecular evolutionary genetics analysis (mega) software version 4.0. Molecular Biology and Evolution **24** (2007) 1596–1599
13. Eisen, M., Spellman, P., Brown, P., Botstein, D.: Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America **95** (1998) 14863–14868
14. Page, R.: Treeview program version 161. Computer Applications in the Biosciences **12** (1996) 357–358
15. da Silva, A.B., Wanderley-Nogueira, A., Silva, R., Berlarmino, L., Soares-Cavalcanti, N., Benko-Iseppon, A.: In silico survey of resistance (r) genes in eucalyptus transcriptome. Genetic Molecular Biology **28**(3) (2005) 562–574
16. Hell, R., Wirtz, M., Berkowitz, O., Droux, M.: The cysteine synthase complex from plants. mitochondrial serine acetyltransferase from arabidopsis thaliana carries a bifunctional domain for catalysis and protein-protein interaction. Europe Journal of Biochemistry **268** (2001) 686–683
17. Müller, J., Aeschbacher, R., Wingler, A., Boller, T., Wiemken, A.: Trehalose and trehalase in arabidopsis. Plant Physiology **125** (2001) 1086–1093

18. Verbruggen, N., Villarroel, R., Montagu, M.: Osmoregulation of a pyrroline-5-carboxylate reductase gene in arabidopsis thaliana. Plant Physiology **103** (1993) 771–781
19. Marsh, N., Adams, M.: Decline of eucalyptus tereticornis near bairnsdale, victoria: insect herbivory and nitrogen fractions in sap and foliage. Australian Journal of Botany **43** (1995) 39–50
20. Adams, M., Richter, A., Hill, A., Colmer, T.: Salt tolerance in eucalyptus spp.: identity and response of putative osmolytes. Plant **and Environment** (2005) 772–787
21. Hazen, S., Wu, Y., Kreps, J.: Gene expression profiling of plant responses to abiotic stress. Journal Functional Integrative Genomics **3**(3) (2003) 105–111
22. Lambais, M.: *In silico* differential display of defense-related ests from sugarcane tissues infected with diazothrophic endophytes. Genetics and Molecular Biology **24**(1-4) (2001) 103–111
23. Initiative, T.A.G.: Analysis of the genome sequence of the flowering plant arabidopsis thaliana. Nature **408**(6814) (2000) 796–815
24. Weston, D., Gunter, L., Rogers, A., Wullschleger, S.: Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants. BMC Systems Biology **2** (2008) 1–17
25. Benko-Iseppon, A., Soares-Cavalcanti, N., Wanderley-Nogueira, A., Berlarmino, L., Silva, R., Almeida, P., Brunelli, K., Kido, L., Kido, E.: Genes associated to biotic and abiotic stresses in cowpea [vigna unguiculata (l.) walp.] and other angiosperms. In R.J.M.C Nogueira, E.L. Araújo, L.W.U.C.E., ed.: Environmental Stresses: damages and benefits to plants, UFRPE University Press Recife (2005) 350–359
26. Ishitani, M., Arakawa, K., Mizuno, K., Kishitani, S., Takabe, T.: Betaine aldehyde in leaves of both betain-accumulating and non-accumulating cereal plants. Plant and Cell Physiology **34** (1993) 493–495
27. Kawashima, C., Berkowitz, O., Hell, R., Noji, M., Saito, K.: Characterization and expression analysis of a serine acetyltransferase gene family involved in a key step of the sulfur assimilation pathway in arabidopsis. Plant Physiology **137** (2005) 220–230
28. Freeman, J., Salt, D.: The metal tolerance profile of thlaspi goesingense is mimicked in arabidopsis thaliana heterologously expressing serine acetyl-transferase. BMC Plant Biology **7** (2007) 63
29. Liu, F., Byung-Chun, Y., Jung-Youn, L., Pan, W., Harmon, A.: Calcium-regulated phosphorylation of soybean serine acetyltransferase in response to oxidative stress. The Journal of Biological Chemistry **281**(37) (2006) 27405–27415
30. Saito, K.: Isolation and characterization of a cdna that encodes a putative mitochondrion-localized isoform of cysteine synthase (o-acetylserine(thiol)-lyase) from spinacia oleracea. The Journal of biological chemistry **269**(28) (1994) 187–192
31. Liszewska, F., Gaganidze, D., Sirko, A.: Isolation of nicotiana plumbaginifolia cdnas encoding isoforms of serine acetyltransferase and o-acetylserine (thiol) lyase in a yeast two-hybrid system with escherichia coli cyse and cysk genes as baits. Acta Biochimica Polonica **52**(1) (2005) 117–128
32. González, E., Ladrera, R., Larrainzar, E., Arrese-Igor, C.: Response of Medicago truncatula to drought stress. In: Medicago truncatula handbook. (2007) 1–16
33. Chandran, D., Sharopova, N., Ivashuta, S., Gantt, J., VandenBosch, K., Samac, D.: Transcriptome profiling identified novel genes associated with aluminum toxicity, resistance and tolerance in medicago truncatula. Planta **228** (2008) 151–166

34. Waldren, R., Teare, I.: Free proline accumulation in drought-stressed plants under laboratory conditions. Plant and Soil **40**(3) (1974) 689–692
35. Weiss, E.: Oilseed crops. LOngman, London (1983)
36. Yoshida, K., Wada, T., Koyama, H., Mizobuchi-Fukuoka, R., Naito, S.: Temporal and spatial patterns of accumulation of the transcript of myo-inositol-1-phosphate synthase and phytin-containing particles during seed development in rice. Plant Physiology **119**(1) (1999) 65–72
37. Kiranmayi, P., Mohan, P.: Metal transportome of neurospora crassa. In Silico Biology **6**(3) (2006) 169–180

SUPPLEMENTARY DATA

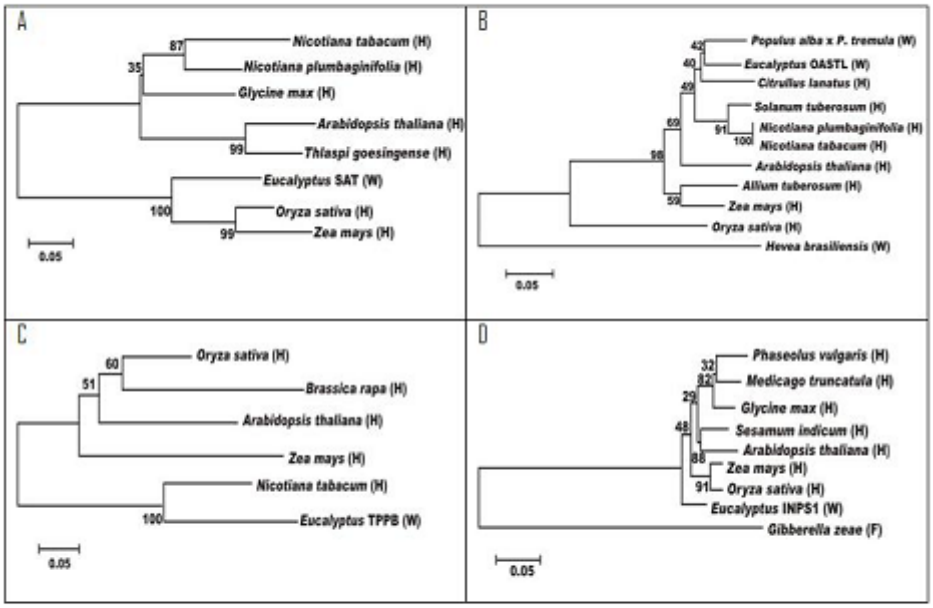


Fig. 3: Dendrograms generated after Neighbor Joining analysis, showing the relationships between the consensus sequences of *A. thaliana*, *Eucalyptus* and other organisms regarding protein sequences: (A) SAT; (B) OASTL; (C) TPPB and (D) INPS1. Bar represents genetic distance, while numbers in the base of clades indicate bootstrap values. The letters in parenthesis mean herbaceous (H) and woody (W) plants or fungi (F).

Capítulo 3

Artigo Científico

In Silico Analysis of Osmoprotectants in Cowpea [*Vigna unguiculata* (L.) Walp.] and Sugarcane (*Saccharum* spp.) Transcriptomes

Artigo a ser enviado para a revista *Genetics and Molecular Research*.

In Silico Analysis of Osmoprotectants in Cowpea [*Vigna unguiculata* (L.) Walp.] and Sugarcane (*Saccharum* spp.) Transcriptomes

Petra Barros dos Santos, Nina M. Soares-Cavalcanti, Gabriela S. Vieira-de-Mello, Valesca Pandolfi, Tercílio Calsa-Junior, Ederson Akio Kido and Ana M. Benko-Iseppon

Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Genética, Laboratório de Genética e Laboratório de Genética Molecular e Biotecnologia Vegetal, Recife, PE, Brazil.

Short running title: Osmoprotectant genes in Cowpea and Sugarcane.

Key words: data mining, salt stress, drought stress, crop evolution, bioinformatics.

Corresponding Author:
Ana Maria Benko-Iseppon, UFPE/CCB/Genética, Laboratório de Genética e Biotecnologia Vegetal, Av. Prof. Moraes Rego, s/ nº; 50732-970, Recife, PE, Brazil. E-mail: ana.benko.iseppon@pesquisador.cnpq.br

Abstract

Environmental stresses such as drought and salinity limit crop productivity in worldwide level. These stresses often lead to the accumulation of compatible osmolytes in most organisms, including plants. Such compounds stabilize cellular macromolecules, reducing water loss, both regarding to the integrity of cell membranes, and the activity of some enzymes. In the present work, osmoprotectants coding gene sequences were selected and used to search for ortologs in the cowpea and sugarcane transcriptome databases using *in silico* procedures. The analyzed sequences were genes whose products are directly responsible for biosynthesis of the osmoprotectant metabolites trehalose, glycine-betaine, *myo*-inositol, proline and cysteine. Alignments revealed that the expression of these genes is quite abundant in higher plants, with 73 clusters found in cowpea database and 56 clusters found in sugarcane, presenting high homology with known osmoprotectant genes. *In silico* expression in sugarcane database revealed that higher expression was in callus tissues and in those infected by *Herbaspirilum rubrisubalbicans* (HR) as well as shoot-root transition zone in plantlets (RZ), confirming the multi-function character of the osmoprotectants that may be active in abiotic and biotic stresses. In cowpea database, most reads were found in the root library from a plant sensitive to salinity after two hours of salt-stress, confirming that these genes are recruited in abiotic stresses such as salinity. As expected, the phylogenetic analysis revealed distinct groups between angiosperms, algae, animals, fungi and bacteria, in almost all dendrograms, revealing that the genic structure was relatively conserved in angiosperms (including cowpea and sugarcane), as also observed in comparative analysis between the ORFs of these cultures and other organisms, suggesting that the accumulation of compatible solutes is a metabolic adaptation found in a great number of organisms, probably regarding very ancient genetic processes.

INTRODUCTION:

Cowpea is one of the most important legume crop in the semi-arid tropics covering Asia, Africa, southern Europe and Central and South America. It is a dicotyledonous belonging to the family Fabaceae, recognized as a drought tolerant and warm weather crop. Cowpea is well-adapted to the drier regions of the tropics, where other food legumes do not perform well. These beans are mostly consumed by rural family farmers, and have multipurpose uses, like food for human beings or fodder for livestock and atmospheric nitrogen fixation (Pengelly and Maass, 2001).

Brazil is the third world producer of cowpea with 11.5 million hectares planted and an average yield of 317 $kg.ha^{-1}$. It is cultivated in North and Northeast regions where it has great economic importance as a source of employment, appearing as one of the main components of the daily food in those regions (Freire-Filho et al., 2005; Simon et al., 2007).

Despite its social and economic importance in the world, cowpea research has received relatively little attention and remains to a large extent an underexploited crop. Among the major goals of cowpea breeding and improvement programs is the stacking of desirable agronomic traits, such as those governing abiotic stress like drought, salinity, heat tolerance, plant growth type, and seed quality with resistances to the numerous biotic stresses like bacterial, fungal, and viral diseases (Singh, 2005; Timko et al., 2007).

The second studied organism, sugarcane, belongs to the family Poaceae (Monocotyledoneae) and is cultivated mainly in tropical and subtropical areas, including more than 80 countries around the globe, being considered one of the world's

most important crop plants. The high sucrose accumulation in the stalk contributes with 60% of the raw sugar production worldwide (Takahashi et al., 2005).

Among the cultivated crops, sugarcane possesses perhaps one of the most complex genomes, since it is thought that *S. officinarum* was originally selected by humans in Papua New Guinea, perhaps from *S. robustum* germplasm and derived hybrids (Grivet and Arruda, 2002). Because of its multispecies origin, sugarcane carries a variable chromosome numbers (generally 2n=70-120) with a commensurately large DNA content, which imposes tremendous difficulties in applying conventional plant breeding techniques (Lu et al., 1994).

Due to the great importance of these crops worldwide it is important have domain of the environmental factors that influence the development of plants. Abiotic stresses greatly influence the growth and development of crop plants. Among several environmental stresses, drought and salinity are the most important factors influencing the growth, survival, yield and natural distribution of plants worldwide (Verslues et al., 2006). They are united by the fact that at least part of their detrimental effect on plant performance is caused by disruption of suitable plant water status. This can cause cellular damages, such as osmotic and oxidative harm (Wang et al., 2003), decrease in the availability of essential nutrients in the soil and often conversely the build-up of toxic ions in the plant (Verslues et al., 2006).

Plants have developed mechanisms like stress avoidance to cope with low water content. Strategies include formation of seeds before drought conditions prevails, morphological adaptations like decreasing transpiration rate, development of specialized leaf surfaces, reduction of leaf area, reduced stomatal conductance or an increase in root length and density to achieve a more efficient water usage (Ramanjulu and Bartels, 2002). As a complex trait, water stress tolerance appears to be the result of coordination

of biochemical and physiological alterations at cellular and molecular level: i.e. the accumulation of late embryogenesis abundant - LEA gene products (Ramanjulu and Bartels, 2002), activation of aquaporins (Javot and Maurel, 2002) and synthesis of osmoprotectant (Yancey et al., 1982) proteins and metabolites. It is important to note that such mechanisms occur both in dehydration-tolerant and non-tolerant plants (Ramanjulu and Bartels, 2002).

The synthesis of compatible osmolytes (osmoprotectants) is one of the main mechanisms that organisms, including plants, have to prevent the harmful effects, caused by abiotic stresses. These molecules are small, electrically neutral and nontoxic at molar concentrations (Yancey, 1994). The accumulation of these substances results in an increase in cellular osmolarity that can drive water influx or reduce its efflux. This provides the turgor that is necessary for cell expansion (Kishor et al., 2005).

Osmoprotectants do not interfere with cellular function and are termed compatible solutes because they can accumulate to high levels without perturbing with metabolism and may also have protective properties (Yancey et al., 1982). They stabilize proteins and membranes against the denaturing effect of high concentrations of salts and other harmful solutes, facilitating the retention of water in the cytoplasm, and raising cellular osmotic pressure and the protection of membranes, protein complex, and cellular constituents (Cherian et al., 2006).

Many organisms, including higher plants, accumulate proline (Pro) in response to environmental stress. This amino acid is not charged at neutral pH and is highly soluble in water. Moreover, at high concentrations, it has no negative effect on macromolecule–solvent interactions (Yancey, 2001). Proline seems to have diverse roles under osmotic stress conditions, such as stabilization of proteins, membranes and subcellular structures (Vanrensburg et al., 1993), protection of cellular functions by

scavenging ROS (Reactive oxygen species; Bohnert and Shen, 1999), as a sink for energy (Phang, 1985), and even as a stress-related signal (Werner and Finkelstein,1995).

The accumulation of proline in plants during water and salt stress is guided by the simultaneous up-regulation of Δ^1-*pyrroline-5-carboxylate synthetase* (*P5CS*) and down-regulation of the *proline dehydrogenase* (*ProDH*) responsible for the proline degradation (Hu et al., 1992; Yoshiba et al., 1997). According to Delauney et al. (1993) the proline is synthesized from L-glutamic acid (l-G1u) via Δ^1-pyrroline-5-carboxylate (P5C) by two enzymes, P5CS (the rate-limiting enzyme in proline biosynthesis) and P5CR (P5C reductase).

Glycinebetaine (GB) is a quaternary ammonium compound that occurs naturally in a wide variety of microorganisms, animals and plants (Jagendorf and Takabe 2001). In almost all cases, it is synthesized by a two-step oxidation: the first mediated by a ferredoxin-dependent choline monooxygenase (CMO) and the second by the enzyme betaine aldehyde dehydrogenase (BADH). Both plant enzymes are chloroplastic (Rontein et al., 2002).

GB appears to be a critical determinant of stress tolerance in plants (Alia et al., 1998). Possible roles for this substance include stabilization of complex proteins and membranes in vivo, protection of the transcriptional and translational machinery and intervention as a molecular chaperone in the refolding of enzymes (Sakamoto and Murata, 2001). In addition, GB might protect electron transport via complex II in mitochondria (Hamilton and Heckathorn, 2001) and might reduce the peroxidation of membrane lipids (Chen et al., 2000)

Trehalose is a disaccharide of glucose, naturally occurring widespread throughout the biological world. It is synthesized in two steps from glucose 6-phosphate

and uridine diphosphoglucose, via trehalose 6-phosphate. The first step is mediated by trehalose 6-phosphate synthase (TPS) and the second by trehalose-6-phosphate phosphatase (TPP) (Rontein et al., 2002). Biochemical studies have shown that trehalose stabilizes proteins and membrane lipids (Ramanjulu and Bartels, 2002) and protect biomolecules against the adverse effects from environmental stresses such as heat, cold, desiccation and anoxia (Penna, 2003), by replacing water through hydrogen bonding to polar residues, trehalose prevents the denaturation of proteins and the fusion of membranes (Wingler, 2002).

Myo-inositol and its derivatives are commonly studied with respect to cell signaling and membrane biogenesis, but they also participate in stress responses in plants and animals, particularly to salt stress, contributing to lower the cytoplasm osmotic potential. Myo-inositol-1-phosphate is formed from glucose-6-phosphate, through an internal reaction of oxyreduction involving NAD^+. The reaction is catalyzed by the enzyme myo-inositol-1-phosphate synthase, which is encoded by *INPS1* gene, with subsequent dephosphorylation in myo-inositol and orthophosphate (Majumder et al., 1997). In some plants, inositol provides substrate for the production and accumulation of *ononitol* and *pinitol* (sugar alcohols) that help to balance sodium accumulation in vacuoles (Nelson et al., 1998).

The cysteine (Cys) synthesis exists in several subcellular compartments as cytosol, plastids and mitochondria. The biosynthesis of Cys is regarded as the exclusive function of sulfur reduction in plants, being a key limiting step in the production of glutathione, a thiol implicated in various cellular functions, including sulfur transport, gene expression, scavenging of ROS and resistance to biotic and abiotic stresses (Youssefian et al., 2001). In a unique mechanism, cysteine synthesis proceeds in two sequential steps catalyzed by serine acetyltransferase (SAT), which converts serine in

O-acetylserine, and O-acetylserine (thiol) lyase (OAS-TL), which catalyzes the formation of Cys from sulfide and O-acetylserine (Campanini et al., 2005).

Many works revealed the role of Cys in osmotic adjustment during abiotic stress. In experiments with leaves from *Arabidopsis thaliana* exposed to NaCl, Romero et al. (2001) showed that the cytosolic isoform of OASTL was increasingly induced and that the level of this enzyme increased gradually for more five days during exposure to salt. The authors suggested that the high rate of cysteine biosynthesis was necessary and essential for the protection and/or adaptation of the plants to higher levels of Na^+, including this amino acid among osmoprotectants in plants kingdom. Furthermore, Ruiz and Blumwald (2002) in experiments with *Brassica napus* observed that the levels of SAT and OASTL increased significantly as a result of exposure to 150 mM NaCl during 15 days.

A key to progress towards breeding better crops under abiotic stress has been to understand the changes in cellular, biochemical and molecular machinery of the plants. It is believed that osmoregulation would be the best strategy for abiotic stress tolerance, especially if osmoregulatory genes could be triggered in response to drought, salinity and high temperature (Bhatnagar-Mathur et al., 2008). To date, various approaches in genetic engineering have allowed the introduction of new pathways for the biosynthesis of various compatible solutes in plants, like glycine betaine, proline, polyamine and trehalose (Chen and Murata, 2002; Cherian et al., 2006).

Several projects aiming to generate EST databases have been developed with the aim to produce comprehensive evidences for breeding purposes in pulse crops like cowpea and sugarcane. This is the case of the project "Functional, structural and comparative genomics of cowpea (*V. unguiculata*) (htpp://www.gentrop.ufpe.br/vigna) that started in 2004, aiming to identify new candidate genes from 100,000 expressed

tags from contrasting genomic libraries including tolerant and sensitive plants and in the presence and absence of different biotic and abiotic stress conditions and tissues (Benko-Iseppon, 2001).

Another EST project that provided cowpea expressed sequences is the HARVEST database, where sequences of many organisms are available, like: citrus, coffee, cowpea, soybean, rice and wheat. This database displays 17 EST libraries from *V. unguiculata* and was developed by the US Department of Energy Joint Genome Institute (141,050 ESTs), the J. Craig Venter Institute (41,505 ESTs) and the University of California Riverside (488 ESTs) and are available to researches by the scientific community worldwide (Harvest, 2008).

The Sugarcane Expressed Sequence Tag (SUCEST) database currently has 237,954 ESTs, which were obtained from 37 cDNA libraries prepared from different tissues and conditions. After assembling, the 237,954 sugarcane expressed sequences resulted in 43,141 clusters (Vettore et al., 2001). Regarding sugarcane osmoprotectants, little information is available for this crop.

The present work performed a data-mining based identification of osmoprotectant genes in the sugarcane and cowpea databases, using well known rice and arabidopsis sequences as templates, and comparing the obtained clusters with sequences from public databases and literature data. Furthermore, the expression profile of the different osmoprotectant categories was evaluated, aiming to identify their role under different conditions in both crops.

MATERIAL AND METHODS

Full length cDNA of *P5CS, P5CR, TPP, TPS, BADH, CMO* and *INPS1* from arabidopsis and rice were obtained at GenBank, from (National Center for

Biotechnology Information; http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and TIGR (The Institute for Genomic Research) databases, respectively (Table 1). Each sequence was compared against the cowpea and SUCEST databases using tBLASTn tool and sequences with e-value equal to e^{-05} or less were annotated.

The selected cowpea sequences were obtained with the assembling of NordEST (18.948) and HARVEST (183.118), both ESTs databases. This assembly was generated to avoid possible further redundancy of results after BLAST. Reads trimming for sequence quality, vector contamination and clustering was performed by using the online version of EGAssembler (http://egassembler.hgc.jp/), with CAP3 default parameters.

Sugarcane and cowpea clusters were translated using the ORF Finder (Open Reading Frame Finder) tool at NCBI and screened for conserved domains using RPS-BLAST CD-search tool (Altschul et al., 1990). Multiple alignments with proteins from arabidopsis, sugarcane, cowpea and other organisms that presented complete domains were generated at the CLUSTALx program (Thompson et al., 1997). The Blink function, an option of GenBank database, was used to obtain similar sequences from other organisms. These multiple alignments allowed the structural analysis of the sequences including conserved and diverging sites. To avoid influence of sequence sizes in the alignments, the non aligned 5' and 3' extremities were excluded, as well as autapomorphic, non informative internal sequence regions. A phylogenetic maximum parsimony analysis using bootstrap function with 2,000 replications was performed, generating a consensus tree with a cut-off of 50 (from 50% more parsimonious trees) using the program MEGA (Molecular Evolutionary Genetic Analysis), Version 4 for Windows (Kumar et al., 2004). For each cladogram the most basal organisms have been manually settled as outgroup, in all cases non angiosperms.

The prevalence of cowpea and sugarcane clusters were verified by direct counting of the reads that composed each cluster, followed by data normalization (considering the total number of reads sequenced in each library) and calculation of the relative frequency (reads per library). To generate an overall picture of gene expression patterns in sugarcane, a hierarchical clustering approach was applied using normalized data, from all identified osmoprotectant transcripts, allowing the generation of a graphic representation with aid of the CLUSTER program (Eisen et al., 1998). The resulting dendrograms including both axes (using the weighted pair group for each gene class and library) were generated by the TreeView program for Windows (Page, 1996). On graphics yellow color indicated no expression and red color means top high of expression.

RESULTS

Results regarding the here analyzed osmoprotectant categories are resumed in Table 2, including cowpea and sugarcane candidates as well as data from best matches from other species. The evaluation of conserved domains and shared features with other taxa are resumed in Table 3.

Cowpea and Sugarcane orthologs

Proline

After trimming redundant clusters, the search for P5CS and P5CR orthologs in cowpea database revealed four and one candidates, respectively. The tBLASTn approach showed high degree of similarity, with best e-value for P5CS ranging from e^{-128} to e^{-73} and the P5CR candidate with an e-value e^{-110}. All clusters involved in the

proline biosynthesis had best matches with their respective protein after BLASTx analysis at the GenBank; the P5CS and P5CR clusters showed similarity with *V. unguiculata* sequences, with exception of the first match from P5CS that had most significant similarity with pyrroline-5-carboxylate synthetase protein from *Phaseolus vulgaris* (Table 2). All *P5CS* candidates obtained from cowpea presented the AA_kinase and ProA conserved domains (CD) incomplete. The same was observed for the *P5CR* candidate, where an incomplete ProC domain was found (Table 3).

Regarding sugarcane database, searches for P5CS and P5CR candidates revealed four clusters to *P5CS* and one to *P5CR* gene. The clusters for P5CS had high degree of similarity with e-values ranging from 0.0 to e^{-49}, while P5CR candidate presented an e-value of e^{-93}. After BLASTx at GenBak the candidates to P5CS gene showed best matches with Poaceae family, except one that presented best alignment with tomato (*Lycopersicon esculentum*, Solanaceae family), while the P5CR sequence showed similarity with the respective protein from *Zea mays*. Considering the searched CDs, the best hit from P5CS presented both complete CDs (AA_kinase and ProA) while the other candidates presented partial CDs. Regarding P5CR candidate the ProC domain was complete.

Glycine-betaine

After tBLASTn with the BADH seed sequence in both databases the results revealed 43 clusters in the cowpea database with e-value varying from 0.0 to e^{-05} and 24 sequences with e-values ranging from 0.0 to e^{-11} at SUCEST. However after reverse alignment (BLASTx) the results revealed that not all candidates to BADH were in fact similar to their respective proteins in GenBank, due to common shared domains. So, only seven cowpea sequences and five sugarcane sequences were in fact similar to

BADH (Table 2). Moreover, the identified cowpea sequences had best similarity with soybean (*Glycine max*) in almost cases and all five BADH candidates to sugarcane had similarity with Poaceae members (two with *Z. mays*, two with *Oryza sativa* and one with *Zoysia tenuifolia*). Regarding the similarity to CMO candidates after tBLASTn at Cowpea and SUCEST databases, the results showed two candidates in both projects. The first hit at the cowpea database showed high degree of similarity (e^{-64}) as well as the first hit at the sugarcane database (e^{-119}). After BLASTx the results confirmed that these clusters were similar to choline monooxigenase proteins. In the case of the best hit at cowpea (NordEST database), the similarity was with *O. sativa* (gi: 33300598, e-value e^{-64}) and in the case of the best hit at SUCEST the similarity was with *Z. mays* (gi: 112790161, e-value e^{-180}).

Considering the searched CDs in cowpea data bank, one cluster from BADH was identified, presenting the ALDH complete CD and one of the two candidates to CMO had only the Rieske_CMO CD complete, while the other HcaE was absent. In the case of sugarcane candidates to BADH two had complete CDs while the CMO clusters presented only a partial HcaE domain (Table 3).

Trehalose

In relation to cowpea TPS and TPP candidates 19 and 11 clusters were obtained. After tBLASTn of TPS clusters candidates with an e-value from 0.0 to e^{-06} were observed, with reverse alignments confirming that all 19 sequences were similar to trehalose-6-phosphate synthase proteins from dicotyledonous plants. Regarding the 11 TPP sequences, the e-value ranged from e^{-122} to e^{-06}, and after BLASTx all clusters belonged to the procured category, being more similar to arabidopsis and *Nicotiana tabacum*.

The search for orthologs, to *TPS* and *TPP* genes in SUCEST database revealed the presence of several clusters with high degree of similarity; 17 and 11 candidates (respectively) were obtained with e-values ranging from e^{-114} to e^{-20}. After reverse alignment most clusters showed similarity to the respective protein from *O. sativa* while only two sequences from TPS were more similar to arabidopsis.

Regarding the CD analysis of cowpea sequences, only one TPS sequence had the searched domains glucosyltransferase-like and HAD-like, although the first was incomplete. Four other sequences had the complete HAD-like domain. The same was observed in sugarcane results, where only one sequence had both domains, although the Glico_transferase was incomplete, while all other five sequences had the complete HAD CD. Two of the TPP candidates in cowpea beard the complete HAD domain, while four presented no CD and the other had incomplete CDs. The same search in sugarcane database revealed three candidates with the complete HAD-like domain.

Cysteine

With respect to *OASTL* and *SAT* genes, our evaluation in cowpea revealed 20 and six clusters, respectively, with orthologs mainly from soybean (*G. max*) after reverse alignment. The prediction of clusters coding regions revealed that ORFs were oriented in both forward and reverse reading frames, with an average of 238 amino acids (aa) in length for OASTL and 292 aa for SAT. In relation to the conserved domains obtained for OASTL, we identified five sequences with a complete PALP domain. Regarding SAT, three sequences presented the domains Satase_N and LbetaH complete while in one, only the Satase_N was complete with the other domain being absent.

The *OASTL* data mining in sugarcane database resulted in 10 orthologs, while the search for SAT showed four sequences. The tBLASTn results for *OASTL* revealed in most cases great similarity with the used seed sequence (OASTL: 12003.m06618) with e-values ranging from e^{-138} to e^{-06}. After BLASTx analysis the identity of all obtained sequences from OASTL and SAT clusters was confirmed, associating them with proteins mainly from *O. sativa*.

The ORF sizes varied from 377 (cluster SCCCLR1065D04 for OASTL) to 100 aa and 318 (cluster SCCCLR1072E08 for SAT) to 145 aa. Regarding the average ORF length for each gene, we observed 243 aa for *OASTL* and 268 aa for *SAT*. Considering the CDs in OASTL candidates, four clusters beard the complete PALP domain, while two sequences presented no CD. Regarding SAT results, two clusters had both complete procured domains (Satase_N and LbetaH).

Myo-inositol

The search for similar sequences to *INPS1* gene revealed four clusters at cowpea transcriptome project, with sequences ranging from 629 to 1967 nt and e-values from e^{-37} to 0.0. The best aligned cluster presented similarity to *V. radiata* and after ORF translation (510 aa) a complete Inos_1-P_synth domain was found, while the other sequences had partial domains.

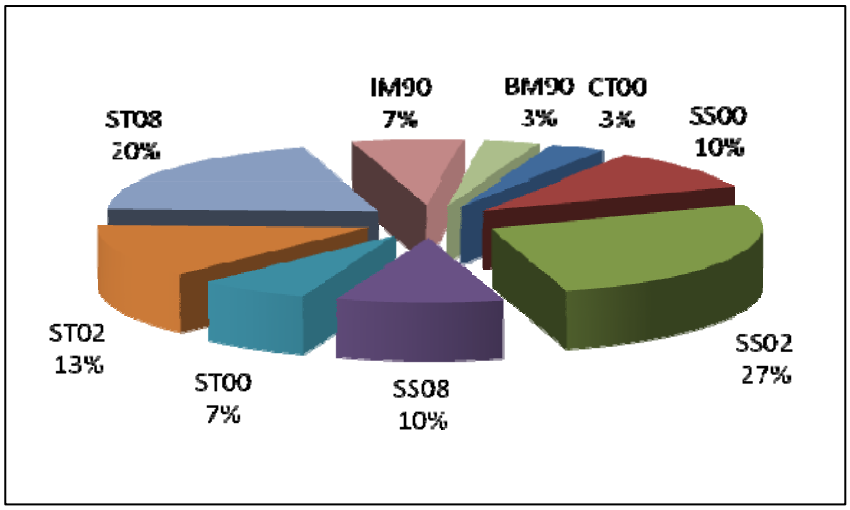
After tBLASTn in SUCEST two *INPS1* candidates were identified with the best hit presenting an e-value of 0.0. The reverse alignments revealed that these two sequences were similar to myo-inositol 1-phosphate synthase protein from *Z. mays*. After RPS-BLAST the procured domain Inos_1-P_synth was complete in the first best hit, being absent at the other sequence.

Expression pattern analysis

Cowpea

The expression profile of cowpea clusters was carried out exclusively with the NordEST database, because many HARVEST libraries (60%) were constructed with mixture of tissues from *V. unguiculata,* preventing localization of specific tissues or situations in which the transcripts of compatible osmolytes were recruited.

Regarding the reads of the NordEST database, the direct counting revealed the presence of 33 transcripts considering all studied genes that participate of osmoprotectant synthesis. Almost all reads were from SS02 library (roots of salinity-sensitive plants after 2 hours of salt-stress) followed by ST08 library (roots of salinity-tolerant plant after 8 hours of salt-stress).



(Figure 1): Predominance of cowpea osmoprotectant genes in the NordEST libraries, showing general distribution of transcripts per library. Abbreviations for libraries: **BM90**=leaves of BR14-Mulato (resistant against cowpea mosaic virus) harvested 30, 60 and 90 minutes after mosaic viruses infection; **IM90**=leaves of IT85F plant (susceptible against this virus) harvested 30, 60 and 90 minutes after mosaic viruses infection; **SS00**=root of salinity sensitive plant without salt stress; **SS02**=root of salinity sensitive plant after 2 hours of stress with 200 mM NaCl; **SS08**=root of salinity sensitive plant after 8 hours of stress with 200 mM NaCl; **ST00**=root of salinity tolerant plant without salt stress; **ST02**=root of salinity tolerant plant after 2 hours of stress with 200 mM NaCl; **ST08**=root of salinity tolerant plant after 8 hours of stress – 200 mM NaCl.

Sugarcane

In a general view, after direct counting of the reads, it was possible to observe that all SUCEST libraries presented at least one osmoprotectant, with exception of a complete absence in the sugarcane NR library (that included all tissues). Interestingly, the candidate clusters to *CMO* and *P5CR* presented the lowest number of reads in sugarcane, with *CMO* being represented only in RT (roots), LR (young leaves) and HR (tissues infected with *Herbaspirilum rubrisublbicans*) libraries, while *P5CR* was expressed in FL (flowers), LR and RZ (root to shoot zone of young plants).

Considering the distribution of the 483 reads (comprised in the 56 clusters selected) from 14 analyzed tissues (comprising 29 libraries), a higher prevalence could be observed in flower tissues (FL=20%), roots and stem-root transition (RZ and RT=12%) followed by seeds (SD=9%) and apical meristem (AM=8%) (Figure 2A).

Regarding correlation of the distribution among reads and osmoprotectant categories, it is clear that reads involved in the trehalose pathway (TPS and TPP) were most abundant in the SUCEST libraries, with 175 reads (26,08% for TPS and 10,14% for TPP), followed by OASTL and SAT (both involved in the cysteine pathway) with 143 reads (22,36% for OASTL and 7,24% for SAT), and by BADH and CMO (both involved in the glycine-betaine pathway) with 75 reads (14,28%) for BADH and only five reads from CMO clusters comprising (1,03%) (Figure 2B).

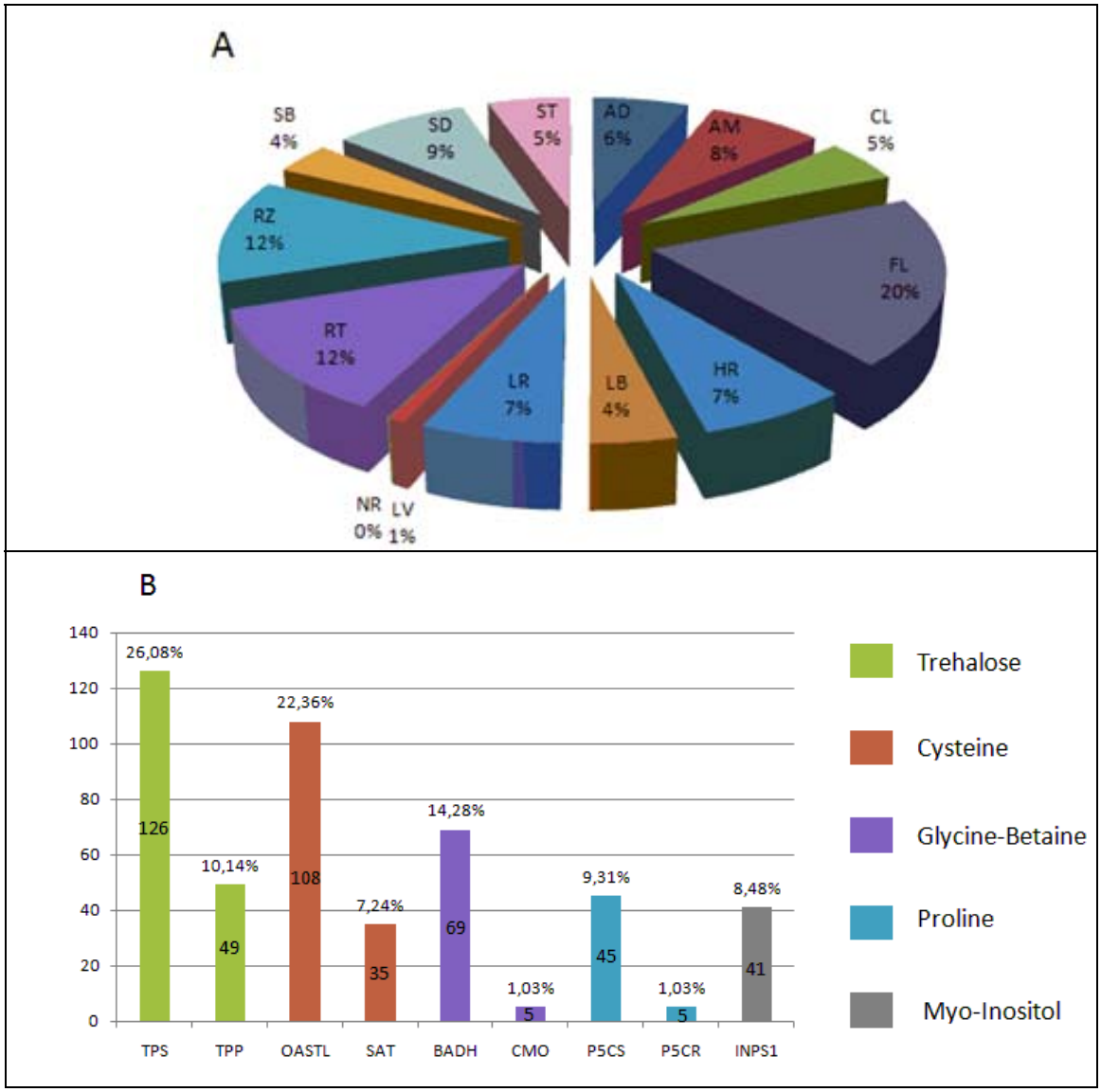


Figure 2: (**A**) General distribution of all osmoprotectant transcripts in the SUCEST libraries. (**B**) Prevalence of reads per osmoprotectant category. Numbers outside columns refer to the percentage of reads that compose each osmoprotectant gene category; Numbers inside columns refer to the absolute number of reads found. Legend for library codes in (A): **AD:** tissues infected by *Gluconacetobacter diazotroficans*, **AM:** Apical meristem; **CL:** Callus; **FL:** Flower; **HR:** tissues infected with *Herbaspirillum rubrisubalbicans*; **LB:** Lateral Bud; **LR:** Leaf Roll; **LV:** Leaves; **NR:** All tissues normalized; **RT**: Root; **RZ:** Stem-Root transition; **SB:** Stalk Bark; **SD:** Seeds; **ST:** Stem.

In silico expression approach for sugarcane using all here identified osmoprotectant transcripts (Figure 3) revealed a higher expression in calli tissues submitted to light/dark and temperature (4^{o}C and 37^{o}C) stresses, more specifically the CL3 and CL4 libraries, followed by the HR1, RZ3 and RT2 libraries. However, the expression was also abundant in other libraries, like SD2 and RT1 (seeds and roots with 10,366 and 8,640 reads, respectively).

Regarding the spatial co-expression among libraries (gray upper dendrogram) it was possible to detect a stronger relation among RZ1/LV1 and CL4/AD1 with LB2. SD2 was correlated with RT3 and RT1. Another correlation regarded ST3/LR2, while the library FL5 was associated with FL4 plus LB1 (Figure 3). It is interesting to note that BADH, TPS, SAT and OASTL had the majority of expression in HR1, RT2 and CL3 and it is clear the co-expression of these libraries in the dendrogram. Concerning the co-expression of genes the analysis revealed two groups INPS1/P5CS and SAT/OASTL + BADH + TPS (Figure 3).

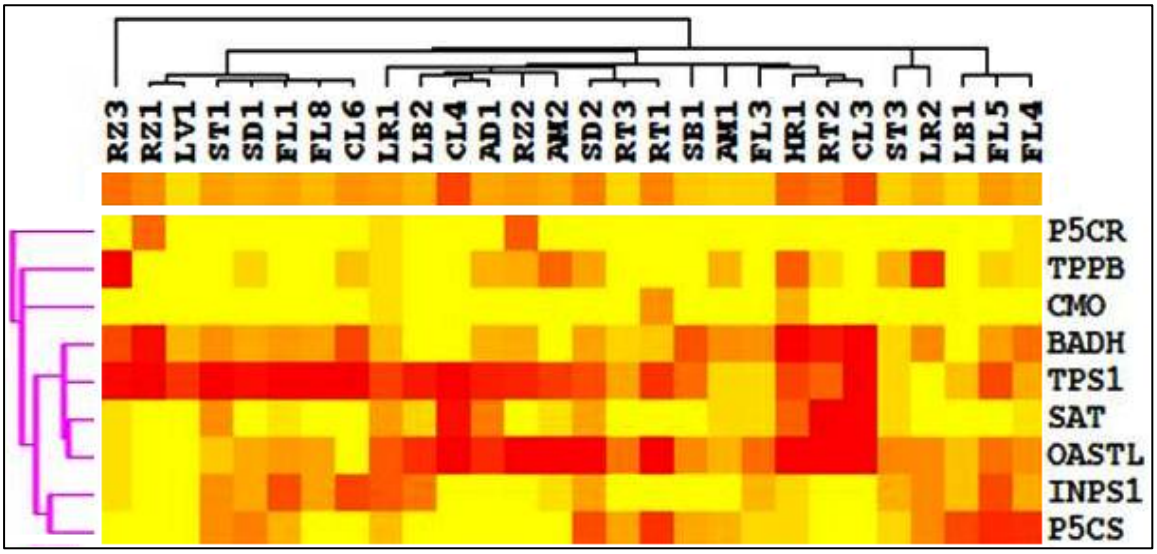


Figure 3: Differential display of standard sugarcane transcripts representing selected osmoprotectant genes. Graphic represents the expression in *P5CS, P5CR, BADH, CMO, TPS, TPP, OASTL, SAT* and *INPS1* clusters. Yellow means no expression and red means all levels of expression. Library codes: **AD1:** tissues infected by *Gluconacetobacter diazotroficans*, **AM1:** Apical meristem from mature plants; **AM2:** Apical meristem from immature plants; **CL3, CL4 and CL6:** Pool of calli treated for 12h at 4º e 37ºC in the dark or light; **FL1**, **FL3**, **FL4**, **FL5** and **FL8:** Flowers harvested at different developmental stages; **HR1:** tissues infected with *Herbaspirillum rubrisubalbicans*; **LB1** and **LB2:** Lateral Bud from mature plants; **LR1:** Leaf Roll from immature plants, large insert; **LR2:** Leaf Roll from immature plants, small insert; **LV1:** Etiolated leaves from plantlets grown *in vitro*; **RT1, RT2** and **RT3:** 0,3 cm length roots from mature plants and root apex; **RZ1**, **RZ2** and **RZ3:** Root to shoot zone transition of young plants zone 1, 2 and 3; **SB1:** Stalk bark from mature plants; **SD1** and **SD2:** Seeds in different stages of development; **ST1:** Stem first internodes; **ST3:** Stem fourth internodes.

Phylogenetic analysis

The multiple alignments made with CLUSTALx program revealed a considerable degree of conservation among bacteria, fungi, protozoan, algae, plants and animals, considering each osmoprotectant gene category. Concerning these classes of

organisms the phylogenetic analysis maintained them in distinct branches in most dendrograms here generated.

For INPS1 the bacteria were determined as an outgroup (branch V), with remaining organisms in four clades, with two of these subdivided with grouping and subgrouping following a clear segregation according to their higher taxonomic grouping. The subclade IA grouped the monocots, IB and IC the dicots, II the amoeba *Dictyostelium discoideum* and III the animals, being this last one subdivided in two subclades: IIIA with the class Insecta and the IIIB with Vertebrata, while the group IV comprised only Fungi with the yeast *Saccharomyces* staying basal to this group (Figure 4).

In relation to P5CS cladogram, a distinct separation of the Angiosperms class was evident, with monocots (IA) and dicots (IB-D) following the traditional taxonomic classification. The dicots subdivision presented three branches: IB with members from Asteridae subclass, IC with Rosidae subclass (IC' with Salicaceae, Celastraceae and Vitaceae and IC'' with only Brassicaceae) and ID with only members from Fabaceae (Figure 5A). Regarding P5CR dendrogram (Figure 5B) it can be seen that Angiosperms remained in an isolated group (clade I), where the Liliopsida (IA) and Magnoliopsida (IB) formed subgroups. Furthermore, the Clorophyta division was separated in clade II (represented by the algae *Chlamydomonas reinhardtii*), with animals in clade III, while fungi and bacteria remained in the clades IV and V, respectively. Additional dendrograms for BADH, SAT, OASTL, TPS and TPP also comprised organisms from various classes and can be seen in supplementary data.

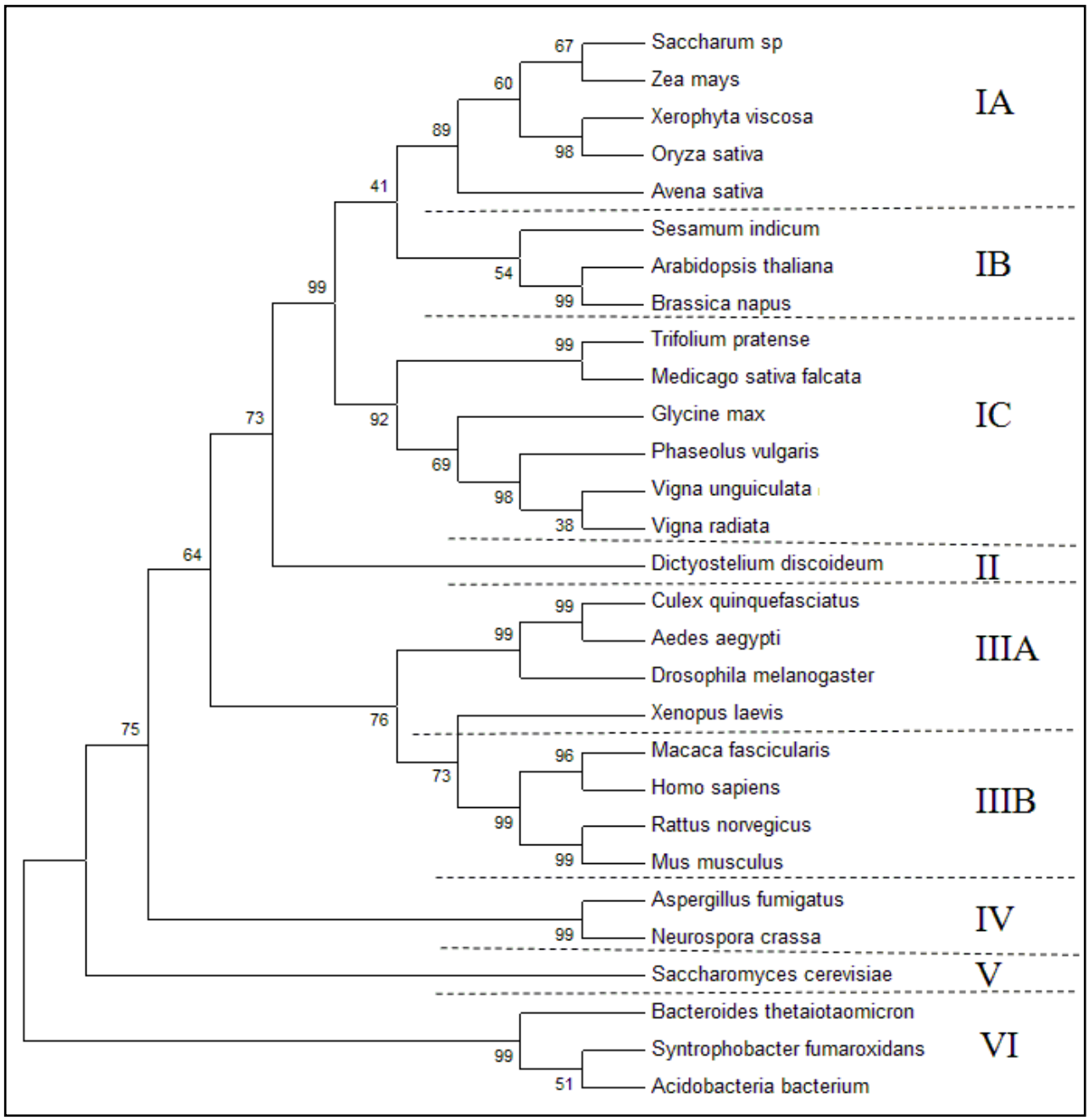


Figure 4: Dendrograms generated after maximum parsimony analysis, showing the relationships between the consensus sequences of *A. thaliana*, *V. unguiculata* and *Saccharum* sp. and other organisms regarding protein sequences from INPS1. Numbers in the base of clades indicate bootstrap values (2,000 replications).

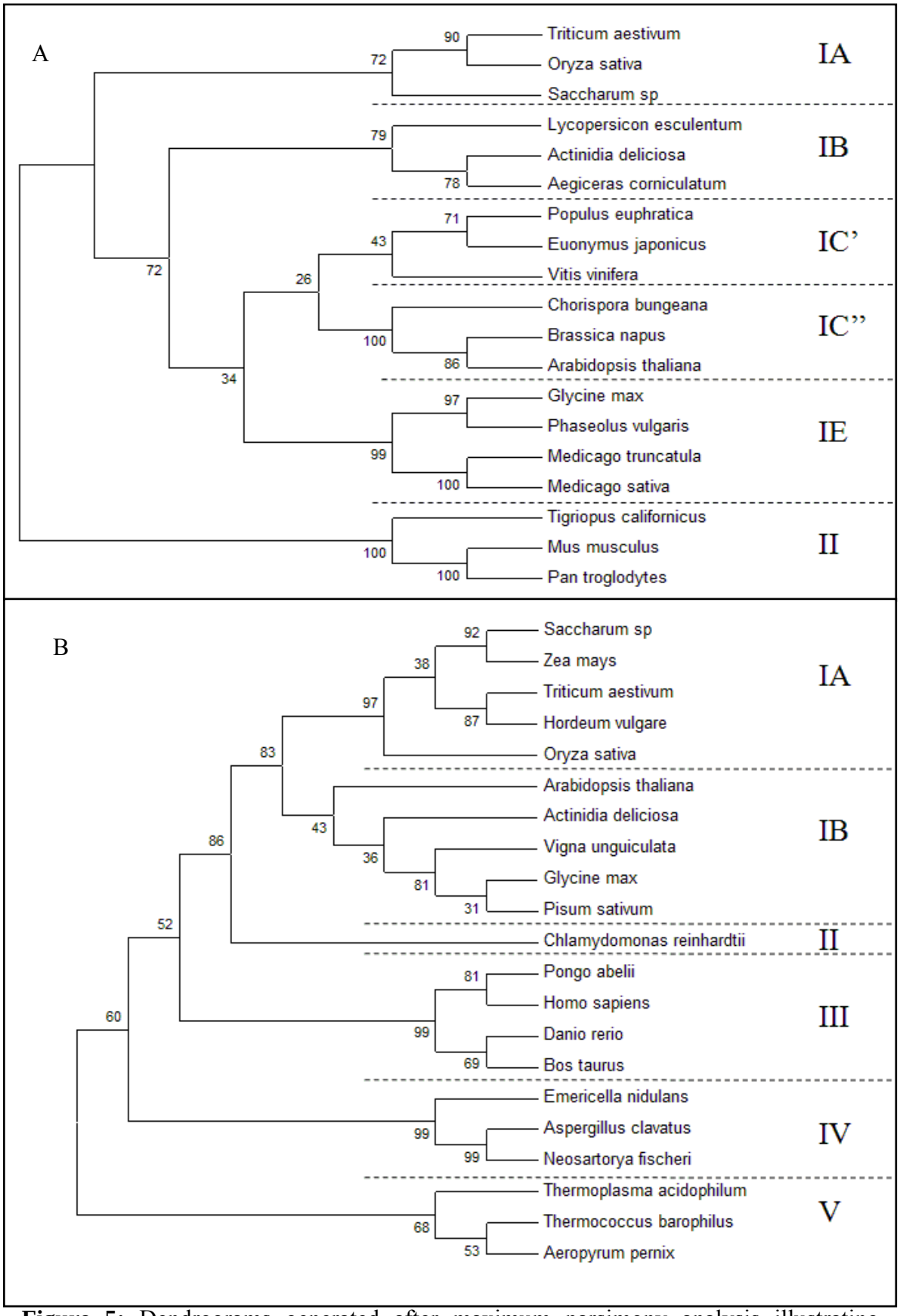


Figure 5: Dendrograms generated after maximum parsimony analysis illustrating relationship revealed by conserved domains similarity in (A) P5CS-like and (B) P5CR-like sequences and putative cowpea and sugarcane orthologs. The numbers in the base of clades regard bootstrap values (2,000 replications).

DISCUSSION

Sugarcane and Cowpea Osmoprotectant Orthologs

Amino acids and their derivatives are the dominant compatible solutes in phylogenetically distant taxa (Yancey et al., 1982). Among these, almost all plants accumulate high concentrations of proline in response to the imposition of a wide range of biotic and abiotic stresses (Hare and Cress, 1997).

In the case of cowpea the great similarity of P5CS and P5CR candidate clusters was in accordance to the classic taxonomic relationships, since all significant alignments occurred within the Fabaceae family. Although none of the candidates has shown all sought domains complete, the existence of proline pathway in this pulse crop is evident, once the synthesis of this amino acid is an ancestral trait that existed before the separation of major taxonomic groups. In fact, Delauney et al. (1993) characterized the gene P5CS from a cDNA library of *V. aconitifolia*, and Hu et al. (1992) in experiments with the same organism indicated that the *P5CS* gene is strongly induced particularly in leaves as response to NaCl treatment. Moreover, Armengaud et al. (2004), studying *Medicago trunculata*, showed that MtP5CS acts as a developmental housekeeping enzymes and may be responsible for the supply of proline in the reproductive organs and appear to be unaffected by osmotic stresses. Thus, it is clear that the regulation of P5CS transcripts varies among species (Stines et al., 1999), being well characterized in legumes. Regarding P5CR, Szoke et al. (1992) found an enzymatic activity in roots, root-nodules and leaves of soybean (*G. max*). These evidences indicated the conservation of the processes involved in the formation of proline in members of the Fabaceae family, a hypothesis also confirmed by the evidences presented in our approach.

With respect to sugarcane P5CS and P5CR orthologs, the best alignments showed high similarity with Poaceae family, also following the taxonomic proximity. Best hits of P5CS and P5CR presented all procured domains complete and despite the low number of candidates, there was a strong presence of their transcripts in roots. This can be supported by the fact that proline accumulation has already been notice in rice (Kavi-Kishor, 1988), maize (Ober and Sharp, 1994) and sugarcane (Kumari et al., 2008) subjected to abiotic stress without any exogenous supply of proline to the medium.

Despite the fact that no candidates to CMO (which together with the BADH is responsible for the synthesis of GB) had the complete CD in both cowpea and sugarcane transcriptomes, the high similarity of the identified sequences to the same gene of rice (in the case of cowpea) and maize (in the case of sugarcane) are enough to indicate the existence of this gene in both transcriptomes. The low number of CMO candidates in both transcriptomes can be explained by diverse factors. First, the rigidity in choline utilization is not uncommon in plant metabolism; the main factor that limits the accumulation of GB is the availability of choline to allow the oxidation reaction, and its transport from chloroplasts to cytosol (Chen and Murata, 2002; Rontein et al., 2002). Second, the catalytic activity of CMO is lower than that of COD and CDH, enzymes present in bacteria that synthesize GB (Burnet et al*.,* 1995). Third, many studies have proved that the number of transcripts of BADH is higher than CMO in almost all organisms (Rhodes et al., 1987; Murata et al., 1992; Jagendorf and Takabe, 2001), a fact also confirmed by our evidences.

Weretilnyk et al. (1989) suggested that the GB pathway may be widely distributed in higher plants, being in many cases only weakly expressed. The main dicot families in which GB accumulation was recognized are the Asteraceae, Caryophyllaceae, Amaranthaceae, Solanaceae, Malvaceae, Plumbaginaceae,

Convovulaceae, Chenopodiaceae, Cuscutaceae and Fabaceae. Thus, the fact that we found a considerable number of BADH orthologs and that in the Fabaceae family a variety of betaines, which includes GB, proline-betaine, have been described (Weretilnyk et al., 1989; Rhodes and Hanson, 1993), are a strong evidence for the existence of GB pathway in the cowpea metabolism.

In the case of monocots, rice is the only important cereal crop that does not accumulate GB. The rice genome includes a nonfunctional gene for *CMO*, which is probably a pseudo-gene, as well as two copies of *BADH* (International Rice Genome Sequencing Project, 2005); one of these BADH isozymes has been reported to be a candidate gene for fragrance, playing a important role in aroma synthesis in this organism (Bradbury et al., 2005). In contrast, maize has been reported as a plant with moderate capacity for betaine accumulation, with maximum levels in the range 5 to 10 μmol/g fresh weight (Rhodes et al., 1987). According to Kumari et al. (2008), sugarcane cells accumulated high amounts of GB when subject to salt stress, proving the existence of GB pathway in sugarcane metabolism.

Orthologs of TPS and TPP, both involved in trehalose synthesis, were found in cowpea and sugarcane transcriptomes. In fact the trehalose pathway may be activated in many organisms when they are subject to some type of stress or serving as an accumulation of storage carbohydrates (Elbein, 1974). In dicots and monocots the trehalose accumulation may occur in the plants with diseases (Keen and Williams 1969) or colonized by microorganisms, for example in mycorrhizal roots (Harley and Smith 1983), nitrogen-fixing nodules (Streeter, 1985) and actinorhizal nodules (Lopez and Torrey 1985). In legumes, Muller et al. (1994) detected large amounts of trehalose in soybean root nodules, being mainly localized in the bacterioid (~70%). Regarding monocots, it is already shown that sugarcane accumulates trehalose (Bosch, 2005); in

rice an isolated cDNA clone of *OsTPP* was highly induced by chilling stress with accumulation of this sugar (Pramanik and Imai, 2005) while in maize an isoform of *TPP* was proposed to regulate inflorescence branching (Satoh-Nagasawa et al., 2006).

According to Hesse et al. (2004) cysteine is required for the production of diverse key metabolites in different pathways: synthesis of protein, glutathione, phytochelatins, etc. SAT and OASTL are directly involved in the cysteine biosynthesis that in plants takes place in cytosol, plastids, and mitochondria. Isoforms for SAT and OASTL have been found in several plants including arabidopsis (Jost et al., 2000), spinach (Noji et al., 2001), phaseolus bean (Bertagnolli and Wedding, 1977) and maize (Clarkson et al., 1999). Expression analysis revealed semi-constitutive activity of the corresponding genes with moderate up-regulation of mRNA steady-state levels in response to external factors such as sulfate deprivation or salt stress (Gutierrez-Alcala et al., 2000; Takahashi et al., 2000; Noctor et al., 2002; Zimmermann et al., 2005).

A great amount of OASTL transcripts was found both in cowpea and sugarcane, more than sequences of SAT. Ruffet et al. (1994) concluded that in plants endogenous levels of OASTL are far in excess as compared to SAT products. In fact the catalytic activity of OASTL is higher than SAT since it is generally found only in association with OASTL, forming the complex CSC (Cysteine Synthase Complex), while OASTL may also be found in its free form in the cell (Hell et al., 2001).

Compounds containing inositol are abundant in plant cells (Loewus and Loewus, 1984). Among these the myo-inositol has a crucial role in plant physiology and development and it is involved in signal transduction, phytic acid biosynthesis, auxin storage and transport, cell wall biosynthesis and osmoprotection under salt-water stress (Keller et al., 1998; Loewus and Murthy, 2000). *INPS1* genes have been reported from more than 70 different organisms so far including a wide variety of prokaryotic,

archaeal, cyanobacterial and eukaryotic sources (Majumder et al., 2003). In plants *INPS1* has been localized at transcriptional level in immature (growing) tissues of rice (Yoshida et al., 1999), wild-rice (Dastidar et al., 2006), soybean (Hegeman et al., 2001) and at protein level in *P. vulgaris* (Johnson and Wang, 1996) what provides strong evidences that the processes involved in the myo-inositol formation are conserved in both monocot and dicot plants.

Together these results point to the presence of all important gene categories for the synthesis of compatible osmolytes in sugarcane and in cowpea transcriptomes, considering the knowledge available from rice and arabidopsis model plants. However, it is expected to find a larger number of transcripts in cowpea transcriptome, since the NordEST project is still underway. With respect to sugarcane, considering the allopolyploid and hybrid nature of this crop and also that there are no libraries under salt or drought stress available in the SUCEST project, an increased number of osmoprotectant sequences can be expected.

Expression Pattern

The effect of NaCl in legumes growth depends greatly on the plant sensitivity and/or tolerance to salinity. In response to abiotic stresses a large amount of genes is activated, conferring tolerance to this kind of stress. Among those genes, one of the most important classes includes those involved in the synthesis of compatible osmolytes (Lopez et al, 2006). Some legumes like *Lotus japonicus,* considered as a salt-sensitive legume, seem to synthesize trehalose in response to salinity. Also in peanut (*Arachis hypogaea*) an increase in proline and glycine betaine content has been described after

exposition to abiotic stress (Muthukumarasamy and Panneerselvan, 1997; Lopez et al, 2006).

The observed majority of osmoprotectant transcripts in the libraries SS02 (roots of cowpea sensitive to salinity after two hours of salt stress) and ST08 (roots of cowpea tolerant to salinity after eight hours of salt stress) were expected, since roots are the first tissues to sense the stress after reduction of water availability, which causes osmotic imbalance (Munns, 2002). The expression of compatible osmolytes in these two contrasting libraries shows the importance of osmoprotectant genes in response to salinity, indicating that in sensitive plants the up regulation of osmoprotectant genes happens in the first hours after stress exposure.

The results from SUCEST database revealed the existence of a higher expression in two libraries of callus culture (CL3 and CL4 = calli submitted to light/dark and temperature stress) confirms that these genes are recruited under additional adverse conditions such absence of light and temperature changes. Moreover, it is important to highlight that the *in vitro* environment is largely known as an abiotic stress condition (Soares-Cavalcanti, 2007); a variety of works has proven the importance of osmoprotectants to tissues of calli in plants. Errabi et al. (2006) have reported higher amounts of proline in sugarcane calli exposed to different osmotic stress intensities. Under stress condition callus cultures of rice were found to accumulate considerable amounts of trehalose, compared with unadapted cells (Garg et al., 2002). In callus of wheat, increased amounts of glycine betaine were reported in tolerant as well as sensitive cultivars in response to water stress (Nayyar and Walia, 2004).

The second sugarcane library that had high number of transcripts was HR1, a library constructed with plantlets inoculated with *H. rubrisubalbicans*, which is an endophytic nitrogen-fixing bacteria that naturally colonizes sugarcane tissues (Lee et al.,

2000). In the SUCEST project the infection with this pathogen occurred during *in vitro* cultivation, so the environmental stress cannot be discarded as an inductive factor of compatible osmolytes expression. Furthermore, Cheong et al. (2002) in studies of gene expression also identified some osmoprotectant genes (P5CS and TPP) up-regulated in response to both abiotic and biotic stresses.

For *P5CR* and *CMO* an almost complete lack of expression was observed in most SUCEST libraries, while in the case of NordEST database none library had the presence of these transcripts, probably due to some reasons already mentioned above.

In *OASTL* and *TPS* results, a higher expression could be detected in most libraries of SUCEST; what is in agreement with the fact that genes encoding *OASTL* isoforms are more or less ubiquitously expressed in all plant cell types analyzed so far in response to external factors (Dominguez-Solis et al., 2001). With respect to TPS, according to Muller et al. (1999), this gene is very closely related to plant development processes, being present in tissues with these characteristics, like seeds and plantlets in different stages of development, flowers, leaves, root-to-shoot transition regions; besides, this last one, is a plant region closely related with sensing of abiotic stress.

For the data of *INPS1* obtained in the NordEST project the expression pattern results were surprising, since larger amounts of reads were expected from roots of cowpea subjected to salt stress, once this organ is the first to sense the environmental changes; however, unlike the expectations, no transcript was found in the project. The essential roles of inositol in many cellular processes, including membrane formation, cell wall biogenesis, stress response and signal transduction, have been well documented in plants (Majumder et al., 1997; Bachhawat and Mande, 2000) and some studies have indicated the light/dark regulation as determinant factor for their expression (Adhikari et al., 1987; Abu-Abied and Holland, 1994). Furthermore, Keller

et al. (1998), studying the INPS1 from *Solanum tuberosum*, identified higher levels of expression in photosynthetic tissues, such as mature leaves, immature leaves, green flower buds and colored flower buds, but down regulation in non-photosynthetic tissues, such as roots. So, the absence of transcripts in cowpea could be explained by the fact that the great majority of NordEST libraries were constructed from root transcripts. Regarding sugarcane, in contrast, it was possible to see a high amount of reads in the library CL6 (calli submitted to temperature changes – 4-37 °C – with a light/dark cycle of 6 hours), FL1 and FL5 (flowers tissues) and LB1 and LB2 (lateral bud tissues), as expected.

Considering both transcriptomes, the expression profile analysis showed the existence of transcripts in almost all libraries, revealing the crucial importance of osmoprotectants for plants. As in the SUCEST none of the 29 sugarcane libraries were constructed under salt or drought stresses, it is supposed that under exposition to these abiotic stresses the amount of genes would increase with possibility of identification of new isoforms. Once the NordEST project was not finalized, it is expected that the number of osmoprotectant sequences from cowpea may also increase.

Some studies using the hierarquical clustering for expression pattern analysis suggested that genes active in the same pathways are expect to have similar patterns of gene expression and could also be physically clustered in the chromosomes (Lambais, 2001; Wanderley-Nogueira et al., 2007). In *A. thaliana* genome, for example, genes that participate of the same pathway (as *TPS-TPP* and *SAT-OASTL*) are localized in the same chromosome (The Arabidopsis Genome Initiative, 2000). In contrast, at the rice genome the *CMO* gene is localized on chromosome 6, and two copies of a gene for *BADH* on chromosome 4 and chromosome 8 (International Rice Genome Sequencing Project, 2005). It is interesting to note that ours results demonstrated co-expression of

some libraries and closer co-expression of *SAT/OASTL* sequences, while *CMO* and *BADH* stayed at different branches, in accordance to the found in model plants.

Dendrograms/Phylogenetic Analysis

According to Benko-Iseppon et al. (2005) dendrograms made with genes related to abiotic stress responses often show groups of sequences in functional arrangements, with common signatures probably reflecting adaptation or selection driven by the environmental pressure. The fact that organisms phylogenetically diverse, as bacteria, unicellular algae, fungi, vascular plants, invertebrates and vertebrates, utilize the same families of organic osmolytes suggests that these genes may be ancient and are subjected to strong selective pressures are associated with convergent evolution (Yancey et al., 1982; Rhodes and Hanson, 1993).

The dendrograms generated from multiple alignments showed high degree of conservation among the candidates of sugarcane and cowpea orthologs with other organisms. Almost all dendrograms reflected, as expected, the evolutionary history of plants. The dendrogram constructed based on the multiple alignment of INPS1 ortholog sequences (Figure 4) showed a clear segregation of these proteins into four different monophyletic groups following the phylogenetic systematic (bacteria, vertebrate, invertebrate and plant). Interesting, it was observed that the yeast (*Saccharomyces cerevisiae*) stayed in a basal position (V) among the fungi subclade (IV), probably due to its unicellular and non filamentous features. The amoeba shared synarqueomorphyc characteristics with flowering-plants, but not with animals and fungi, as expected. The dicot clade appeared divided into two conspicuous subclades. The subclade IC was formed by legume species, possibly reflecting ancestral characters shared by these

Fabaceae plants, which are able to fix nitrogen. Besides their N-fixing capacity, medicago is considered a salt sensitive plant (Veatch et al., 2004) as well as *Trifolium pratense*, a plant affected by moderate salt levels (Winter and Lauchli, 1982). The bean species were grouped, as expected, since they are more tolerant to water deficit (Omae et al., 2006), while soybean seems to be more tolerant to salt stress (Waldren and Teare, 1974).

In studies conducted by Majumder et al. (1997) the alignments of INPS1 amino acid sequences from diverse organisms including Arabidopsis and bean (*P. vulgaris*), brewer's yeast (*Saccharomyces cerevisiae*) and *Entamoeba histolytica* revealed remarkable evolutionary conservation considering the proteins primary structure. According to Styer (2000), analysis of INPS1 indicated that the plant proteins are closely related, not exceeding 8% of divergence and the yeast INPS1 protein is closely related to the plant INPS1 (38.5% divergence). In our dendrogram fungi grouped in a separate clade as plants and animals that presented a more complex life cycle and adaptive traits, needing better strategies to cope with environmental stresses. It is clear that this protein is found in diverse organisms, both eukaryotic and prokaryotic, suggesting that the pathway for inositol 1-phosphate biosynthesis from Glucose 6-phosphate arose early in the evolution of life (Majumder et al., 1997; Bachhawat and Mande, 2000).

Regarding the P5CS dendrogram (Figure 5A) the presence of a symplesiomorphic character among the monophyletic clades of plants and animals was detected. In fact, Fujita et al. (1998) revealed that P5CS proteins from higher eukaryotes are clearly monophyletic. The observed plant topology clearly showed the existence of two groups, separating the P5CS proteins from monocots and dicots. In bacteria two genes are responsible for proline biosynthesis (*proB* and *proA*), with the two enzymatic

domains of P5CS corresponding to the ProB and ProA proteins of *Escherichia coli* (Kishor et al., 2005). It has been proposed that the corresponding plant genes may have fused and originated the bifunctional enzyme present in plant genomes (Hu et al., 1992). In animal systems, a similar event of domain fusion must have occurred since P5CS activity has been detected in mammalian cells and a single-gene encodes both functional enzymatic activities (Turchetto-Zolet et al., 2009).

The P5CR dendrogram (Figure 5B) also revealed a great conservation among protein sequences including different eukaryotic groups in distinct monophyletic branches. Regarding angiosperm group it was possible to note that monocots and dicots formed a monophyletic group, subdivided in two groups and the Fabaceae members together in the same subbranch.

Regarding OASTL dendrogram (Figure I – Supplementary Data) it was possible to see that the sequences from bacteria were placed as an outgroup (V), sharing synarqueomorphic characters with the rest of the organisms. The Clorophyta and Embryophyta organisms presented a common arqueomorphyc character, as expected since they belong to the Archaeplastida, the major line of eukaryotes. In respect to the angiosperms, it is clear that different isoforms of OASTL were represented in two distinct groups (IA+IB and IC+ID), both formed by Liliopsida and Magnoliopsida species. In relation to the monocots in the IA subgroup, the *Saccahrum*_01 sequence grouped to the OASTL citosolic isoform from *Z. mays* (gi: 758353). In the IB subgroup almost all sequences represented the citosolic isoform, showing that the cowpea sequences (Vigna 1 and Vigna 2) probably corresponded to the citosolic isoform of OASTL. The second group of monocots (IC) aligned the *Saccharum*_02 sequence with the OASTL plastid isoform from *O. sativa* (gi: 57899533), which also was the best match in the reverse alignment. Finally, the ID subgroup the sequences from *S. oleracea*

(gi: 303902) as well as from *Knorringia sibirica* (gi: 186688080) are described as a plastid isoform of OASTL. Thus, it is interesting to note that the organisms from the Angiospermae division were not separated based on their phylogenetic relation, but according to their expression site, showing that theses two isoforms present significant differences.

The multiple alignments with SAT proteins showed great degree of conservation among different organisms, revealing its ancestral character. In the resulted dendrogram was noted that Bacteria, Chlorophyta (green algae) and Angiospermae formed distinct monophyletic groups (Figure II – Supplementary Data). Regarding the angiosperms, the dicots were separated into two different levels, with the three Fabaceae species emerging in a basal position relative to a clade that included two subclades with remaining dicots and all monocots, respectively.

The allosteric feedback inhibition of SATase activity by L-Cys is an important regulatory mechanism for OAS levels. Other mechanism is the modulation of SATase activity through reversible formation of a protein complex with Cys synthase (Saito, 2000; Kawashima et al., 2005). The presence of feedback regulation in SATase isoforms differs within plant species and subcellular compartments (Kawashima et al., 2005) and has been reported in watermelon (*Citrullus vulgaris*; Saito et al., 1995), arabidopsis (Noji et al., 1998), *Allium tuberosum* (Urano et al., 2000), *Spinacea oleracea* (Noji et al., 2001) and soybean (Chronis and Krishnan, 2004). It is important to note that all these organisms grouped together in the dendrogram. In contrast, Droux (2003) showed that the cytosolic and mitochondrial isoforms from pea shows insensitivity to Cys inhibition. It is interesting that most legumes (with exception of soybean) appeared as a special subclade separated from other angiosperms, possibly reflecting these features. Moreover, in other phylogenetic studies conducted by

McManus et al. (2005) the SAT isoform 1 grouped sequences from *S. oleracea*, *G. max*, *Beta vulgaris* and *A. thaliana* in an isolated clade, similarly to our SAT dendrogram, where the SAT isoform 1 grouped these organisms together with *Citrullus lanatus*.

In the TPP dendrogram (Figure III – Supplementary Data) two distinct groups were formed, regarding the flowering plants (both including dicots and monocots species), probably due to the different types of TPP isoforms. Supporting this idea, Lunn (2007) had shown that many plants (arabidopsis, rice, poplar) presented until 10 types of TPP isoforms. Moreover, Lunn (2007), in phylogenetic analyses of the plant TPP protein sequences, found that the angiosperm sequences were consistently divided into two major groups, providing robust support for a fundamental dichotomy within the angiosperm TPP family. In the branch IA, separated groupings could be observed in the monocot subbranch: the *Saccharum* spp. 1 sequence presented more similarity with the *Eragrostis tef* sequence, while the *Saccharum* spp. 2 shared a larger number of characters with a sequence from maize; it is important to stick out that both *E. tef* and maize are recognized by their active TPP enzyme (designated as Ramosa3) that is involved in control of inflorescence branching by modification of a sugar signal that moves into axillary meristems (Satoh-Nagasawa et al., 2006). The *V. unguiculata* 1 was grouped with *Vitis vinifera* in the same branch were most sequences are a RAR3-like which provides strong evidence that these candidates presented the same TPP isoform.

Comparing the sequences within branch IB it was possible to see the sequences of sugarcane and rice (gi: 45544517) sharing the same branch. According to Pramanik and Imai (2005), this rice TPP isoform is highly induced by chilling stress in shoot and root tissues of seedlings and also by salt and drought stress. Looking for the reads that made the cluster (*Saccharum* spp. 3 used in the alignment) it was possible to see that two of the four reads came from root to shoot zone transition of young plants (RZ3

library) a tissue very sensitive to abiotic factors, in consistence with the information given by the authors. In *A. thaliana*, salt stress induces expression of several TPP genes in roots (Lunn, 2007) and it is important to note that most cowpea reads from NordEST project came from roots subjected to salt stress. The sequence from *N. tabacum* that grouped with these last two organisms came from studies conducted by Wang et al. (2005), that showed the expression of TPP genes during heat, NaCl and low-temperature stresses. Thus, it is possible to infer that these two branches represented two distinct isoforms, the first (IA) with proteins involved with inflorescence architecture and the second (IB) involved in stress responses.

Regarding the TPP from the *Mycobacterium tuberculosis*, it was observed that it stayed as an out-group, presenting less common traits with the remaining organisms. Despite that, Lunn (2007) suggested that the plant TPP sequences are most closely related to those from bacteria. According to Brown et al. (2001) the plant TPP genes may have originated from the endosymbiotic ancestor of mitochondria, which is thought to be similar to the actual gene present in some types of bacteria. With respect to the branch II, it is interesting to note that the group included *Physcomitrella patens* and *Selaginella moellendorffii*, two species strongly resistant to adverse environmental conditions. According Frank et al. (2005), studies with *P. patens* (Bryophyta) showed that this species presents a high salt and drought tolerance, while the *Selaginella* species (Lycopodiophyta) are known as resurrection ferns that can recover from almost complete desiccation.

The complete sequencing of the arabidopsis genome revealed a family of 11 genes (AtTPS1–11) encoding TPS enzymes (The Arabidopsis Initiative, 2000), which are formed by the glucosyltransferase-like and HAD-like domains. The generated TPS dendrogram (Figure IV – Supplementary Data), using sequences that had these two

domains, in contrast with the other analyzed osmoprotectant genes, did not followed the current phylogenetic systematic. The main plant clade (I) presented three distinct subclades, one with monocots (IA) and other with dicots (IB), with *L. japonicus* (IC) in a basal position relative to both groups (IA/IB). Moreover, it is interesting to note that the other angiosperm group (III) (including rice, sugarcane and cowpea) was placed out of the group I, while the expected was that this clade would appear as sister group. Avonce et al. (2006) explained that these inconsistencies are evident only for genes displaying multiple paralogous copies in a single organism, probably due to either lateral gene transfer or differential loss of paralogs. They also showed that all plant TPS genes are under selection pressure suggesting that all of them have a particular function, which could probably be related to other processes, not necessarily related to osmoprotection.

In the BADH dendrogram (Figure V – Supplementary Data) the main groups (IA - Monocots, IB - Dicots, II - Bacteria and III - Archea) appeared as monophyletic clades. Regarding the dicots, the subclades IB' and IB'' comprised only members of the Rosidae subclass (Brassicaceae and Fabaceae families), while the subclade IB''' presented only members of the Caryophyllidae subclass, which is known to accumulate far more GB than do other plants in response to osmotic stresses (Weretilnyk et al., 1989). With respect to the monocotyledonous plants, studies suggested that BADH is localized in peroxisomes (Nakamura et al, 1997) while, differently, in the dicots the BADH is localized in the chloroplasts (Rontein et al., 2002). Another difference between monocots and dicots regards the RNA processing pattern; the sequencing data from cDNA clones from diverse monocots demonstrated that all sequences tested had an unusual posttranscriptional processing resulting in deletion(s) of the 5′ exonic sequences (Niu et al., 2007).

The synthesis of glycine betaine involving BADH genes seems to be universal in the three domains of life, from members of Bacteria and Archaea (with diverse ecophysiological characteristics) to plants and algae. It's a rare phenomenon in heterotrophic bacteria but common in phototrophic bacteria and methanogenic archaea (like *Methanosaeta thermophila*), which exhibit moderate tolerance to high salt concentrations (Empadinhas and Da Costa, 2008).

Concluding Remarks

Many crops lack the ability to efficiently synthesize some types of osmoprotectants that are naturally accumulated by stress-tolerant plants. The production of transgenic plants that can accumulate osmoprotectants, in particular using model plants, such as arabidopsis and rice, will allow the transference of this defense mechanism to important crop plants, such as sugarcane and cowpea. In addition, more studies involving transgenic plants tolerant to stress conditions will help to verify their utility potential in crop-breeding programs.

In conclusion, this survey of osmoprotectant-related genes in the sugarcane and cowpea transcriptomes can provide new insights into the study of the genetic and also synthetic pathways of compatible osmolytes, biosynthesis regulation of these compounds and phylogenetic relationship among their natural producers, continuing to expand our knowledge on the evolution of flowering-plants adaptations to abiotic stresses and providing a framework on which future studies into the function(s) of the osmoprotectants in the plants metabolism may be based.

Acknowledgements

The authors thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FACEPE (Fundação de Amparo à Pesquisa do Estado de Pernambuco) for the concession of fellowships. To FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), SUCEST and HARVEST coordination for the access to the Sugarcane and Cowpea EST data bank, respectively. We also thank all who were involved in the cowpea EST sequencing of the NordEST Project.

Literature Cited

Abu-Abied M and Holland D (1994). Two newly isolated genes from citrus exhibit different patterns of diurnal expression and light response. *Plant Mol. Biol.* 26: 165–173.

Adhikari J, Majumder AL, Bhaduri TJ, Das Gupta S et al. (1987). Chloroplast as a locale of L-myo-inositol- 1-phosphate synthase. *Plant Physiol*. 85: 611–614.

Alia, Chen THH and Murata N (1998). Transformation with a gene for choline oxidase enhances the cold tolerance of Arabidopsis during germination and early growth. *Plant Cell Environ*. 21:232-239.

Altschul SF, Gish W, Miller W, Myers E, et al. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.

Armengaud P, Thiery L, Buhot N, March GG, et al. (2004). Transcriptional regulation of proline biosynthesis in *Medicago truncatula* reveals developmental and environmental specific features. *Physiol. Plant*. 120: 442–450.

Avonce N, Mendoza-Vargas A, Morett E and Iturriaga G (2006). Insights on the evolution of trehalose biosynthesis. *BMC Evol. Biol*. 6:109.

Bachhawat N and Mande SC (2000). Complex evolution of the inositol-1-phosphate synthase gene among archaea and eubacteria. *Trends Genet*. 16: 111–113.

Bhatnagar-Mathur P, Vadez V and Sharma KK (2008). Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. *Plant Cell Rep.* 27:411-424

Benko-Iseppon AM (2001). Estudos moleculares no caupi e em espécies relacionadas: Avanços e perspectivas. *Embrapa Docs.* 56: 327-332.

Benko-Iseppon AM, Soares-Cavalcanti NM, Wanderley-Nogueira AC, Berlarmino LC, et al. (2005). Environmental Stresses: damages and benefits to plants. In: Genes associated to biotic and abiotic stresses in cowpea [*Vigna unguiculata* (L.) *Walp.*] and other angiosperms (Nogueira RJMC, Araújo EL, Willadino LC, Cavalcante UMT, eds.). UFRPE University Press Recife, 350–359.

Bertagnolli BL and Wedding RT (1977). Purification and initial kinetic characterization of different forms of O-acetyl-L-serine sulfhydrylase from seedlings of two species of phaseolus. *Plant Physiol*. 60: 115–121.

Bohnert HJ and Shen B (1999). Transformation and compatible solutes. *Sci. Hortic*. 78:237-260.

Bosch S (2005). Trehalose and carbon partitioning in sugarcane. *Doctoral Thesis*, University of Stellenbosch, South African.

Bradbury LMT, Fitzgerald TL, Henry RJ, Jin Q, et al. (2005). The gene for fragrance in rice. *Plant Biotechnol. J.* 3:363-370.

Brown JR, Douady CJ, Italia MJ, Marshall WE, et al. (2001) Universal trees based on large combined protein sequence data sets. *Nature Genet.* 28:281-285.

Burnet M, Lafontaine PJ and Hanson AD (1995). Assay, purification, and partial characterization of choline monooxygenase from spinach. *Plant Physiol.* 108:581–588.

Campanini B, Speroni F, Salsi E, Cook PF, et al. (2005). Interaction of serine acetyl-transferase with O-acetylserine sulfhydrylase active site: Evidence from fluorescence spectroscopy. *Protein Science*. 14:2115-2124.

Clarkson DT, Diego E and Amancio S (1999). Uptake and assimilation of sulphate by sulphur-deficient *Zea mays* cells: The role of O-acetyl-L-serine in the interaction between nitrogen and sulphur assimilatory pathways. *Plant Physiol. Biochem.* 37: 283–290.

Chen WP, Li PH and Chen THH (2000). Glycinebetaine increases chilling tolerance and reduces chilling-induced lipid peroxidation in *Zea mays* L. *Plant Cell Environ*. 23:609-618.

Chen TH and Murata N (2002). Enhancement of tolerance of abiotic stress by metabolic engineering of betaines and other compatible solutes. *Curr. Opin. Plant Biol.* 5:250–257.

Cheong YH, Chang HS, Gupta R, Wang X, et al. (2002). Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. *Plant Physiol.* 129:661-677.

Cherian S, Reddy MP and Ferreira RB (2006). Transgenic plants with improved dehydration stress tolerance: progress and future prospects. *Biol. Plant.* 50(4):481-495.

Chronis D and Krishnan HB (2004). Sulfur assimilation in soybean (**Glycine max** [L.] Merr.): molecular cloning and characterization of a cytosolic isoform of serine acetyltransferase. *Planta.* 218: 417–426.

Dastidar KG, Maitra S, Goswami L, Roy D, et al. (2006). An insight into the molecular basis of salt tolerance of L-myo-Inositol 1-P synthase (PcINO1) from *Porteresia coarctata* (Roxb.) Tateoka, a halophytic wild rice. *Plant Physiol*. 140: 1279–1296.

Delauney AJ, Hu CA, Kavi-Kishor PB and Verma DPS (1993). Cloning of ornithine & aminotransferase cDNA from *Vigna aconitifolia* by transcomplementation in *Escherichia coli* and regulation of proline biosynthesis. *J. Biol. Chem*. 268(25): 18673–18676.

Dominguez-Solis JR, Gutierrez-Alcala G, Vega JM, Romero LC, et al. (2001). The cytosolic O-acetylserine(thiol)lyase gene is regulated by heavy metals and can function in cadmium tolerance. *J. Biol. Chem*. 276:9297–9302.

Droux M (2003). Plant serine acetyltransferase: new insights for regulation of sulphur metabolism in plant cells. *Plant Physiol. Biochem.* 41: 619–627.

Elbein AD (1974). The metabolism of alpha,alpha-trehalose. *Adv. Carb. Chem. Biochem*. 30:227-256.

Eisen MB, Spellman PT, Brown PO and Botstein D (1998). Cluster analysis and display of genomic-wide expression pattern. *Proc. Natl. Acad. Sci. USA*. 95:14863-14868.

Empadinhas N and da Costa MS (2008). Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes. *Int. Microbiol*. 11:151–161.

Errabii T, Gandonou CB, Essalmani H, Abrini J, et al. (2006). Growth, proline and íon accumulation in sugarcane callus cultures under drought-induced osmotic stress and its subsequent relief. *African J. Biotec*. 5:1488-1493.

Frank W, Ratnadewi D and Reski R (2005). *Physcomitrella patens* is highly tolerant against drought, salt and osmotic stress. *Planta*. 220(3): 384–394.

Freire-Filho FR, Rocha MM, Ribeiro VQ and Lopes ACA (2005). Adaptabilidade e estabilidade produtiva de feijão-caupi. *Cien. Rural*. 35:24-30.

Fujita T, Maggio A, Garcia-Rios M, Bressan RA, et al. (1998). Comparative analysis of the regulation of expression and structures of two evolutionarily divergent genes for Δ^1-Pyrroline-5-carboxylate synthetase from tomato. *Plant Physiol.* 118:661-674.

Garg AK, Ju-Kon K, Owens TG, Ranwala AP, et al. (2002). Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. *Proc. Natl. Acad. Sci. USA*. 99(25):15898-15903.

Grivet L and Arruda P (2002). Sugarcane genomics: depicting the complex genome of an important tropical crop. *Curr. Opin. Plant. Biol.* 5: 122–127.

Gutierrez-Alcala G, Gotor C, Meyer AJ, Fricker, M, et al. (2000). Glutathione biosynthesis in Arabidopsis trichome cells. *Proc. Natl. Acad. Sci. USA.* 97: 11108-11113.

Hamilton EW and Heckathorn SA (2001). Mitochondrial adaptations to NaCl. Complex I is protected by anti-oxidants and small heat-shock proteins, whereas complex II is protected by proline and betaine. *Plant Physiol.* 126:1266-1274.

Hare PD and Cress WA (1997). Metabolic implications of stress-induced proline accumulation in plants. *Plant Growth Regul*. 21:79-102.

Harley JL and Smith SE (1983). Mycorrhizal symbiosis. Academic Press, London.

Hegeman CE, Good LL and Grabau EA (2001). Expression of D-*myo*-inositol-3-phosphate synthase in soybean. Implications for phytic acid biosynthesis. *Plant Physiol*. 125: 1941–1948.

Hell R, Wirtz M, Berkowitz O and Droux M (2001). The cysteine synthase complex from plants. Mitochondrial serine acetyltransferase from *Arabidopsis thaliana* carries a bifunctional domain for catalysis and protein-protein interaction. *Eur. J. Biochem.* 268: 686–683.

Hesse H, Nikiforova V, Gakière B and Hoefgen R (2004). Molecular analysis and control of cysteine biosynthesis: integration of nitrogen and sulphur metabolism. *J. Exp. Bot*. 55(401): 1283–1292.

Hu CA, Delauney AJ and Verma DP (1992). A bifunctional enzyme (delta^1-pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants. *Proc. Natl. Acad. Sci. USA.* 89:9354-9358.

International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. *Nature*. 436:793-800.

Jagendorf AT and Takabe T (2001). Inducers of glycinebetaine synthesis in barley. *Plant Physiol*. 127:1827-1835.

Javot H and Maurel C (2002). The role of aquaporins in root water uptake. *Ann. Bot.* 90:301-313.

Johnson M and Wang X (1996). Differentially expressed forms of 1l-myo-inositol-1-phosphate synthase in *Phaseolus vulgaris*. *J. Biol. Chem.* 271:17215–17218.

Jost J, Berkowitz O, Wirtz M, Hopkins L, et al. (2000). Genomic and functional characterization of the oas gene family encoding O-acetylserine (thiol) lyases, enzymes catalyzing the final step in cysteine biosynthesis in *Arabidopsis thaliana*. *Gene.* 253: 237–247.

Kavi-Kishor PB (1988). Effect of salt stress on callus cultures of *Oryza sativa* L. *J. Exp. Bot*. 39: 235–240.

Kawashima CG, Berkowitz O, Hell R, Noji M, et al. (2005). Characterization and expression analysis of a serine acetyltransferase gene family involved in a key step of the sulfur assimilation pathway in Arabidopsis. *Plant Physiol.* 137(1): 220–230.

Keller R, Brearley C, Trethewey R and Muller-Rober B (1998). Reduced inositol content and altered morphology in transgenic potato plants inhibited for 1D-myo-inositol 3-phosphate synthase. *Plant J*. 16:403–410.

Keen NT and Williams PH (1969), Synthesis and degradation of starch and lipids following infection of cabbage by *Plasmodiophora brassicae*. *Phytopathol.* 59: 778-785.

Kishor PBK, Sangam S, Amrutha RN, Laxmi PS, et al. (2005).Regulation of proline biosynthesis, degradation,uptake and transport in higher plants: Its implications in plant growth and abiotic stress tolerance. *Curr. Sci*. 88(3):424-438.

Kumar S, Tamura K and Nei M (2004). MEGA 3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinf*. 5:150-163.

Kumari A, Patade VY and Suprasanna P (2008). In silico analysis of *P5CS* gene evolution in plants. *Online J. Bioinfo.* 9(1): 1–11.

Lambais MR (2001). In silico differential display of defense-related ESTs from sugarcane tissues infected with diazothrophic endophytes. *Genet. Mol. Biol.* 24:103-111.

Lee S, Reth A, Meletzus D, Sevilla M, et al. (2000). Characterization of a major cluster of *nif, fix*, and associated genes in a sugarcane endophyte, *Acetobacter diazotrophicus. J. Bacteriol.* 182:7088-7091.

Loewus F and Loewus MW (1984). Myo-inositol: Its biosynthesis and metabolism. *Annu. Rev. Plant Physiol*. 34:137-161.

Loewus F and Murthy P (2000). Myo-inositol metabolism in plants. *Plant Sci.* 150:1–19.

Lopez MF and Torrey JG (1985). Purification and properties of trehalase in *Frankia* ArI3. *Arch. Microbiol*. 143:209-215.

Lopez M, Herrera-Cervera JA and Lluch C (2006). Trehalose metabolism in root nodules of the model legume *Lotus japonicus* in response to salt stress. *Plant Physiol.* 128: 701–709.

Lu YH, D'Hont A, Paulet F, Grivet L, et al. (1994). Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica.* 78: 217–226.

Lunn JE *(*2007). Gene families and evolution of trehalose metabolism in plants. *Funct. Plant Biol*. 34:550-63.

Majumder AL, Johnson MD and Henry SA (1997). 1l-myo-inositol-1-phosphate synthase. *Biochim. Biophys. Acta.* 1348: 245–256.

Majumder AL, Chatterjee A, Ghosh-Dastider K and Majee M (2003). Diversification and evolution of L-myo-inositol 1-phosphate synthase. *FEBS Lett.* 533: 3–10.

McManus MT, Leung S, Lambert A, Scott RW, et al. (2005). Molecular and biochemical characterisation of a serine acetyltransferase of onion, *Allium cepa* (L.). *Phytochem.* 66(12):1407–1416.

Müller J, Xie ZP, Staehelin C, Mellor RB, et al. (1994). Trehalose and trehalase in root nodules from various legumes. *Physiol. Plant.* 90:86-92.

Müller J, Wiemken A and Aeschbacher R (1999). Trehalose metabolism in sugar sensing and plant development. *Plant Sci.* 147(1): 37–47.

Munns R (2002). Comparative physiology of salt and water stress. *Plant Cell Environ*. 28: 239–250.

Murata N, Mohanty PS, Hayashi H and Papageorgiou GC (1992). Glycinebetaine stabilizes the association of extrinsic proteins with the photosynthetic oxygen-evolving complex. *FEBS Lett.* 296: 187–189.

Muthukumarasamy M and Panneerselvam R (1997). Amelioration of NaCl stress by triadimefon in peanut seedlings. *Plant Growth Regul*. 22: 157–162.

Nakamura T, Yokota S, Muramoto Y, Tsutsui K, et al. (1997). Expression of a betaine aldehyde dehydrogenase gene in rice, a glycinebetaine nonaccumulator, and possible localization of its protein in peroxisomes. *Plant J*. 11(5):1115-1120**.**

Nayyar H and Walia DP (2004). Genotypic variation in wheat in response to water stress and abscisic acid induced accumulation of osmolytes in developing grains. *J. Agron. Crop. Sci* 190: 39–45.

Nelson DE, Rammesmayer G and Bohnert HJ (1998). Regulation of cell-specific inositol metabolism and transport in plant salinity tolerance. *Plant Cell*. 10:753-764.

Niu X, Zheng W, Lu B, Ren G, et al. (2007). An unusual posttranscriptional processing in two *Betaine Aldehyde Dehydrogenase* loci of cereal crops directed by short, direct repeats in response to stress conditions. *Plant Physiol.* 143:1929-1942.

Noctor G, Gomez L, Vanacker H and Foyer CH (2002). Interactions between biosynthesis, compartmentation and transport in the control of glutathione homeostasis and signalling. *J. Exp. Bot.* 53:1283–1304.

Noji M, Inoue K, Kimura N, Gouda A, et al. (1998). Isoform-dependent differences in feedback regulation and subcellular localization of serine-acetyltransferase involved in cysteine biosynthesis from *Arabidopsis thaliana***.** *J. Biol. Chem.* 273: 32739–32745.

Noji M, Takagi Y, Kimura N, Inoue K, et al. (2001). Serine acetyltransferase involved in cysteine biosynthesis from spinach: molecular cloning, characterization and expression analysis of cDNA encoding a plastidic isoform. *Plant Cell Physiol.* 42: 627–634.

Ober E and Sharp R (1994). Proline accumulation in maize (*Zea mays* L.) primary root at low water potentials. *Plant Physiol.* 105(3):981-987.

Omae H, Kumar A, Egawa Y, Kashiwaba K, et al. (2006). Adaptation to heat and drought stress in snap bean (*Phaseolus vulgaris*) during the reproductive stage of development. *Jpn. Agric. Res. Q*. 40(3): 213–216.

Page RDM (1996). TreeView: An application to display phylogenetic trees on personal computers. *Comp. Applic. Biosci.* 12: 357–358.

Phang JM (1985). The regulatory functions of proline and pyrroline-5-carboxylic acid. *Curr. Top. Cell Regul*. 25:91-132.

Penna S (2003). Building stress tolerance through over-producing trehalose in transgenic plants. *Trends Plant Sci*. 8:355-357.

Pengelly BC and Maass BL (2001). *Lablab purpureus* (L.) Sweet – diversity, potential use and determination of a core collection of this multi-purpose tropical legume. *Genet. Resour. Crop Evol.* 48:261-272.

Pramanik MH and Imai R (2005). Functional identification of a trehalose 6-phosphate phosphatase gene that is involved in transient induction of trehalose biosynthesis during chilling stress in rice. *Plant Mol. Biol*. 58(6):751-762.

Ramanjulu S and Bartels D (2002). Drought and desiccation-induced modulation of gene expression in plants. *Plant Cell Environ*. 25:141-151.

Rhodes D, Rich PJ, Myers AC, Reuter CR, et al. (1987). Determination of betaines by fast atom bombardment mass spectrometry: identification of glycine betaine deficient genotypes of *Zea mays*. *Plant Physiol.* 84: 781–788.

Rhodes D and Hanson AD (1993). Quaternary ammonium and tertiary sulfonium compounds in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 44: 357–384.

Romero LC, Domínguez-Solís JR, Gutiérrez-Alcalá G and Gotor C (2001). Salt regulation of O-acetylserine(thiol)lyase in *Arabidopsis thaliana* and increased tolerance in yeast. *Plant Physiol. Biochem.* 39:643−647.

Rontein D, Basset G and Hanson AD (2002). Metabolic engineering of osmoprotectant accumulation in plants. *Metab. Eng*. 4(1):49-56.

Ruffet ML, Droux, M and Douce R (1994). Purification and kinetic properties of serine acetyltransferase free of O-acetylserine(thiol) lyase from spinach chloroplasts. *Plant Physiol.* 104:597–604.

Ruiz JM and Blumwald E (2002). Salinity-induced glutathione synthesis in *Brassica napus*. *Planta*. 214, 965–969.

Saito K, Kurosawa M, Tatsuguchi K, Takagi Y, et al. (1995). Modulation of cysteine biosynthesis in chloroplasts of transgenic tobacco overexpressing cysteine synthase [*O*-acetylserine(thiol)-lyase] *Plant Physiol*. 106:887–895.

Saito K (2000). Regulation of sulfate transport and synthesis of sulfur-containing amino acids. *Curr. Opin. Plant Biol.* 3:188–195.

Sakamoto A and Murata N (2001). The use of bacterial choline oxidase, a glycinebetaine-synthesizing enzyme, to create stress-resistant transgenic plants. *Plant Physiol*. 125:180-188.

Satoh-Nagasawa N, Nagasawa N, Malcomber S, Sakai H, et al. (2006). A trehalose metabolic enzyme controls inflorescence architecture in maize. *Nature.* 441: 227–230.

Simon MV, Benko-Iseppon AM, Resende LV, Winter P, et al. (2007). Genetic diversity and phylogenetic relationships in *Vigna* Savi germplasm revealed by DNA amplification fingerprinting. *Genome*. 50: 538-547.

Singh BB (2005). Genetic resources, chromosome engineering and crop improvement In: Cowpea *Vigna unguiculata* (L.) Walp. (Singh RJ, Jauhar PP eds.). Boca Raton: CRC Press, 117–162.

Soares-Cavalcanti NM (2007). Estudo in silico de genes que codificam fatores de transcrição responsivos à seca, salinidade e congelamento nos genomas do eucalipto, cana e arroz, *Master's thesis*, UFPE, Recife.

Stines AP, Naylor DJ, Hoj PB and Heeswijck RV (1999). Proline accumulation in developing grapevine fruit occurs independently of changes in the levels of D1-pyrroline-5-carboxylate synthetase mRNA or protein. *Plant Physiol.*120: 923–931.

Streeter JG (1985). Nitrate inhibition of legume nodule growth and activity. Short term studies with high nitrate supply. *Plant Physiol*. 77(2):325-328.

Styer JC (2000). Regulating inositol biosynthesis in plants: Myo-inositol phosphate synthase and myo-Inositol monophosphate. *Master's thesis*, Faculty of Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Szoke A, Miao GH, Hong Z and Verma DPS (1992). Subcellular location of delta^{1}-pyrroline-5-carboxylate reductase in root/nodule and leaf of soybean. *Plant Physiol*. 99:1642-1649.

Takahashi S, Furukawa T, Asano T. Terajima Y, et al. (2005). Very close relationship of the chloroplast genomes among *Saccharum* species. *Theor. Appl. Genet.* 110: 1523–1529.

Takahashi H, Watanabe-Takahashi A, Smith FW, Blake-Kalff M, et al. (2000). The roles of three functional sulphate transporters involved in uptake and translocation of sulphate in *Arabidopsis thaliana*. *Plant J*. 23: 171-182.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature.* 408(6814):796–815.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin J, et al. (1997) The Clustal_X windows interface: flexible strategies formultiple sequencealignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876-4882.

Turchetto-Zolet AC, Margis-Pinheiro M and Margis R (2009). The evolution of pyrroline-5-carboxylate synthase in plants: a key enzyme in proline synthesis. *Mol. Genet. Genomics*. 281:87–97.

Timko MP, Ehlers JD and Roberts PA (2007). Genome mapping and molecular breeding in plants, pulses, sugar and tuber crops in cowpea. (Kole C ed.). Springer-Verlag, Berlin, 49–68.

Urano Y, Manabe T, Noji M and Saito K (2000). Molecular cloning and functional characterization of cDNAs encoding cysteine synthase and serine acetyltransferase that may be responsive for high cellular cysteine content in *Allium tuberosum. Gene.* 257: 269–277.

Veatch ME, Smith SE and Vandemark G (2004). Shoot biomass production among accessions of *Medicago truncatula* exposed to NaCl. *Plant Genet. Resour.* 44:1008–1013.

Vanrensburg L, Kruger GHJ and Kruger RH (1993). Proline accumulation as drought tolerance selection criterion: Its relationship to membrane integrity and chloroplast ultra structure in *Nicotiana tabacum* L. *J. Plant Physiol*. 141:188-194.

Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, et al. (2006). Techniques for molecular analysis. Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *Plant J.* 45:523-539.

Vettore AL, Da Silva FR, Kemper EL and Arruda P (2001). The libraries that made SUCEST. *Genet. Mol. Biol*. 24: 1–7.

Yancey PH, Clark ME, Hand SC, Bowlus RD, et al. (1982). Living with water stress: Evolution of osmolyte systems. *Science.* 217(24):1214-1222.

Yancey PH (1994). Cellular and molecular physiology of cell. In: Compatible and counteracting solutes. (Strange K, ed.) CRC Press, Boca Raton, 81–109.

Yancey PH (2001). Water Stress, Osmolytes and Proteins. *Amer. Zool.* 41:699–70.

Yoshiba Y, Kiyosue T, Nakashima K, Yamaguchi-Shinozaki K, et al. (1997). Regulation of levels of proline as an osmolyte in plants under water stress. *Plant Cell Physiol*. 38:1095-1102.

Youssefian S, Nakamura M, Orudgev E and Kondo N (2001). Increased cysteine biosynthesis capacity of transgenic tobacco overexpressing an *O*-Acetylserine(thiol) lyase modifies plant responses to oxidative stress. *Plant Physiol.* 126(3): 1001–1011.

Waldren RP and Teare ID (1974). Free proline accumulation in drought-stressed plants under laboratory conditions. *Plant Soil.* 40(3): 689–692.

Wanderley-Nogueira AC, Soares-Cavalcanti MN, Lima-Morais D, Silva LCB, et al. (2007). Abundance and diversity of resistance (R) genes in the sugarcane transcriptome. *Genet. Mol. Res*. 6(4):866-889.

Wang W, Vinocur B and Altman A (2003). Plant responses to drought salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta.* 218:1-14.

Wang YJ, Hao YJ, Zhang ZG, Chen T, et al. (2005). Isolation of trehalose-6-phosphate phosphatase gene from tobacco and its functional analysis in yeast cells. *J. Plant Physiol*. 162(2):215–223.

Weretilnyk EA, Bednarek S, McCue KF, Rhodes D, et al. (1989). Comparative biochemical and immunological studies of the glycine betaine synthesis pathway in diverse families of dicotyledons. *Planta.* 178: 342-352.

Werner JE and Finkelstein RR (1995). Arabidopsis mutants with reduced response to NaCl and osmotic stress. *Physiol. Plant*. 93:659-666.

Wingler A (2002). Molecules of Interest: The function of trehalose biosynthesis in plants. *Phytochem*. 60:437-440.

Winter E and Lauchli A (1982). Salt tolerance of *Trifolium alexandrinum* L. I. comparison of the salt response of *T. alexandrinum* and *T. pretense*. *Aust. J. Plant Physiol.* 9(2):221-226.

Zimmermann T, Zeizinger M and Burda JV (2005). Cisplatin interaction with cysteine and methionine, a theoretical DFT study. *J. Inorg. Biochem*. 99(11):2184-2196.

Internet Resources

Online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments (EGAssembler), http://egassembler.hgc.jp/

Eucalyptus EST Database (FORESTs), https://forests.esalq.usp.br

Harvest Database (HARVEST), harvest.ucr.edu/

National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov

Sugarcane EST Genome Project (SUCEST), http://sucest.lad.dcc.unicamp.br

Table 1: Type and features of osmoprotectants genes used as ***seed sequences*** against the Cowpea and Sugarcane Databases. The genes are grouped in Cowpea and Sugarcane transcriptome with respective gene name, accession number at NCBI and TIGR, size of the protein in aminoacids (aa), organism, with respective domains.

Seed sequences				Number Identification	Conserved Domain 1				Conserved Domain 2			
Gene class	Transcriptome	Organism Database	Gene name		CD Name	Size (aa)	Begin	End	CD Name	Size (aa)	Begin	End
PROLINE	Cowpea	Arabidopsis	*P5CS*	gi|2642162|	AAK	284	7	281	ProA	417	293	701
			P5CR	gi|166815|	P5CR	248	12	258	-	-	-	-
	Sugarcane	Rice	*P5CS*	12001.m12410	AAK	284	26	300	ProA	417	315	728
			P5CR	12001.m13237	P5CR	268	20	284	-	-	-	-
TREHALOSE	Cowpea	Arabidopsis	*TPS1*	gi|15218422|	Glyco_ transferase	471	92	559	TPP_ase	235	604	814
			TPPB	gi|15218205|	TPP_ase	235	121	354	-	-	-	-
	Sugarcane	Rice	*TPS1*	12001.m11619	Glyco_ transferase	470	79	579	TPP_ase	234	628	862
			TPPB	12010.m06822	TPP_ase	234	120	365	-	-	-	-
GLYCINE-BETEINE	Cowpea	Arabidopsis	*BADH*	gi|21759093|	Aldedh	464	16	486	-	-	-	-
			CMO	gi|23978945|	HcaE	367	69	406	Rieske_ RO_Alpha_CMO	118	95	212
	Sugarcane	Rice	*BADH*	12007.m09101	Aldedh	459	61	527	-	-	-	-
			CMO	12006.m09377	HcaE	367	67	399	Rieske_ RO_Alpha_CMO	118	84	201
CYSTEINE	Cowpea	Arabidopsis	*OASTL*	gi|2244845|	PALP	298	9	298	-	-	-	-
			SAT	gi|608677|	SATase_N	105	48	152	LbetaH	101	186	267
	Sugarcane	Rice	*OASTL*	12003.m06618	PALP	300	51	442	-	-	-	-
			SAT	12003.m06459	SATase_N	105	54	154	LbetaH	101	187	272
MYO-INOSITOL	Cowpea	Arabidopsis	*INPS1*	gi|1161312|	Inos-1-P_synth	388	76	495	-	-	-	-
	Sugarcane	Rice	*INPS1*	12003.m06430	Inos-1-P_synth	388	62	494	-	-	-	-

Table 2: Main Cowpea and Sugarcane clusters similar to known Osmoprotectants-genes. tBLASTn results and sequence evaluation of Cowpea and Sugarcane Osmoprotectants-genes including the **two best** matches of each gene class: (I) Features and evaluation results with cowpea and sugarcane cluster number, cluster size in nucleotides (n), ORF (Open Reading Frame) size in amino-acids (aa), e-value; score and frame. (II) Data about BLASTx best alignment: NCBI gi|-number, e-value; score;

(I) Cluster Features and Evaluation								(II) BLASTx Information					
Gene class	Transcriptome	Gene name	Cluster Nr.	Size (n)	E-value	ORF (aa)	# HITs	NCBI gi|-Nr.	E-value	Score	aa Positives (%)	Frame	Plant Species
PROLINE	Cowpea	*P5CS*	VU_UP123390	1072	e-128	189	4	gi:164564394	e-130	467	98	2	*Phaseolus vulgaris*
			VU_UP121380	1112	e-109			gi: 13161405					
		P5CR	VU_UP1215945	1272	e-110	159	1	gi:13161407	e-134	293	98	1	*Vigna unguiculata*
			-	-	-	-		-	-	-	-	-	-
	Sugarcane	*P5CS*	SCJFLR1073H12	2627	0.0	729	4	gi:34908290	0.0	1199	94	2	*Oryza sativa*
			SCCCLR2001F05	1753	0.0	419		gi:53749354	0.0	771	95	-3	*Oryza sativa*
		P5CR	SCEPRZ1011H03	1050	e-93	243	1	gi:66356280	e-135	483	95	-1	*Zea mays*
			-	-	-	-		-	-	-	-	-	-
TREHALOSE	Cowpea	*TPS1*	VU_UP1212258	1417	0.0	405	19	gi:55792822	0.0	681	89	2	*Gossypium hirsutum*
			VU_UP1216388	1494	e-157	411		gi:2245136	0.0	655	86	1	*Arabidopsis thaliana*
		TPPB	VU_UP122655	1639	e-122	369	11	gi:30683008	e-121	439	81	1	*Arabidopsis thaliana*
			VU_UP123871	1659	e-97	389		gi:51458330	e-163	577	86	3	*Nicotiana tabacum*
	Sugarcane	*TPS1*	SCQSFL1122C11	410	e-114	201	17	gi:52353687	e-123	441	98	-3	*Oryza sativa*
			SCVPFL1065E04	664	e-88	184		gi:4836875	e-89	328	92	-1	*Arabidopsis thaliana*
		TPPB	SCEZRZ3019C12	1287	e-96	262	11	gi:50945643	e-135	485	91	2	*Oryza sativa*
			SCEZHR1055G12	872	e-96	274		gi:50945643	e-147	524	97	-2	*Oryza sativa*
GLYCINE-BETEINE	Cowpea	*BADH*	VU_Contig2642	1908	0.0	364	7	gi:167962392	0.0	726	97	2	*Glycine max*
			VU_UP1211024	1996	e-58	503		gi:167962392	0.0	914	98	1	*Glycine max*
		CMO	VU_UP1227033	629	e-64	141	1	gi: 33300598	e-64	248	85	-2	*Oryza sativa*

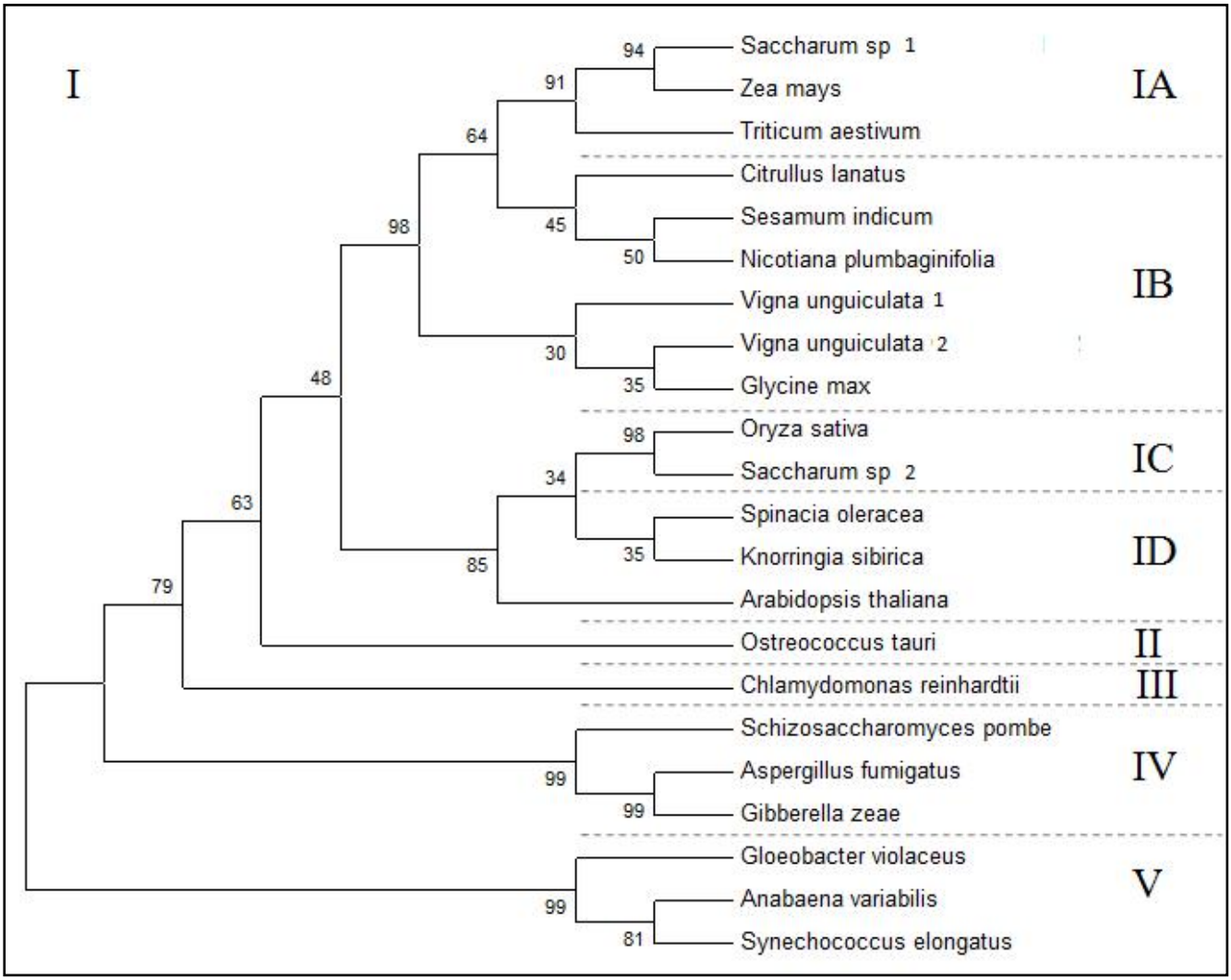
BETEINE			-	-	-	-		-	-	-	-	-	-
	Sugarcane	*BADH*	SCCCCL3002F02	2101	0.0	506	5	gi: 50950101	0.0	946	96	1	*Zoysia tenuifolia*
			SCJFRZ1007A10	1477	0.0	393		gi: 50086699	0.0	881	98	1	*Zea mays*
		CMO	SCJFHR1030B02	1114	e-119	97	2	gi: 112790161	e-180	634	96	2	*Zea mays*
			SCCCRT1C04E12	539	e-31			gi: 11279016	e-60	470	89	2	*Zea mays*
CYSTEINE	Cowpea	*OASTL*	VU_UP124226	1614	e-165	375	20	gi: 148562451	0.0	643	94	2	*Glycine max*
			VU_UP1213129	1667	e-122	387		gi: 148562457	0.0	650	94	1	*Glycine max*
		SAT	VU_UP1217378	1338	e-121	355	6	gi: 39841350	e-141	505	90	3	*Glycine max*
			VU_UP121066	1381	e-116	349		gi: 39841350	e-111	405	83	1	*Glycine max*
	Sugarcane	*OASTL*	SCCCCL3140C10	1322	e-138	325	10	gi:758353	e-162	575	98	2	*Zea mays*
			SCCCLR1072D06	1534	e-123	305		gi: 57899533	e-154	548	93	1	*Oryza sativa*
		SAT	SCCCLR1072E08	1468	e-90	318	4	gi:34910686	e-99	363	94	-1	*Oryza sativa*
			SCCCCL3080C06	1133	e-90	311		gi:25991547	e-103	377	96	3	*Zea mays*
MYO-INOSITOL	Cowpea	*INPS1*	VU_UP125689	1967	0.0	510	4	gi: 189169790	0.0	998	99	1	*Vigna radiata*
			VU_UP126259	778	e-120	240		gi: 84311237	e-128	464	98	2	*Glycine max*
	Sugarcane	*INPS1*	SCCCLR1C02F07	2527	0.0	510	2	gi:11762100	0.0	1011	99	-2	*Zea mays*
			SCJLLR1105B03	464	e-29	-		gi:11762100	e-31	135	100	-1	*Zea mays*

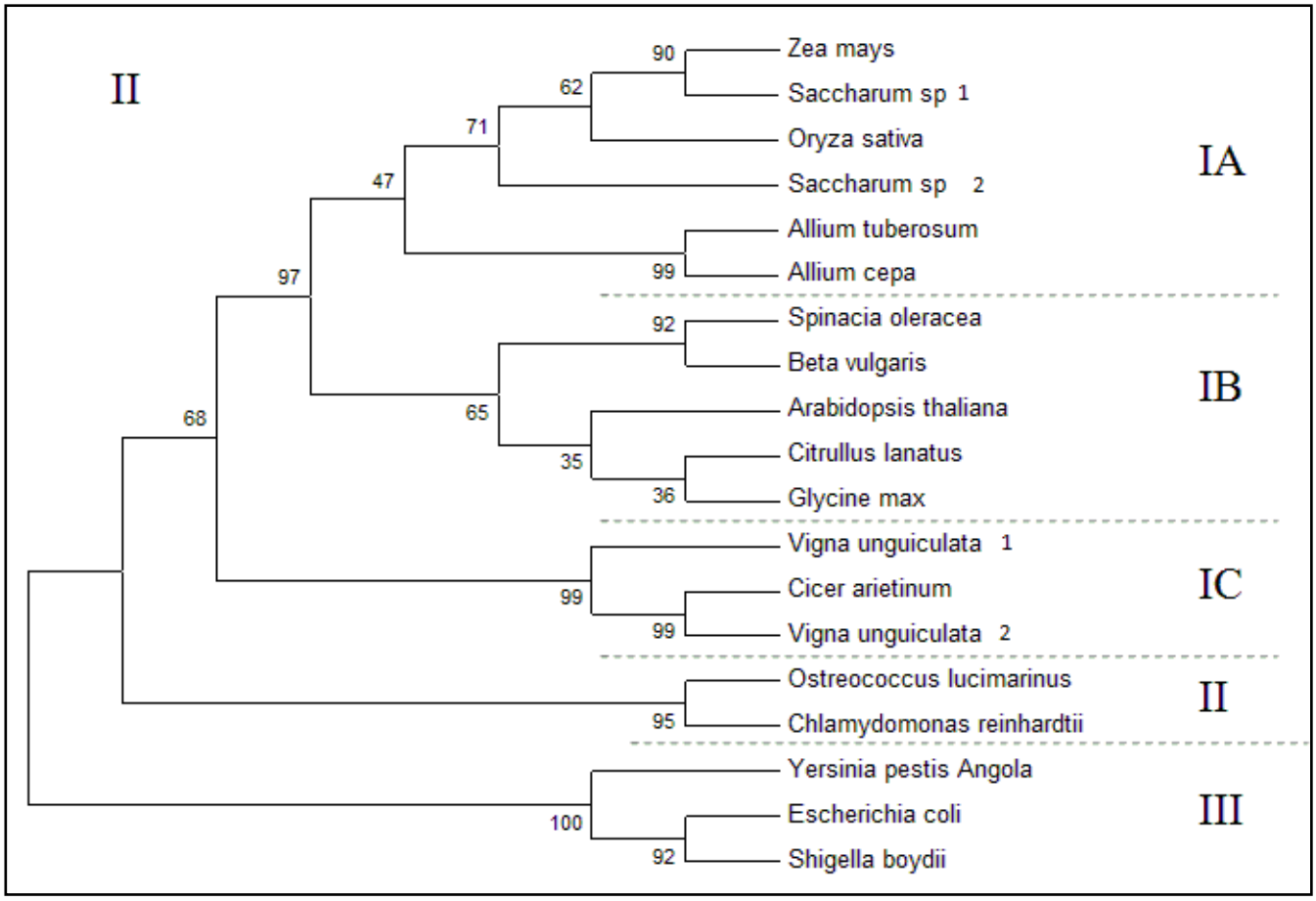
Table 3. Conserved Domains description of the best hit in Cowpea and Sugarcane Database of each osmoprotectant type, including gene name, cluster identification, Conserved Domain name (CD name) size of the CD in amino-acids (aa), ORF size in amino-acids (aa), alignment of protein (Ptn), integrity; complete and incomplete domains (CP/IC) and number (#) of hits with Complete Domain (#CP).

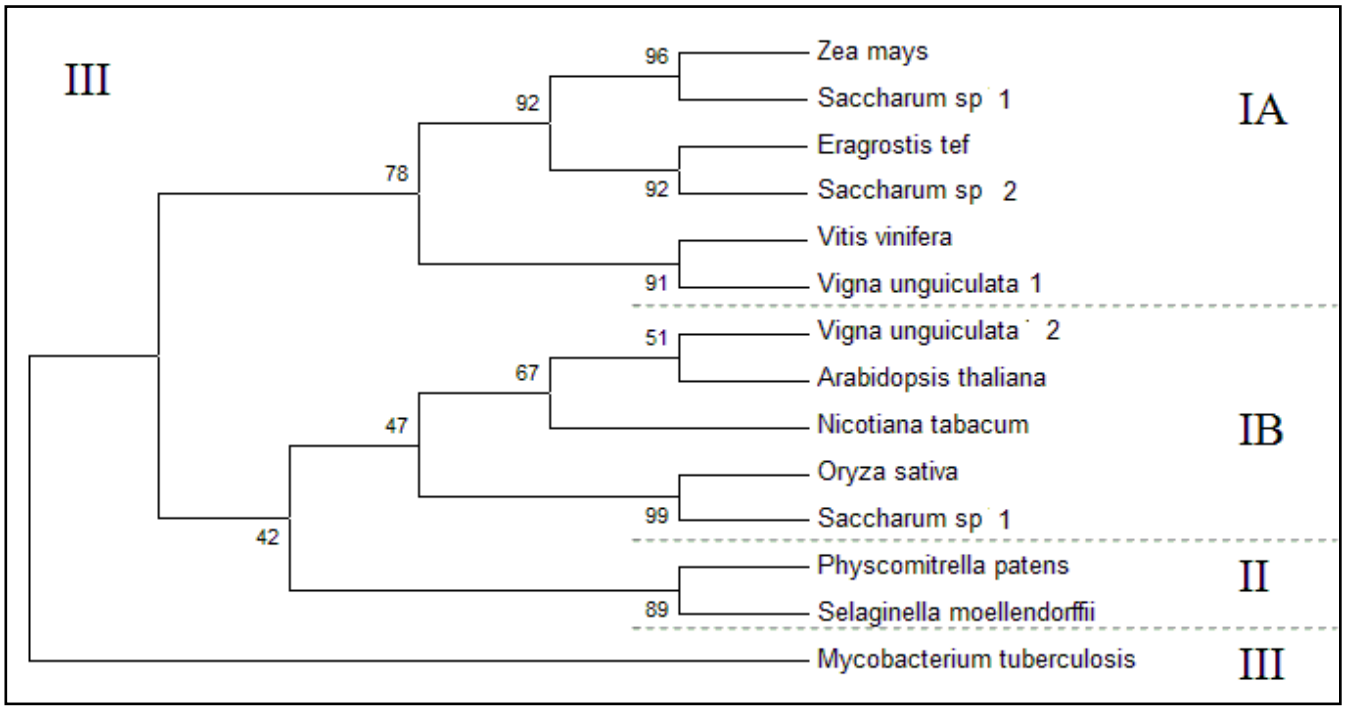
Conserved Domains (CD)				Conserved Domain 1						Conserved Domain 2				
Gene class	Transcriptome	Gene name	Cluster Identification	CD Name	Size (aa)	ORF (aa)	Alignment Ptn/CD	CP/ IC	# CP	CD Name	Size (aa)	Alignment Ptn/CD	CP/ IC	# CP
PROLINE	Cowpea	*P5CS*	VU_UP1222455	AAK	284	264	61/1 – 265/215	IC	0	ProA	417	-	-	0
		P5CR	VU_UP1215945	P5CR	248	159	2/108 – 159/266	IC	0	-	-	-	-	-
	Sugarcane	*P5CS*	SCJFLR1073H12	AAK	284	729	21/2 – 294/284	CP	1	ProA	417	309/6 – 712/415	CP	2
		P5CR	SCEPRZ1011H03	P5CR	268	243	23/2 – 243/222	CP	1	-	-	-	-	-
TREHALOSE	Cowpea	*TPS1*	VU_UP1216388	Glyco_ transferase	471	411	1/381 – 104/470	IC	0	TPP_ase	235	153/1 – 386/234	CP	4
		TPPB	VU_UP122655	TPP_ase	235	369	114/1 - 346/233	CP	2	-	-	-	-	-
	Sugarcane	*TPS1*	SCCCCL3002E04	Glyco_ transferase	470	440	1/335 – 121/460	IC	0	TPP_ase	234	171/1 – 406/234	CP	5
		TPPB	SCEZRZ3019C12	TPP_ase	234	262	2/1 - 239/235	CP	3	-	-	-	-	-
GLYCINE-BETEINE	Cowpea	*BADH*	VU_UP1211024	Aldedh	464	503	18/2 – 485/459	CP	1	-	-	-	-	-
		CMO	VU_UP12_5545	HcaE	367	260	-	-	0	Rieske_CMO	118	71/1 - 203/134	CP	1
	Sugarcane	*BADH*	SCCCCL3002F02	Aldedh	459	506	11/02 – 489/466	CP	2	-	-		-	-
		CMO	SCJFHR1030B02	HcaE	367	97	15/275 – 88/349	IC	0	Rieske_CMO	118	-	-	0
CYSTEINE	Cowpea	*OASTL*	VU_Contig1089	PALP	298	239	1/50 - 234/279	CP	5	-	-	-	-	-
		SAT	VU_UP1217378	SATase_N	105	355	87/1 - 193/105	CP	4	LbetaH	72	212/10 - 284/72	CP	3
	Sugarcane	*OASTL*	SCCCCL3140C10	PALP	300	325	8/1 - 307/299	CP	4	-	-	-	-	-
		SAT	SCCCLR1072E08	SATase_N	105	318	51/1 - 155/105	CP	2	LbetaH	101	179/1 – 270/101	CP	2
MYO-INOSITOL	Cowpea	*INPS1*	VU_UP125689	Inos-1-P_synth	376	510	61/5 – 494/361	CP	2	-	-	-	-	-
	Sugarcane	*INPS1*	SCCCLR1C02F07	Inos-1-P_synth	388	510	62/1 - 494/388	CP	1	-	-	-	-	-

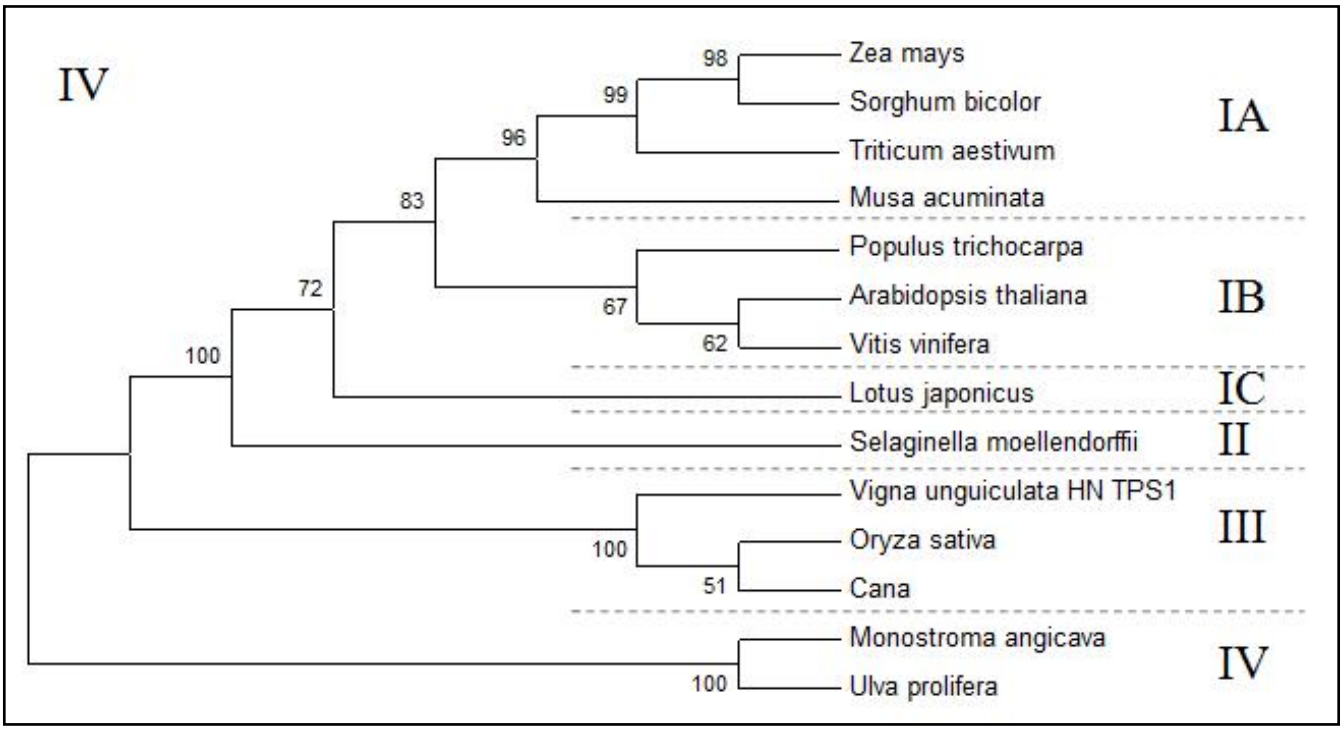
Supplementary Data

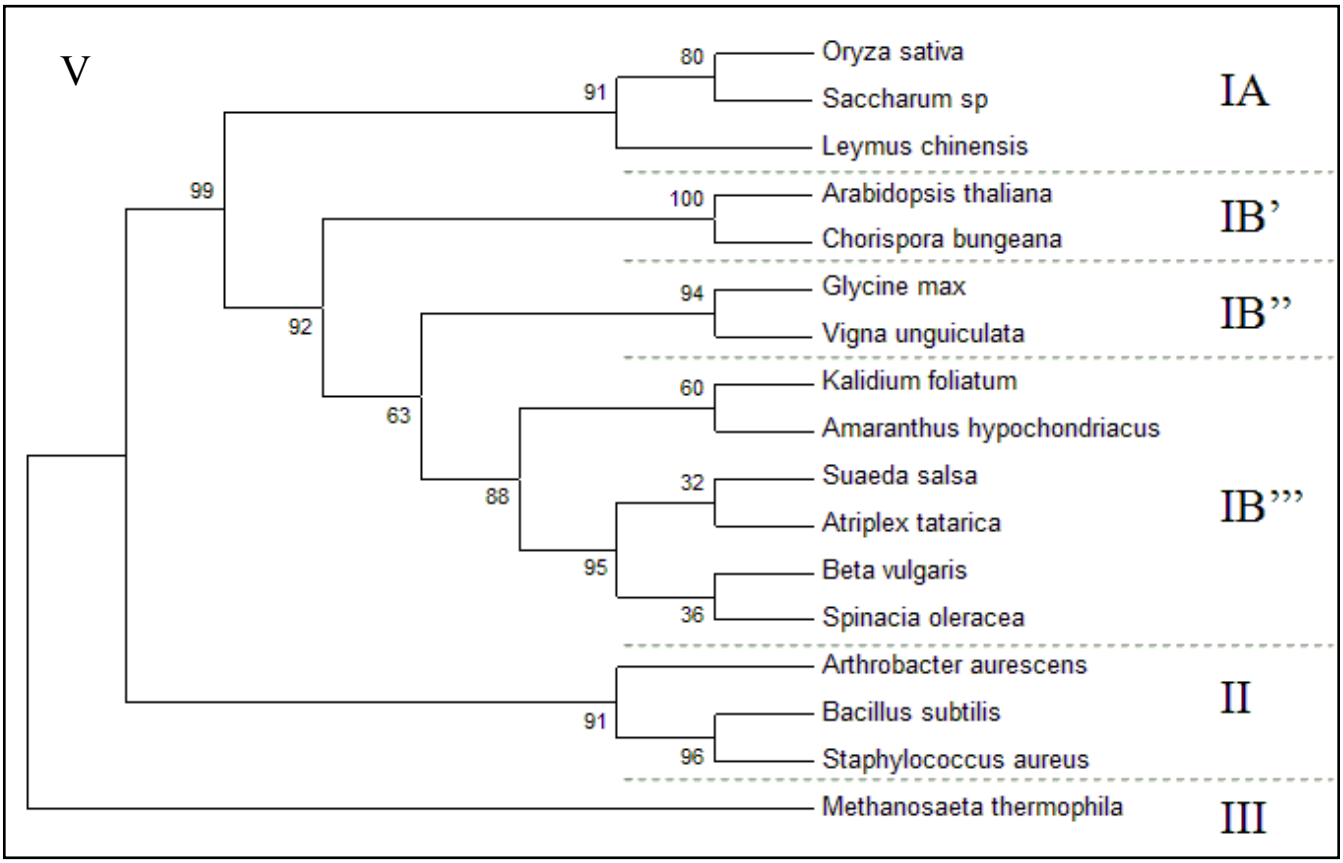
Dendrograms generated after maximum parsimony analysis illustrating relationship revealed by conserved domains similarity in (I) OASTL-like (II) SAT-like (III) TPP-like (IV) TPS-like and (V) BADH-like sequences and putative cowpea and sugarcane orthologs. The numbers in the base of clades regard bootstrap values (2,000 replications).




II
90
62
71
47
97
68
65
35
36
99
92
99
95
100
92
Zea mays
Saccharum sp 1
Oryza sativa
Saccharum sp 2
Allium tuberosum
Allium cepa
Spinacia oleracea
Beta vulgaris
Arabidopsis thaliana
Citrullus lanatus
Glycine max
Vigna unguiculata 1
Cicer arietinum
Vigna unguiculata 2
Ostreococcus lucimarinus
Chlamydomonas reinhardtii
Yersinia pestis Angola
Escherichia coli
Shigella boydii
IA
IB
IC
II
III


III
96
92
78
92
91
51
67
47
99
42
89
Zea mays
Saccharum sp 1
Eragrostis tef
Saccharum sp 2
Vitis vinifera
Vigna unguiculata 1
Vigna unguiculata 2
Arabidopsis thaliana
Nicotiana tabacum
Oryza sativa
Saccharum sp 1
Physcomitrella patens
Selaginella moellendorffii
Mycobacterium tuberculosis
IA
IB
II
III


IV
98
99
96
83
72
100
67
62
100
51
100
Zea mays
Sorghum bicolor
Triticum aestivum
Musa acuminata
Populus trichocarpa
Arabidopsis thaliana
Vitis vinifera
Lotus japonicus
Selaginella moellendorffii
Vigna unguiculata HN TPS1
Oryza sativa
Cana
Monostroma angicava
Ulva prolifera
IA
IB
IC
II
III
IV


V
80
91
99
100
94
92
60
63
32
88
95
36
91
96
Oryza sativa
Saccharum sp
Leymus chinensis
Arabidopsis thaliana
Chorispora bungeana
Glycine max
Vigna unguiculata
Kalidium foliatum
Amaranthus hypochondriacus
Suaeda salsa
Atriplex tatarica
Beta vulgaris
Spinacia oleracea
Arthrobacter aurescens
Bacillus subtilis
Staphylococcus aureus
Methanosaeta thermophila
IA
IB'
IB''
IB'''
II
III

CONCLUSÕES GERAIS

- As culturas do eucalipto, cana-de-açúcar e feijão-caupi apresentam em seus transcriptomas, representantes de todas as categorias de osmoprotetores estudados, apontando para existência de mecanismos de resistência/tolerância a estresses bióticos e abióticos similares aos processos amplamente conhecidos em plantas modelo, como *Arabidopsis thaliana*.

- Considerando cada categoria de osmólitos compatíveis, a maioria das proteínas apresenta alto grau de conservação de caracteres, especialmente na região dos motivos e domínios, confirmando a ancestralidade da osmoproteção nos principais grupos de organismos.

- Em eucalipto, a expressão de osmoprotetores ocorre com maior intensidade em tecidos de plântulas cultivadas no escuro, caules de mudas sob déficit hídrico e além de calos formados no escuro, confirmando a hipótese de que diferentes tipos de estresse abiótico induzem a expressão destas proteínas.

- No feijão-caupi a maior expressão de osmoprotetores se dá em raiz de plantas sensíveis à salinidade após duas horas de exposição ao estresse, tecido este que primeiro percebe e responde às mudanças nas condições ambientais.

- Em contrapartida ao observado nos transcriptomas já descritos, cana-de-açúcar expressa os genes envolvidos na osmoproteção tanto sob condições de estresse biótico, quanto abiótico; com grande abundância de transcritos em tecidos de calo submetidos a ciclos de luz/escuro e baixas temperaturas e tecidos infectados por *Herbaspirilum rubrisubalbicans*.

- O número de *reads* encontradas no NordEST foi inferior ao encontrado nos outros projetos, e isso se deve ao fato do projeto ainda não estar completo, apesar disso pôde-se perceber que o estudo com indivíduos controle e àqueles submetidos a estresse é satisfatório visto que alguns genes ligados a osmoproteção são mais expressos em tecidos submetidos a estresse, nesse caso salino.

- A topologia dos dendrogramas gerados com os representantes do eucalipto e outros organismos mostram que a evolução dos genes envolvidos na osmoproteção aconteceu de acordo com processos seletivos e adaptativos em resposta as pressões ambientais.

- Os dendrogramas incluindo cana e caupi, que agruparam os organismos de acordo com suas proximidades taxonômicas, sugerem que estes genes além de sujeitos a fortes pressões seletivas também estão associados à evolução convergente.

ANEXOS

Instruções para Autores

Instructions for Authors

Genetics and Molecular Research (GMR) publishes research articles, research reports, technical notes, scientific commentaries, news, views, and review articles with regard to genetics, evolution, molecular biology, and bioinformatics. Review articles are normally received by invitation only. If you would like for us to consider a review article, please consult the editor first; send a proposed title, a brief outline and a list of papers relevant to the review submitted by the author(s). GMR is an exclusively online journal.

The journal is maintained by the not-for-profit scientific foundation Ribeirão Preto Foundation for Scientific Research (FUNPEC-RP). **The fee per accepted submission is R$ 800.00 for Brazilian authors and US$450.00 for authors from other countries. The US dollar amount reflects the approximate current foreign exchange rate and is subject to change.** This fee covers part of the expenses for final language and technical revision, for page setup, and for publishing online.

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

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

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

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

Contributions should be sent either by e-mail as attachments to [gmr@funpecrp.com.br] or by regular mail to: Prof. Dr. Francisco A. Moura Duarte, Editor. Av. Presidente Vargas, 2627, 2º andar - Condomínio Itamarati, 14020-260 Ribeirão Preto, SP, Brazil.

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

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

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

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

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

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

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

Submission information

Authors are required to provide the following information with their electronic submissions:

Author submitting the article; article title; authors (full list); article type and session; status of article (e.g., new, revised, etc.); postal address; e-mail address; phone number; fax number; names and types of the files sent.

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

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

Acknowledgment of electronic submissions: Successful receipt and processing of the author's submission will be acknowledged by e-mail when the submitted manuscript has been

checked. If no reply has been received within one week, the author should contact the editor at [gmr@funpecrp.com.br].

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

A review article should contain: an abstract of 250 words or less, no more than 6 key words, a running title and no more than 60 references. It should be divided into sections with appropriate tittles and subtitles.

Preparation of the manuscript

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

Title Page: The title page should include the title of the article, authors' names, and authors' affiliation. The affiliation should comprise the department, institution (usually university or company), city, and state (or nation). The title page should include the name and complete mailing address, telephone number, fax number, and e-mail address of the author designated to review proofs. The title page should start below the top margin, be single-spaced, and have no space left before the Summary/Abstract. A running title of no more than 60 characters (including spaces) should be provided.

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

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

Text

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

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

Tables/Charts: Special care should be taken to ensure that all tables are properly formatted. Scientific symbols used should be in Symbol or Times New Roman. Tables should

be on a separate page, numbered consecutively (with Arabic numerals) referred to by number in the text and designed to fit the column or page size of the journal. Use tables with cells to separate columns. Do not use spaces, tabs or vertical lines. Left justify the title above the table. Indicate each table's location within the manuscript.

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

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

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

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

Reference, under the heading "References", should include only works referred to in the text. It should be arranged in alphabetical order under the first author's last name. References should be cited as follows: journal papers - names and initials of the first four authors (after that using et al.), year, full title, journal abbreviated according to PubMed or Web of Science, volume number, first and last page numbers; books - names of authors, year, full title, edition, publishers, address (city); articles published in symposia - names of authors, year, full title of book, name(s) of editor(s) in parentheses, publisher, address (city), first and last page numbers.

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

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

Examples of reference style

Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Examples of bibliographic style

JOURNAL (Standard article). Ex.:

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

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

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

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

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

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

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

REPORT. Ex.:

CIAT (International Center for Tropical Agriculture) (1980). Annual Report. CIAT, Cali.
FAO (Food and Agricultural Organization) (2004). Production Yearbook. FAO, Rome.

THESIS. Ex.:

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

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

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

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

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

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

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

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

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

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

HTML. HyperText Markup Language. http://www.w3.org/TR/html401/. Accessed August 23, 2007.

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

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