UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS–GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS
CENTRO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE GENÉTICA

SuperSAGE: identificação e análise de genes super e sub regulados em Soja (*Glycine max*) sob condições de desidratação radicular

Recife, PE
Fevereiro, 2014

JOSÉ RIBAMAR COSTA FERREIRA NETO

SuperSAGE: identificação e análise de genes super e sub regulados em Soja (*Glycine max*) sob condições de desidratação radicular

Tese para obtenção do grau de Doutor(a) em Ciências Biológicas, área de concentração "Biologia Molecular", da Universidade Federal de Pernambuco .

Orientador: Prof. Dr. Ederson Akio Kido, Depto. de Genética, Centro de Ciências Biológicas UFPE.
Coorientadora: Profª. Drª. Ana Maria Benko-Iseppon, Depto. de Genética, Centro de Ciências Biológicas UFPE.

Recife
Fevereiro, 2014

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

Ferreira Neto, José Ribamar Costa
SuperSAGE: identificação e análise de genes super e sub regulados em soja (*Glycine max*) sob condições de desidratação radicular / José Ribamar Costa Ferreira Neto. – Recife: O Autor, 2014.

69 folhas: il.
Orientadores: Ederson Akio Kido, Ana Maria Benko-Iseppon
Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Ciências Biológicas. Programa de Pós-graduação em Ciências Biológicas, 2014.
Inclui bibliografia

1. Plantas – Melhoramento genético 2. Soja I. Melo Neto, Osvaldo (orient.) II. Benko-Iseppon, Ana Maria (coorient.) III. Título.

631.52 CDD (22.ed.) UFPE/CCB-2014-142

José Ribamar Costa Ferreira Neto

SuperSAGE: identificação e análise de genes super e sub regulados em Soja (*Glycine max*) sob condições de desidratação radicular

Comissão Julgadora da Tese para obtenção do grau de Doutor em Ciências Biológicas

_____

Ederson Akio Kido
Departamento de Genética – Universidade Federal de Pernambuco
Presidente da Banca

_____

Francismar Corrêa Marcelino-Guimaraes
EMBRAPA. Centro Nacional de Pesquisa de Soja

_____

Sergio Crovella
Departamento de Genética – Universidade Federal de Pernambuco

_____

Péricles de Albuquerque Melo Filho
Departamento de Agronomia **–** Universidade Federal Rural de Pernambuco

_____

Rejane Jurema Mansur Custódio Nogueira
Departamento de Biologia **–** Universidade Federal Rural de Pernambuco

Recife, 28 de Fevereiro, 2014

Aos meus pais (Maria Alice de Oliveira Costa Ferreira e José Ribamar Costa Ferreira Júnior), aos meus avós [Raimunda Araújo Costa Ferreira e José Ribamar Costa Ferreira (in memorian)] e a meus irmãos (Marcell de Oliveira Costa Ferreira e Ana Alice de Oliveira Costa Ferreira), ***dedico***.

AGRADECIMENTOS

A Deus e à Nossa Senhora pela ajuda e pela graça concedida;

À minha avó (Raimunda Araújo Costa Ferreira) e ao meu avô (José Ribamar Costa Ferreira; in memoriam) pelo direcionamento aos estudos, incentivo, auxílio e presença nos momentos cruciais de minha vida;

Aos meus pais (José Ribamar Costa Ferreira Júnior e Maria Alice de Oliveira Costa Ferreira), por todo suporte ao longo dessa árdua caminhada;

Aos meus irmãos (Marcell de Oliveira Costa Ferreira e Ana Alice de Oliveira Costa Ferreira), pelos diálogos, atenção e força;

Ao meu grande amigo, Mauro Sérgio Pavão Soares, pela amizade, atenção, longas conversas e pelo meu afilhado (Pedro Lucas);

Ao meu orientador, Éderson Akio Kido, pela amizade ao longo desses oito anos;

À professora Ana Maria Benko-Iseppon pelo auxilio e habitual presteza ao longo do doutorado;

À Valesca Pandolfi, pela amizade e por seu altruísmo;

À Maria da Conceição Teixeira, pela atenção e carinho;

À professora Francismar Corrêa Marcelino-Guimaraes e à Cynara Romero pela ajuda e aprendizado no que se refere à PCR quantitativa em tempo real;

À Erika Araújo pelas ótimas e longas conversas durante o atribulado período de finalização dos detalhes desta Tese.

Meus sinceros agradecimentos!

Sumário

Lista de Figuras

Página

Estado da Arte

Capítulo I "Early Transcriptional Response of Soybean Contrasting Accessions to Root Dehydration".

Lista de Tabelas

Página

Estado da Arte

Lista de abreviações

ABA	Abscisic acid
ABRE	ACGT-containing ABA response elements
ATP	Adenosine 5'-triphosphate
BADH	Glycine betaine
BMY	β-amylase
CBL	Calcineurin B-like protein
CDPK	Calcium-dependent protein kinase
DR	Down-regulated
DREB	Dehydration-responsive element binding
EREBP	Ethylene-responsive element binding protein
ERF	Ethylene response factor
EST	Expressed sequence tag
FC	Fold change
GO	Gene ontology
GPX5	Glutathione Peroxidase 5
HGMR	3-hydroxy-3-methylglutaryl coenzyme A reductase
MIPS	L- Myo-Inositol 1-Phosphate Synthase
n.s.	Not-significant ($p < 0.05$)
NAC3	NAM, ATAF, and CUC (NAC) transcription factors
NGS	Next Generation Sequencing
P5CR	Pyrroline-5-carboxylate reductase
P5CS	Δ 1-pyrroline-5-carboxylate synthetase
PCR	Polymerase chain reaction
PER	Peroxidase

PIP	Plasma intrinsic protein
RAP2-4	Ethylene-responsive transcription factor RAP2-4
ROS	Reactive oxigen species
RT-qPCR	Real time quantitative PCR
SNPs	Single nucleotide polymorphisms
TPS1	Trehalose-6-phosphate synthase 1
TSM	Tolerating a single mismatch
UR	Up-regulated
UTR	Untranslated region
XET	Xyloglucan endotransglycosylase

Resumo

A soja (*Glycine max*) é o principal produto do agronegócio nacional. A despeito de sua importância estratégica e dos progressos de seus programas de melhoramento, essa cultura ainda sofre severas perdas devido à ocorrência de condições adversas, com ênfase à seca. Desta forma, a adição de tecnologias que propiciem o uso de novas ferramentas no melhoramento, se faz imperativo. Os trabalhos aqui apresentados tiveram como objetivos analisar os transcriptomas de acessos contrastantes (tolerante e sensível) de soja, submetidos à desidratação radicular, visando o seu entendimento e mineração de possíveis alvos moleculares para aplicação no melhoramento da referida espécie. Análises globais dos transcriptomas produzidos indicaram que tais acessos expressaram 120.770 *tags* DeepSuperSAGE, dos quais 1.127 foram induzidos somente no acesso tolerante, enquanto que 1.557 foram induzidos em ambos, quando comparados aos acessos seus respectivos controles. A categorização do transcriptoma via ferramenta *Gene Ontology*, indicou que CDPK e CBL foram as famílias proteicas com maior número de representantes induzidos para a categoria "Função Molecular". Adicionalmente, a validação de nove transcritos (HMGR, XET, WRKY20, RAP2-4, EREBP, NAC3, PER, GPX5 e BMY) indicou que a partir de 25 minutos de submissão do estresse os transcriptomas dos acessos analisados já são submetidos à reorganização para enfrentar a nova condição imposta. Em outro trabalho, foi realizada mineração de dados nas bibliotecas analisadas, buscando-se transcritos específicos associados à osmoproteção. Foram encontrados 35 unitags diferencialmente expressas associadas a quatro classes de osmoprotetores, sendo: prolina (P5CS: 4 transcritos, P5CR: 2 transcritos), trealose (TPS1: 9, TPPB: 1), glicina betaina (BADH: 4) e myo-inositol (MIPS: 7, INPS1: 8)], os quais foram mapeados em 25 loci no genoma de soja. A análise da genômica estrutural dos genes ancoradores das referidas unitags indicou que os mesmos estão presentes em todos os cromossomos da espécie analisada, com alta densidade para algumas regiões subterminais e sintenia observada entre alguns pares cromossômicos. Tais genes e loci são importantes alvos para futuros ensaios, visando obtenção de maior tolerância ao estresse aqui analisado. Dessa forma os trabalhos aqui apresentados fornecem recursos inéditos de genômica funcional para aplicação no melhoramento de soja, sendo que, o desdobramento biotecnológico dos dados aqui obtidos auxiliará na manutenção da estabilidade da cadeia produtiva dessa cultura e diminuição de perdas econômicas.

Palavras-chave: transcriptômica, tolerância, estresse, seca.

Abstract

Soybean (*Glycine max*) is the main crop of the Brazilian agribusiness. Despite the success of their breeding programs, this crop still suffers severe losses due to adverse conditions, with emphasis on drought. Thus, the adoption of new technologies enabling the breeding effort is essential. This work aimed to analyze the transcriptomes of contrasting soybean accessions (tolerant and sensitive to drought), undergoing root dehydration, trying to identify potential molecular targets for use in breeding programs. Global transcriptomes analyzes indicated 120,770 unique DeepSuperSAGE tags (26 bp) expressed, of which 1,127 were induced in the tolerant accession while 1,557 were induced in both accessions when compared to their respective controls. The Gene Ontology categorization of the transcripts anchoring unitags indicated CDPK and CBL as the protein families with the largest number of representatives in the "Molecular Function" category. Additionally, nine transcripts (HMGR, XET, WRKY20, Rap2-4, EREBP, NAC3, PER, and GPX5, BMY) validated by RT-qPCR indicated that after 25 minutes of stress submission the transcriptomes of both accessions were undergoing reorganization responding to the new condition imposed. The data mining in the annotated unitags, searching for transcripts related to osmoprotectants, resulted in 35 differentially expressed unitags in four groups: proline (P5CS: 4 unitags; P5CR: 2), trehalose (TPS1: 9; TPPB: 1), glycine betaine (BADH: 4) and myo-inositol (MIPS: 7; INPS1: 8). Those unitags were mapped at 25 loci on the soybean genome (Phytozome assembly). The structural genomics analysis comprising the genes anchoring unitags indicated their location all over the soybean chromosomes, with high density for some subterminal regions and strong synteny between legume chromosomes. Such genes and loci are valuable targets for future trials, aiming to achieve higher stress tolerance. Thus, the works presented here provide novel functional genomics insights and resources for future use in soybean breeding programs, helping the agribusiness productive sector with reduction of potential financial losses.

Keywords: transcriptomics, tolerance, stress, drought.

1. Introdução

A soja (*Glycine max* L. Merrill) é um vegetal de reconhecida importância mundial, do qual se obtém proveitos em diversos segmentos. É amplamente utilizada na alimentação humana, por conter macro e micronutrientes essenciais, além de responder por mais da metade da produção global de óleo vegetal e representar uma fonte potencial para a produção de biodiesel (Wilson, 2008; USDA, 2013).

Em nosso país, a soja é considerada o principal produto do agronegócio, com negociações anuais que ultrapassam US$ 20 bilhões (MAPA, 2013). Entretanto, apesar desse *status* e de ser uma das leguminosas mais bem estudadas, o agronegócio do complexo soja (óleo, farelo e farinha) tem sofrido perdas consideráveis devido a estresses abióticos, com ênfase para a seca (Valliyodan e Nguyen, 2008). Nos EUA, existem relatos de perdas ao redor de 40% provocadas pela estiagem (Specht, 1999). No Brasil, a safra de 2004/2005 foi severamente prejudicada na região sul (detentora de 40% da produção de soja no país) devido a esse fenômeno, levando a uma redução de 25% na produção dessa região, acarretando em prejuízos de 2,32 bilhões de dólares (Martins *et al.,* 2008). Dados recentes de órgãos oficiais, como Ministério da Agricultura e Pecuária (MAPA) e Instituto Brasileiro de Geografia e Estatística ainda não estão disponíveis, entretanto há relatos em periódicos não científicos, que abordam temas associados à produção agropecuária, indicando que, na safra de soja 2011/2012, essa mesma região foi novamente afetada por períodos de seca, sendo que os 75,3 milhões de toneladas registradas na safra passada (2010/2011) foram reduzidos a 65,2 milhões de toneladas (Revista Globo Rural, 2012). Infelizmente, a tendência deste panorama é cada vez pior, tendo em vista as perspectivas de mudanças climáticas (aquecimento global; Manalavan *et al*., 2009).

Devido ao exposto, o melhoramento visando uma maior adaptação da cultura da soja às condições de déficit hídrico se faz necessário. Consequentemente, estudos relacionados à genética, fisiologia e biologia molecular dos mecanismos de defesa, que sustentam as plantas sob essa condição, são fundamentais para o desenvolvimento de variedades adaptadas a condições não ideais, uma vez que as características associadas à tolerância / resistência a estresses são controladas por múltiplos genes (ou, do inglês, *Quantitative Trait Loci* - *QTLs*), sendo dificultosas para o melhoramento tradicional (Pathan *et al*., 2007; Valliyodan e Nguyen, 2008).

Análises em alta escala do transcriptoma figuram entre as alternativas mais utilizadas para um melhor entendimento molecular de questões biológicas, se tornando uma abordagem de elevada importância aos pesquisadores que anseiam estudar como os organismos elaboram suas respostas e se comportam sob determinadas situações. Dentre as diversas metodologias disponíveis, a técnica de SuperSAGE (e seus derivados) representa uma das mais eficientes, por apresentar arquitetura aberta e alto rendimento, especificidade e sensibilidade (Matsumura *et al*., 2008; Molina *et al*., 2008; Terauchi *et al.*, 2008). Assim, o emprego dessa técnica para identificação de genes de soja que confiram melhor desempenho frente à seca, poderá proporcionar informações valiosas as quais contribuirão, sobremaneira, para um melhor entendimento dos mecanismos envolvidos, gerando informações inéditas para aplicação no melhoramento da referida espécie.

2. Objetivos

2.1 Objetivo Geral

Analisar padrões de transcrição de diferentes acessos de soja (tolerante e sensível), sob condições de desidratação radicular, propiciando o estabelecimento de estratégias para incorporação de genes importantes a cultivares comerciais.

2.2. Objetivos específicos

2.2.1. Identificação de *unitags* (sequencias expressas de 26 pares de bases) diferencialmente transcritas, a partir de bibliotecas SuperSAGE, utilizando-se de métodos estatísticos ($p < 0,05$);

2.2.2. Anotar *unitags* das diferentes bibliotecas por meio de ferramentas de bioinformática, utilizando diversos bancos de dados públicos;

2.2.3. Categorizar os transcriptomas expressos, visando o entendimento da orquestração de seus transcritos constituintes;

2.2.4. Identificar genes-candidatos relacionados ao estresse de desidratação radicular e que possam estar envolvidos nas vias metabólicas principais e responsivas ao estresse aqui analisado;

2.2.5. Validar a expressão de *unitags* diferencialmente transcritas via RT-qPCR;

3. Estado da Arte

3.1. Soja – Importância de um vegetal multiuso

A soja (*G. max*) é um vegetal de grande importância econômica, pertencente à família das leguminosas (Fabaceae). A demanda por essa cultura é alta e continua a crescer devido à sua utilização na fabricação de gêneros alimentícios e outros produtos industrializados (Wilson, 2008). É amplamente utilizada na alimentação humana e animal por ser fonte de proteínas de alta qualidade, aminoácidos essenciais, vitaminas, ácidos graxos essenciais (como o ômega 3), fitoquímicos e fármacos, além de auxiliar na redução do colesterol e no combate a diabetes, obesidade e certos tipos de câncer (Singh *et al.*, 2007). Possui empregabilidade, também, na fabricação de corantes e cosméticos, além da obtenção de biodiesel, tornando-se uma alternativa renovável e menos poluidora ao uso de combustíveis fósseis (Singh *et al.*, 2007). Atualmente, responde por 56,7% da produção mundial de óleo vegetal (USDA, 2013); além disso, segundo a *Global Industry Analysts, Inc* (2013), tal cultura alcançará a movimentação anual de US$ 43,2 bilhões de dólares, no ano de 2015, somente considerando derivados alimentícios da soja.

O Brasil é o maior produtor de soja do globo, exportando cerca de 34,8 milhões de toneladas métricas anuais, seguido pelos Estados Unidos da América, Argentina e Paraguai, respectivamente (Soystats, 2013). Em nosso país, a soja é a cultura que mais cresceu nas últimas três décadas e corresponde a 49% da área plantada em grãos (MAPA, 2013). O complexo soja (grão, farelo e óleo) é o principal gerador de divisas cambiais, com negociações anuais que ultrapassam US$ 20 bilhões (MAPA, 2013a). Previsões indicam que em 2019, a produção nacional deve representar 40% do comércio

mundial do grão e 73% do óleo de soja (MAPA, 2013a). Segundo previsões do Ministério da Agricultura, Pecuária e Abastecimento (2013b), o Valor Bruto da Produção (VBP) das 20 principais culturas agrícolas do país deverá alcançar o recorde de R$ 277,2 bilhões (98% do total), em 2019, sendo o complexo da soja responsável por 89,3 bilhões (32,2% do total).

Com o *status* de principal cultura vegetal produzida no país, e com índices elevados de crescimento previsto para os próximos anos, a soja é uma das leguminosas de maior interesse para a realização de pesquisas. Para tanto, consideráveis investimentos foram e continuam sendo feitos por diversos países com o intuito de obter informações - incluindo o sequenciamento de todo o genoma (Schmutz *et al*., 2010) – para aplicação direta no melhoramento dessa espécie. No caso do Brasil, a EMBRAPA – Soja, em associação com outras instituições, dentre as quais a UFPE, é responsável pelo consórcio GENOSOJA, o qual se apresenta como uma contraparte brasileira a um esforço internacional para estudos de genômica do referido vegetal, incluindo estudos do genoma da soja em nível estrutural e funcional, sequenciando e caracterizando importantes regiões e seus produtos. Deste modo, contribui para a identificação de genes diferencialmente expressos no transcriptoma, considerando, especialmente, estresses bióticos e abióticos que afetam essa cultura no hemisfério sul (Benko-Iseppon *et al*., 2012).

3.2. Soja e seca: problemática e desenvolvimento científico atual

Os vegetais são sujeitos a uma gama de estresses bióticos e abióticos que afetam seu desenvolvimento, com árduas consequências à produção agrícola, aos produtores e à

economia (Valliyodan e Nguyen, 2008; Manalavan *et al.,* 2009), sendo a seca o principal fator limitante da produção agrícola em nosso planeta (Valliyodan e Nguyen, 2008). Globalmente, a área cultivada com soja ocupa mais de 100 milhões de hectares, sendo que Brasil e Estados Unidos, países frequentemente submetidos a esse fenômeno, são responsáveis por cerca de metade desse quantitativo (FAO, 2012). Estudos realizados em campo e em ambientes com condições controladas demonstram que tal situação reduz severamente a produção de grãos (de 25 a 50%). Revisão aprofundada sobre tais dados pode ser vista em Ku *et al*. (2013).

No Brasil, a seca na região Sul afetou severamente a safra de soja 2011/2012, sendo que os 75,3 milhões de toneladas registradas na safra passada foram reduzidos a 65,2 milhões de toneladas, naquele ano. A produtividade média nacional, que na safra anterior foi de 51,9 sacas por hectare, caiu para 43,3 sacas por hectare (Revista Globo Rural, 2012). Nos anos de 2004/2005, esses dados chegaram a ter proporções maiores devido a períodos de seca mais significativos, levando a perdas de cerca de 25% e a prejuízos de 2,32 bilhões de dólares na região Sul, detentora de 40% da produção brasileira de soja, à época (Martins *et al.,* 2008).

Um terço da população mundial reside em regiões com déficit hídrico e, devido aos elevados níveis de CO_2 na atmosfera e às mudanças climáticas (como o aquecimento global), prevê-se que a seca se tornará mais frequente e severa no futuro (Valliyodan e Nguyen, 2008). Devido a essa perspectiva, o melhoramento visando melhor adaptação às condições de déficit hídrico se torna uma boa opção para redução de perdas (Martins *et al.,* 2008). Apesar de largos recursos serem destinados ao melhoramento de soja, poucos progressos têm sido feitos em relação à tolerância à seca. Para tanto, o entendimento de como as plantas respondem ao déficit hídrico a nível molecular (orquestração gênica), celular e fisiológico torna-se fundamental para o

desenvolvimento de variedades tolerantes (Bray, 2004; Shinozaki e Yamaguchi-Schinozaki, 2007; Pathan *et al.*, 2007).

3.3. Métodos para análise de transcriptoma

O genoma das células eucariotas contém milhares de genes cuja expressão é precisamente regulada para mantê-las vivas e funcionais. A detecção de mudanças na atividade gênica é usada para definir a identidade, função e estado de proliferação das células, sendo o perfil de expressão dos genes frequentemente correlacionado com a presença ou ausência de seus RNAs mensageiros (RNAm) correspondentes nas células. Desta forma, métodos para a apuração dos níveis de RNAm revelam o padrão transcricional temporal e espacial de genes e permitem a correlação da atividade gênica com processos biológicos e doenças. Um total de 27 métodos para análise da expressão gênica em larga escala (chamado também de Transcriptoma) são descritos por Shimkets *et al*. (2004). Tais metodologias podem ser divididas em duas categorias principais (Terauchi *et al.,* 2008):

a. Baseadas em hibridização de fitas de nucleotídeos complementares a sequencias alvos imobilizadas (tais como cDNAs, amplicons ou oligonucleotídeos);

b. Baseadas no sequenciamento e contagem de transcritos.

O principal representante do primeiro grupo é a técnica de *Microarray* (Schena *et al.,* 1995). Resumidamente, esse método, baseado em "chips" de DNA, analisa a hibridização de alvos fluorescentes (representados por RNAs mensageiros das células sob análise) a microscópicos grupos de sondas de cDNAs de genes específicos, colados quimicamente a uma região fixa chamada *spot*. A análise da hibridização sonda-alvo é

realizada por meio de um sistema informativo baseado em cores capturadas por meio de um leitor (ou scanner) que consiste de alguns lasers, um microscópio especial e uma câmera.

A técnica de *microarray* ao longo dos anos tem contribuído grandemente para análises de transcriptoma em espécies-modelo sendo ainda considerada uma técnica padrão para análise de expressão gênica. Entretanto, análises utilizando essa metodologia, em espécies cultivadas, as quais constituem a maioria das espécies, ainda é considerada dificultosa uma vez que para o desenho de tal ensaio é requerido um trabalho laborioso e financeiramente oneroso para obtenção de ESTs (Matsumura *et al.*, 2008). Muitas outras desvantagens existem em relação a essa técnica, a citar: pobre correlação entre diferentes plataformas de *microarray*; resultado semi-quantitativo; tem como pré-requisito a necessidade de grandes quantidades de RNA para a obtenção de respostas robustas; hibridização cruzada de sondas com diferentes alvos; detecção dificultosa de transcritos com número de cópias abaixo de 50 unidades por célula; ambiguidade na análise de dados e interpretação dos resultados, dentre outras (Matsumura *et al.*, 2008; Murphy, 2002).

No que tange ao segundo grupo, os principais representantes são as análises de bibliotecas ESTs (*Expressed Sequence Tags*; Adams *et al*., 1991), SAGE (*Serial Analyses of Gene Expression*; Velculescu *et al*., 1995), MPSS (*Massively Parallel Signature Sequencing*; Brenner *et al.*, 2000) e, recentemente, RNA-Seq (Morin *et al*., 2008). A análise de bibliotecas ESTs é um método comum para averiguar o padrão de expressão gênica, embora o objetivo primário embutido na obtenção dessas seja a geração de sequencias gênicas (Meyers *et al.*, 2004). Com o sequenciamento de milhares de clones, o número de transcritos de diferentes genes pode ser enumerado através de ferramentas de bioinformática, sendo sua frequência a representação do perfil

de expressão desses na amostra biológica em estudo (Terauchi *et al.*, 2008). Apesar da ampla aplicabilidade dessa técnica em estudos genômicos, essa metodologia apresenta diversas limitações que impedem seu uso em certas aplicações. Devido ao seu protocolo de obtenção, sua qualidade é baixa, sendo que a porcentagem de ESTs com alto escore em busca de similaridades pode ser muito restrita (Lorkowski e Cullen, 2003). Adicionalmente, dependendo do tipo de biblioteca de cDNA (primária ou normalizada) usada para gerar essas sequencias, a redundância pode ser muito alta reduzindo a possibilidade de encontrar transcritos raros (Lorkowski e Cullen, 2003).

Ensaios baseados na tecnologia SAGE têm sido amplamente utilizados para estudos de expressão gênica. Em tal metodologia, desenvolvida por Velculescu *et al.*, 1995, um fragmento (*tag*) de 15 pares de bases de extensão é isolado de cada cDNA que constitui a biblioteca sob análise, sendo que diversas *tags* são clonadas e concatenadas em um mesmo vetor plasmidial. Após sequenciamento, as *tags* são anotadas via análises de bioinformática e, pela combinação da frequência com sua anotação, um compreensivo e quantitativo perfil de expressão gênica pode ser derivado (Terauchi *et al.*, 2008).

Em relação à metodologia baseada em ESTs, a tecnologia SAGE possui maior rendimento uma vez que cada inserto pode gerar informação para até 50 *tags*, resultando em aproveitamento de cerca de 50 vezes maior que o sequenciamento de ESTs (Terauchi *et al.*, 2008). Adicionalmente, certas características inerentes à metodologia SAGE, e às técnicas que compõem o grupo "b", se sobrepõem às observadas em *microarray*. Como exemplo, em contraste aos grupos de dados análogos obtidos por métodos baseados em hibridização, os dados SAGE, e de suas versões otimizadas, são digitais e fáceis de manusear utilizando-se ferramentas de bioinformática, facilitando a comparação direta do perfil de expressão de diferentes genes e a comparação entre

bancos de expressão gênica (Terauchi *et al.,* 2008). Somando-se a isso, destaca-se o fato de SAGE ser considerada uma técnica de "arquitetura aberta", ou seja, não necessita de conhecimento prévio sobre as sequências de transcritos, além de ter a capacidade de descobrir transcritos inéditos. Tal característica não está disponível para a técnica de *microarray*, a qual é considerada uma metodologia de "arquitetura fechada", onde se pode avaliar somente os genes imobilizados no chip (Meyers *et al.,* 2004; Terauchi *et al.,* 2008; Matsumura *et al.*,2008)

À despeito das amplas vantagens supracitadas, a tecnologia SAGE apresenta sérias limitações. A extensão de sua *tag*, 15 pb, é muito curta para a identificação inequívoca do gene correspondente. Adicionalmente, para organismos desprovidos de bancos de dados, essa metodologia se mostra quase inócua, uma vez que buscas através do BLAST necessitam de banco de dados para serem executadas. Para contornar tal problema, diversos métodos para obter sequências de maior extensão a partir de tags SAGE vêm sendo reportados (Matsumura *et al*., 2008), muitos baseados principalmente em amplificações via PCR de fragmentos de cDNA, utilizando-se tais *tags* como *primers*. Entretanto, essas tags são consideradas muito curtas para amplificação de fragmentos únicos de cDNA via 3' RACE (Matsumura *et al.,* 2008; Terauchi *et al.*, 2008).

Outra técnica, denominada *Massively Parallel Signature Sequencing* (MPSS), criada por Brenner *et al.* (2000), constitui-se em uma das mais novas tecnologias para análise de transcriptoma. Baseia-se na clonagem individual de moléculas de cDNA em microesferas e sequenciamento de *tags*, utilizando um método não convencional e engenhoso capaz de sequenciar um grande número de sequências em paralelo. O resultado final é apresentado na forma de milhares de tags distintas, constituídas de 16 a 20 nucleotídeos (nt) de extensão. Essa característica (extensão da tag) é uma das duas

vantagens que essa metodologia apresentada sobre SAGE, uma vez que diminui a ambiguidade no processo de anotação (Reinartz *et al.*, 2002). A outra está associada à natureza automatizada da MPSS, a qual torna possível a produção de milhões de *tags*, fornecendo uma melhor amostragem e possibilitando a identificação de genes biologicamente importantes e com baixos níveis de expressão (Reinartz *et al.*, 2002). Tal fato se contrapõe aos dados SAGE, cujos bancos de dado são constituídos de 20.000 a 60.000 *tags*, sendo que ainda não está claro se tal montante é significativo para análises de todos os genes em uma amostra (Reinartz *et al.*, 2002). Entretanto, diferentemente de SAGE, MPSS possui uma metodologia bastante complexa não podendo ser realizada em laboratórios individuais (Meyers *et al.*, 2004).

Por sua vez, a técnica de RNA-Seq, é beneficiada pelo uso de metodologias de sequenciamento desenvolvidas recentemente. Em geral, uma população de transcritos [total ou fracionada, tal como poli $(A)^+$] é convertida em uma biblioteca de fragmentos de cDNA com adaptadores anexados a ambas extremidades (Wang *et al*., 2009). Tal técnica representa uma das mais modernas atualmente. Revisão aprofundada pode ser vista em Wolf (2013).

3.3.1 SuperSAGE e derivados

Atualmente, dentre as versões melhoradas da tecnologia SAGE e das diversas metodologias para análise em alto escala de transcriptoma, SuperSAGE (Matsumura *et al.*, 2003) é uma das mais eficientes e modernas (Terauchi *et al.*, 2008). Nessa técnica, as *tags* possuem 26 pb e são obtidas a partir da fragmentação de cDNAs por meio de uma enzima de restrição do tipo III, chamada EcoP15I (ver esquema na **Figura 1A e 1B**). Geralmente, a extensão das *tags* depende da habilidade de uma determinada

enzima clivar sítios distantes de seus motivos de reconhecimento na molécula de DNA, sendo que a enzima supracitada apresenta a maior distância entre esses pontos (Matsumura *et al.,* 2008). O aumento na extensão da *tag* melhora significativamente o processo de anotação, ou seja, identificação das tags correspondentes aos genes. Para demonstrar tal propriedade, Matsumura *et al.*, (2005) selecionaram, aleatoriamente, 50 tags SuperSAGE, obtidas em um experimento anterior (Matsumura *et al.,* 2003), e as utilizaram como *queries* (sequências de consulta) para a busca de sequências de DNA que apresentassem um pareamento perfeito com essas. As tags foram "quebradas" a partir da extremidade 3' para formar fragmentos com a extensão de 20 pb ou 18 pb (correspondente a tags de LongSAGE ou MPSS, respectivamente) e 15 pb (correspondente a SAGE) e então alinhadas via BLAST contra o GenBank, banco este constituído por milhões de sequências de cerca de 130 mil espécies, à época. O número de espécies apresentando pareamento com um determinado tipo de tag foi enumerado. Os resultados mostraram que as tags SAGE parearam, em média, com sequências de DNA pertencentes a 7,5 espécies diferentes, sendo que todas as 50 tags desse grupo parearam com duas ou mais espécies, nenhuma *tag* apresentou especificidade. As *tags* de 20 pb parearam, em média, com sequências de DNA pertencentes a 1,3 espécies com um máximo de quatro espécies. Por sua vez, as tags SuperSAGE parearam, em média, com 1,1 espécies, com um máximo de apenas duas espécies. Devido a essa característica, SuperSAGE apresenta significativa vantagem sobre as demais metodologias, podendo ser utilizada, com eficiência, para estudar amostras biológicas nas quais dois ou mais eucariotos estão mixados (parasita – hospedeiro, patógeno-hospedeiro, etc.), possibilitando o estudo do transcriptoma de interação (Terauchi *et al.*, 2008; Matsumura *et al.,* 2005). Adicionalmente, para organismos desprovidos de sequências em bancos de dados, as *tags* SuperSAGE se mostram adequadas e atuam

eficientemente na recuperação das sequências adjacentes, através da técnica 3' RACE, facilitando o processo de anotação via BLASTX (Matsumura *et al.*, 2005; Terauchi *et al.*, 2008; Matsumura *et al.*, 2003b). Outro fator relevante a ser mencionado, se relaciona a adequação dessa técnica às novas tecnologias de sequenciamento, proporcionando uma maior amostragem (milhões de *tags*) do transcriptoma, diminuindo custos e tempo de produção e originando atualizações com maior rendimento e eficiência nas análises, como HT-SuperSAGE (Matsumura *et al.*, 2010) e DeepSuperSAGE (**Figura** 1; Matsumura *et al*., 2012).

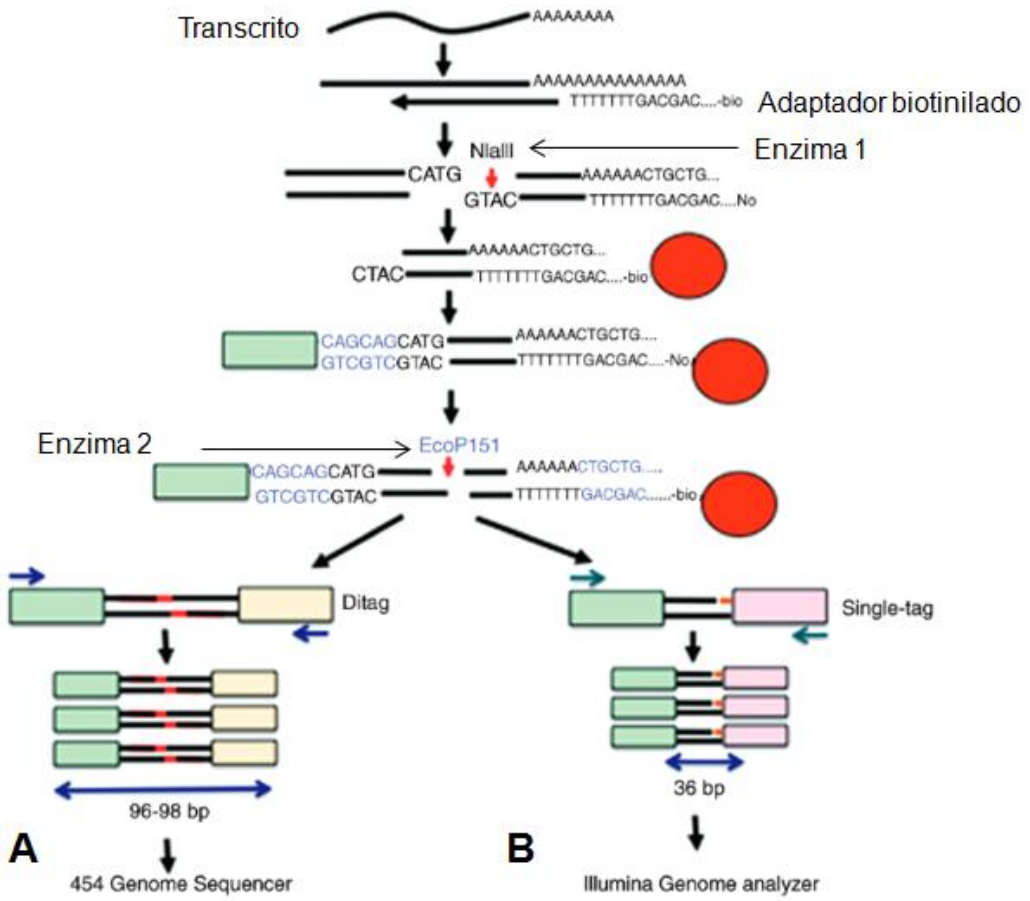


Figura 1. Esquema de obtenção de bibliotecas SuperSAGE (A) e DeepSuperSAGE (B), utilizando-se das técnicas de sequenciamento de nova geração (*Next Generation Sequence*) *454 Genome Sequencer* e *Illumina Genome Analyzer*, respectivamente. Adaptado de Matsumura *et al*. (2012).

A era da transcriptômica é uma realidade para soja, bem como para muitas outras culturas de importância econômica. Recentemente, foi desenvolvida uma base de dados genômicos que inclui tal espécie, no qual estão alocadas as sequencias componentes de seu genoma, bem como informações sobre expressão, composição gênica e proteica (Goodstein *et al*., 2012). Adicionalmente aos esforços realizados para o sequenciamento total do seu genoma, outras fontes de informação têm sido desenvolvidas, incluindo banco de dados de ESTs (*Expressed Sequence Tags*), *full-lenght* cDNA e plataformas microarranjos (Valliyodan e Nguyen, 2008). Tais recursos fornecem uma gama de oportunidades para aplicação no melhoramento de soja através do ganho de informação por meio do entendimento da orquestração gênica, durante períodos de condições não favoráveis, desenvolvimento de marcadores moleculares, transgenia, etc.

3.4. Transcriptômica de soja: ferramentas, disponibilidade de dados e aplicações

Na última década, o conhecimento nas áreas de genética, genômica e transcriptômica de soja tem avançado consideravelmente. Muitos dos recursos disponíveis previamente somente para espécies consideradas modelo, agora estão disponíveis para essa cultura. Tal espécie possui uma mapa genético bastante detalhado (Cregan, 2008), um mapa físico completado recentemente (Shoemaker *et al*., 2008) e disponibilidade de informações de genética reversa para o estudo da função dos genes (Bilyeu, 2008).

Especificamente, em relação ao campo da transcriptômica, dados de soja obtidos a partir de vários projetos resultaram em 35986 unigenes alocados na base de dados pública GenBank, do NCBI, (*National Center for Biotechnology Information*) até

Setembro de 2013. O projeto "*The Gene Index*" (do *Computational Biology Laboratory, Harvard University*) inclui um total de 137174 unigenes consistindo de 73178 TCs (sequencias do tipo "*Tentative Consensus*") e 63866 *singletons* (sequências não clusterizadas), além de 130 ET (*singletons mature transcripts*; Quackenbush *et al*., 2001). Adicionalmente, três plataformas de microarranjos estão disponíveis: duas consistindo de 18432 genes derivados de conjuntos de cDNA de baixa redundância (Vodkin *et al*., 2008). e uma plataforma mista denominada "*Soybean GeneChip*" (http://www.affymetrix.com). Essa última possui cerca de 37500 transcritos de *G. max*, 15800 do fungo *Phytophtora sojae* (que geralmente ataca culturas de soja) e 7500 transcritos do nematóide *Heterodera glycines* (um nematódeo danoso a plantações de soja; Valliyodan e Nguyen, 2008). Uma quarta plataforma de microarranjo está comercialmente disponível, sendo denominada de *66 K Affymetrix Soybean Array GeneChip* e é utilizada para análises de maior abrangência.

Uma recente análise (Dezembro/2013) no banco de dados PubMed (NCBI) apresentou 37 trabalhos relacionados aos termos "soja e transcriptoma" (**Tabela** 1). Filtrando-se os associados à seca, observa-se que Le *et al*. (2012; **Tabela** 1) utilizaram a plataforma de *microarray GeneChip Soybean Genome Arrays* (GCSGA) para análises de expressão global em tecidos foliares, da cultivar Williams 82, em dois estádios diferentes de desenvolvimento (V6 e R2) submetidos ao estresse acima mencionado (na condição de: umidade do solo a 5% e conteúdo hídrico da folha a 32±2%). Ha *et al*. (2013; **Tabela** 1), por sua vez, utilizaram GCSGA e técnica de sequenciamento de alto rendimento (sequenciamento Illumina) para analisar a orquestração gênica da família de fatores de transcrição ARF (*Auxin Responsive Factor*) em diferentes tecidos (em estádios de desenvolvimento pré-determinados) de plantas de soja expostas a déficit hídrico (0, 2 e 10 h), observando que os transcriptomas produzidos em resposta ao

estresse eram tecido-específicos. Marcolino-Gomes *et al*. (2013; **Tabela** 1) analisaram a expressão de dez fatores de transcrição do tipo *APETALA2/Ethylene Responsive Element Binding-like* (AP2/EREB-like), em dois acessos fiosiologicamente contrastantes para resposta à déficit hídrico. Tais autores observaram a indução de nove desses após a aplicação do estresse, além de constatarem diferenças transcricionais que foram dependentes da extensão do tratamento a que os acessos foram expostos e do tecido analisado. Rodrigues *et al*. (2012; **Tabela** 1) utilizaram bibliotecas subtrativas (SSH) para analisar o padrão de genes diferencialmente expressos em soja submetida a déficit hídrico nos tempos de 0 a 150 mim (com intervalo de 25 min entre os tratamentos). Enquanto que Li *et al*. (2011; **Tabela** 1) analisaram a resposta de uma linhagem chinesa (HJ-1) submetida a estresse salino (120 mM NaCl) ou alcalino (70 mM NaCl e 50 mM $NaHCO_3$) ou seca (2% PEG 8000), por 48 h, focando em seu trabalho, identificar, principalmente, genes responsivos somente ao estresse alcalino o qual, segundo os autores, é fator limitante ao cultivo de soja na China. Por fim, Le *et al*. (2011; **Tabela** 1) utilizaram GCSGA e sequenciamento Illumina para analisar a orquestração gênica da família de fatores de transcrição NAC, em folhas de soja, quando da exposição da plantas soja ao déficit hídrico (conteúdo hídrico foliar de 60%) durante estádios de desenvolvimento pré-determinados (V6 e R2). A análise sistemática desses autores identificou importantes alvos tecido-específicos responsivos à desidratação, os quais serão utilizados para caracterização molecular aprofundada e futuro desenvolvimento de soja transgênica tolerante à seca.

Tabela 1. Apresentação dos trabalhos disponíveis abordando os termos "soja" e "transcriptômica" na base de dados PubMed (NCBI), em dezembro de 2013, evidenciando o tecido ou órgão estudado, condição analisada, bem como a tecnologia empregada e o foco da análise.

Tecido /órgão	Condição	Tecnologia	Foco	Autores
Nódulos, Meristema Apical do Caule (MAC), Raiz e Vagem Verde	Estresse por Déficit Hídrico e Tecidos Não Estressados em Estádios Específicos de Desenvolvimento	Microarray (GCSGA) e Sequenciamento Illumina	Família ARF de Fatores de Transcrição	Ha *et al*. (2013)
Raiz	Micorrização	Microarray (GCSGA)	Fatores de Transcrição GmNF-YA1a/b	Schaarschmidt *et al*. (2013)
14 Diferentes Tecidos (Raiz, Flor, Vagem, etc..)	Tecidos Não Estressados em Estádios Específicos de Desenvolvimento	Microarray (GCSGA) e Sequenciamento Illumina	Transportadores de Silício	Deshmukh *et al*. (2013)
Folha e MAC	Floração Após Dia Curto	Sequenciamento Illumina	Análise Global	Wong *et al*. (2013)
Semente	Desenvolvimento da semente (Fertilização até a Maturidade)	Sequenciamento Illumina	Análise Global	Jones e Vodkin (2013)
Folha	Seca	Microarray (GCSGA)	Análise Global	Le *et al*. (2012)
Raiz e Folha	Estresse Alcalino	Microarray (GCSGA)	Família TIFY de Fatores de Transcrição	Zhu *et al*. (2013)
Folha	Infecção por *Aphis glycines*	Microarray (GCSGA)	Análise Global	Studham e MacIntosh *et al*. (2013)
Raiz e Folha	7 Condições Diferentes, Incluindo Deficiência de Ferro e Infecção por Agentes Patogênicos, etc..	Microarray (GCSGA)	Fatores de Transcrição	Zhu *et al*. (2012)
Raiz e Folha	Déficit hídrico	SSH**	Análise Global	Rodrigues *et al*. (2012)
Raiz	Infecção por *Meloidogyne javanica*	Sequenciamento de ESTs	Análise Global	Sá *et al*. (2012)
Cotilédones e Tegumento	Tecidos Não Estressados em Estádios Específicos de Desenvolvimento	Sequenciamento Illumina	Análise de MicroRNAs	Shamimuzzaman e Vodkin (2012)
Raízes e brotos	Deficiência de Potássio	Sequenciamento Illumina	Análise Global	Wang *et al*. (2012)
Folha	Infecção por *B. japonicum*	Sequenciamento Illumina	Análise Global	Reid *et al*. (2012)
Semente	Vagem, Grão de 2 mm e 5 mm, além do Grão Maduro	Microarray (GCSGA)	Análise Global	Asakura *et al*. (2012)
Raiz	Seca, Salinidade e Estresse Alcalino	Sequenciamento Illumina	Análise de MicroRNAs	Li *et al*. (2011)
Raiz	Estresse por Alumínio	Microarray (GCSGA)	Análise Global	You *et al*. (2011)
Folha e MAC	Comparação entre Genótipos (Selvagem e Mutante)	Sequenciamento Illumina	Análise Global	Hunt *et al*. (2011)

*GCSGA: GeneChip Soybean Genome Arrays – Affimetrix; **SSH: Suppressive Subtractive Hybridization (obtida via Sequenciamento Illumina)

Continuação da **Tabela 1.**

Tecido /órgão	Condição	Tecnologia	Foco	Autores
Raiz e Parte Superior	Crescimento sob Disponibilidade Limitante de Nitrogênio	Sequenciamento Illumina	Genes Associados à Eficiência no Uso do Nitrogênio	Hao *et al*. (2011)
Raiz	Infecção por *Heterodera glycines*	Microarray (GCSGA)	Análise Global	Matsye *et al*. (2011)
Tegumento	Comparação entre Genótipos de Cores Diferentes	Microarray (GCSGA)	Genes Associados à Síntese de Pigmentos	Kovinich *et al*. (2011)
Raiz e Parte Superior	Plantas Após 12 Dias de Germinação e Estresse de Desidratação Radicular	Microarray (GCSGA) e Sequenciamento Illumina	Família NAC de Fatores de Transcrição	Le *et al*. (2011)
Cotilédones	Comparação entre Genótipos (Selvagem e pRNAiSP−)	Microarray (GCSGA) e Sequenciamento Illumina	Análise Global	Schmidt *et al*. (2011)
Folha	Infecção por *Xanthomonas axonopodis* pv. *glycines*	Sequenciamento Illumina	Análise Global	Kim *et al*. (2011)
Raiz e Folha	Déficit hídrico	SSH** e Sequenciamento Illumina	Fatores de Transcrição AP2/EREB-like	Marcolino-Gomes *et al*. (2013)
9 Diferentes Tecidos (Nódulos, Raiz, MAC, vagens, etc.)	Tecidos Não Estressados em Estádios Específicos de Desenvolvimento	Microarray (GCSGA) e Sequenciamento Illumina	Citocromo P450 Monoxigenase	Guttikonda *et al*. (2010)
Raiz	Estresse por Alumínio	Microarray (GCSGA)	Análise Global	Duressa *et al*. (2010a)
Raiz	Estresse por Alumínio	Microarray (GCSGA)	Análise Global	Duressa *et al*. (2010b)
Folha	Estresse Alcalino	Microarray (GCSGA)	Análise Global	Ge *et al*. (2011)
Flores, folhas, Vagens, Sementes, Raiz e Nódulos	Tecidos Não Estressados em Estádios Específicos de Desenvolvimento	Sequenciamento Illumina	Análise Global	Severin *et al*. (2010)
Raiz, Folhas, Sementes, Vagem, Flores e Caule	Tecidos Não Estressados em Estádios Específicos de Desenvolvimento	Sequenciamento Illumina	Análise Global	Libaut *et al*. (2010a)
Raiz	Infecção por *Bradyrhizobium japonicum*	Microarray (GCSGA) e Sequenciamento Illumina	Análise Global	Libaut *et al*. (2010b)
Pólen	Plantas Após 10 Dias de Germinação	Microarray (GCSGA)	Análise Global	Haerizadeh *et al*. (2009)
Raiz	Infecção por *Phytophthora sojae*	Microarray (GCSGA)	Análise Global	Zhou *et al*. (2009)
Folha	Infecção por *Phakopsora pachyrhizi*	Microarray (GCSGA)	Análise Global	Panthee *et al*. (2009)
Folhas, Caule e Cotilédones	Resposta a Glifosato	Microarray (GCSGA)	Análise Global	Zhu *et al*. (2008)
Folha	Comparação entre Genótipos (Selvagem e Transgênico)	Microarray (GCSGA)	Análise Global	Cheng *et al*. (2008)
Folha	Infecção por *Pseudomonas syringae*	Microarray (GCSGA)	Via dos Fenilpropanóides	Zabala *et al*. (2006)

*GCSGA: GeneChip Soybean Genome Arrays – Affimetrix; **SSH: Suppressive Subtractive Hybridization (obtida via Sequenciamento Illumina)

4. Referências Bibliográficas

Adams, MD, *et al*. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252: 1651–1656.

Asakura T, Tamura T, Terauchi K, Narikawa T, Yagasaki K, *et al*. (2012) Global gene expression profiles in developing soybean seeds. Plant Physiol Biochem 52: 147–153.

Benko-Iseppon AM, Nepomuceno AL, Abdelnoor R V (2012) GENOSOJA – The Brazilian Soybean Genome Consortium : High throughput omics and beyond. Genet Mol Biol 35: i–iv.

Bilyeu, KD. Forward and Reverse Genetics in Soybean. In: Stacey, G (ed.). **Genetics and Genomics of soybean.** 1ª ed. Nova Iorque: Springer, cap. 6, p. 135-139, 2008.

Bray EA (2004) Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. J Exp Bot 55: 2331–2341.

Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, *et al*. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18: 630–634.

Cheng KC, Beaulieu J, Iquira E, Belzile FJ, Fortin MG, *et al*. (2008) Effect of transgenes on global gene expression in soybean is within the natural range of variation of conventional cultivars. J Agric Food Chem 56: 3057–3067.

Cregan, PB; Schlueter JA; Jackson SA. Soybean Genome Structure and Organization. In: Stacey, G (ed.). **Genetics and genomics of soybean.** 1ª ed. Nova Iorque: Springer, cap. 6, p. 91-99, 2008.

de Sá MEL, Lopes MJC, de Araújo MC, Paiva LV, Dos Santos RMA, *et al*. (2012) Transcriptome analysis of resistant soybean roots infected by *Meloidogyne javanica*. Genet Mol Biol 35: 272–282.

Deshmukh RK, Vivancos J, Guérin V, Sonah H, Labbé C, *et al*. (2013) Identification and functional characterization of silicon transporters in soybean using comparative genomics of major intrinsic proteins in Arabidopsis and rice. Plant Mol Biol 83: 303–315.

Duressa D, Soliman K, Chen D (2010) Identification of Aluminum Responsive Genes in Al-Tolerant Soybean Line PI 416937. Int J Plant Genomics 2010: Article ID 164862.

Duressa D, Soliman KM, Chen D (2010) Mechanisms of magnesium amelioration of aluminum toxicity in soybean at the gene expression level. J Exp Bot 797: 787–797.

FAO (Food and Agriculture Organization), 2012: http://goo.gl/HQniuJ. Acessado em 25 deNovembro, 2013.

Ge Y, Li Y, Lv D-K, Bai X, Ji W, *et al*. (2011) Alkaline-stress response in *Glycine soja* leaf identifies specific transcription factors and ABA-mediated signaling factors. Funct Integr Genomics 11: 369–379.

Global Industry Analysts, Inc, (2013): http://goo.gl/qoph6a. Acessado em 30 de Novembro, 2013.

Globo Rural, 2012: http://goo.gl/TKW11P; Acessado em 30 de Novembro, 2013.

Goodstein DM, Shu S, Howson R, *et al*. (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40(Database issue): D1178-1186.

Guttikonda SK, Trupti J, Bisht NC, Chen H, An Y-QC, *et al*. (2010) Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. BMC Plant Biol 10: 243–262.

Ha C Van, Le DT, Nishiyama R, Watanabe Y, Sulieman S, *et al*. (2013) The auxin response factor transcription factor family in soybean: genome-wide identification and expression analyses during development and water stress. DNA Res: 1–14.

Haerizadeh F, Wong CE, Bhalla PL, Gresshoff PM, Singh MB (2009) Genomic expression profiling of mature soybean (*Glycine max*) pollen. BMC Plant Biol 9: 25–37.

Hao QN, Zhou X a, Sha AH, Wang C, Zhou R, *et al*. (2011) Identification of genes associated with nitrogen-use efficiency by genome-wide transcriptional analysis of two soybean genotypes. BMC Genomics 12: 525–540.

Hunt M, Kaur N, Stromvik M, Vodkin L (2011) Transcript profiling reveals expression differences in wild-type and glabrous soybean lines. BMC Plant Biol 11: 145–161.

Jones SI, Vodkin LO (2013) Using RNA-Seq to profile soybean seed development from fertilization to maturity. PLoS One 8: e59270.

Kim KH, Kang YJ, Kim DH, Yoon MY, Moon J-K, *et al*. (2011) RNA-Seq analysis of a soybean near-isogenic line carrying bacterial leaf pustule-resistant and -susceptible alleles. DNA Res 18: 483–497.

Kovinich N, Saleem A, Arnason JT, Miki B (2011) Combined analysis of transcriptome and metabolite data reveals extensive differences between black and brown nearly-isogenic soybean (*Glycine max*) seed coats enabling the identification of pigment isogenes. BMC Genomics 12: 381–399.

Ku Y-S, Au-Yeung W-K, Yung Y-L, Li M-W, Wen C-Q, Liu X, Lam H-M: Drought Stress and Tolerance in Soybean. In: Board J (ed.). **A comprehensive survey of international soybean research - genetics, physiology, agronomy and nitrogen relationships.** 1ª ed. Intech, cap. 10, p. 209-237, 2013.

Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K, *et al*. (2011) Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. DNA Res 18: 263–276.

Le DT, Nishiyama R, Watanabe Y, Tanaka M, Seki M, *et al*. (2012) Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. PLoS One 7: e49522.

Li H, Dong Y, Yin H, Wang N, Yang J, *et al*. (2011) Characterization of the stress associated microRNAs in *Glycine max* by deep sequencing. BMC Plant Biol 11: 170–182.

Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, *et al*. (2010) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to Bradyrhizobium japonicum infection. Plant Physiol 152: 541–552.

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, *et al*. (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. Plant J 63: 86–99.

Lorkowski S, Cullen P. 2003. **Analysing gene expression**. Weinheim, Germany: Wiley-VCH.

Manavalan LP, Guttikonda SK, Tran L-S, Nguyen HT (2009) Physiological and molecular approaches to improve drought resistance in soybean. Plant Cell Physiol 50: 1260–1276.

MAPA (Ministério da Agricultura e Pecuária), 2013a: http://goo.gl/QPrkHL. Acessado em 25 de Novembro, 2013.

MAPA (Ministério da Agricultura e Pecuária), 2013b: http://goo.gl/0t1U9L. Acessado em 25 de Novembro, 2013.

Martins PK, Jordão BQ, Yamanaka N, Farias JRB, Beneventi MA, *et al*. (2008) Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill Fabales, Fabaceae) cultivar MG/BR46 (Conquista) under two water deficit induction systems. Genet Mol Biol 31: 512–521.

Matsumura H, Ito A, Saitoh H, Winter P, Kahl G, *et al*. (2005) SuperSAGE. Cell Microbiol 7: 11–18.

Matsumura H, Krüger DH, Kahl G, Terauchi R (2008) SuperSAGE: A modern platform for genome-wide quantitative transcript profiling. Curr Pharm Biotechnol 9: 368–374.

Matsumura H, Reich S, Ito A, Saitoh H, Kamoun S, *et al*. (2003) Gene expression analysis of plant host-pathogen interactions by SuperSAGE. Proc Natl Acad Sci USA 100: 15718–15723.

Matsumura H, Yoshida K, Luo S, Kimura E, Fujibe T, *et al*. (2010) High-throughput SuperSAGE for digital gene expression analysis of multiple samples using next generation sequencing. PLoS One 5: e12010.

Matsumura H, Molina C, Krüger DH, Terauchi R, Kahl G. DeepSuperSAGE: High-throughput transcriptome sequencing with now- and next-Generation sequencing technologies. In: Habers M, Kahl G (eds.). **Tag-based next generation sequencing**. 1ª ed. Wiley-Blackwell, cap. 1, p. 3-19, 2012.

Matsye PD, Kumar R, Hosseini P, Jones CM, Tremblay A, *et al*. (2011) Mapping cell fate decisions that occur during soybean defense responses. Plant Mol Biol 77: 513–528.

Meyers BC, Galbraith DW, Nelson T, Agrawal V (2004) Methods for transcriptional profiling in plants. Be Fruitful and Replicate. Plant Physiol 135: 637–652.

Molina C, Rotter B, Horres R, Udupa SM, Besser B, *et al*. (2008) SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. BMC Genomics 9: 553.

Morin R, Bainbridge M, Fejes A, *et al*. (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques 45: 81-94.

Murphy D (2002) Gene expression studies using microarrays: principles, problems and prospects. Adv Physiol Educ 26: 256–270.

Panthee DR, Marois JJ, Wright DL, Narváez D, Yuan JS, *et al*. (2009) Differential expression of genes in soybean in response to the causal agent of Asian soybean rust (*Phakopsora pachyrhizi* Sydow) is soybean growth stage-specific. Theor Appl Genet 118: 359–370.

Pathan MDS, *et al*. Recent advances in breeding for drought and salt stress tolerance in soybean. In: Jenks, M. A.; Hasegawa, P. M.; Jain, S. M. (eds.). **Advances in molecular breeding toward drought and salt tolerant crops**. 1ª ed. Dordrecht: Springer, cap. 30, p. 739-773, 2007.

Quackenbush J, Cho J, Lee D, Liang F, Holt I, *et al*. (2001) The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Research 29: 159–164.

Reid DE, Hayashi S, Lorenc M, Stiller J, Edwards D, *et al*. (2012) Identification of systemic responses in soybean nodulation by xylem sap feeding and complete transcriptome sequencing reveal a novel component of the autoregulation pathway. Plant Biotechnol J 10: 680–689.

Reinartz J, Bruyns E, Lin J, Burcham T, Brenner S, *et al*. (2002) Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Briefings Funct Genomics Proteomics 1: 95–104.

Rodrigues FA, Marcolino-Gomes J, de Fátima Corrêa Carvalho J, do Nascimento LC, Neumaier N, *et al*. (2012) Subtractive libraries for prospecting differentially expressed genes in the soybean under water deficit. Genet Mol Biol 35: 304–314.

Schaarschmidt S, Gresshoff PM, Hause B (2013) Analyzing the soybean transcriptome during autoregulation of mycorrhization identifies the transcription factors GmNF-YA1a/b as positive regulators of arbuscular mycorrhization. Genome Biol 14: R62.

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-70.

Schmidt M a, Barbazuk WB, Sandford M, May G, Song Z, *et al*. (2011) Silencing of soybean seed storage proteins results in a rebalanced protein composition

preserving seed protein content without major collateral changes in the metabolome and transcriptome. Plant Physiol 156: 330–345.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, *et al*. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178–183.

Severin AJ, Woody JL, Bolon Y-T, Joseph B, Diers BW, *et al*. (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol 10: 160–176.

Shamimuzzaman M, Vodkin L (2012) Identification of soybean seed developmental stage-specific and tissue-specific miRNA targets by degradome sequencing. BMC Genomics 13: 310–324.

Shimkets RA. **Gene expression profiling**. 1º ed. Totowa: Humana Press, Inc, 2004.

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. J Exp Bot 58: 221–227.

Singh R. J.; Chung G. H.; Nelson R. L (2007) Landmark research in legumes. Genome 50: 525-537.

Soystats, 2013: http://soystats.com/. Acessado em 25 de Novembro, 2013.

Specht, J. E.; Hume D. J.; Kumudini, S. V (1999) Soybean yield potential – a genetic and physiological perspective. Crop Science 39: 1560-1570.

Studham ME, MacIntosh GC (2013) Multiple phytohormone signals control the transcriptional response to soybean aphid infestation in susceptible and resistant soybean plants. Mol Plant-Microbe Iteractions (MPMI) 26: 116–129.

Terauchi R, Matsumura H, Krüger DH, Kahl G. SuperSAGE: The most advancedtranscriptoma technology for functional genomics. In: Kahl, Günter; Meksem, Khalid (eds.). **The handbook of plant functional genomics**. 1ª ed. Weinheim: Wiley-VCH Verlag GmbH & Co, cap. 3, p. 37-54, 2008.

USDA, 2013: http://goo.gl/nwn1CF; Acessado em 23 de Dezembro, 2013.

Valliyodan B, Nguyen HT. Genomics of Abiotic Stress in Soybean. In: Stacey, G (ed.). **Genetics and Genomics of soybean.** 1ª ed. Nova Iorque: Springer, cap. 18, p. 343-372, 2008.

Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler K. W (1995) Serial analysis of gene expression. Science 270: 484-487.

Vodkin L, *et al*. Genomics of soybean seed development. In: Stacey, G (ed.). **Genetics and Genomics of soybean.** 1ª ed. Nova Iorque: Springer, cap. 11, p. 163-184, 2008.

Wang C, Chen H, Hao Q, Sha A, Shan Z, *et al*. (2012) Transcript profile of the response of two soybean genotypes to potassium deficiency. PLoS One 7: e39856

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57–63.

Wilson RF. Soybean: market driven research needs. In: Stacey, Gary (ed.). **Genetics and genomics of soybean**. 1ª ed. Heidelberg: Springer, 2008. cap. 1, p. 3-15.

Wolf JBW (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. Mol Ecol Resour 13: 559–572.

Wong CE, Singh MB, Bhalla PL (2013) The dynamics of soybean leaf and shoot apical meristem transcriptome undergoing floral initiation process. PLoS One 8: e65319.

You J, Zhang H, Liu N, Gao L, Kong L (2011) Transcriptomic responses to aluminum stress in soybean roots. 933: 923–933.

Zabala G, Zou J, Tuteja J, Gonzalez DO, Clough SJ, *et al*. (2006) Transcriptome changes in the phenylpropanoid pathway of *Glycine max* in response to *Pseudomonas syringae* infection. BMC Plant Biol 18: 26–43

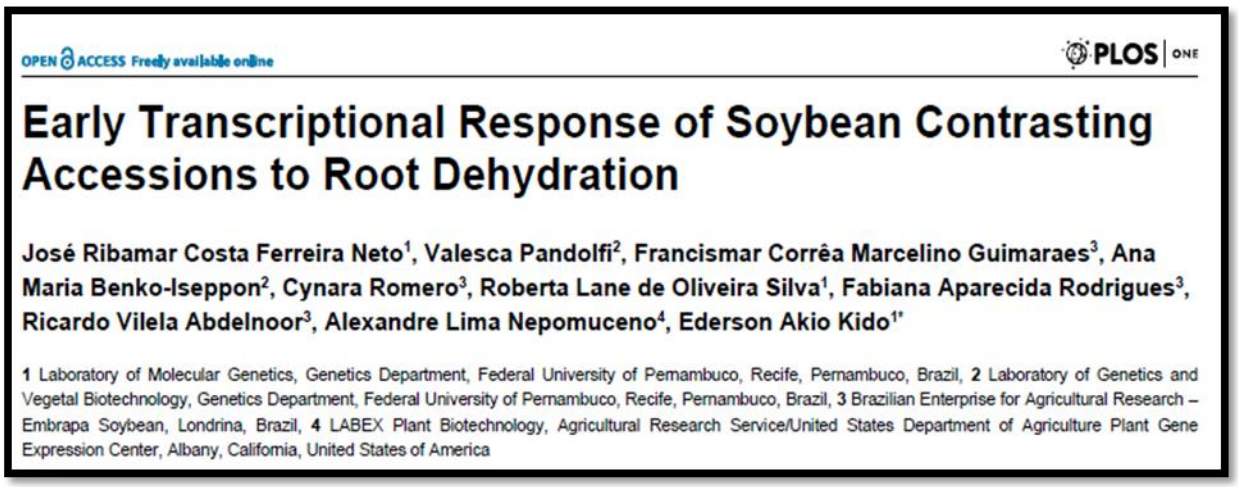
Zhou L, Mideros SX, Bao L, Hanlon R, Arredondo FD, *et al*. (2009) Infection and genotype remodel the entire soybean transcriptome. BMC Genomics 10: 49–67.

Zhu D, Bai X, Luo X, Chen Q, Cai H, *et al*. (2013) Identification of wild soybean (*Glycine soja*) TIFY family genes and their expression profiling analysis under bicarbonate stress. Plant Cell Rep 32: 263–272.

Zhu J, Patzoldt WL, Shealy RT, Vodkin LO, Clough SJ, *et al*. (2008) Transcriptome response to glyphosate in sensitive and resistant soybean. J Agric Food Chem 56: 6355–6363.

Zhu M, Deng X, Joshi T, Xu D, Stacey G, *et al*. (2012) Reconstructing differentially co-expressed gene modules and regulatory networks of soybean cells. BMC Genomics 13: 437–450.

Capítulo I



OPEN ACCESS Freely available online

PLOS ONE

Early Transcriptional Response of Soybean Contrasting Accessions to Root Dehydration

José Ribamar Costa Ferreira Neto[1], Valesca Pandolfi[2], Francismar Corrêa Marcelino Guimaraes[3], Ana Maria Benko-Iseppon[2], Cynara Romero[3], Roberta Lane de Oliveira Silva[1], Fabiana Aparecida Rodrigues[3], Ricardo Vilela Abdelnoor[3], Alexandre Lima Nepomuceno[4], Ederson Akio Kido[1*]

1 Laboratory of Molecular Genetics, Genetics Department, Federal University of Pernambuco, Recife, Pernambuco, Brazil, 2 Laboratory of Genetics and Vegetal Biotechnology, Genetics Department, Federal University of Pernambuco, Recife, Pernambuco, Brazil, 3 Brazilian Enterprise for Agricultural Research – Embrapa Soybean, Londrina, Brazil, 4 LABEX Plant Biotechnology, Agricultural Research Service/United States Department of Agriculture Plant Gene Expression Center, Albany, California, United States of America

Classificação WebQualis CAPES: A2

Fator de Impacto: 3,73

PLOS ONE

Early Transcriptional Response of Soybean Contrasting Accessions to Root Dehydration

José Ribamar Costa Ferreira Neto[1], Valesca Pandolfi[2], Francismar Corrêa Marcelino Guimaraes[3], Ana Maria Benko-Iseppon[2], Cynara Romero[3], Roberta Lane de Oliveira Silva[1], Fabiana Aparecida Rodrigues[3], Ricardo Vilela Abdelnoor[3], Alexandre Lima Nepomuceno[4], Ederson Akio Kido[1*]

1 Laboratory of Molecular Genetics, Genetics Department, Federal University of Pernambuco, Recife, Pernambuco, Brazil, **2** Laboratory of Genetics and Vegetal Biotechnology, Genetics Department, Federal University of Pernambuco, Recife, Pernambuco, Brazil, **3** Brazilian Enterprise for Agricultural Research – Embrapa Soybean, Londrina, Brazil, **4** LABEX Plant Biotechnology, Agricultural Research Service/United States Department of Agriculture Plant Gene Expression Center, Albany, California, United States of America

Abstract

Drought is a significant constraint to yield increase in soybean. The early perception of water deprivation is critical for recruitment of genes that promote plant tolerance. DeepSuperSAGE libraries, including one control and a bulk of six stress times imposed (from 25 to 150 min of root dehydration) for drought-tolerant and sensitive soybean accessions, allowed to identify new molecular targets for drought tolerance. The survey uncovered 120,770 unique transcripts expressed by the contrasting accessions. Of these, 57,610 aligned with known cDNA sequences, allowing the annotation of 32,373 unitags. A total of 1,127 unitags were up-regulated only in the tolerant accession, whereas 1,557 were up-regulated in both as compared to their controls. An expression profile concerning the most representative Gene Ontology (GO) categories for the tolerant accession revealed the expression "protein binding" as the most represented for "Molecular Function", whereas CDPK and CBL were the most up-regulated protein families in this category. Furthermore, particular genes expressed different isoforms according to the accession, showing the potential to operate in the distinction of physiological behaviors. Besides, heat maps comprising GO categories related to abiotic stress response and the unitags regulation observed in the expression contrasts covering tolerant and sensitive accessions, revealed the unitags potential for plant breeding. Candidate genes related to "hormone response" (LOX, ERF1b, XET), "water response" (PUB, BMY), "salt stress response" (WRKY, MYB) and "oxidative stress response" (PER) figured among the most promising molecular targets. Additionally, nine transcripts (HMGR, XET, WRKY20, RAP2-4, EREBP, NAC3, PER, GPX5 and BMY) validated by RT-qPCR (four different time points) confirmed their differential expression and pointed that already after 25 minutes a transcriptional reorganization started in response to the new condition, with important differences between both accessions.

Citation: Ferreira Neto JRC, Pandolfi V, Guimaraes FCM, Benko-Iseppon AM, Romero C, et al. (2013) Early Transcriptional Response of Soybean Contrasting Accessions to Root Dehydration. PLoS ONE 8(12): e83466. doi:10.1371/journal.pone.0083466

Editor: Szabolcs Semsey, Niels Bohr Institute, Denmark

Received July 24, 2013; **Accepted** November 4, 2013; **Published** December 12, 2013

Funding: This work was supported by the following Brazilian funding agencies: CNPq (National Council for Scientific and Technological Development, Brazil; http://www.cnpq.br/; Edital CT-Agronegócio/MCT/CNPq-Genosoja-no 38/2007), responsible for the financial support for all the experimental assays and fellowship (JRCFN); FACEPE (The Foundation for Science and Technology of Pernambuco State; http://www.facepe.br) for the RLOS's fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: kido.ufpe@gmail.com

Introduction

Soybean [*Glycine max* (L.) Merr.] is recognized as a relevant global crop with an annual contribution to the world economy around US$ 48,6 billion dollars [1], and increasing importance due to its multiple uses in food, feed and industrial applications, such as oil and biodiesel production. In Brazil, soybean represents the main agribusiness product; the country is the largest producer in the world [2]. Despite this status and the fact that soybean is one of the most studied legumes, the soy complex agribusiness has suffered significant losses due to abiotic stresses, with emphasis on drought [3]. In USA, there are reports of around 40% losses caused by water deficit [4], whereas in Brazil, in 2004-2005, soybean severely damaged by drought resulted in approximately 25% yield reduction [5], in an area (southern region) responsible for about 40% of this yield. Last year (2012), in a less severe drought, the production reduced in almost 11% in that same region [6].

Unfortunately, this scenario is increasingly uncertain, considering the climate change perspectives [7]. Therefore, breeding programs looking for effective soybean plants adapted to water deficit are crucial. Studies regarding genetics,

physiology and molecular biology of tolerance mechanisms sustaining plant growth and yield under water deficit are essentials for the development of new varieties. In general, features associated with tolerance controlled by many genes make conventional plant breeding more difficult [3,8]. Transcriptome analysis is one of the widest alternatives adopted to identify the repertoire of genes and their biological responses to certain stimuli. Soybean data from various transcriptome projects resulted in a set of 35,986 unigenes [9] stored in GenBank at NCBI (National Center for Biotechnology Information) until May, 2013. Similarly, The Gene Index Project (The Computational Biology Laboratory, Harvard University) includes a total of 137,174 unigenes, consisting of 73,178 TC (Tentative Consensus) sequences, 63,866 singletons and 130 singletons mature transcripts (ET) [10]. Additionally, two microarray slide sets are available; each one consisting of 18,432 single-spotted PCR products derived from the low redundancy cDNA sets [11]. A mixed Soybean GeneChip (http://www.affymetrix.com) is commercially available with ~37,500 *G. max* transcripts, 15,800 *Phytophtora* root and stem root transcripts, and over 7,500 soybean cyst nematode transcripts [3]. Another commercially available microarray platform is the 66 K Affymetrix Soybean Array GeneChip. Despite having high performance, affordable price and still be widely used, microarray technology has serious limitations. Some of them including the cross-hybridization of probes with different potential targets, semi-quantitative results, uncertainty in analysis and interpretation of data, as well as the inability to analyze and discover new genes (only restricted to those immobilized on chips) [12].

A recent survey (May, 2013) at PubMed database (NCBI) showed 81 reports related to “soybean and transcriptome”, most of them using microarray approaches, as in the case of Le et al. [13] that used the 66 K Affymetrix Soybean Array GeneChip for genome-wide expression profiling of leaf tissues (soybean cv. Williams 82) subjected to drought stress (soil moisture content of 5% and leaf relative water content = 32±2%) from two stages (V6 and R2). Concerning the reports using high-throughput sequencing methods, Libaut et al. [14] studied the transcriptome of root hair cells under *Bradyrhizobium japonicum* infection. After that, Libault et al. [15] tried to generate a transcriptome atlas using various soybean tissues; Le et al. [16] focused on the NAC transcription factor family in soybean during development and dehydration stress; Li et al. [17], otherwise, looked for stress associated microRNAs in *G. max* by deep sequencing, while Hao et al. [18] searched for soybean genes associated with nitrogen-use efficiency, and Kido et al. [19] looked for plant antimicrobial peptides in soybean transcriptome after *P. pachirizy* induction. Moreover, Fan et al. [20] analyzed the late expression (48 h after stress) to different conditions including drought (2% PEG 8000), in leaves and roots of seedlings (two-leaf stage) of the soybean inbred line HJ-1. Using RNA-Seq method, specifically, Severin et al. [21] searched for a high-resolution gene expression in a collection of fourteen different tissues; Hunt et al. [22] tried to characterize the transcriptional profiles of a wild-type and glabrous soybean lines while Reid et al. [23] looked for transcript abundance changes that occur during AON (autoregulation of nodulation), and Peiffer et al. [24] attempted to identify candidate genes underlying an iron efficiency quantitative trait locus.

Thus, it is clear that there is still a gap in regard to reliable information on transcriptomics to recognize the initial response to water deficit response in soybean. Also, no previous transcriptome approaches evaluated contrasting (tolerant/sensitive) soybean accessions. Thus, the aim of this study was to fill this gap using DeepSuperSAGE (26 bp tags), a highly sensitive transcriptome method, comparing contrasting accessions under root dehydration stress (25-150 min), aiming to identify tolerance-associated gene candidates, especially regarding the early response not evaluated up to date.

Results and Discussion

Qualitative and Quantitative Analysis of the DeepSuperSAGE Libraries

The DeepSuperSAGE libraries based on the total number of sequenced tags [2,551,286, of which 1,030,443 for ‘Embrapa 48’ (tolerant accession) and 1,520,843 for ‘BR 16’ (sensitive accession)] allowed a comprehensive evaluation of the soybean transcriptome under root dehydration stress. Thus, after singlets exclusion from the total number of tags, 120,770 unitags (unique tags) followed for further analysis. Comparing the contrasts between two libraries, the unitag number ranged from 73,807 to 89,205 (Table 1). It should be highlighted that the estimated number of protein-coding loci for soybean is 66,153 [25]. Thus, the high number of unitags (120,770 for the four libraries) could be justified by the presence of sister unitags (those with a single base difference in a given position and not grouped in a consensus unitag), possibly constituting potential SNPs, alternative transcripts or (less probably) artifacts.

The number of differentially expressed up- (UR) and down-regulated (DR) unitags and those not differentially expressed (n.s.), at the level of $p < 0.05$ (see Material and Methods), for some contrasting libraries can be seen in Table 1. The n.s. unitags accounted for more than 70% of the total, regardless of the considered contrast (Table 1), and probably regard housekeeping genes or genes associated to other physiological processes. Otherwise, the number of UR unitags was higher than the DR in all contrasts (Table 1), also when comparing both accessions under stress (ET1-6 *vs* BT1-6) and even both negative controls (ET0 *vs* BT0; Table 1).

Primary Annotation of DeepSuperSAGE Unitags

After annotation (BLASTn) of the 120,770 unitags against different EST databases, 57,610 (47.7%) of them presented ESTs matches tolerating a single mismatch (TSM) maximum in the alignments (Table 2). From those TSM alignments, 32,373 unitags (56.2%) could be annotated based on previous characterized ESTs (Table 2), disregarding the “unknown” hits (ESTs, cDNAs or mRNAs) or just clones or chromosomes annotations with no given function. Concerning the annotated unitags, 14,903 (46.0%) of them showed 100.0% identity (26 bp of the unitag) in perfect BLASTn alignments with ESTs (Table 2), which 14,545 of them with *G. max* ESTs (data not

Table 1. Number of differentially expressed soybean unitags (UR: u-regulated; DR: down-regulated; n.s.: non-significant at $p < 0.05$) based on SuperSAGE libraries contrasts.

	ET1-6 *vs* ET0		BT1-6 *vs* BT0		ET1-6 *vs* BT1-6		ET0 *vs* BT0	
	Tags	%	Tags	%	Tags	%	Tags	%
UR	13,532	18.1	10,751	12.0	12,347	16.7	6,468	7.9
DR	7,423	9.9	5,587	6.3	7,634	10.3	3,135	3.8
n.s.	53,878	72.0	72,867	81.7	53,826	73.0	73,067	88.3
Unitags	74,833	100.0	89,205	100.0	73,807	100.0	82,67	100.0

*ET0 (tolerant accession 'Embrapa 48'; unstressed control); BT0 (sensitive accession 'BR 16'; unstressed control); ET1-6 ('Embrapa 48' after root dehydration stress); BT1-6 ('BR 16' after root dehydration stress).

doi: 10.1371/journal.pone.0083466.t001

Table 2. Summary of primary annotation of the unitags.

Features	Alignment unitag-EST		Total	%
	Single mismatch	Perfect		
Unitags	-	-	120,770	100.0
Unitags with no hit	-	-	63,160	52.3
Unitags with hits	26,911	30,699	57,610	47.7
With descriptions	17,470	14,903	32,373	56.2*
Without description	9,441	15,796	25,237	43.8*
With GO terms	18,619	17,366	35,985	62.5*

* In relation to 57,610 (unitags with hits).

doi: 10.1371/journal.pone.0083466.t002

shown). Such ESTs, when related to differentially expressed unitags, are potentially useful for primer and probe design, aiming RT-qPCR validation and, at the same time, avoiding following sequencing for unitag identification. Alternatively, from those unitags with appropriate ESTs (57,610), it was possible to characterize 35,985 unitags by GO (Gene Ontology), i.e., more expressive than those 32,373 unitags associated with ESTs with appropriate annotations (Table 2). Thus, for those unitags aligned to ESTs without a decent gene/function description, the GO characterization was a valuable reference and information source.

Regarding perfectly aligned (100% identity) unitags with ESTs (Table 2), 15,796 remained non-annotated. Those unitags and appropriate ESTs can be a valuable source of candidates for further evaluations and inferences on their function, especially concerning those differentially expressed and responsive against root dehydration stress. On the other hand, besides the 14,903 unitags presented ESTs descriptions, others 17,366 unitags showed EST-GO terms (Table 2). The best characterized set of unitags (i.e., adequate annotation and carriage GO terms) comprised 24,924. Another appealing group (9,441 unitags with a single mismatch; Table 2) showed ESTs with no informative descriptions, requiring further characterization. A third useful group comprised 63,160 unitags with "no hit" after BLASTn (Table 2). These numbers emphasize the importance of the DeepSuperSage open architecture technology, allowing access to new gene-candidates.

Anchoring of Unitags in Soybean Genome

The unitags aligned via BLASTn against soybean transcripts and genome, both from the Phytozome database (http://www.phytozome.net/), allowed the identification of potential non-annotated genes. The BLASTn analysis involved in TSM alignments of unitags - ESTs comprised 71,171 unitags and 44,204 ESTs, which was restricted to 27,190 unique ESTs, when only the best hits were considered. On the other hand, the BLASTn analysis against the soybean genome included TSM alignments, ending up with 77,163 anchored unitags in 20 chromosomes and some scaffolds (data not shown). In an effort to determine which unitags were present in each group (ESTs, chromosomes or scaffolds), a Venn diagram (Figure 1) showed that, from the 71,171 aligned unitags with the soybean ESTs, 78 were also anchored in scaffolds, while 69,645 were anchored in chromosomes, as well. This result is consistent with what it was expected since the DeepSuperSAGE tags are generated mainly from the 3'UTRs presented in both genomic and transcripts sequences. Moreover, 1,448 unitags (Figure 1) aligned only with ESTs. When analyzing in which transcript region these alignments took place, almost all of them (1,290) showed match (TSM) with coding regions (CDS; data not show). This was not a predicted outcome. Once the *Nla*III is a frequent cutting enzyme, it was expected the digestion in 3'UTR of each expressed transcript, as mentioned before. A possible explanation for these results could be a partial digestion of cDNAs by the *Nla*III enzyme. In an attempt to minimize it, the cDNAs underwent a process of double digestion by the enzyme. Nevertheless, a lack of the *Nla*III

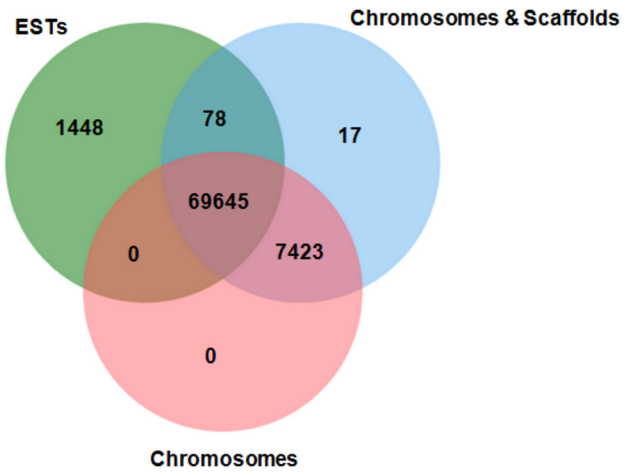


Figure 1. Venn diagram for sets of soybean unitags aligned* with soybean ESTs and genomic sequences.** * Via BLASTn (tolerating maximum of one mismatch).
** Soybean genome browser (Phytozome database: http://www.phytozome.net/).
doi: 10.1371/journal.pone.0083466.g001

enzyme restriction site in the 3'UTR or an unforeseen gene transcript sampling or even an alternative, non-described, transcript for an already predicted gene, could all explain these generated unitags.

Besides, 7,440 unitags exclusively anchored in genomic regions, without any ESTs alignments, with almost all of them (7,423 unitags) anchoring in the predicted chromosomes (only 17 anchored in scaffolds; Figure 1). From this total of chromosomes anchored unitags, 1,865 were differentially expressed in the stressed *vs* control contrast involving tolerant or sensitive accession (data not shown). To almost all of them (1,667) it was observed the perfect match of the unitags (26 bp) with the genomic sequences (data not shown). These results can indicate the presence of genes in those regions or of new transcripts that were sampled, or even of alternative, non-described, already predicted genes, but all these possibilities would include transcripts that have significant responses to the applied stress.

A more detailed analysis comprising 296 unitags anchored to chromosome 1, using the tool genome browser at the Phytozome site, showed 82 unitags anchored at introns and another 35 at the exon/intron borders (Figure 2). To the majority of these anchored sites, gene expression was reinforced by available RNA-Seq data, as indicated in the genome browser (Figure 2). Additionally, 179 unitags anchored in regions without predicted genes in their surroundings, notwithstanding 106 of those unitags presented in their respective loci RNA-Seq data covering it (Figure 2). In those sites covered by the RNA-Seq, differentially expressed unitags ($p < 0.05$) were observed after the stress stimulus (Figures 3A, 3B and 3C). Regarding this differential gene expression response, Embrapa 48 showed more induced unitags than its counterpart BR 16 (Figures 3A, 3B and 3C). Thus, the DeepSuperSAGE data, in association with the RNA-Seq data mentioned for those unannotated regions of the soybean genome, suggest that those regions may play important roles in the plant physiology, acting in response to root dehydration and assisting in plant homeostasis maintenance. Meanwhile, more studies are needed to determine the real importance of these sequences in the analyzed stress response.

Distribution of the Differentially Expressed DeepSuperSAGE Unitags

For a better understanding of the contrasts between libraries, it is necessary to understand the effects included in each comparison. The ET1-6 *vs* ET0 contrast (approach I) addressed the drought-tolerant response to root dehydration; BT1-6 *vs* BT0 (II), the drought-sensitive response to the stress; ET0 *vs* BT0 (III), the differences between the accessions under normal conditions (controls), and ET1-6 *vs* BT1-6 (IV), the differences when both accessions were under stress.

Considering the UR unitags, a Venn diagram (Figure 4A) isolated 1,127 unitags only observed in the drought-tolerant accession response to the stress (approach I), in contrast to 3,773 unitags only observed in the sensitive accession response (approach II), and while 1,557 unitags showed induction in both accessions. These exclusive UR unitags from the drought-tolerant accession probably included those transcripts and genes responsible for a better performance of this accession under the stress applied. The annotation of these tolerant-exclusive UR unitags showed 484 with informative descriptions (gene/function) and GO terms associated while 162 presented only descriptions, 209 only GO terms and 272 with no information regarding their role (Table 3).

Another useful set regarded the 4,141 UR unitags shared by the approaches I and IV (Figure 4A) that highlight the differentially induced expression of the drought-tolerant accession under stress as compared with the appropriate negative control or the sensitive accession also under stress. Considering these unitags, 1,734 presented informative descriptions and GO terms associated, while 561 only

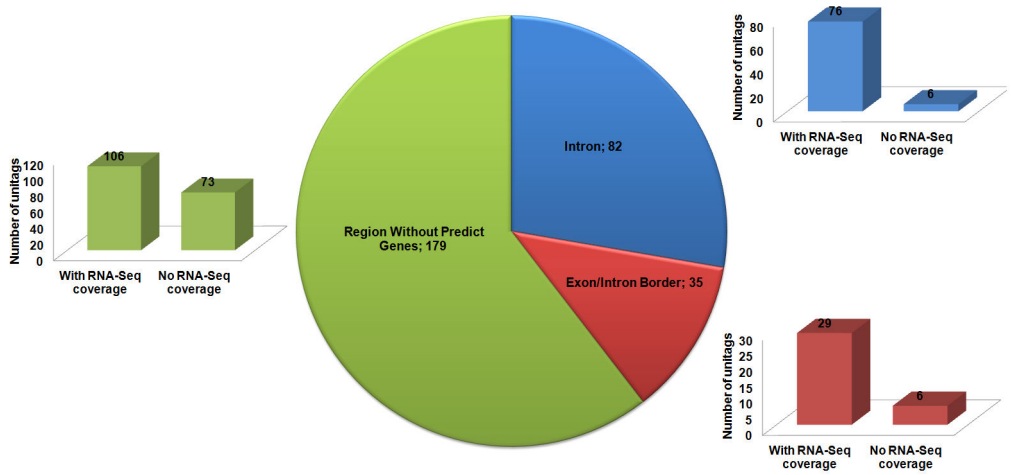


Figure 2. Number of unitags anchored in different soybean genomic regions*, with or without the coverage of RNA-Seq data*. * According to the soybean genome browser (Phytozome database: http://www.phytozome.net/).

doi: 10.1371/journal.pone.0083466.g002

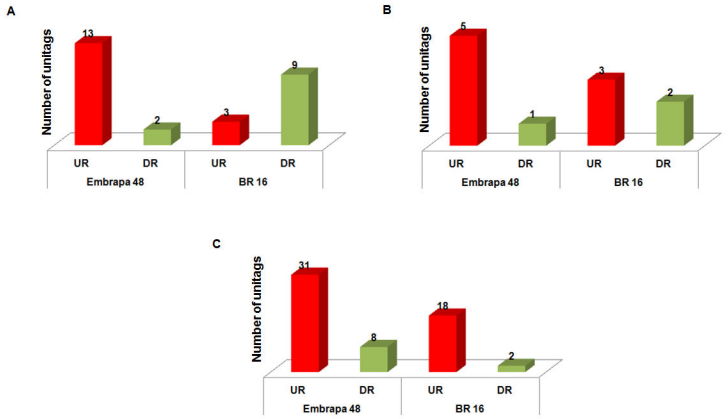


Figure 3. Number of unitags differentially expressed for each accession, and mapped in genomic regions* with predicted coverage by RNA-Seq*. (A) unitags mapped in introns; (B) unitags mapped in exon/intron borders; (C) unitags mapped in genomic regions without any predicted gene. * According to the soybean genome browser (Phytozome database: http://www.phytozome.net/).

UR: up-regulated; DR: down-regulated.

doi: 10.1371/journal.pone.0083466.g003

presented descriptions, 809 only GO terms and 1,037 lacked any knowledge (Table 3). A third relevant set included 290 unitags over-expressed in I, II and IV (Figure 4A), regarding UR unitags in the respective approaches: tolerant, sensitive, and both accessions under stress response.

The same evaluation may be carried out in the DR unitags (Figure 4B). In this analysis, 1,812 unitags showed suppression exclusively in the drought-tolerant accession response to root dehydration (approach I) while 1,798 presented in approaches I and IV (Figure 4B). From these 3,610 DR unitags (1,812 + 1,798), 1,691 presented adequate descriptions and GO terms, while 421 presented only descriptions; 766 only GO terms, whilst 732 remained uncharacterized (data not shown). Another group (339 DR unitags, Figure 4B) showed DR unitags in the approaches I (tolerant), II (sensitive), and IV (both accessions under stress). The real meaning of these suppressed sets should be investigated.

The high number of promising candidates based on unitags highlights the potential of the DeepSuperSAGE technology in the disclosure of relevant transcripts responding to the applied stress. The first step to understand the functional background relies on the use of bioinformatic tools and the effective annotation and functional categorization of the differentially expressed unitags.

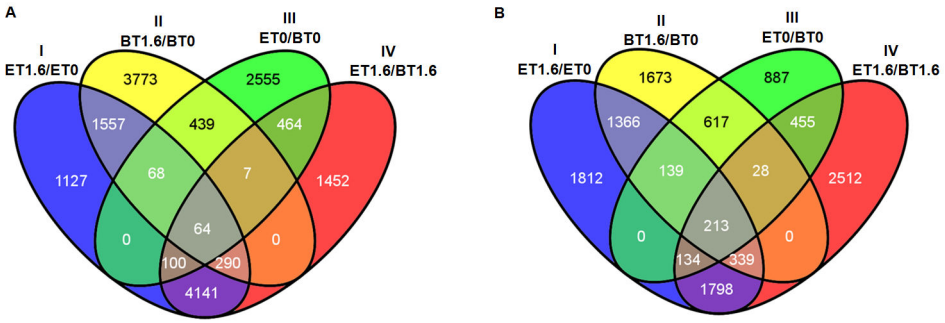


Figure 4. Venn diagram of the UR (A) and DR (B) unitags from soybean DeepSuperSAGE libraries. UR: up-regulated; DR: down-regulated; I-Tolerant accession under stress *versus* respective control (ET1.6, tolerant accession under stress library; ET0, tolerant accession control library); II-Sensitive accession under stress *versus* respective control (BT1.6, sensitive accession under stress library; BT0, sensitive accession control library); III-Tolerant accession control library *versus* sensitive accession control library; IV-Tolerant accession under stress library *versus* sensitive accession under stress library.

doi: 10.1371/journal.pone.0083466.g004

Table 3. Number of soybean UR unitags presented in different comparisons.

	Exclusive UR unitags	Common UR unitags
Descriptions / GO terms	**(ET1.6 *vs* ET0*)1**	**(ET1.6 *vs* ET0*)[1] & (ET1.6 *vs* BT1.6*)[2]**
With description / with GO	484	1,734
With description / no GO	162	561
No Description / with GO	209	809
No Description / no GO	272	1,037
Total	1,127	4,141

UR (up-regulated); *ET0 (tolerant accession 'Embrapa 48'; unstressed control); BT0 (sensitive accession 'BR 16'; unstressed control); ET1.6 ('Embrapa 48' after root dehydration stress); BT1.6 ('BR 16' after root dehydration stress).

doi: 10.1371/journal.pone.0083466.t003

Functional Categorization of ESTs Anchoring DeepSuperSAGE Unitags

The GO categorization [26] of 42,042 ESTs related to the unitags resulted in 179,670 different terms, including the three main categories: Biological Process (BP; 67,459), Molecular Function (MF; 61,568) and Cellular Component (CC; 50,643). The categorization of the ESTs related to the drought-tolerant accession (Figure 5) considered the GO terms regarding the 8,634 differentially expressed unitags (65.7% of all UR and DR unitags, approaches I, Figure 4A and 4B, respectively). The CC category refers to the place in the cell where the gene products are working [27]. The most represented CC subcategories were: "nucleus" (GO: 0005634; 575 UR and 444 DR tags), "cytoplasm" (GO: 0005737; 580 UR and 399 DR unitags) and "plasma membrane" (GO: 0005886; 321 UR and 329 DR unitags) (Figure 5). The expected prevalence of these cell compartments represent the lodging site of the genetic material responsible by the coordination of their cellular functions and reactions, and also because cell membranes are the first stress receptors, protecting the cell from modifications affecting both stress perception and rigidity of the cell structure [28]. For instance, a change in the fluidity of the plasma membrane might induce a conformational change in a receptor that activates a downstream kinase cascade [29]. Furthermore, cellular membranes threatened by reactive oxygen species (ROS) during cell metabolism, as a result of stress [30-32], produce lipid peroxides that can be used as a stress indicator [28].

Regarding the MF categories, the terms most represented was "protein binding" (GO: 0005515756; 756 UR and 711 DR unitags), "metal ion binding" (GO: 0046872; 310 UR and 276 DR unitags) and "ATP binding" (GO: 0005524; 306 UR and 241 DR unitags) (Figure 5). "Protein binding" represents selective and non-covalent interactions with any protein or protein complex, including binding to calcium-dependent proteins, calmodulin receptors, and transcription factors, among others. Considering those descriptions, the importance of this category can be seen in a brief summary where abiotic stresses (mainly drought and salinity) induce changes in cytosolic Ca^{2+} levels [33]. Ca^{2+}-binding proteins [calcium-dependent protein kinase (CDPK), calmodulin (CaM), and calcineurin B-like protein (CBL)] serve as transducers of the Ca^{2+} signal, leading to the activation of the signaling pathways, resulting in plant responses to those stresses [34-36]. These Ca^{2+}-binding

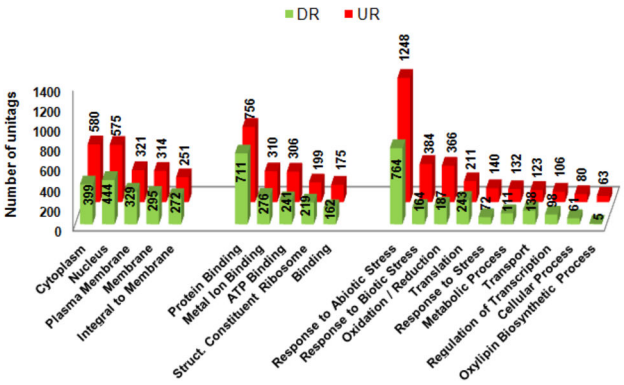


Figure 5. Gene Ontology categorization of the tolerant accession transcripts based on the UR* and DR* soybean DeepSuperSAGE unitags. UR: up-regulated; DR: down-regulated ; *Unitags from tolerant accession under stress versus respective control.

doi: 10.1371/journal.pone.0083466.g005

protein classes were expressed in all the evaluated contrasts (data not shown), whereas CDPKs and CBLs presented a higher number of up-regulated unitags when comparing both stressed accessions in relation to the appropriate negative controls (Table S1). CDPKs have recognized participation in abiotic stress tolerance, especially in the modulation of ABA signaling to reduce ROS [37]. In turn, CBLs showed an association with drought tolerance and osmotic stress in Arabidopsis. Loss-of-function Arabidopsis mutants lacking CBL1, CBL9, or CIPK1 found to be more sensitive to drought and osmotic stress than the wild-type plants [38,39].

In turn, “metal ion binding” represents proteins that interact selectively in a non-covalent way with any metal ion. In this way, dehydrins, considered effective in the tolerance process to different stresses [40], present metal binding properties to Fe^{+3}, Co^{+2}, Ni^{+2}, Cu^{+2} and Zn^{+2} [41]. Considering the contrasting accessions after stress, a total of 101 unitags associated with this protein family could be identified. Interestingly, the sensitive accession had an increased number of up-regulated unitags (70) than its tolerant counterpart (50) (Table S2). Considering UR unitags in the tolerant accession, 24 of them were n.s. or non-observed in the sensitive accession (Table S2), becoming potential targets for further studies.

“ATP binding” includes proteins that interact selectively and in a non-covalent way with ATP (adenosine 5'-triphosphate), a universally relevant coenzyme and enzyme regulator. Among them, ABC transporters stand out, also for their involvement in the abscisic acid (ABA) transport [42]. This stress-related hormone plays a key role in the tolerance process against abiotic stresses, especially regarding drought and salinity [43]. In the present evaluation, 23 possible ABC transporters found to be differentially expressed in the analyzed accessions, being eight induced only in the tolerant accession (Table S3).

In general, the analyzed transcripts in both accessions showed similar isoforms regulation based on unitags, but some presented contrasting regulation (e.g. UR in the tolerant and DR/n.s. in the sensitive) or accession-specific unitags (Table S1, S2 and S3), and these may act in their physiological differentiation when the drought stress is applied. Concerning the BP categories (biological processes in which the gene products are involved [27]), the two most depicted subcategories were “response to abiotic stress” (GO: 0009628; 1248 UR; 764 DR unitags; Figure 5) and “response to biotic stress” (GO: 0009607; 387 UR; 164 DR unitags; Figure 5). Further details will be address in the next topic, due to the importance and pertinence of the “response to abiotic stress” to the current evaluated subject (root dehydration). The second well represented subcategory was “response to biotic stress” that highlights the crosstalk mechanism, i.e., the co-activation of genes among both biotic and abiotic stress types. For example, the interaction of transcriptional regulation of environmental challenges, such as heavy metal ($CuSO_4$) stress, with incompatible necrotrophic pathogen infection revealed significant overlap between biotic and abiotic stress responses [44]. Also, large-scale microarray transcriptome data strongly supported the existence of such interaction between signaling networks [45]. Moreover, the reactive oxygen species (ROS) generation is as a key process shared between biotic and abiotic stress responses [46,47]. Thus, a growing number of evidences supports the notion that plant signaling pathways consist of complex networks with some crosstalk, thereby allowing plants to regulate both abiotic stress tolerance and disease resistance.

Analysis of the GO Subcategory “Response to Abiotic Stress”

The expression patterns by heats maps of differentially expressed unitags related to “abiotic stress response” GO category, considered different contrasts and the modulation expression values (FC) of such unitags. This GO category included “response to hormone stimulus” (GO: 0009725), “response to water” (GO: 0009415), “response to salt stress” (GO: 0009651) and “response to oxidative stress” (GO: 0006979).

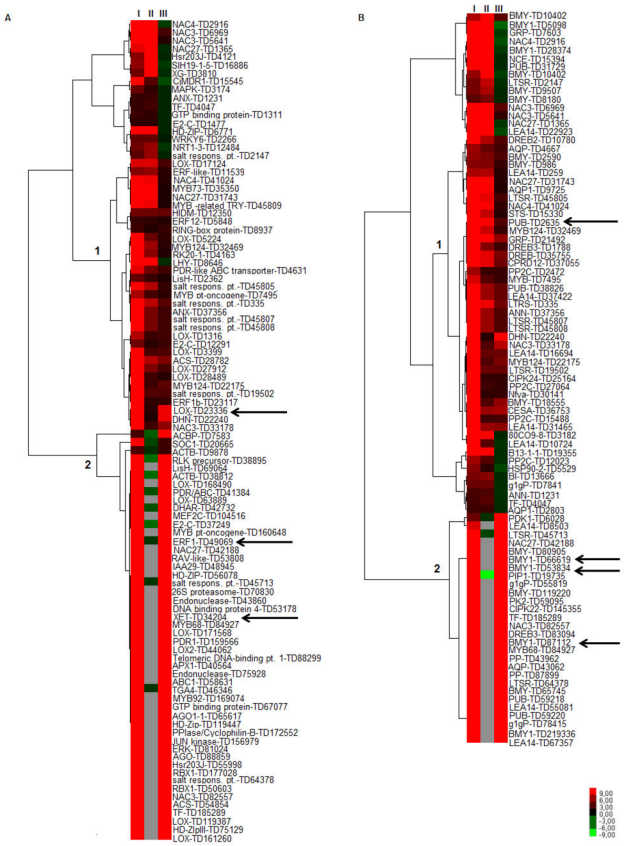


Figure 6. Hierarchical clusterization[1] regarding GO categories [(A) Response to hormones; (B) Response to water], and several contrasts[2]. [1]Gray spots: no expressed unitags; Red: up-regulated unitags; green: down-regulated unitags; black: constitutive expression [2]. Tolerant accession under stress vs. respective control (I); Sensitive accession under stress vs. respective control (II); and Tolerant accession vs. Sensitive accession, both after root dehydration stress (III). Arrows indicate transcripts mentioned in the discussion.

doi: 10.1371/journal.pone.0083466.g006

Response to Hormone Stimulus

Hormones are chemical messengers that trigger different processes in animal development, being also present in the vegetal kingdom controlling various aspects of plant growth and development [48]. Plant hormones [salicylic acid (SA), jasmonic acid (JA), ethylene (ET) and ABA] form a complex system that plays key roles in disease resistance and response to abiotic stresses, including drought [49,50].

The UR unitags clusterization covering "response to hormone stimulus" and the tolerant response (approach I) compared with the sensitive one (approach II) or even both accessions under stress (approach III) resulted in a heat map (Figure 6A; Table S4) where clusters 1 and 2 (left side of the heat map) might be highlighted. The Cluster 1 (Figure 6A; Table S4) regards unitags presented in both accessions and co-induced mainly in approaches I and II, also some unitags in III. Such a similar expression even in contrasting accessions may represent a key role of such genes in the process of acclimatization to the additional condition imposed. Example of this group was a lipoxygenase, with meaningful FCs [LOX, TD23336; FC_I = 36.0, FC_{II} = 1.5, FC_{III} = 8.5] (Figure 6A). LOX is an enzyme implicated with developmental processes and responses to stress and hormones in plants. Bell and Mullet [51] observed water deficit response associated with overexpression of some LOX isoforms in soybean (*G. max*) and pea (*P. sativum*).

The Cluster 2 (Figure 6A; Table S4) covered UR unitags in the tolerant response (I and III) and DR or absent in the sensitive response (II). Such unitags may be associated with genes whose regulation helps in the distinction of physiological behavior among accessions. Representative of this group was an Ethylene-responsive transcription factor 1b (ERF1b, TD49069; FC_I = 9.6, FC_{II} = -1.9 and FC_{III} = 9.6), a gene known for its activation in response to ethylene hormone. The expressions of ethylene-related genes occur through transduction of the ethylene signal from receptors to dedicated transcription factors [52]. ERFs (restricted to plants), an AP2/EREBP-type transcription factors, which serve as trans-acting factors at the last step of transduction [53], presented implications on stress tolerance against abiotic stress. An ERF protein, JERF3, overexpressed in tobacco was responsible for a better adaptation to stresses, such as water deficit, freezing

and high salinity [54]. The same authors observed the transcription factor activity on the control of genes involved in the oxidative stress regulation. The overexpression of another ERF gene in rice (TSRF1) also increased the tolerance against drought [55]. In the same way, a putative transcription factor ERF1b (UR in our soybean libraries), usually correlated with basal metabolic processes (development and fruit ripening in plum) [56], showed results indicating a possible involvement in soybean root dehydration response. Also, two AP2/EREBP transcripts, with expressions validated by RT-qPCR (see the specific item), reinforced this association. Another highly modulated unitag was associated to the xyloglucan endotransglycosylase family (XET, TD34204; FC_I = 26.4, FC_{III} = 26.4). Enzymes of this family have the potential to enzymatically modify wall components modulating the degree of cross-linking in the cell wall to allow cells to expand during development [57]. The first molecular genetic evidence that connects the cell wall and plant stress tolerance was provided after overexpression of a cell wall peroxidase in tobacco, improving the seed germination of transgenic plants under osmotic stress [58]. Other putative XET transcript isoform unitag (TD31210) analyzed by RT-qPCR, in the present work, validated the contrasted expression showed by the accessions (see the results along this article).

Response to Water

Unitags associated to the GO "response to water" and up-regulated in the approach I, when compared with the *in silico* expression in the approaches II and III, regarding their modulation of expression, presented, as in the previous situation, two clusters (1 and 2, Figure 6B). The modulation of the respective genes highlighted in that clusters may be explained in accordance with the reasoning presented in Figure 6A.

The Cluster 1 (Figure 6B; Table S5) showed unitags available in both accessions and up-regulated in most of the three approaches (I, II, and some of the III). This was the case of the putative U-box E3 ubiquitin ligase (PUB, TD2635) that showed one of the highest frequency modulation (FC_I = 45.0, FC_{II} = 17.3, FC_{III} = 2.2). Such protein is a part of the ubiquitin-proteasome (Ub-26S) pathway, a cascade mediated by three sequential ubiquitination enzymes that modify the selective ubiquitin ligation. About more than 5% (> 1,300 genes) of the Arabidopsis genome encodes main components that operate in the Ub-26S pathway, where about 1,200 genes encode for E3 ubiquitin ligase components [59]. This abundance illustrates how valuable this protein degradation process is in plants. The large number of E3 ubiquitin ligase genes relative to the Ub pathway-related genes in Arabidopsis and other eukaryotes is indicative of the importance of the E3 ubiquitin ligase step during the selectivity of the ubiquitin-proteasome pathway. Some induced isoforms present intimate relationship with abiotic stress, especially water stress, acting as negative regulators in Arabidopsis, coordinately controlling a drought signaling pathway by ubiquitinating cytosolic RPN12a [60].

In turn, the cluster 2 (Figure 6B; Table S5) includes unitags potentially valuable in the tolerance response. The three most expressed unitags in this group (TD87112, TD53834 and TD66619), annotated as β-amylase (BMY) enzymes, showed FCs (FC_I and FC_{III}) ranged from 19.2 to 24.0 (Table S5). BMY expression and activity is affected by abiotic stress including osmotic stress and drought. Exposure of barley [61], pearl millet and maize [62] to osmotic stress (300 mM sorbitol for four days) resulted in the increase of vacuolar BMY activity and BMY protein levels. Similarly, when cucumber cotyledons treated with 30 or 50% polyethylene glycol for up to one day, BMY activity increased followed by increases in sucrose and maltose [63]. Yang et al. [64], in turn, observed that both α- and β-amylase activities were enhanced by water stress, with the former enhanced more than the latter, and were significantly correlated with the concentrations of soluble sugars in the stems. It has been suggested that these sugars work in the osmotic adjustment process in plants [65].

Response to Salinity

The exposure to drought or salt stress triggers many common reactions in plants. Both stresses lead to cellular dehydration, which causes osmotic stress and water removal from the cytoplasm into the extracellular space, resulting in a reduction of the cytosolic and vacuolar volumes [66]. Early responses to water deficit and salt stresses are largely identical, except for the ionic component. These similarities include metabolic processes, such as photosynthesis [67] and hormonal processes, like rising levels of the plant hormone ABA [68]. Those processes include genes potentially involved in the crosstalk response. The heat map, comprising unitags associated to the GO category "response to salinity" and up-regulated in the tolerant response, when compared with those observed in the approaches II and III, highlighted two distinct clusters (Figure 7A).

The Cluster 1 (Figure 7A; Table S6) encloses unitags presented in both accessions, but UR in the approaches I, II and III, or even n.s. in the approach III. The up-regulation in approach III regarded unitags with a higher expression in the tolerant after stress, when compared with the sensitive one. This situation regarded TD1254 unitag (FC_I = 6.1, FC_{II} = 3.1 and FC_{III} = 1.5), a possible WRKY transcription factor. According to Eulgem et al. [69] members of this family were overexpressed responding to various stress types. Among 72 WRKY genes (Arabidopsis), 49 presented differential expression in response to hormones (salicylic acid treatment) or biotic stress (infection by a bacterial pathogen) [70]. Also, these genes were implicated in responses to wounding (*A. thaliana* [45]), drought and heat (tobacco [71]) and cold (*Solanum dulcamara* [72]). At least 64 soybean SuperSAGE unitags were possible WRKY transcription factors transcripts (data not show).

The Cluster 2 (Figure 7A; Table S6) contain unitags induced in approaches I and III (absent or n.s. in the approach II). One of the most expressed modulated unitags in this group was TD22175 (FC_I = 28.8, FC_{III} = 3.4), a possible MYB transcription factor. Members of the MYB family are abundant in all eukaryotes, being the most frequent transcription factor family (TF) in plants [73]. In the present data, 425 unitags annotated as MYB TFs (data not show). MYB TFs are key factors in the regulation pathways that control development, metabolism and

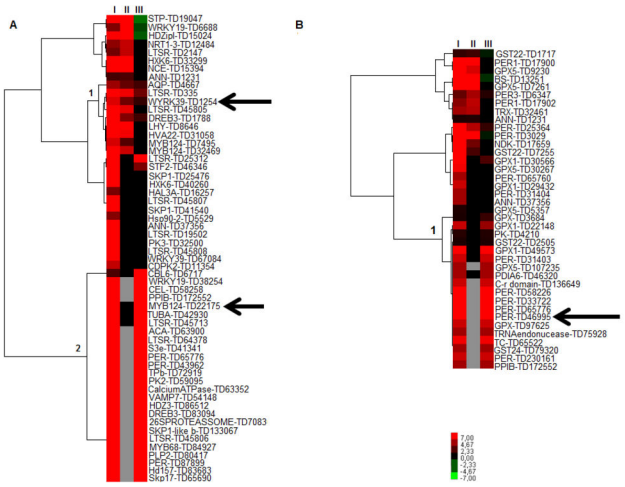


Figure 7. Hierarchical clusterization[1] regarding GO categories [(A) Response to salinity, (B) Response to oxidative stress], and several contrasts[2]. [1]Gray spots: no expressed unitags; Red: up-regulated unitags; green: down-regulated unitags; black: constitutive expression [2]. Tolerant accession under stress vs. respective control (I); Sensitive accession under stress vs. respective control (II); and Tolerant accession vs. Sensitive accession (both under stress; III). Arrows indicate transcripts mentioned in the discussion.

doi: 10.1371/journal.pone.0083466.g007

response to biotic and abiotic stress [74]. Concerning their role in the drought tolerance, Seo et al. [75] reported that a R2R3-type MYB TF (MYB96) regulated drought stress response by integrating ABA and auxin signals. The putative MYB124 observed in the DeepSuperSAGE data (MYB124_TD22175) could be, along with MYB88, generating regular stomatal patterning, as in Arabidopsis [76], optimizing gas exchange and guard cell ion transport.

Response to Oxidative Stress

Reactive oxygen species (ROS) production is a unifying commonality in a large number of abiotic stresses [77]. The redox-modulated changes are main events in cellular responses since ROS may help stress perception, but also damage the cell due to oxidation of membranes and other cellular components [78]. Responsive genes in such situations are, therefore, relevant to the maintenance of cellular homeostasis in adverse situations. The heat map based on the "response to oxidative stress" GO category and UR unitags in the approach I, compared with the approaches II and III, presented a set of UR unitags (approaches I and III), probably acting in the physiological behavior differentiation showed by the accessions, since these unitags were absent or n.s. in the sensitive accession (approach II, Cluster 1, Figure 7B; Table S7). One of the highest expressed modulated unitag (TD46995; FC_I = 110.4, FC_{III} = 110.4) was a putative peroxidase (PER). PER is an enzyme with oxidoreductase function that oxidizes a vast array of compounds (hydrogen donors) in the presence of H_2O_2. Like other enzymes from the ROS group, PER is a "ROS Scavenging Enzyme". ROS scavenging increases the level of antioxidant enzymes, contributing to salt tolerance in different plants, including soybean [79]. This is in consonance with Zhang and Kirkhan [80] that observed an increase of peroxidase activity associated to the water deficit response.

Differential Response of the Accessions Based on Biological Processes (BP)

A sample of the differential behavior between the studied accessions can be observed in the Figure 8, representing some BP subcategories with UR unitags observed in the approaches I, II and III. Considering six among 10 analyzed subcategories ["translation" (GO: 0006412), "metabolic process" (GO: 0008152), "response to water deprivation" (GO: 0009414), "regulation of transcription" (GO: 0045449), "response to wounding" (GO: 0009611), "transmembrane transport" (GO: 0055085)], the number of UR unitags in the tolerant accession (approach I) was larger than that in the sensitive one (approach II), indicating that, in those subcategories, the tolerant accession recruited and differentially expressed a larger number of unique transcripts. Also based on those subcategories, from the approach III it was clear that some UR unitags belonging to the tolerant accession were also up-regulated in relation to the sensitive accession, both under stress. From the quantitative point of view, the approach III revealed how many unitags were up-regulated in the tolerant accession, in relation to the sensitive one (both under stress). Here, the subcategory "transmembrane transport" had the largest number of UR unitags in the approach III, as compared to approach I. At a first glance, the results seemed to be incoherent. However, to the set of common UR unitags from approaches I and III it is necessary to include the constitutively expressed unitags from the approach I, as these tags are indeed up-regulated, when compared to the contrast III (both accessions under stress).

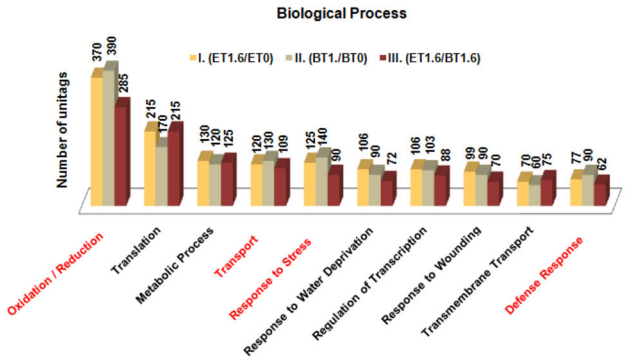


Figure 8. Number of UR unitags characterized by Gene Ontology ("Biological Process" subcategories), considering several contrasts*. UR: up-regulated; *Tolerant accession under stress versus respective control (I); Sensitive accession under stress versus respective control (II) and Tolerant accession vs. Sensitive accession (both under stress; III).
doi: 10.1371/journal.pone.0083466.g008

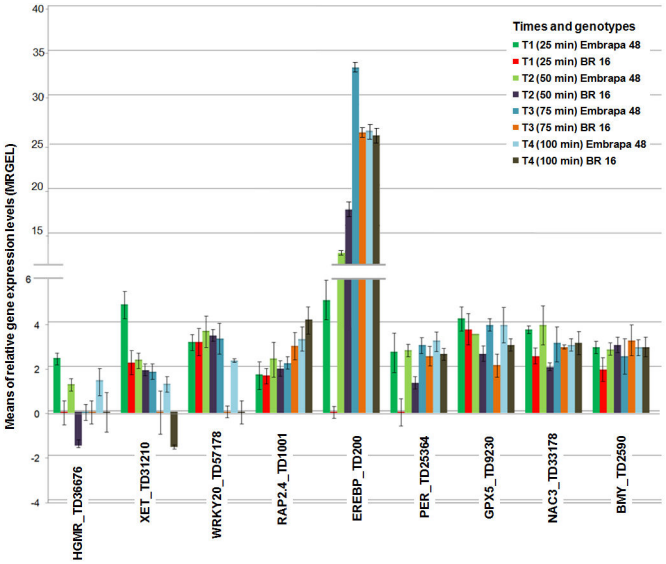


Figure 9. RT-qPCR of the unitags measured at the appropriate sample time using REST2009 software.
doi: 10.1371/journal.pone.0083466.g009

On the other hand, in four subcategories ["oxidation reduction" (GO: 0055114), "transport" (GO: 0006810), "response to stress" (GO: 0006950), "defense response" (GO: 0006952)], a larger absolute number of UR unitags expressed by the sensitive accession (approach II), when compared to the tolerant one (approach I), demonstrated for those subcategories that the sensitive accession recruited and induced more unitags (Figure 8). Despite of this higher number of inducted unitags by the sensitive accession, in these subcategories, the overexpression by the tolerant accession in relation to the counterpart sensitive one, when both under stress, were demonstrated (approach III, Figure 8), pointing to a higher transcriptional efficiency of the tolerant over the sensitive after the stress. In short, the transcripts pool from the tolerant comparing with the sensitive accession, varied in both, quantitative and qualitatively aspects.

Expression Analysis of Unitags in Contrasting Accessions by RT-qPCR

The strategy to generate two DeepSuperSAGE libraries for each accession [negative control and bulk of samples gathering different times of stress imposition (25, 50, 75, 100, 125 and 150 min) reduced the number of libraries and became economically more realistic the project, but in turn also became more difficult to analyze the expression over the times sampled. The use of the RT-qPCR method provided the opportunity to integrate the differential expression of the candidate gene to the temporal variant opening of the bulked sample, based on expression in the times: 25, 50, 75, 100 min. The hereby studied nine genes (Table S8), covering contrasting and similar accession responses, present a concise overview of how transcriptional orchestration works in the analyzed condition, helping the understanding of the plant

physiological behavior of each accession, and presenting the way that a transcript population changes over time in the addressed situation.

Regarding the unitags showing different response between the accessions (UR in the tolerant and DR or n.s. in the sensitive contrasts in relation to the appropriate negative control), those validated by RT-qPCR were:

a) 3-hydroxy-3-methylglutaryl coenzyme A reductase [HMGR, EC 1.1.1.34; soybean gene model Glyma11g09330.1; unitag TD36676 ($FC_{tolerant}$ = 24.0, $FC_{sensitive}$ = -9.8)].

The HMGR enzyme acts in the metabolism of isoprenoids, also called terpenoids. In plants, terpenoids show variation in structure and function, covering besides isoprenols (essential to biomembranes), also hormones, carotenoids and clorophyllins (photosynthetic pigments), among others [81]. Terpenoids can be synthesized by two mechanisms: desoxyxylulose 5-phosphate/2-C-methyl-D-erythritol 4 phosphate pathway (also called DPX pathway), widespread in eubacteria, and Mevalonate (MVA) pathway, prevalent in archaea and eukaryotes [82]. The HMGR enzyme catalyzes a key regulatory step of the MVA pathway, being modulated by various endogenous and external stimuli [83], including phytohormones, calcium, calmodulin, light, wounding, elicitor treatment, and pathogen attack [84]. Recently, Yang et al. [85] demonstrated in *Savia miltiorrhiza* hairy roots that HMGR mRNA levels and the enzyme activity were stimulated by abscisic acid (ABA) and methyl jasmonate (MJ), hormones known to be involved in the water deficit response, as well as polyethylene glycol (PEG), a compound that mimic drought stress effects in plants. The studied accessions differed regarding the transcriptional regulation of the unitag TD36676, in the course of the tested times (Figure 9), with the tolerant accession showing overexpression ($p < 0.05$) at all sampled times (MRGEL: 1.2 to 2.4; Figure 9), compared to the appropriate negative control (T0). Exception occurred for T3, in which the expression did not change (Figure 9). The sensitive accession, in turn, did not show differential expression at times T1, T3, T4, whereas, in T2, the expression decreased in regard to the sensitive negative control (MRGEL: - 1.5; Figure 9). Therefore, a role of this gene together with the MVA pathway in root dehydration differential response showed by the tolerant accession is likely.

b) Xyloglucan endotransglycosylase [XET, EC 2.4.1.207; soybean gene model Glyma13g38040.1; unitag TD31210 ($FC_{tolerant}$ = 28.8, $FC_{sensitive}$ = -6.8)].

The XET or XTH enzyme acts in processes covering plant cell wall modifications. During cellular expansion, XET enhances the extensibility of the cell wall by cleaving xyloglucan at the xyloglucan–cellulose network presented in the plant cell wall [86,87]. Under water deficit conditions, the plant upper part growth inhibition and maintenance of root growth are often associated, in a well adaptive mechanism. However, in dry soils conditions, plant roots tend to grow seeking water richer zones. Despite XET relationship to plant cell wall strengthening processes [88], studies have shown a positive correlation with root elongation [89,90] and also of other plant organs [91]. The RT-qPCR results related to TD31210 unitag showed different behavior in both contrasting accessions as shown in Figure 9. The tolerant accession presented overexpression at all the analyzed stress times (MRGEL: 1.3 to 4.8 times, relative to T0; Figure 9). On the other hand, the sensitive one failed to keep this overexpression during the same studied times. Expression modulation for this accession (MRGEL: 1.9 to 2.2; Figure 9) occurred only in the first two times (25 and 50 min; Figure 9). In the following times (75 and 100 min), the expression was not significant or showed suppression (MRGEL: -1.6; Figure 9). Thus, the results suggest that radicular growth and soil remodeling may be involved in the tolerance observed in Embrapa 48 (tolerant accession), enabling the reestablishment of the proper functioning of its physiology.

c) Transcription Factor WRKY20 [soybean gene model Glyma05g36970.1; TD57178 unitag ($FC_{tolerant}$ = 21.6, $FC_{sensitive}$ = -4.4)].

The main steps in plant tolerance to adverse environmental conditions process are stress condition perception, signal transduction, activation and regulation of stress responsive genes. The two previous steps requiring greater efforts at the transcriptional level, with a large portion of the plant genomics capacity driven by TFs. Soybean has 5,671 putative TFs, distributed in 63 families, which equates to 12.2% of the 46,430 predicted soybean protein-coding loci [25], while Arabidopsis and rice genomes code more than 2,100 and 2,300 TFs respectively [92]. Among TFs, WRKY is one of the largest families of plant transcriptional regulators modulating plant processes [93], also in plant stress responses. In rice, for instance, OsWRKY11 overexpression (under the control of HSP101 promoter) led to enhance drought tolerance [94]. Recently, Luo et al. [95] observed that the expression of wild soybean WRKY20 in Arabidopsis enhances drought tolerance and regulates ABA signalling. The differential behavior of the analyzed accessions for the expression (RT-qPCR) based on the TD57178 unitag was evident. The tolerant accession kept its induction over all the stress times tested (25, 50, 75, 100 min) modulating the expression (MRGEL) 2.3 to 3.6 times in relation to T0 (Figure 9); the sensitive accession, in turn, showed overexpression only in the early time points (25 and 50, MRGEL: 3.1 and 3.4, respectively, Figure 9). According to Chen et al. [96] the strict control and fine-tuning of WRKY proteins during plant stress responses contribute to the installation of complex signaling networks, highlighting the importance of WRKY proteins in plant abiotic stress response.

The unitags with similar regulation in both accessions under stress (Figure 9) allowed six genes to be RT-qPCR validated:

a) NAC3 transcription factor [NAC3; soybean gene model Glyma06g38410.1; unitag TD33178 ($FC_{tolerant}$ = 33.6, $FC_{sensitive}$ = 10.6); cluster 1 (gene expression heat map Figure 6A); Table S4].

Under stress conditions, plants do not induce only gene transcriptions that operate in cellular protection, namely, enzyme coding genes and other functional proteins, but they also produce the regulatory transcripts that act in the transduction of signals from their perception organs. In this context are the transcription factors (TF) coding genes. Among the plant-specific transcription factors, NAC (NAM, ATAF,

CUC) proteins constitute one of the largest families, present in a wide range of land plants [97]. The NAC was the most represented in the expression cluster "Response to hormonal stimulus" (Figure 6A). Specifically, the TD33178 unitag (NAC3 isoform) showed large modulation for both accessions (Figure 6A; Table S4); this being induced at all analyzed time intervals [tolerant accession (MRGEL: 3.0 to 3.9; sensitive accession: MRGEL: 2.0 to 3.1; Figure 9). This TF participation in the tolerance process to abiotic stresses has been demonstrated. Liu et al. [98] obtained tobacco transgenic lines transformed with *AhNAC3* (from peanut), and those showed hyper-resistance to dehydration and drought stresses and accumulated more proline and less superoxide anion (O_2^-) than wild type under dehydration and drought conditions. They also observed that four functional genes, superoxide dismutase, pyrroline-5-carboxylate synthetase, late embryogenic abundant proteins, and early response to drought 10, were induced in the transgenic lines, been suggested that NAC3 improves water stress tolerance by increasing superoxide scavenging and promoting the accumulation of various protective molecules.

b) AP2 (Apetala2) / ERF family, also called AP2/EREBP [99,100], presented the soybean gene models Glyma13g01930.1 and Glyma16g27950.1 associated, respectively, to the unitags TD1001 ($FC_{tolerant}$ = 7.2; $FC_{sensitive}$ = 6.0) and TD200 ($FC_{tolerant}$ = 22.0; $FC_{sensitive}$= 21.0).

The AP2/ERF (EREBP) is a plant-specific TF large family that shares a well-conserved DNA-binding domain, comprising AP2, RAV, EREBP subfamilies, with the EREBP subfamily subdivided into DREB (Dehydration-responsive element-binding) or A subgroup and the ERF (Ethylene response factor) or B subgroup [101]. The up-regulation of those unitags observed in both accessions, in relation to the expression in the appropriate unstressed controls, suggests a conservative action even in contrasting accessions. Concerning the TD1001 unitag, the level of expression was similar for both accessions, considering each time evaluated, showing differential expression since 25 min after stress (Figure 9). Additionally, the unitag expression level in both accessions was smaller than those observed in TD200 (Figure 9). BLASTn analysis based on the RefSeq_RNA database (NCBI) revealed that Glyma13g01930.1 represents, specifically, an FT-type RAP2-4 (data not shown). RAP2-4 is a TF AP2/DREB-type, which belongs to EREBP subfamily. This TF was down-regulated by light but up-regulated by salt and drought stresses, in Arabidopsis [102]. Recently, Rae et al. [103] investigated the expression and function of RAP2-4B and RAP2-4 (both DREB TFs) using microarray-based transcriptional profiling of double knockout and overexpression lines. Expression analysis of stressed and control plants revealed both genes highly expressed in stems and roots and differentially induced in response to cold, dehydration and osmotic stress. The same authors also concluded that RAP2-4 is a probable significant aquaporin co-expression network regulator during the early phase of dehydration response. During that study, six aquaporin genes – from which three (*AtPIP2; 1*, *AtPIP2;2* and *AtPIP2;3*) from the PIP group and three (*AtTIP1;1*, *AtTIP2;2* and *At*TIP 2;3) from the TIP group – were down-regulated in the double knockout line and consequently up-regulated in the appropriate overexpression line [103]. In relation to the TD200 unitag (Glyma16g27950.1; annotated as a TF AP2/ERF, EREBP subfamily), the tolerant accession response revealing a faster response (25 min) than the sensitive accession (50 min; Figure 9). Also, in general, the average gene expression level presented by the tolerant accession was higher than the observed in the sensitive (Figure 9) also considering each analyzed time point. BLAST2Seq analysis performed to gather similarity between Glyma16g27950.1and Glyma13g01930.1 transcripts, since they belong to the same EREBP subfamily, did not show significance (data not shown). As mentioned before, this TF subfamily comprises DREBs and ERFs [101]. In soybean, the overexpression of a DREB homologous gene (GmDREB2) activated expression of downstream genes in transgenic Arabidopsis, resulting in enhanced tolerance to drought and high-salt stresses, without plant growth retardation [104]. Besides, its overexpression in tobacco resulted in higher proline content rates compared to wild type plants under drought condition [104]. ERFs also respond to drought tolerance. In soybean, GmERF3, a member of this subfamily, showed its expression induced by biotic stress [soybean mosaic virus, SMV] and abiotic stresses, such as high salinity, drought and hormones (ABA, SA, JA and ET) [105]. Additionally, osmoregulation is among the known ERF-associated mechanisms. The overexpression of GmERF3 in transgenic tobacco led to higher levels of free proline and soluble carbohydrates compared to wild-type plants under drought conditions [105].

c) B-amilase [BMY; soybean gene model Glyma15g10480.1; unitag TD2590 ($FC_{tolerant}$ = 4.0, $FC_{sensitive}$ = 3.5); cluster 1, gene expression heat map Figure 6B; Table S5].

As mentioned before, it has been suggested that β-amylases act in the cellular osmotic regulation, when the plant is exposed to drought [65]. According to Ocampo and Robles [106] osmotic adjustment is the plant capacity to increase its solute concentration in leaves, roots and other organs responding to dehydration. This leads to the maintenance of the turgor pressure when the plant water potential declines, being crucial to the support of several biochemical and physiological processes [107]. In this study, one of the most abundant transcript classes in the gene expression heat map "Response to water" comprised 16 induced β-amylase isoforms in the tolerant accession and repressed or n.s. in the sensitive one (Figure 3B; Table S5). EST anchoring the TD2590 unitag after primers design and the respective RT-qPCR validation confirmed induction expression by both accessions, since the beginning of stress imposition (25 min) until the end time (100 min) [tolerant accession MRGEL: 2.5 to 2.9; sensitive accession: MRGEL: 1.9 to 3.2; Figure 9]. Such similar regulation in contrasting accessions, suggests the BMY importance in soybean root dehydration stress response.

d) Glutathione Peroxidase 5 [GPX5, EC 1.11.1.9; soybean gene model Glyma11g02630.1; unitag TD9230 ($FC_{tolerant}$ = 16.9, $FC_{sensitive}$= 12.8); gene expression heat map Figure 7B; Table S7].

Plant survival in non-favorable growth conditions depends on its ability of stress perception, stimulus propagation and its physiology adaptation to new situations. As it was mentioned in

the previous section ("Response to Oxidative Stress") modulation of genes related to the cell redox status modulation is of utmost importance once such molecules act in the stress perception; however their excess presence damage cell structure [78]. By analyzing the active components of this response to oxidative stress and the gene expression heat map "Response to Oxidative Stress" (Figure 7B, Table S7), it was noticed that GPX and PER (see next target) were the most abundant enzymes. The TD9230 unitag, a potential GPX5 (Figure 7B; Table S7), after RT-qPCR validation, confirmed induction in both accessions, for all the analyzed time intervals, it being more modulated in the tolerant accession than in the sensitive (Figure 9). Yoshimura et al. [108] generated transgenic tobacco plants expressing a GPX-like protein in the cytosol (TcGPX) or chloroplasts (TpGPX). The transgenic plants showed increased tolerance to oxidative stress caused by application of methylviologen (MV: 50 μM) under moderate light intensity (200 μE m^{-2} sec^{-1}), chilling stress under high light intensity (4 °C, 1000 μE m^{-2} sec^{-1}), or salt stress (250 mM NaCl). In the transgenic plants the capacity of the photosynthetic and antioxidative systems remained higher than those of wild-type plants under chilling or salt stress.

e) Peroxidase [PER, EC 1.11.1.7; soybean gene model Glyma20g31190.1; unitag TD25364 ($FC_{tolerant}$ = 33.6, $FC_{sensitive}$= 10.6); gene expression heat map Figure 7B; Table S7].

As mentioned before, PER was one of the most active components responding to oxidative stress (Figure 7B, Table S7). The unitag TD25364, a potential PER, presented induction (RT-qPCR; Figure 9) in both accessions. However, the tolerant accession response was faster, inducing PER since the time interval of 25 min, than the sensitive accession that only began to respond at the 50 min time interval (Figure 9). The quick response of the ROS scavenging associated machinery is of utter importance to plant organisms exposed to non-favorable growth conditions since it confers adaptative advantages to the organisms that behave adopting such transcriptional strategy. In this way, the tolerant accession would be more effective in this transcriptional response to the generated oxidative stress.

Conclusions

This work provides novel genomic resources to support soybean approaches aiming to increase drought tolerance. A global evaluation of the soybean transcriptome under root dehydration stress using DeepSuperSAGE and high throughput sequencing allowed the identification of 1,127 unitags exclusively overexpressed in the stress-tolerant accession, many of them with considerable expression fold changes as compared to the tolerant negative control. Some of these were non-annotated unitags (209) only characterized by gene ontology terms using the EST anchoring the unitag. Other non-annotated induced unitags (272) showed "no hits"; these unknown transcripts were probably not yet associated to drought. Both groups comprise potential targets for further evaluation, validation and transgenesis. Also, some up-regulated unitags could be associated with important categories recognized by their role in plant abiotic stress response (e.g. "response to hormone stimulus", "response to water", "response to salt stress" and "response to oxidative stress"), revealing that the response to water deficit in both accessions recruited a repertoire of different genes, from the quantitative and qualitative point of views, with a higher number of induced genes as compared to those repressed. Additionally, data validation by RT-qPCR revealed an accession-specific transcriptome reprogramming detected 25 minutes after stress imposition, highlighting not only the effective responses associated to the tolerant accession, but also the non-efficient responses considering the sensitive accession.

Materials and Methods

Biological Material, Experimental Design and Stress Application

For root dehydration treatment, soybean (*G. max*) accessions 'Embrapa 48' (drought-tolerant) and 'BR 16' (drought-sensitive) [109] were grown in a greenhouse at Embrapa-Soybean station (Londrina, Brazil) using an aerated hydroponic system in 30 L plastic containers with pH 6.6-balanced nutrient solution as described by Kulcheski et al. [110]. Briefly, seeds were pre-germinated on moist filter paper in the dark at 25°C ± 1°C and in 65% ± 5% relative humidity. Plantlets were then placed in polystyrene supports, so the roots of the seedlings were fully immersed in the nutrient solution. Each seedling tray was maintained in a greenhouse at 25°C ± 2°C and in 60% ± 5% relative humidity under natural daylight (photosynthetic photon flux density (PPFD) = 1.5 × 10^3 μmoles m^{-2} s^{-1}, equivalent to 8.93 × 10^4 lux) for 12 h/day. After 15 days, seedlings with the first trifoliate leaf fully developed (V2 developmental stage) [111] were submitted to different root dehydration periods, when the nutrient solution was removed from each plastic container where the roots were kept, in the tray, in the dark, without nutrient solution or water for 0 minutes (negative control) or 25 (T1), 50 (T2), 75 (T3), 100 (T4), 125 (T5), 150 minutes (T6). At the end of each period, the roots of the seedlings were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction. The experimental design was a factorial (accession × root dehydration times) with three replicates. Each replicate composed of five plantlets sampled in bulk. To avoid the impact of volatile compounds, each treatment was carried out in isolated spaces presenting the same growing conditions.

RNA Extraction and Generation of DeepSuperSAGE Libraries

Total RNA was extracted of each treatment using the Plant RNeasy (Qiagen) kit, taking equimolar RNA quantities of each sample for bulk composition. Four DeepSuperSAGE libraries were generated with the bulks or the control RNA samples: ET1-6 (root dehydration-tolerant accession after stress – bulk of six times), BT1-6 (root dehydration-sensitive accession after stress – bulk of six times), ET0 (tolerant accession, negative control) and BT0 (sensitive accession, negative control). DeepSuperSAGE libraries were generated according to the procedures described by Matsumura et al. [112], under the guidance of GenXPro GmbH (Frankfurt, Germany) technical

staff, with posterior SOLEXA sequencing of the tags. The data presented here can be downloaded from the Genosoja project (http://bioinfo03.ibi.unicamp.br/soybean/) [113].

Statistical Analysis and Unitag-Gene Annotation

Tags (26 bp) were analyzed to identify unique tags (unitags) and those differentially expressed ($p < 0.05$), based on Poisson statistics developed by Audic and Claverie [114], as implemented in DiscoverySpace (v.4.01) software [115]. The singlets (tags sequenced only once) were excluded from the present evaluation. Unitags were annotated by BLASTn [116] against nucleotide sequences from following databases: (1) NCBI (the Plant Reference Sequence Database – RefSeq, and a limited dbEST file with ESTs from genera *Cicer* and *Pisum*; National Center for Biological Information, accessed in October 2012 [117]); (2) Kyoto Encyclopedia of Genes and Genomes, KEGG (ESTs from *Lotus japonicus*; *G. max*; *Vigna unguiculata*; *Phaseolus vulgaris*; *P. coccineus*; *Medicago truncatula*; *Arachis hypogaea* and *A. thaliana*; accessed in October 2012 [118]; (3) Resource for Plant Comparative Genomics, PlantGDB [119] (plant mRNAs multifasta file); (4) Plant Gene Indices / Gene Index Project (PHVGI, release 3-1; PCGI, release 1; GMGI, release 15; MTGI, release 9; LJGI, release 5 [10]); (5) Soybean Phytozome V5.0 (Glyma1 cDNA dataset [27]); (6) Nord*EST*: cowpea ESTs from Brazilian Nord*EST* network clustered with ESTs from the HarvEST-cowpea project [120]. The clusters and singlets were previously annotated by BLASTx (e-value cut off e^{-10}) against the UniProtKB/Swiss-Prot database [121]. The BLASTn alignments (unitag-hit) with e-values of 0.001 or less and scores higher than 42, reflecting unitag-EST alignments tolerating a maximum of a single mismatch (TSM) were identified among the plus/plus alignments without mismatches regarding the four first bases CATG, to guarantee the integrity of the unitag. Besides the BLASTn analysis of unitags against ESTs, the soybean genome available in the Phytozome database (http://www.phytozome.org/) was also used in order to anchor unitags (TSM alignments) and to complement the analysis.

Gene Ontology of ESTs Anchoring DeepSuperSAGE Unitags

Multifasta file comprising the ESTs related to the unitags alignments (TSM) was analyzed by a local BLASTx using the UniProtKB/Swiss-Prot database and *e-value* cut-off e^{-10}. The result imported by the software BLAST2GO v.2.4.4 [122] allowed the GO-mapping step. The GO terms in a data matrix together with the previously annotation results enabled data filtering and searches by keyword in a spreadsheet file.

Keyword Search and Tag-Gene Annotation

Keyword searches performed on the original EST annotations included all BLASTn results and databases. Searches carried out on GO terms tried to confirm identities. The choice for best unitag-hit (EST) considered three consecutive rounds of redundancy elimination: (i) hits with inadequate/limited gene description and no GO term available; (ii) hits with only adequate description or only GO terms available and (iii) hits with adequate description and GO terms available. In each elimination round, only the best alignment (higher score, alignment size and identity) for each unitag, presenting (i) an adequate described soybean hit or (ii) in the absence of that, the best described hit from a soybean related species, suggested by Doyle and Luckow [123] or (iii) in the absence of both, an adequate described hit from another angiosperm, remained as the most informative ones.

The Fold Change Estimation, the Heat Maps and Venn Diagram

Values reflecting expression data (*p-value* and up- or down-regulation regarding each unitag) were associated to the data matrix together with the respective unitag annotation, GO terms, the normalized frequencies in the libraries and the fold change values (FC). FC values comprised the ratio (R) of the normalized frequencies of one unitag in the contrast of two libraries, where the "zero" frequency was replaced by "one". When $R > 1$ the FC were immediately considered and when $R < 1$ the $FC = -1/R$. Negative FC values indicated repressed unitags. To generate heat maps considering different comparisons, differentially expressed unitags were hierarchically clustered (HAC) with support of the Cluster 3.0 (v.1.1.4r3) software [124] using default parameters and FC values as input data. The lateral dendrograms were generated using the TreeView software [125]. Finally, the Venn diagrams were generated with assistance of the software Venny [126].

RT-qPCR Analyses

In order to substantiate the DeepSuperSAGE expression, nine selected unitags were validated by RT-qPCR. Unitags were selected based in their annotation, expression differentially regulated by the accessions and expressive FC values. Then, cDNAs related to the selected unitags were used for primers development, using the tool QuantPrime (http://www.quantprime.de/) and default parameters. The selected transcripts involved transcription factors [AP2 / ERF (EREBP), WRKY and NAC family], xyloglucan endotransglycosylase,3-hydroxy-3-methylglutaryl coenzyme A reductase 4, peroxidase, glutathione peroxidase 5 and β-amilase (Table S8).

Considering that DeepSuperSAGE and RT-qPCR are different methods, the expression levels observed with these procedures were not expected to be similar. So, to validate the DeepSuperSAGE data, the samples were not pooled for the RT-qPCR analysis, as they were for DeepSuperSAGE, and it was considered an agreement between the two approaches, when at least in one time point in RT-qPCR, similar results to DeepSuperSAGE ($p < 0.05$) were demonstrated. For this purpose, cDNA synthesis was achieved using total RNA extracted with Trizol® reagent (Invitrogen) and the QuantiTec® Reverse Transcription Kit (Qiagen); both according to the manufacturer's instructions. RT-qPCR analyses were performed in a 7300 Real Time System (Applied Biosystems) thermocycler and the Platinum® SYBR® Green qPCR SuperMix UDG (Invitrogen). The reactions conditions were 50°C for 2 min, 95°C for 2 min, 45 cycles at 95°C for 15 s, 62°C for 30 s, and 72°C for 30 s; data were collected in the exponential phase of the RT-qPCR. The formula $E = [10^{-1/slope}] - 1$ was applied to calculate the reaction efficiency. For each time point (0, 25, 50,

75, 100 min under root dehydration stress, three biological replicates, each with three technical replicates, were analyzed. Results were captured by the Sequence Detection program (Perkin Elmer) and analyzed by the Relative Expression Software Toll (REST) version 2.0.7 [127]. *Gmβ-actin* and *GmRNA18S* were used as reference genes for normalization [128]. Primers sequences, their efficiencies in RT-qPCR reactions and expected amplicons (bp) for the selected target genes are showed in Table S8.

All relative quantification was assessed using REST software 2009 [127,128], REST Standard, using the pair-wise fixed randomization test with 2,000 permutations.

Supporting Information

Table S1. Ca^{2+}-binding proteins [calcium-dependent protein kinase (CDPK), calmodulin (CaM), and calcineurin B-like protein (CBL)] differentially expressed in at least one treatment. Unitags associated to CDPKs, CaM and CBLs, their normalized frequencies, *p*-value [114], fold changes (FC_I: tolerant accession under stress *vs* tolerant accession control; FC_{II}: sensitive accession under stress *vs* sensitive accession control), regulation (in different approaches), identification, appropriate annotated EST, reference data bank, gene acronym and score regarding each unitag-EST alignment. *GMGI: *Glycine max* Gene Index (Plant Gene Indices / Gene Index Project Database); GDB: PlantGDB database; KEGG: Kyoto Encyclopedia of Genes and Genomes Database; PHVGI: *Phaseolus vulgaris* Gene Index (Plant Gene Indices / Gene Index Project Database) [1].ET1.6, tolerant accession under stress library; ET0, tolerant accession control library [2].BT1.6, sensitive accession under stress library; BT0, sensitive accession control library. UR: up-regulated; DR: down-regulated; n.s.: not significant ($p < 0.05$); (−) unitag not expressed.
(XLS)

Table S2. Dehydrin transcripts differentially expressed in at least one treatment. Unitags associated to dehydrins, their normalized frequencies, *p-value* [114], fold changes (FC_I: tolerant accession under stress *vs* tolerant accession control; FC_{II}: sensitive accession under stress *vs* sensitive accession control), their regulation (in different approaches), identification, appropriate annotated EST, reference data bank, gene acronym and score regarding each unitag-EST alignment. *GMGI: *Glycine max* Gene Index (Plant Gene Indices / Gene Index Project Database); GDB: PlantGDB database; MTGI: *Medicago truncatula* Gene Index (Plant Gene Indices / Gene Index Project Database) [1].ET1.6, tolerant accession under stress library; ET0, tolerant accession control library [2].BT1.6, sensitive accession under stress library; BT0, sensitive accession control library. UR: up-regulated; DR: down-regulated; n.s.: not significant ($p < 0.05$); (−) unitag not expressed.
(XLS)

Table S3. ABC transporters differentially expressed in at least one treatment. Unitags associated to ABC transporters, their normalized frequencies, *p-value* [114], fold changes (FC_I: tolerant accession under stress *vs* tolerant accession control; FC_{II}: sensitive accession under stress *vs* sensitive accession control), their regulation (in different approaches), identification, appropriate annotated EST, reference data bank, gene acronym and score regarding each unitag-EST alignment. *GMGI: *Glycine max* Gene Index (Plant Gene Indices / Gene Index Project Database); GDB: PlantGDB database; MTGI: *Medicago truncatula* Gene Index (Plant Gene Indices / Gene Index Project Database) [1].ET1.6, tolerant accession under stress library; ET0, tolerant accession control library [2].BT1.6, sensitive accession under stress library; BT0, sensitive accession control library. UR: up-regulated; DR: down-regulated; n.s.: not significant ($p < 0.05$); (−) unitag not expressed.
(XLS)

Table S4. Data on the corresponding unitags presented in the heatmap regarding the GO category "Response to hormones" (Figure 6A). Unitags associated to the term "Response to hormones", their normalized frequencies, *p-value* [114], fold changes (FC_I: tolerant accession under stress *vs* tolerant accession control; FC_{II}: sensitive accession under stress *vs* sensitive accession control; FC_{III}: tolerant accession under stress *vs* sensitive accession under stress), their regulation (in different approaches), identification, appropriate annotated EST, reference data bank, gene acronym and score regarding each unitag-EST alignment. *GMGI: *Glycine max* Gene Index, MTGI: *Medicago truncatula* Gene Index, *Phaseolus coccineus* Gene Index (Plant Gene Indices / Gene Index Project Database); GDB: PlantGDB database; KEGG: Kyoto Encyclopedia of Genes and Genomes Database [1].ET1.6, tolerant accession under stress library; ET0, tolerant accession control library [2].BT1.6, sensitive accession under stress library; BT0, sensitive accession control library. UR: up-regulated; DR: down-regulated; n.s.: not significant ($p < 0.05$); (−) unitag not expressed.
(XLS)

Table S5. Data on the corresponding unitags presented in the heatmap regarding the GO category "Response to water" (Figure 6B). Unitags associated to the term "Response to water", their normalized frequencies, *p-value* [114], fold changes (FC_I: tolerant accession under stress *vs* tolerant accession control; FC_{II}: sensitive accession under stress *vs* sensitive accession control; FC_{III}: tolerant accession under stress *vs* sensitive accession under stress), their regulation (in different approaches), identification, appropriate annotated EST, reference data bank, gene acronym and score regarding each unitag-EST alignment. *GMGI: *Glycine max* Gene Index (Plant Gene Indices / Gene Index Project Database); GDB: PlantGDB database; KEGG: Kyoto Encyclopedia of Genes and Genomes Database [1].ET1.6, tolerant accession under stress library; ET0, tolerant accession control library [2].BT1.6, sensitive accession under stress library; BT0, sensitive accession control library. UR: up-regulated; DR: down-regulated; n.s.: not significant ($p < 0.05$); (−) unitag not expressed.
(XLS)

Table S6. Data on the corresponding unitags presented in the heatmap regarding the GO category "Response to salinity" (Figure 7A). Unitags associated to the term "Response to salinity", their normalized frequencies, *p-value* [114], fold changes (FC_I: tolerant accession under stress *vs* tolerant accession control; FC_{II}: sensitive accession under stress *vs* sensitive accession control; FC_{III}: tolerant accession under stress *vs* sensitive accession under stress), their regulation (in different approaches), identification, appropriate annotated EST, reference data bank, gene acronym and score regarding each unitag-EST alignment. *GMGI: *Glycine max* Gene Index (Plant Gene Indices / Gene Index Project Database); GDB: PlantGDB database; KEGG: Kyoto Encyclopedia of Genes and Genomes Database [1].ET1.6, tolerant accession under stress library; ET0, tolerant accession control library [2].BT1.6, sensitive accession under stress library; BT0, sensitive accession control library. UR: up-regulated; DR: down-regulated; n.s.: not significant ($p < 0.05$); (−) unitag not expressed.
(XLS)

Table S7. Data on the corresponding unitags presented in the heatmap regarding the GO category "Response to oxidative stress" (Figure 7B). Unitags associated to the term "Response to oxidative stress", their normalized frequencies, *p-value* [114], fold changes (FC_I: tolerant accession under stress *vs* tolerant accession control; FC_{II}: sensitive accession under stress *vs* sensitive accession control; FC_{III}: tolerant accession under stress *vs* sensitive accession under stress), their regulation (in different approaches), identification, appropriate annotated EST, reference data bank, gene acronym and score regarding each unitag-EST alignment. *GMGI: *Glycine max* Gene Index (Plant Gene Indices / Gene Index Project Database); MTGI: *Medicago truncatula* Gene Index (Plant Gene Indices / Gene Index Project Database); KEGG: Kyoto Encyclopedia of Genes and Genomes Database [1].ET1.6, tolerant accession under stress library; ET0, tolerant accession control library [2].BT1.6, sensitive accession under stress library; BT0, sensitive accession control library. UR: up-regulated; DR: down-regulated; n.s.: not significant ($p < 0.05$); (−) unitag not expressed.
(XLS)

Table S8. Target transcripts selected based on the DeepSuperSAGE differential expression for quantitative real-time amplification (RT-qPCR) including their gene acronym, primers sequences, amplicon length and primers efficiencies.
(XLS)

Acknowledgements

The authors thank Prof. Dr. Günter Kahl (University of Frankfurt, Germany), Dr. Peter Winter, Dr. Björn Rotter, Nicolas Krezdorn and Ruth Jungmann (GenXPro GmbH) for scientific and technical advices.

Author Contributions

Conceived and designed the experiments: EAK AMBI RVA ALN FCMG. Performed the experiments: JRCFN VP CR RLOS FAR EAK. Analyzed the data: JRCFN VP EAK AMBI RLOS CR FAR. Contributed reagents/materials/analysis tools: JRCFN EAK FCMG ALN RVA. Wrote the manuscript: JCRFN VP EAK AMBI.

References

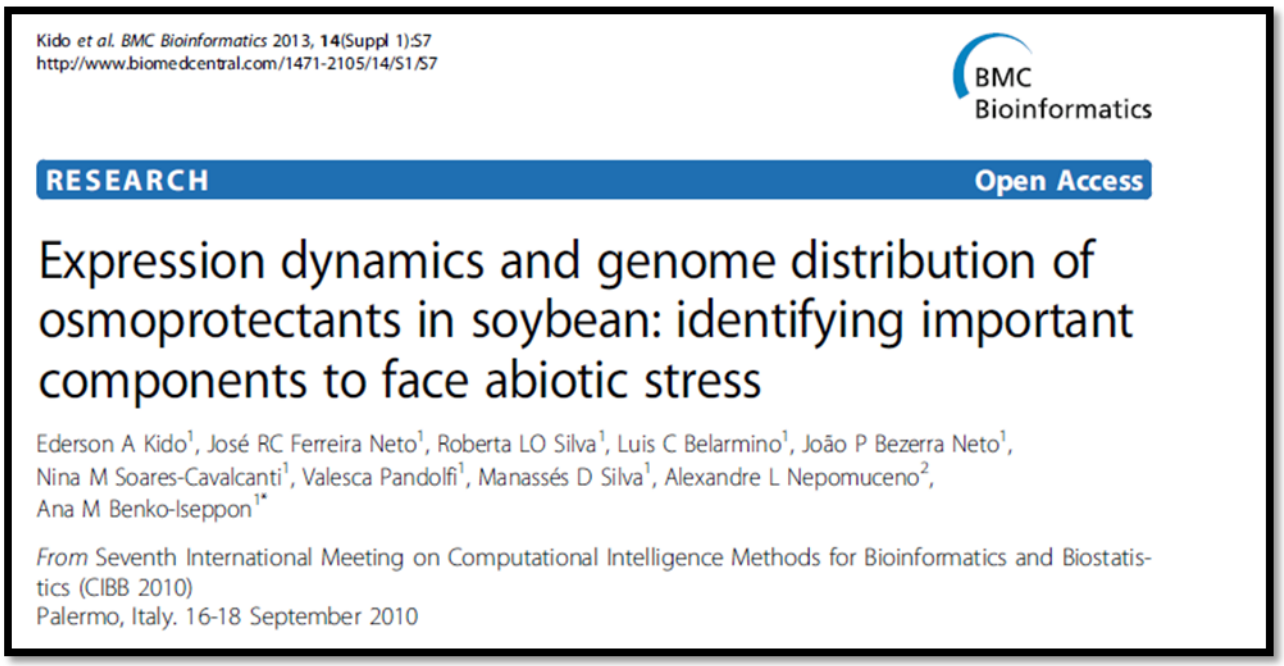
1. Wilson RF (2008) Soybean: market driven research needs. In: GV Stacey. Genetics and genomics of soybean. New York: Springer. pp. 3-15.
2. Stats Soy. Available: http://soystats.com/2012-soybean-highlights/. Accessed 09 October 2013
3. Valliyodan B, Nguyen H (2008) Genomics of abiotic stress in soybean. In: GV Stacey. Genetics and genomics of soybean. New York: Springer. pp. 343-372.
4. Specht JE, Hume DJ, Kumudini SV (1999) Soybean Yield Potential - A Genetic and Physiological. Perspectives - Crop Sci 39: 1560–1570.
5. Martins PK, Jordão BQ, Yamanaka N, Farias JRB, Beneventi MA, et al. (2008) Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill Fabales, Fabaceae) cultivar MG/BR46 (Conquista) under two water deficit induction systems. Genet Mol Biol 31: 512–521.
6. Zimmerman J (2012) Brazil: oilseeds and products annual. Available: http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Oilseeds%20and%20Products%20Annual_Brasilia_Brazil_3-30-2012.pdf. Accessed 15 October 2012
7. Manavalan LP, Guttikonda SK, Tran L-S, Nguyen HT (2009) Physiological and molecular approaches to improve drought resistance in soybean. Plant Cell Physiol 50: 1260-1276. doi:10.1093/pcp/pcp082. PubMed: 19546148.
8. Pathan MDS, Lee J-D, Shannon JG, Nguyen HT (2007) Recent advances in breeding for drought and salt stress tolerance in soybean. In: MA JenksPM HasegawaSM Jain. Advances in Molecular Breeding Toward Drought and Salt Tolerant Crops. New York: Springer. pp. 739-773.
9. UniGene Build Available #42. Available: http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=3847. Accessed 15 October 2012
10. Quackenbush J, Cho J, Lee D, Liang F, Holt I et al. (2001) The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res 29: 159–164. doi:10.1093/nar/29.1.159. PubMed: 11125077.
11. Vodkin L, Jones S, Gonzalez DO, Thibaud-Nissen F, Zabala G et al. (2008) Genomics of soybean seed development. In: GV Stacey. Genetics and genomics of soybean. New York: Springer. pp.163-184.
12. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57–63. doi:10.1038/nrg2484. PubMed: 19015660.
13. Le DT, Nishiyama R, Watanabe Y, Tanaka M, Seki M et al. (2012) Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. PLOS ONE 7: e49522. doi:10.1371/journal.pone.0049522. PubMed: 23189148.
14. Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ et al. (2010) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. Plant Physiol 152: 541–552. PubMed: 19933387.
15. Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ et al. (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. Plant J 63: 86–99. PubMed: 20408999.
16. Le DT, Nishiyama R, Watanabe Y, Mochida K, Yamaguchi-Shinozaki K et al. (2011) Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. DNA Res 18: 263–276. doi:10.1093/dnares/dsr015. PubMed: 21685489.
17. Li H, Dong Y, Yin H, Wang N, Yang J et al. (2011) Characterization of the stress associated microRNAs in *Glycine max* by deep sequencing.

BMC Plant Biol 11: 170. doi:10.1186/1471-2229-11-170. PubMed: 22112171.
18. Hao QN, Zhou XA, Sha AH, Wang C, Zhou R et al. (2011) Identification of genes associated with nitrogen-use efficiency by genome-wide transcriptional analysis of two soybean genotypes. BMC Genomics 12: 525-540. doi:10.1186/1471-2164-12-525. PubMed: 22029603.
19. Kido EA, Pandolfi V, Houllou-Kido LM, Andrade PP, Marcelino FC et al. (2010) Plant antimicrobial peptides: an overview of SuperSAGE transcriptional profile and a functional review. Curr Protein Pept Sci 11: 220-230. doi:10.2174/138920310791112110. PubMed: 20088771.
20. Fan X-D, Wang J-Q, Yang N, Dong Y-Y, Liu L et al. (2013) Gene expression profiling of soybean leaves and roots under salt, saline-alkali and drought stress by high-throughput Illumina sequencing. Gene 512: 392–402. doi:10.1016/j.gene.2012.09.100. PubMed: 23063936.
21. Severin AJ, Woody JL, Bolon Y-T, Joseph B, Diers BW et al. (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol 10: 1-16. doi:10.1186/1471-2229-10-1. PubMed: 20047654.
22. Hunt M, Kaur N, Stromvik M, Vodkin L (2011) Transcript profiling reveals expression differences in wild-type and glabrous soybean lines. BMC Plant Biol 11: 145-186. doi:10.1186/1471-2229-11-145. PubMed: 22029708.
23. Reid DE, Hayashi S, Lorenc M, Stiller J, Edwards D et al. (2012) Identification of systemic responses in soybean nodulation by xylem sap feeding and complete transcriptome sequencing reveal a novel component of the autoregulation pathway. Plant Biotechnol J 10: 680–689. doi:10.1111/j.1467-7652.2012.00706.x. PubMed: 22624681.
24. Peiffer GA, King KE, Severin AJ, May GD, Cianzio SR et al. (2012) Identification of candidate genes underlying an iron efficiency quantitative trait locus in soybean. Plant Physiol 158: 1745–1754. doi: 10.1104/pp.111.189860. PubMed: 22319075.
25. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178–183. doi:10.1038/nature08670. PubMed: 20075913.
26. Camon E, Barrell D, Brooksbank C, Magrane M, Apweiler R (2003) The Gene Ontology Annotation (GOA) Project--Application of GO in SWISS-PROT, TrEMBL and InterPro. Comp Funct Genomics 4: 71–74. doi:10.1002/cfg.235. PubMed: 18629103.
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H et al. (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25: 25-29. doi:10.1038/75556. PubMed: 10802651.
28. Toumi I, Gargouri M, Nouairi I, Moschou PN, Salem-Fnayou AB et al. (2008) Water stress induced changes in the leaf lipid composition of four grapevine genotypes with different drought tolerance. Biol Plant 52: 161–164. doi:10.1007/s10535-008-0035-2.
29. Bartels D, Souer E (2004) Molecular responses of higher plants to dehydration. In: H HirtK Shinozaki. Berlin. Heidelberg: Springer. pp. 9-38.
30. Vassilev A, Lidon F, Scotti P, Da Graca M, Yordanov I (2004) Cadmium-induced changes in chloroplast lipids and photosystem activities in barley plants. Biol Plant 48: 153–156. doi:10.1023/B:BIOP.0000024295.27419.89.
31. Pospíšilová J, Vágner M, Malbeck J, Ková ATČ, PBAŤ Ková (2005) Interactions between abscisic acid and cytokinins during water stress and subsequent rehydration. Biol Plant 49: 533–540. doi:10.1007/s10535-005-0047-0.
32. Koca H, Ozdemir F, Turkan I (2006) Effect of salt stress on lipid peroxidation and superoxide dismutase and peroxidase activities of *Lycopersicon esculentum* and L. pennellii. Biol Plant 50: 745–748. doi: 10.1007/s10535-006-0121-2.
33. Knight H, Brandt S, Knight MR (1998) A history of stress alters drought calcium signalling pathways in Arabidopsis. Plant J 16: 681–687. doi: 10.1046/j.1365-313x.1998.00332.x. PubMed: 10069075.
34. Luan S, Kudla J, Rodriguez-Concepcion M, Yalovsky S, Gruissem W (2002) Calmodulins and calcineurin B-like proteins: calcium sensors for specific signal response coupling in plants. Plant Cell 14: 389-400. PubMed: 12045290.
35. Yang T, Poovaiah BW (2003) Calcium/calmodulin-mediated signal network in plants. Trends Plant Sci 8: 505–512. doi:10.1016/j.tplants.2003.09.004. PubMed: 14557048.
36. Kim K-N (2012) Stress responses mediated by the CBL calcium sensors in plants. Plant Biotechnology Reports 7: 1–8.
37. Asano T, Hayashi N, Kikuchi S, Ohsugi R (2012) CDPK-mediated abiotic stress signaling. Plant Signal Behav 7: 817–821. doi:10.4161/psb.20351. PubMed: 22751324.
38. Pandey GK, Cheong H, Kim KN, Grant JJ, Li L et al. (2004) The calcium sensor calcineurin B-like 9 modulates abscisic acid sensitivity and biosynthesis in Arabidopsis. Plant Cell 16: 1912-1924. doi:10.1105/tpc.021311. PubMed: 15208400.
39. D'Angelo C, Weinl S, Batistic O, Pandey GK, Cheong YH et al. (2006) Alternative complex formation of the Ca-regulated protein kinase CIPK1 controls abscisic acid-dependent and independent stress responses in Arabidopsis. Plant J 48: 857–872. doi:10.1111/j.1365-313X.2006.02921.x. PubMed: 17092313.
40. Hanin M, Brini F, Ebel C, Toda Y, Takeda S et al. (2011) Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. Plant Signal Behav 6: 1503–1509. doi:10.4161/psb.6.10.17088. PubMed: 21897131.
41. Hara M, Fujinaga M, Kuboi T (2005) Metal binding by citrus dehydrin with histidine-rich domains. J Exp Bot 56: 2695–2703. doi:10.1093/jxb/eri262. PubMed: 16131509.
42. Kang J, Park J, Choi H, Burla B, Kretzschmar T et al. (2011) Plant ABC Transporters. The Arabidopsis Book.. Rockville: American Society of Plant Biologists. 25 pp.
43. Zhang J, Jia W, Yang J, Ismail AM (2006) Role of ABA in integrating plant responses to drought and salt stresses. Field Crops Res 97: 111–119. doi:10.1016/j.fcr.2005.08.018.
44. Narusaka Y, Narusaka M, Seki M, Umezawa T, Ishida J et al. (2004) Crosstalk in the responses to abiotic and biotic stresses in Arabidopsis: analysis of gene expression in cytochrome P450 gene superfamily by cDNA microarray. Plant Mol Biol 55: 327–342. doi:10.1007/s11103-004-0685-1. PubMed: 15604685.
45. Cheong YH, Chang HS, Gupta R, Wang X, Zhu T et al. (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. Plant Physiol 129: 661-677. doi:10.1104/pp.002857. PubMed: 12068110.
46. Apel K, Hirt H (2004) Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu Rev Plant Biol 55: 373–399. doi: 10.1146/annurev.arplant.55.031903.141701. PubMed: 15377225.
47. Torres MA, Dangl JL (2005) Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. Curr Opin Plant Biol 8: 397–403. doi:10.1016/j.pbi.2005.05.014. PubMed: 15939662.
48. Hay A, Craft J, Tsiantis M (2004) Plant hormones and homeoboxes: bridging the gap? BioEssays 26: 395–404. doi:10.1002/bies.20016. PubMed: 15057937.
49. Hassanein RA, Hassanein AA, El-din AB, Salama M, Hashem HA (2009) Role of jasmonic acid and abscisic acid treatments in alleviating the adverse effects of drought stress and regulating trypsin inhibitor production in soybean plant. Aust J Basic Appl Sci 3: 904-919.
50. Rock CD, Sakata Y, Quatrano RS (2010) Stress signaling I: The role of abscisic acid (ABA). In: A PareekSK SoporivyHJ BohnertGovindjee. Abiotic Stress Adaptation Plants. Dordrecht: Springer. pp. 33–73.
51. Bell E, Mullet JE (1991) Lipoxygenase gene expression is modulated in plants by water deficit, wounding, and methyl jasmonate. Mol Gen Genet 230: 456–462. PubMed: 1766441.
52. Giovannoni JJ (2004) Genetic regulation of fruit development and ripening. Plant Cell 16: 170–181. doi:10.1105/tpc.019158.
53. Ohme-takagi M, Shinshi H (1995) Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell 7: 173–182. doi:10.1105/tpc.7.2.173. PubMed: 7756828.
54. Wu L, Zhang Z, Zhang H, Wang X-C, Huang R (2008) Transcriptional modulation of ethylene response factor protein JERF3 in the oxidative stress response enhances tolerance of tobacco seedlings to salt, drought, and freezing. Plant Physiol 148: 1953–1963. doi:10.1104/pp.108.126813. PubMed: 18945933.
55. Quan R, Hu S, Zhang Z, Zhang H, Zhang Z et al. (2010) Overexpression of an ERF transcription factor TSRF1 improves rice drought tolerance. Plant Biotechnol J 8: 476–488. doi:10.1111/j.1467-7652.2009.00492.x. PubMed: 20233336.
56. El-Sharkawy I, Sherif S, Mila I, Bouzayen M, Jayasankar S (2009) Molecular characterization of seven genes encoding ethylene-responsive transcriptional factors during plum fruit development and ripening. J Exp Bot 60: 907–922. doi:10.1093/jxb/ern354. PubMed: 19213809.
57. Campbell P, Braam J (1999) Xyloglucan endotransglycosylases: Diversity of genes, enzymes and potential wall-modifying functions. Trends Plant Sci 4: 361–366. doi:10.1016/S1360-1385(99)01468-5. PubMed: 10462769.
58. Amaya I, Botella MA, de la Calle M, Medina MI, Heredia A et al. (1999) Improved germination under osmotic stress of tobacco plants overexpressing a cell wall peroxidase. FEBS Lett 457: 80-84. doi: 10.1016/S0014-5793(99)01011-X. PubMed: 10486568.
59. Smalle J, Vierstra RD (2004) The ubiquitin 26S proteasome proteolytic pathway. Annu Rev Plant Biol 55: 555-590. doi:10.1146/annurev.arplant.55.031903.141801. PubMed: 15377232.
60. Cho SK, Ryu MY, Song C, Kwak JM, Kim WT (2008) Arabidopsis PUB22 and PUB23 are homologous U-box E3 ubiquitin ligases that

play combinatory roles in response to drought stress. Plant Cell 20: 1899-1914. doi:10.1105/tpc.108.060699. PubMed: 18664614.
61. Dreier W, Schnarrenberger C, Börner T (1995) Light- and stress-dependent enhancement of amylolytic activities in white and green barley leaves: β-Amylases are Stress-Induced. Proteins - J Plant Physiol 145: 342-348. doi:10.1016/S0176-1617(11)81900-0.
62. Datta R, Selvi MT, Seetharama N, Sharma R (1999) Stress mediated enhancement of beta-amylase activity in pearl millet and maize leaves is dependent on light. J Plant Physiol 154: 657-664. doi:10.1016/S0176-1617(99)80241-7.
63. Todaka D, Matsushima H, Morohashi Y (2000) Water stress enhances beta-amylase activity in cucumber cotyledons. J Exp Bot 51: 739-745. doi:10.1093/jexbot/51.345.739. PubMed: 10938866.
64. Yang J, Zhang J, Wang Z, Zhu Q (2001) Activities of starch hydrolytic enzymes and sucrose-phosphate synthase in the stems of rice subjected to water stress during grain filling. J Exp Bot 52: 2169–2179. PubMed: 11604456.
65. Mohammadkhani N, Heidari R (2008) Drought-induced accumulation of soluble sugars and proline in two maize varieties. World App Sci J 3: 448-453
66. Bartels D, Sunkar R (2005) Drought and Salt Tolerance in Plants. CRC. Crit Rev Plant Sci 24: 23-58. doi:10.1080/07352680590910410.
67. Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: An overview. Arch Biochem Biophys 444: 139–158. doi:10.1016/j.abb.2005.10.018. PubMed: 16309626.
68. Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K et al. (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol 133: 1755–1767. doi:10.1104/pp.103.025742. PubMed: 14645724.
69. Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends Plant Sci 5: 199–206. doi:10.1016/S1360-1385(00)01600-9. PubMed: 10785665.
70. Dong J, Chen C, Chen Z (2003) Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. Plant Mol Biol 51: 21-37. PubMed: 12602888.
71. Rizhsky L, Liang H, Mittler R (2002) The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiol 130: 1143-1151. doi:10.1104/pp.006858. PubMed: 12427981.
72. Huang T, Duman JG (2002) Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade, *Solanum dulcamara*. Plant Mol Biol 48: 339-350. doi:10.1023/A:1014062714786. PubMed: 11905961.
73. Cominelli E, Tonelli C (2009) A new role for plant R2R3-MYB transcription factors in cell cycle regulation. Cell Res 19: 1231–1232. doi:10.1038/cr.2009.123. PubMed: 19881525.
74. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C et al. (2010) MYB transcription factors in Arabidopsis. Trends Plant Sci 15: 573-581. doi:10.1016/j.tplants.2010.06.005. PubMed: 20674465.
75. Seo PJ, Xiang F, Qiao M, Park J-Y, Lee NL (2009) The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in Arabidopsis. Plant Physiol 151: 275–289. doi:10.1104/pp.109.144220. PubMed: 19625633.
76. Lai LB, Nadeau JA, Lucas J, Lee E-K, Nakagawa T et al. (2005) The Arabidopsis R2R3 MYB proteins FOUR LIPS and MYB88 restrict divisions late in the stomatal cell lineage. Plant Cell 17: 2754-2767. doi:10.1105/tpc.105.034116. PubMed: 16155180.
77. Brosché M, Overmyer K, Wrzaczek M, Kangasjärvi J (2010) Stress Signaling III: Reactive Oxygen Species (ROS). In: A PareekSK SoporyHJ BohnertGovindjee. Abiotic stress adaptation in plants: physiological, molecular and genomic foundation. Dordrecht: Springer. pp. 91-102.
78. Choudhury S, Panda P, Sahoo L, Panda SK (2013) Reactive oxygen species signaling in plants under abiotic stress. Plant Signal Behav 8: e23681. doi:10.4161/psb.23681. PubMed: 23425848.
79. Sekmen AH, Türkan I, Takio S (2007) Differential responses of antioxidative enzymes and lipid peroxidation to salt stress in salt-tolerant *Plantago maritime* and salt-sensitive Plantago media. Physiol Plant 131: 399–411. doi:10.1111/j.1399-3054.2007.00970.x. PubMed: 18251879.
80. Zhang J, Kirkhan MB (1994) Drought-stress-induced changes in activities of superoxide dismutase, catalase, and peroxidase in wheat species. Plant Cell Physiol 35: 785-791.
81. Liao Z-H, Chen M, Gong Y-F, Miao Z-Q, Sun X-F et al. (2006) Isoprenoid Biosynthesis in Plants: Pathways, Genes, Regulation and Metabolic. Engineering - J Biol Sci 6: 209-219.
82. Kirby J, Keasling JD (2009) Biosynthesis of plant isoprenoids: perspectives for microbial engineering. Annu Rev Plant Biol 60: 335–355. doi:10.1146/annurev.arplant.043008.091955. PubMed: 19575586.
83. Leivar P, Antolín-Llovera M, Ferrero S, Closa M, Arró M et al. (2011) Multilevel control of Arabidopsis 3-hydroxy-3-methylglutaryl coenzyme A reductase by protein phosphatase 2A. Plant Cell 23: 1494-1511. doi:10.1105/tpc.110.074278. PubMed: 21478440.
84. Stermer BA, Bianchini GM, Korth KL (1994) Regulation of HMG-CoA reductase activity in plants. J Lipid Res 35: 1133–1140. PubMed: 7964176.
85. Yang D, Ma P, Liang X, Wei Z, Liang Z et al. (2012) PEG and ABA trigger methyl jasmonate accumulation to induce the MEP pathway and increase tanshinone production in *Salvia miltiorrhiza* hairy roots. Physiol Plant 146: 173-183. doi:10.1111/j.1399-3054.2012.01603.x. PubMed: 22356467.
86. Fry SC, Smith RC, Renwick KF, Martin DJ, Hodge SK et al. (1992) Xyloglucan endotransglycosylase, a new wallloosening enzyme activity from plants. Biochemical J 282: 821–828.
87. Nishitani K, Tominaga R (1992) Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyzes transfer of a segment of xyloglucan molecule to another xyloglucan molecule. J Biol Chem 267: 21058-21064. PubMed: 1400418.
88. Antosiewicz DM, Purugganan MM, Polisensky DH, Braam J (1997) Cellular localization of Arabidopsis xyloglucan endotransglycosylase-related proteins during development and after wind stimulation. Plant Physiol 115: 1319–1328. doi:10.1104/pp.115.4.1319. PubMed: 9414546.
89. Vissenberg K, Martinez-Vilchez IM, Verbelen J-P, Miller JG, Fry SC (2000) *In vivo* colocalization of xyloglucan endotransglycosylase activity and its donor substrate in the elongation zone of Arabidopsis roots. Plant Cell 12: 1229–1238. doi:10.2307/3871267. PubMed: 10899986.
90. Wu YJ, Cosgrove DJ (2000) Adaptation of roots to low water potentials by changes in cell wall extensibility and cell wall proteins. J Exp Bot 51: 1543–1553. doi:10.1093/jexbot/51.350.1543. PubMed: 11006305.
91. Shin YK, Yum H, Kim ES, Cho H, Gothandam KM, et al. (2006) BcXTH1, a Brassica campestris homologue of Arabidopsis XTH9 is associated with cell expansion. Planta 224: 32–41.
92. Riaño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B (2007) PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics 8: 42-52. doi:10.1186/1471-2105-8-42. PubMed: 17286856.
93. Rushton PJ, Somssich IE, Ringler P, Shen QJ (2010) WRKY transcription factors. Trends Plant Sci 15: 247-258. doi:10.1016/j.tplants.2010.02.006. PubMed: 20304701.
94. Wu X, Shiroto Y, Kishitani S, Ito Y, Toriyama K (2009) Enhanced heat and drought tolerance in transgenic rice seedlings overexpressing OsWRKY11 under the control of HSP101 promoter. Plant Cell Rep 28: 21-30. doi:10.1007/s00299-008-0614-x. PubMed: 18818929.
95. Luo X, Bai X, Sun X, Zhu D, Liu B et al. (2013) Expression of wild soybean WRKY20 in Arabidopsis enhances drought tolerance and regulates ABA signalling. J Exp Bot 64: 2155-2169. doi:10.1093/jxb/ert073. PubMed: 23606412.
96. Chen L, Song Y, Li S, Zhang L, Zou C, Yu D (2012) The role of WRKY transcription factors in plant abiotic stresses. Biochim Biophys Acta 1819: 120-128. doi:10.1016/j.bbagrm.2011.09.002. PubMed: 21964328.
97. Olsen AN, Ernst HA, Leggio LL, Skriver K (2005) NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci 10: 79-87. doi:10.1016/j.tplants.2004.12.010. PubMed: 15708345.
98. Liu X, Liu S, Wu J, Zhang B, Li X et al. (2013) Overexpression of *Arachis hypogaea NAC3* in tobacco enhances dehydration and drought tolerance by increasing superoxide scavenging. Plant Physiol Biochem 70: 354-359. doi:10.1016/j.plaphy.2013.05.018.
99. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2012) AP2/ERF family transcription factors in plant abiotic stress responses. Biochim Biophys Acta 1819: 86–96. doi:10.1016/j.bbagrm.2011.08.004. PubMed: 21867785.
100. Dietz K-J, Vogel MO, Viehhauser A (2010) AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signaling. Protoplasma 245: 3–14. doi:10.1007/s00709-010-0142-8. PubMed: 20411284.
101. Oñate-Sánchez L, Anderson JP, Young J, Singh KB (2007) AtERF14, a member of the ERF family of transcription factors, plays a nonredundant role in plant defense. Plant Physiol 143: 400–409. PubMed: 17114278.
102. Lin RC, Park HJ, Wang HY (2008) Role of Arabidopsis RAP2-4 in regulating light- and ethylene-mediated developmental processes and drought stress tolerance. Mol Plant 1: 42–57. doi:10.1093/mp/ssm004. PubMed: 20031913.
103. Rae L, Lao NT, Kavanagh TA (2011) Regulation of multiple aquaporin genes in Arabidopsis by a pair of recently duplicated DREB

transcription factors. Planta 234: 429–444. doi:10.1007/s00425-011-1414-z. PubMed: 21509693.
104. Chen M, Wang Q-Y, Cheng X-G, Xu Z-S, Li L-C et al. (2007) GmDREB2, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. Biochem Biophys Res Commun 353: 299-305. doi:10.1016/j.bbrc.2006.12.027. PubMed: 17178106.
105. Zhang G, Chen M, Li L, Xu Z, Chen X et al. (2009) Overexpression of the soybean GmERF3 gene, an AP2/ERF type transcription factor for increased tolerances to salt, drought, and diseases in transgenic tobacco. J Exp Bot 60: 3781–3796. doi:10.1093/jxb/erp214. PubMed: 19602544.
106. Ocampo ETM, Robles RP (2000) Drought tolerance in mungbean. I. Osmotic adjustment in stressed mungbean. Philipp J Crop Sci 25: 1-5.
107. Hsiao TC (1973) Plant response to water stress. Annual Rev Plant Physiol 24: 519-570. doi:10.1146/annurev.pp.24.060173.002511.
108. Yoshimura K, Miyao K, Gaber A, Takeda T, Kanaboshi H et al. (2004) Enhancement of stress tolerance in transgenic tobacco plants overexpressing Chlamydomonas glutathione peroxidase in chloroplasts or cytosol. Plant J 37: 21-33. doi:10.1046/j.1365-313X.2003.01930.x. PubMed: 14675429.
109. Oya T, Nepomuceno AL, Neumaier N, Farias JRB, Tobita S et al. (2004) Drought tolerance characteristics of Brazilian soybean cultivars - evaluation and characterization of drought tolerance of various Brazilian soybean cultivars in the field. Plant Prod Sci 7: 129–137. doi: 10.1626/pps.7.129.
110. Kulcheski FR, Oliveira LF, Molina LG, Almerão MP, Rodrigues FA et al. (2011) Identification of novel soybean microRNAs involved in abiotic and biotic stresses. BMC Genomics 12: 307-324. PubMed: 21663675.
111. Fehr WR, Caviness CE (1977) Stages of soybean development. Ames: Iowa State University of Science and Technology. 11p.
112. Matsumura H, Molina C, Krüger DH, Terauchi R, Kahl G (2012) DeepSuperSAGE: High-throughput transcriptome sequencing with now- and next-Generation sequencing technologies, in tag-based next generation sequencing. In: M HarbersG Kahl. Tag-Based Next Generation Sequencing. Germany: Wiley-VCH Verlag GmbH & Co.. pp. 3-21.
113. Nascimento LC, Costa GGL, Binneck E, Pereira GAG, Carazzolle MF (2012) A web-based bioinformatics interface applied to the GENOSOJA Project: Databases and pipelines. Genet Mol Biol 35 (1 Suppl): 203–211. doi:10.1590/S1415-47572012000200002. PubMed: 22802706.
114. Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Res7: 986–995. PubMed: 9331369.
115. Robertson N, Oveisi-Fordorei M, Zuyderduyn SD, Varhol RJ, Fjell C et al. (2007) DiscoverySpace: an interactive data analysis application. Genome Biol 8: 1: R6. doi:10.1186/gb-2007-8-1-r6. PubMed: 17210078.
116. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410. doi:10.1016/S0022-2836(05)80360-2. PubMed: 2231712.
117. National Center for Biological Information. Available: http://www.ncbi.nlm.nih.gov/dbEST/. Accessed 15 October 2012
118. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28: 27-30. doi:10.1093/nar/28.7.e27. PubMed: 10592173.
119. Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD et al. (2008) PlantGDB: a resource for comparative plant genomics. Nucleic Acids Res 36: D959-D965. PubMed: 18063570.
120. Close TJ, Wanamaker S, Roose ML, Lyon M (2008) HarvEST: An EST Database and Viewing Software. In: D Edwards. Plant Bioinformatics: Methods and Protocols. New Jersey: Humana Press. pp. 161-177.
121. UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res 38: D142-D148. doi:10.1093/nar/gkq368. PubMed: 19843607.
122. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M et al. (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674-3676. doi: 10.1093/bioinformatics/bti610. PubMed: 16081474.
123. Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol 131: 900–910. doi:10.1104/pp.102.018150. PubMed: 12644643.
124. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863-14868. doi:10.1073/pnas.95.25.14863. PubMed: 9843981.
125. Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12: 357-358. PubMed: 8902363.
126. Oliveros JC (2007) VENNY: An interactive tool for comparing lists with Venn diagrams. Available: http://bioinfogp.cnb.csic.es/tools/venny/index.html. Accessed 15 October 2012
127. Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30: e36. doi:10.1093/nar/30.9.e36. PubMed: 11972351.
128. Stolf-Moreira R, Lemos EGM, Abdelnoor RV, Beneventi MA, Rolla AAP et al. (2011) Identification of reference genes for expression analysis by real-time quantitative PCR in drought-stressed soybean. Pesq Agropec Bras 46: 58-65. doi:10.1590/S0100-204X2011000100008.

Capítulo II



Kido *et al. BMC Bioinformatics* 2013, **14**(Suppl 1):S7
http://www.biomedcentral.com/1471-2105/14/S1/S7

BMC Bioinformatics

RESEARCH Open Access

Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress

Ederson A Kido[1], José RC Ferreira Neto[1], Roberta LO Silva[1], Luis C Belarmino[1], João P Bezerra Neto[1], Nina M Soares-Cavalcanti[1], Valesca Pandolfi[1], Manassés D Silva[1], Alexandre L Nepomuceno[2], Ana M Benko-Iseppon[1*]

From Seventh International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB 2010)
Palermo, Italy. 16-18 September 2010

Classificação WebQualis CAPES: B1

Fator de Impacto: 3,02

Kido et al. BMC Bioinformatics 2013, 14(Suppl 1):S7
http://www.biomedcentral.com/1471-2105/14/S1/S7

BMC Bioinformatics

RESEARCH Open Access

Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress

Ederson A Kido[1], José RC Ferreira Neto[1], Roberta LO Silva[1], Luis C Belarmino[1], João P Bezerra Neto[1], Nina M Soares-Cavalcanti[1], Valesca Pandolfi[1], Manassés D Silva[1], Alexandre L Nepomuceno[2], Ana M Benko-Iseppon[1*]

From Seventh International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB 2010)
Palermo, Italy. 16-18 September 2010

Abstract

Background: Despite the importance of osmoprotectants, no previous *in silico* evaluation of high throughput data is available for higher plants. The present approach aimed at the identification and annotation of osmoprotectant-related sequences applied to short transcripts from a soybean HT-SuperSAGE (High Throughput Super Serial Analysis of Gene Expression; 26-bp tags) database, and also its comparison with other transcriptomic and genomic data available from different sources.

Methods: A curated set of osmoprotectants related sequences was generated using text mining and selected seed sequences for identification of the respective transcripts and proteins in higher plants. To test the efficiency of the seed sequences, these were aligned against four HT-SuperSAGE contrasting libraries generated by our group using soybean tolerant and sensible plants against water deficit, considering only differentially expressed transcripts ($p \leq 0.05$). Identified transcripts from soybean and their respective tags were aligned and anchored against the soybean virtual genome.

Results: The workflow applied resulted in a set including 1,996 seed sequences that allowed the identification of 36 differentially expressed genes related to the biosynthesis of osmoprotectants [Proline (*P5CS*: 4, *P5CR*: 2), Trehalose (*TPS1*: 9, *TPPB*: 1), Glycine betaine (*BADH*: 4) and *Myo*-inositol (*MIPS*: 7, *INPS1*: 8)], also mapped *in silico* in the soybean genome (25 loci). Another approach considered matches using Arabidopsis full length sequences as seed sequences, and allowed the identification of 124 osmoprotectant-related sequences, matching ~10.500 tags anchored in the soybean virtual chromosomes. Osmoprotectant-related genes appeared clustered in all soybean chromosomes, with higher density in some subterminal regions and synteny among some chromosome pairs.

Conclusions: Soybean presents all searched osmoprotectant categories with some important members differentially expressed among the comparisons considered (drought tolerant or sensible *vs*. control; tolerant *vs*. sensible), allowing the identification of interesting candidates for biotechnological inferences. The identified tags aligned to corresponding genes that matched 19 soybean chromosomes. Osmoprotectant-related genes are not regularly distributed in the soybean genome, but clustered in some regions near the chromosome terminals, with some redundant clusters in different chromosomes indicating their involvement in previous duplication and rearrangements events. The seed sequences, transcripts and map represent the first transversal evaluation for osmoprotectant-related genes and may be easily applied to other plants of interest.

* Correspondence: ana.iseppon@gmail.com
[1]Departament of Genetics/Biological Sciences Center, Federal University of Pernambuco, Recife, Pernambuco, CEP 50.670-420, Brazil
Full list of author information is available at the end of the article

Background

Osmoprotectants figure among the most fundamental solutes in living organisms, being present from bacteria and fungi to higher plants and animals [1]. Main plant osmoprotectants are chemically composed by amino acids or carbohydrates, but share common features as low molecular weight and nontoxic character even at high concentrations, playing vital roles during abiotic stresses in plants as salinity, drought and chilling [2].

To face such constraints many plants accumulate organic osmolytes, or compatible solutes, in response to the resulting osmotic stress, maintaining cell turgor and therefore the driving gradient for water uptake. They include sugars, mainly fructose and sucrose, sugar alcohols (like *Myo*-inositol), complex sugars (like trehalose and fructans) and charged metabolites (like glycinebetaine, proline and ectoine) [1,3].

Osmolytes can also act as free-radical scavengers or chemical chaperones by directly stabilizing membranes and/or proteins [4]. Moreover, the accumulation of compatible solutes may also protect plants against damage by scavenging of reactive oxygen species, and by their chaperone-like activities in maintaining protein structures and functions [5]. Plant cells defend against stresses by modulating their expression according to the type and severity of stress and developmental stage of the plant [6].

Most previous works focused on expression assays regarding a single osmoprotectant as in Chen *et al.* [7] or searches in EST databases as in Barros *et al.* [8] or even their expression evaluation in transgenic plants [9,10]. No previous appreciation regarding in deep evaluation of transcriptomics databases generated under stress with Next Generation Sequencing (NGS) was carried out up to date. In the present work an 'in silico' annotation workflow was carried out integrating high throughput transcriptomics in soybean (*Glycine max*) plants under water deficit and biotic stress using HT-SuperSAGE, as compared with traditional transcriptomics and genome distribution of plant osmoprotectants.

The present approach focused on seven genes related to the biosynthesis of four classes of the most important plant osmoprotectants: Proline (genes *P5CS* and *P5CR*), Trehalose (*TPS1* and *TPPB*), Glycine betaine (*BADH* and *CMO*) and *Myo*-inositol (*INPS1*).

Proline - Comprises a proteinogenic amino acid, essential for primary metabolism in plants during drought and salt stresses, presenting a molecular chaperone role due to its stabilizing action either as a buffer to maintain the pH of the cytosolic redox status of the cell [11] or as antioxidant through its involvement in the scavenging of free highly reactive radicals [12] or still acting as a singlet oxygen quencher [13]. In higher plants, proline biosynthesis may proceed either via glutamate, by successive reductions catalyzed by Delta(1)-pyrroline-5-carboxylate synthase (P5CS) and Delta(1)-pyrroline-5-carboxylate reductase (P5CR) or by ornithine pathway, by ornithine d-aminotransferase (OAT), representing generally the first activated osmoprotectant after stress perception [14,15].

Trehalose - In plants this sugar participates mainly in the response to dehydration being first described in the so called resurrection plants *Myrothamnus flabellifolius* [16] and *Selaginella tamariscina* [17] both able to recover after almost complete dehydration. Such ability to act in the stabilization of proteins and membranes [18], as well as its role in ROS scavenging process [19] are the possible features of its cellular function during non-ideal conditions encountered by plants, where it's synthesis normally occurs by the formation of the trehalose-6-phosphate (T6P) from the UDP-glucose and glucose-6-phosphate, a reaction catalyzed by the trehalose 6-phosphate synthase (TPS). Afterwards the T6P is dephosphorylated by the trehalose-6-phosphate phosphatase (TPP) resulting in the formation of free trehalose [20]. A transgenic assay using *Agrobacterium*-mediated gene transfer allowed the insertion of the gene *TPS1* from yeast to tomato plants and resulted in higher content of chlorophyll and starch, besides pronounced tolerance to drought, salinity and oxidative stress, despite some pleiotropic changes [21].

Glycine betaine (GB) - Regards a quaternary ammonium compound (QAC) occurring in plants, animals and microorganisms. According to Chen and Murata [22] GB accumulates in chloroplasts and plastids especially in halotolerant plants, but also in other plants under high salinity, drought and cold stresses [23], with a recognized role associated to antioxidative responses [24]. In most organisms GB is synthesized either by the oxidation (or dehydrogenation) of choline or by the N-methylation of glycine. However, the pathway from choline to GB has been the main GB-accumulation pathway in plant species [25]. In this pathway choline is converted to betaine aldehyde by choline monooxygenase (CMO) [26], which is then converted to GB by betaine aldehyde dehydrogenase (BADH) [27].

***Myo*-inositol** - This osmoprotectant is an important cellular component forming the basis of a significant number of lipid signaling molecules involved in diverse pathways, including stress responses. *Myo*-inositol is the most abundant stereoisomer among the nine existing in nature, composed by a cyclohexanehexol, which is a cyclic carbohydrate with six hydroxyl groups, one on each carbon ring [28], acting as substrate in the biosynthesis of many compounds, especially the raffinose family oligosaccharides (RFOs) [29] that accumulate in plants under stress conditions [30]. In multicellular eukaryotes,

Myo-inositol becomes incorporated into phosphatidylinositol phosphate (PtdInsP), *Myo*-inositol phosphate (InsP), and certain sphingolipid signalling molecules that act in diverse processes, including regulation of gene expression [31]. It is synthesized by a two-step pathway, including: (1) conversion of D-glucose-6-P to D-*Myo*-inositol (1)-Monophosphate, 1D-MI-1-P, which is catalyzed by a L-*Myo*-inositol 1-phosphate synthase (MIPS) [32], and (2) specific dephosphorylation to free *Myo*-inositol by the Mg^{++} dependent L-*Myo*-inositol 1-phosphate phosphatase (IMP) [33]).

Considering the potential of these molecules for plant biotechnological approaches, the present work generated a curated list of osmoprotectants, osmoprotectant-related sequences and important regulatory elements, indicating most adequate tools for their identification and annotation. To evaluate the sensitivity of the proposed approach, the generated seed sequences and the proposed workflow were used to search of osmoprotectant-related sequences in short sequences (26 bp) generated from HT-SuperSAGE [34] deposited in the GENOSOJA (Brazilian Soybean Genome Consortium) data Bank [35]. A significant number of tags matched to known osmoprotectant-related sequences showing the effectiveness of the present approach useful for searches in other (actually very abundant) databanks comprising second generation sequences associated to the high performance sequencing approaches [e.g. Pyrosequencer (454 Roche®), Solexa (Illumina®) and SOLiD (Applied Biosystems®)] regarding genomic and transcriptomic libraries.

The present work also represents the first overall evaluation of the osmoprotectants in a higher plant comparing the prevalence of genes encoding enzymes of osmoprotectants biosynthetic pathways in sequence databanks with different backgrounds considering tissues, stages, stress conditions and also molecular approaches used to generate transcripts (ESTs, subtractive, cDNA full length, HT-SuperSAGE, BACs, etc.). In this aspect soybean offers one of the most abundant data sources for such an evaluation in legumes (see Benko-Iseppon *et al.* [36]), due to its importance as a source of food and oil in our planet.

Results and discussion

Seed sequences and annotation routine

The strategy regarding the use of seed sequences to find relevant literature and posterior mining and curation (Figure 1) was very effective, allowing the identification of 1,996 seed-sequences (Additional file 1) related to the procured osmoprotectants (proline, trehalose, *Myo*-inositol and glycine betaine). The sequences were aligned (BLASTx, cutoff e^{-10}) against the soybean peptide database at Phytozome v8.0 [37], also permitting the identification of the respective transcripts from soybean transcriptome used to associate with the transcripts from HT-SuperSAGE libraries.

Selection of HT-SuperSAGE tags for expression evaluation

After exclusion of the singlets, 2,551,286 tags from four libraries were selected for further evaluation, concerning 120,770 unitags. Considering the contrasts between any pair of libraries compared the number of unitags per library ranged from 73,807 to 89,205. The numbers of differentially expressed tags [upregulated (UR) and downregulated (DR) at the level of $p \leq 0.05$] for each compared pair of libraries are presented in the Table 1. In all analyzed contrasts, the number of tags differentially expressed overruled the observed amount in the comparison among the controls. The same situation was observed in relation to the UR and DR tags, highlighting the effect of the stress application in the gene activation in both accessions, indicating a richness of analyzable transcripts in the present approach.

Identification of osmoprotectant-related genes and differential expression in soybean

The carried approach was very successful, allowing the identification of 36 differentially expressed HT-SuperSAGE tags associated to 65 osmoprotectant-related sequences anchored in 25 loci (Glyma sequences; Additional file 2) based on the generated seed sequence bank (Additional file 1). Many of them regard interesting candidates for a posterior in deep evaluation, as further discussed.

Betaine aldehyde dehydrogenase (BADH, EC 1.2.1.8)

A total of 77 osmoprotectants OSMTL sequences presented significant similarity (BLASTx, e-value cut-off e^{-10}) to *G. max* sequences from four loci coding BADHs and annotated as Aldehyde Dehydrogenase Family 10A (Additional file 1). Out of four loci, one (Glyma06g19820) was associated with four HT-SuperSAGE tags in BLASTn alignments, tolerating at most a single mismatch (Additional file 2). From these, two (GmDr_44 and 3640) were induced after stress in both tolerant and sensible accessions (Embrapa 48 and BR-16), being mapped in the 3'UTRs of all three alternative transcripts of the locus Glyma06g19820 (Additional file 2). Other two tags (GmDr_2643 and 55655) were induced only in the drought sensible accession BR-16 after stress regarding the same three transcripts, whereas one of them (GmDr_2443) was mapped in the 3'UTR and another in the CDS (GmDr_55655) (Additional file 2). The tag GmDr_55655 also mapped in the transcript Glyma11g27100.1 with a mismatch in the CDS region, but no 3'UTR was identified for this transcript (Additional file 2). Despite its induction in the sensible accession in relation to the control, the

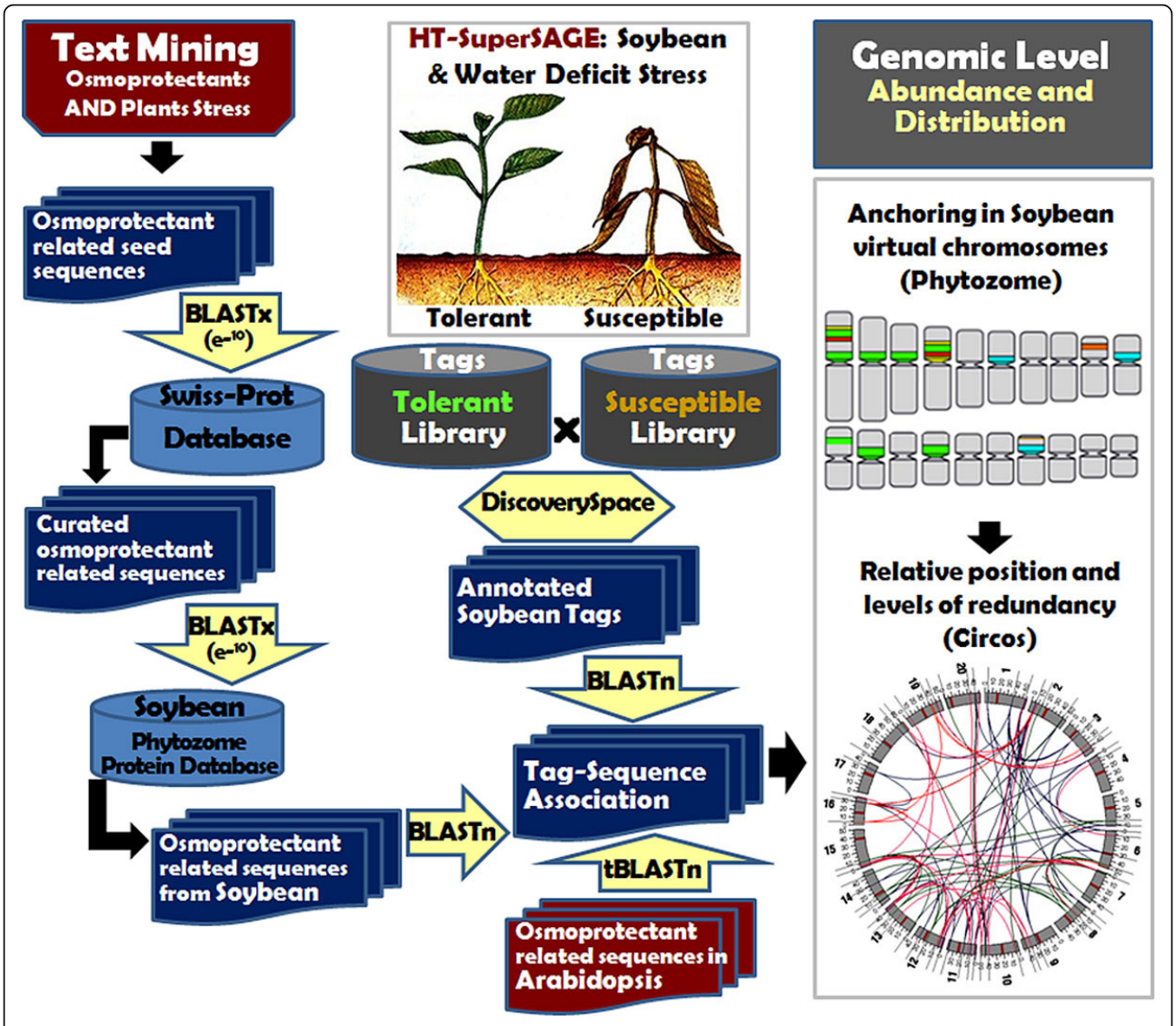


Figure 1 Schematic representation of the *in silico* steps for the generation of the local bank of seed sequences, their annotation, SuperSAGE tag-gene association, expression evaluation and anchoring of osmoprotectant candidates and their matched tags in the soybean virtual chromosomes.

normalized frequency was only six tmp (tags per million; Additional file 2). Thus, the locus Glyma06g19820 emerged as a likely BADH candidate gene induced in response to the water deficit stress in the studied libraries.

The members of the ALDH (aldehyde dehydrogenase) gene superfamily here identified in soybean genome were also categorized by Kotchoni *et al.* [38] that provided a unified nomenclature for the soybean ALDH

Table 1 Tags from soybean drought tolerant and sensible accessions, considering contrasting libraries and their expression profile.

Contrast	TS *vs.* TC	SS *vs.* SC	TS *vs.* SS	TC *vs.* SC	Total
UR	13,532	10,751	12,347	6,468	43,098
DR	7,423	5,587	7,634	3,135	23,779
n.s.	53,878	72,867	53,826	73,067	253,638
Unitags	74,833	89,205	73,807	82,670	320,515

UR: upregulated HT-SuperSAGE tags ($p \leq 0.05$, DiscoverySpace 4.01); DR: downregulated tags; n.s.: not significant; TS: drought-tolerant Embrapa 48 roots under water deficit; TC: drought-tolerant Embrapa 48 root without stress; SS: drought-sensible BR-16 roots under water deficit; SC: drought-sensible BR-16 roots without stress).

members, including the ALDH family 10, also described as putative BADH. A previous work [39] also observed the induction of BADH (almost 8-fold and 2-fold increase) under salinity and its accumulation in response to water stress or drought, indicating a common response of the plant to osmotic changes that affect its water status. The importance of identifying different candidates of this enzyme was highlighted by Nakamura *et al.* [40] that isolated two BADH transcripts (BBD1 and BBD2) from barley, one of them (BBD2) more similar to previously reported BADH genes from dicots. Both barley BADH genes showed different expression patterns. While BBD1 transcript was more abundant in roots and was induced to higher levels under salinity, drought and abscisic acid (ABA) treatment, the BBD2 transcript was more abundant in leaves after induction by salt, drought, PEG and ABA treatments, showing the potential of both genes for breeding purposes.

Delta(1)-pyrroline-5-carboxylate synthase (P5CS, EC 1.5.1.12)

Polypeptides regarding seven transcripts of delta(1)-pyrroline-5-carboxylate synthase 2 were similar to 19 OSMTL sequences (Additional file 1). Considering the transcripts, only Glyma18g40770.1 was not linked to a SuperSAGE tag. Seven tags matched with transcripts of the remaining six loci. From these, four were differentially expressed in the stressed library as compared with the negative control: one DR (downregulated GmDr_18680 mapped in the 3'UTR of Glyma01g24530.1) in both tolerant and sensible accessions; two DR tags in the sensible accession (tag GmDr_4918 mapped in the last CATG of the CDS of Glyma02g41850.1 and in the CDS of Glyma14g07120.1, and also tag GmDr_20800 at the 3'UTR of Glyma07g16510.1), besides a UR tag (FC = 9.6) only in the tolerant accession (tag GmDr_57499 at the CDSs of Glyma02g41850.1 and Glyma14g07120.1) (Additional file 2). The fact that both tags were associated to the CDS of Glyma02g41850.1 (Additional file 2) may be justified by the absence of the CATG sequence in the 3'UTR region. Even in the absence of expressive induction, the most prevalent tag (19-40 tpm; GmDr_4918) was observed in all libraries (Additional file 2).

Significant upregulation (RTqPCR) in leaves of PvP5CS (from common bean *Phaseolus vulgaris*) was demonstrated with transcription increase after 4d drought stress (2.5 times the control level), 2 h post-treatment (200 mM NaCl) of salt stress (about 16.3 times the control) and 2 h after of cold stress (11.7-fold). Another P5CS (PvP5CS2) also from common bean [41] presented predicted amino acid sequence showing 83.7% identity with PvP5CS and an overall 93.2% identity with GmP5CS [*G. max* P5CS], suggesting PvP5CS2 represented a soybean P5CS homolog gene. Indels (insertion and deletion events) and SNPs (single nucleotide polymorphisms) were found in the cloned PvP5CS2 genome sequence when the authors compared different accessions, helping in the development of a molecular marker in the chromosome b01. The association of molecular markers and phenotypes, in this case Pro accumulation is highly applicable for genetic improvement of plants and germplasm screening.

Delta(1)-pyrroline-5-carboxylate reductase (P5CR, EC 1.5.1.2)

The seed sequences OSMTL431, 432 and 434 were similar to P5CR polypeptides from two soybean loci (Glyma19g31230 and Glyma03g28480; Additional file 1), which transcripts were associated to SuperSAGE tags (Additional file 2). Of these one was repressed in the sensible accession (GmDr_4445, mapped at the 3'UTR region of the transcript Glyma19g31230.1) after stress (Additional file 2). A second one (GmDr_42728, mapped in the 3'UTR of Glyma03g28480.1 and CDS of Glyma03g28480.1) was not significantly modulated in the tolerant accession under stress as compared with the respective control, but presented a significant difference when compared to the sensible accession under stress (fold change of 12,0) (Additional file 2).

Previous genomic analysis indicated that there are only two to three copies of the P5CR gene in the soybean genome [42], similar to the proposed for pea [43]. Besides, the primary structure of pea P5CR is 85% identical with that of soybean isolated by Delauney and Verma [42]. The mentioned pea P5CR exhibited significant homology to human, yeast, and *E. coli* P5CR [43], a conservation that favours the here used approach in the search of orthologs using seeds sequences.

The suggestion that P5CR gene is osmoregulated was confirmed after subjecting soybean seedlings to osmotic stress (400 mM NaCl solution), resulting in an almost six-fold increase in the level of root P5CR mRNA [42]. An interesting aspect in association with proline overexpression and accumulation regards its influence on the concentration of other amino acids, suggesting a coordinated regulation of distinct metabolic pathways [44]. Free amino acid levels were compared in wild type and transgenic soybean (*G. max* cv. Ibis) transformed with P5CR in sense and antisense directions. The most rapid increase in Pro content was found in the sense transformants that exhibited the least water loss, while the slowest elevation of Pro levels was detected in the antisense transformants that exhibited the greatest water loss during stress. Correspondingly, the level of the Pro precursors Glu and Arg was higher in sense transformants and lower in antisense ones compared to the wild type plants during the initial exposure to stress (drought and heat) [44].

Myo-inositol 1-phosphate synthase (MIPS, EC 5.5.1.4)

A total of 13 OSMTL seed sequences (Additional file 1) presented similarities to polypeptides from three soybean *Myo*-inositol sequences. With exception of the transcript Glyma08g14670.1 that matched with MIPS1 the other two transcripts, matching MIPS2 and MIPS3, were associated to tags (Additional file 2). The tag GmDr_37 (mapped at the 3'UTR in all four alternative transcripts of Glyma18g02210) was the most frequent tag (615-1446 tpm) being DR in both stressed accessions (Additional file 2). The other tag (GmDr_3907) presented a perfect match with the 3'UTR of all three alternative transcripts of Glyma05g31450, with DR expression in the sensible accession under stress (Additional file 2). Another tag, GmDr_5821 (mapped at the 3'UTR in four alternative transcripts of Glyma18g02210) was induced (UR) in the tolerant accession Embrapa 48 under stress when compared with the respective control (Additional file 2). Considering all transcripts identified the locus Glyma18g02210 (MIPS2) seems to be the most interesting candidate for future validation and transgenic expression (in detriment to Glyma05g31450, MIPS3).

The confirmation of such a differential expression regarding MIPS is useful for plant breeding as highlighted by Kaur *et al.* [45] that observed two divergent genes encoding MIPS1 and MIPS2 (isolated from a drought-tolerant plant) in chickpea with differential expression but discrete overlapping roles, despite their pronounced divergence in respect to their íntrons composition, at the same time retaining 85% identity to their exons. Expression analysis showed both genes being expressed in all organs except seed, where only MIPS2 transcript was detected. Under environmental stresses (high temperature and salinity), only MIPS2 was induced whereas MIPS1 expression remained the same. Also, in those conditions of high temperature and salinity MIPS2 retained higher activity than MIPS1.

Myo-inositol monophosphatase (IMP, EC 3.1.3.25)

A total of 12 seed sequences (OSMTL61-66, OSMTL331-335 and OSMTL94) presented similarities with annotated IMP polypeptides regarding 10 *G. max* loci (Additional file 1), for those 19 SuperSAGE tags were identified. From the differentially expressed tags (Additional file 2), GmDr_3452 mapped at the last CATG of the Glyma08g19430.1, with bases in the CDS and 3'UTR and was induced after stress in both accessions. Similarly other tags were induced in the tolerant accession under stress (GmDr_23844, at the 3'UTR of Glyma16g28310.1 and GmDr_32375 at the 3'UTRs of both Glyma07g30110.1 and Glyma08g07200.1) (Additional file 2). By the other hand, the tag GmDr_5543 was mapped at the 3'UTR of three alternative transcripts of Glyma04g01170, being upregulated in the sensible accession and downregulated in the tolerant accession under stress (Additional file 2). Also the tag GmDr_25343 (mapped at the 3'UTR of Glyma15g07240.1) was downregulated in the tolerant accession after stress (Additional file 2). The abundance and differential expression of various IMP candidates in diverse comparisons indicate an important role in soybean water deficit. Despite of that and of the known role of these osmoprotectant-related genes, it is interesting that few expression essays or transgenic approaches have been carried using these candidates up to date.

In Arabidopsis transformants [46], two IMP candidate genes, IMPL1 and IMPL2 were expressed in a similar manner both in the vegetative and reproductive organs. The expression of IMP genes in a promoter-GUS assay on developing seeds was not coupled with the expression of the genes encoding MIPSs, which supply the substrate for IMPs in a 'de novo' synthesis pathway. Instead, IMP expression was correlated with SAL1 expression (encoding *Myo*-inositol polyphosphate 1-phosphatase), which is involved in the *Myo*-inositol salvage pathway.

Trehalose-6-phosphate synthase (TPS, EC:2.4.1.15)

After BLASTx 53 TPS OSMTL sequences were associated with 26 soybean transcripts of 21 loci (Additional file 1). From these, tags matched 22 transcripts and 17 loci, including TPS5, TPS7, TPS9 and TPS11 (Additional file 1). Among the differentially expressed tags (Table S2), three (GmDr_1203, GmDr_3893 and GmDr_9994) mapped at Glyma01g03870.1, Glyma06g19590.1 and Glyma17g07530, respectively and were considered induced in both accessions under stress. In turn, tag GmDr_62319 (Glyma04g35190.1, 3'UTR) was induced only in the sensible accession, while tag GmDr_25843 (Glyma01g03870.1, 3'UTR) was repressed under stress in the tolerant accession (Additional file 2).

Other two tags (GmDr_48598 and GmDr_57367, both mapping in Glyma01g03870.1, 3'UTR) were also DR in the tolerant accession under stress (Additional file 2). These two tags with different expression behavior for the same transcript could be considered as a possible annotation mistake, but further analysis showed that they regard sister tags, differing by a SNP, both mapping to Glyma01g03870.1 in an upstream site when compared to the mapped GmDr_25843 tag. Therefore, this last tag could be the result of a partial *Nla*III digestion, with the DR expression being questionable and therefore demanding validation. By the other hand, this possibility is quite unlikely, since a double digestion with *Nla*III was carried out prior to generation of HT-SuperSAGE libraries.

A similar situation was observed for two tags (GmDr_169137 and GmDr_198028, mapped both at Glyma06g19590.1) considered UR in the tolerant accession, while other two UR tags (GmDr_53228 and GmDr_61653) aligned to the same transcript with a single mismatch

(Additional file 2). A careful analysis revealed that the tags GmDr_169137, GmDr_198028 and GmDr_53228 mapped to CDS region, while GmDr_61653 mapped at the 3'UTR, in a CATG near the Poli-A tail, as expected for most SuperSAGE tags (Additional file 2). Thus, the most valid representative of this transcript seems to be GmDr_61653, induced in the tolerant accession under stress (Additional file 2).

Additional differentially expressed tags included GmDr_80395 (Glyma10g41680, 3'UTR) considered UR in the tolerant accession under stress; GmDr_66719 (mapped with two alternative transcripts of Glyma17g07530 at 3'UTR) UR in the sensible accession; GmDr_9508 (Glyma06g42820 and Glyma12g15500, both at CDS region), DR in the tolerant accession under stress (Additional file 2).

Such abundance and induction of TPS were also observed in other species. For example rice (*Oryza sativa*) contains 11 OsTPS genes, but only OsTPS1 showed TPS activity [47]. To demonstrate the physiological function of OsTPS1 the authors used the respective gene to transform rice plants and found that OsTPS1 overexpression improved the tolerance of seedling to cold, high salinity and drought conditions without other significant phenotypic changes.

Trehalose-phosphatase family protein (TPP; EC 3.1.3.12)

Contrasting with the results generated for TPS, a single transcript was observed for TPP (Glyma04g41640.1) in a locus associated to sixteen available OSMTL sequences (Additional file 1). This transcript was associated to only two differentially expressed tags (GmDr_43033 and GmDr_108104), both mapped at the 3'UTR region (Additional file 2), with discrete expression (2-5 tpm) in two out of four libraries. As in our case, few examples in the literature associated TPP expression with water deficit stress in plants, maybe due to their restricted prevalence in previously analyzed libraries. Despite the scarce number of reports the work of Ge *et al.* [48] revealed the transient upregulation of OsTPP1 (rice) after salt, osmotic and abscisic acid (ABA) treatments, with discrete upregulation under cold stress. Also, the overexpression lines analysis revealed that OsTPP1 triggered abiotic stress response genes, suggesting a possible transcriptional regulation pathway in stress induced reprogramming initiated by OsTPP1.

Tag-gene anchoring in the soybean genome

The search for osmoprotectant-associated homologs in soybean genes that matched with SuperSAGE tags recovered 179 sequences. However, these sequences were anchored in only 124 loci in the soybean genome (Additional file 3), what indicates the occurrence of alternative splicing of primary transcripts, suggesting an important role of osmoprotectants in vital processes during abiotic stresses in plants, probably inducing specific transcripts for particular environmental conditions. The anchoring of these osmoprotectant-related sequences and their respective SuperSAGE tags in the soybean virtual chromosomes revealed that osmoprotectant-related genes are present in 19 out of 20 soybean chromosomes (Figure 2). Most osmoprotectant-related sequences presented syntenic regions among non homologous chromosomes, often forming gene clusters mainly on both arms of chromosomes 2, 6 and 8, followed by chromosomes 5 and 17. On the opposite, some chromosomes presented few copies, as in the case of chromosomes 11-16 and 19-20, whereas chromosome 10 presented no match. The great number of osmoprotectant-related members in the short arm of chromosome 2 is also accompanied by a preferential distribution in the subterminal region, a phenomenon also observed in a lesser extent in the chromosomes 1, 5, 6, 7, 8, 9 and 17. Less frequently, gene-rich regions were also in the pericentromeric regions (e.g. in chromosomes 2, 3, 11 and 18) or intercalary regions.

A similar distribution was observed in regard to aquaporin genes, another gene family associated to drought stress in soybean [49]. Besides redundancies among chromosomes, aquaporins were also prevalent in terminal and subterminal gene clusters. As for aquaporins, the observed redundancy of osmoprotectant-related gene clusters corroborates previous suggestions of the soybean octoploid nature [50].

Another previous approach anchoring 59 soybean defense genes (two super-families: *R* resistance and *PR* pathogen related genes) in the virtual chromosomes of the legume *Medicago truncatula* revealed 1,253 sites, most of them clustered in subterminal or terminal positions. The 59 sequences were distributed in all nine medicago chromosomes, whereas 58 genes presented similarities with distinct segments in the same chromosome or appeared twice in distinct chromosomes [51]. Similar clustering was described for arabidopsis [52], indicating that such a distribution may occur in regard to different gene families and plant groups.

The redundancies observed probably reflect past duplication events, increasing the number of osmoprotectant-related genes in soybean genome [53,54]. The observed clustering and prevalence in some chromosomes, especially those combining different gene categories (as in the short arm of chromosomes 2, 6, and 7 or in the long arms of chromosomes 6, 8 and 9) indicate that these regions probably regard QTLs (Quantitative Trait Loci) useful for mapping approaches and marker assisted selection.

Conclusions

High throughput sequencing is generating a huge amount of sequences in given tissues and under contrasting conditions. In the present case we evaluated

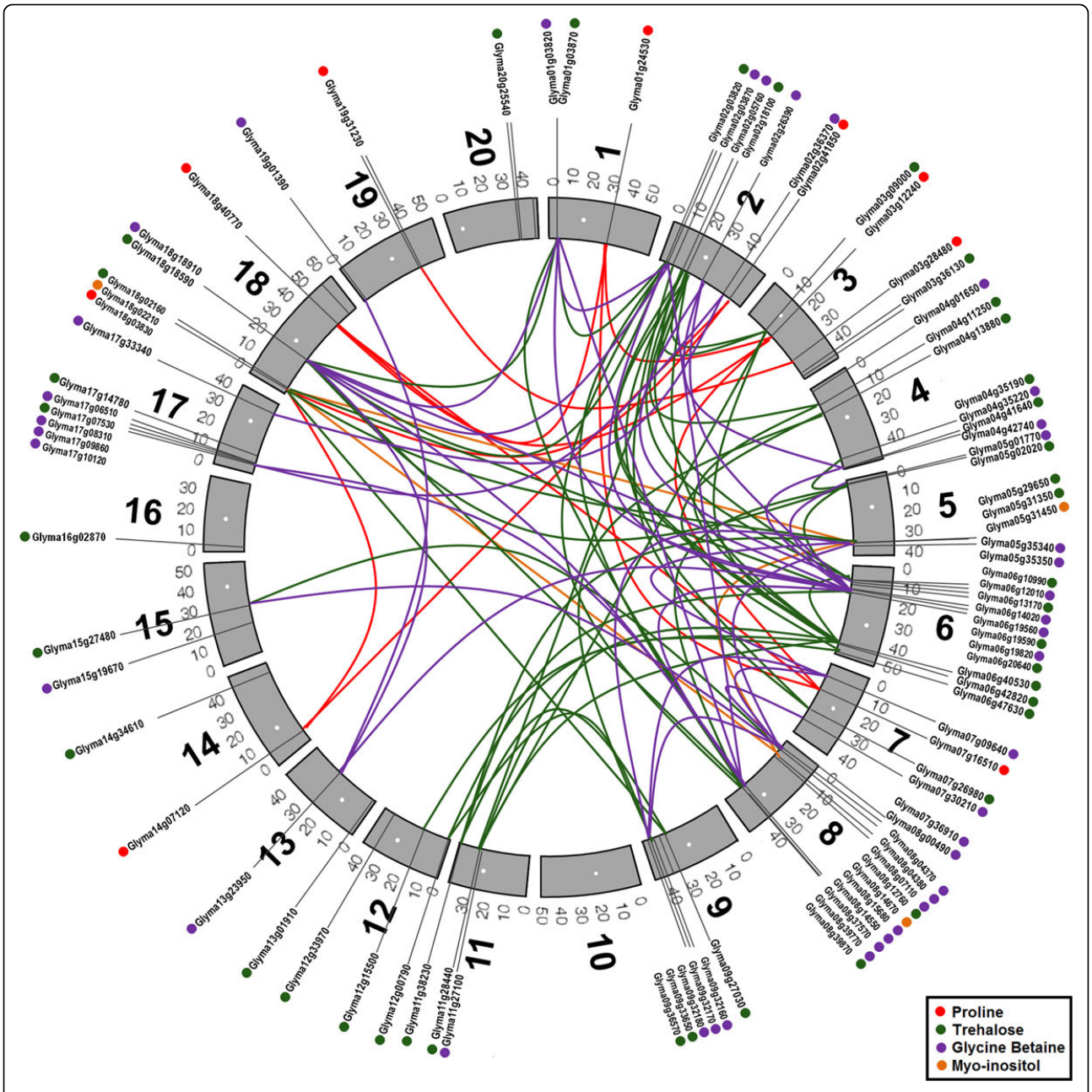


Figure 2 Distribution and syntenic relationships of 124 osmoprotectants in soybean genome. Chromosomes are marked at centromere (white circle) in their expected position. Each osmoprotectants name is shown outside in described position of SoyBase. Syntenic relationships are shown as color lines between chromosome regions. Red line - Proline (*P5CS* and *P5CR*); Green line - Trehalose (*TPS1* and *TPPB*); Purple line - Glycine Betaine (*BADH* and *CMO*); Blue line - Cysteine (*OASTL* and *SAT*); Orange line - *Myo*-inositol (*INPS1*).

osmoprotectant-related sequences in 26-bp tags from HT-SuperSAGE libraries from soybean coupled with Solexa/Illumina® sequencing in a digital gene expression profile. The approach permitted tags identification and annotation and their association with sequences from different sources (genomic regions, transcripts and proteins); identifying 36 differentially expressed osmoprotectant-related transcripts relative to 25 loci potentially active comprising four osmoprotectants classes. The 1,996 seed sequences and the workflow are also applicable to evaluate other angiosperms. Their clustering observed in soybean may be prevalent in other plant groups (or at least in legumes) and may be associated to interesting QTLs for breeding purposes or still for metabolic engineering in association with drought and salinity and chilling tolerance.

Methods

Seed sequences and annotation routine

The selection of seed sequences (Additional file 1) was based in a literature search in the PubMed database [55] using the key words "Osmoprotectants" AND "Plant Stress". In the selected articles the NCBI [56] descriptors for posterior mining were selected and retrieved from the Uniprot SwissProt (cutoff e^{-10}) using BLASTx. In order to confirm their involvement in the biosynthesis of osmoprotectants (proline, trehalose, *Myo*-inositol and glycine betaine) the sequences were aligned (BLASTx, cutoff e^{-10}) against the soybean peptide database at Phytozome v. 8.0 [37], also allowing the identification of the respective transcripts from soybean transcriptome used to associate with the available SuperSAGE tags.

Biological material, experimental design and stress application - Soybean HT-SuperSAGE libraries were generated according to the procedures described by Matsumura *et al.* [57] at GenXPro GmbH, with posterior SOLEXA sequencing of the tags. The generated tags are distributed into four libraries (Additional file 2) including root tissues subjected to dehydration: two libraries from the drought tolerant cultivar Embrapa 48 [Tolerant after stress (TS) and negative control (TC)] and two libraries from a drought sensible cultivar BR-16 [Sensible after Stress (SS) and negative control (SC)]. The conditions for the generation of the mentioned libraries, time frame experiments, and laboratory protocols used are described in Soares-Cavalcanti *et al.* [58]. The generated sequences are available at the GENOSOJA database (Brazilian Soybean Genome Consortium) [35].

Statistical analysis, tag-gene annotation and the tag fold change estimation - The in silico procedures are illustrated in Figure 1. Initially 26 bp-tags were analyzed with the DiscoverySpace (v.4.01) software [59] aiming to identify unique tags (unitags) and those unitags differentially expressed ($p \leq 0.05$) considering a contrast among two libraries. Tags counted only once (singlets) were excluded from the present evaluation. Unitags were annotated by BLASTn [60] against nucleotide sequences from the soybean Phytozome database v8.0 (Glyma1 cDNA dataset) [37,50]. BLASTn alignments (tag-hit) with e-values of 0.0001 or less and tolerating a single mismatch maximum (TSM) were taken into account. Moreover, only plus/plus alignments without mismatches regarding the four first bases CATG were accepted, in order to guarantee the integrity of the SuperSAGE tag. Specific keyword searches on the original glyma annotations were performed looking for the transcripts and tags candidates. Values reflecting expression data (p-value and up- or down-regulation regarding each tag) were associated to the data matrix including the respective tag annotation, the normalized frequencies in the libraries and the fold change values (FC). FC estimative were based on the ratio (R) of the normalized frequencies of the tag in the contrast of the two libraries, where the 'zero' frequency was replaced by 'one'. When $R > 1$ the FC were directly considered and when $R < 1$ the $FC = -1/R$. Negative FC values indicated repressed tags.

Tag-gene identification and anchoring in the soybean genome

A further approach consisted in the identification and generation of a curated list consisting of seven genes related to the biosynthesis of four classes of osmoprotectants [i.e. Proline (genes *P5CS* and *P5CR*), Trehalose (*TPS1* and *TPPB*), Glycine betaine (*BADH* and *CMO*) and *Myo*-inositol (*INPS1*)]. For this purpose a initial list was generated based on well known data from *Arabidopsis thaliana* (Additional file 3) used to identify corresponding sequences at SoyBase available on Phytozome [37,50], allowing the construction of a local database comprising complete soybean osmoprotectants for the alignment with the previously identified SuperSAGE tags and posterior anchoring in the SoyBase web server (consisting of pseudochromosomes from genome sequences including mainly BACs and molecular markers).

Circos mapping

Sequence matches for the nine selected osmoprotectant-related genes were aligned against the SoyBase pseudochromosomes aiming to infer about their distribution in the virtual chromosomes available at SoyBase. BLAST algorithm parameters (score, e-value and percentage of identity) were adjusted to allow the anchoring of soybean sequences position along the soybean virtual chromosomes. Afterwards the identified anchoring positions were submitted to the Circos program [61] and so edited to generate a picture of higher resolution. This approach allowed the generation of a graph based on a circular organization of the soybean chromosomes ($n = 20$), allowing the identification of a virtual ideogram with linear distribution of the osmoprotectants identified, the associated SuperSAGE tags, as well as redundant portions.

Additional material

Additional file 1: Table S1. BLASTx (Identity % and e-value) results regarding nucleotide sequences involved in the biosynthesis pathway and corresponding best hit (cut-off e^{-10}) in the Uniprot-SwissProt and Phytozome (v.8 *Glycine max*) databases, represented by the respective loci/transcripts in soybean (Glyma) as well as their annotation.

Additional file 2: Table S2. BLASTn results (Identity %, e-value) using HT-SuperSAGE 26-bp tags against *Glycine max* seed sequences given in Table S1. Libraries consisted of soybean cultivar Embrapa 48, tolerant (T) against drought, with roots submitted to water deficit after 1-6 h of stress submission (TS) as compared with the tolerant non stressed

control (TC). The same treatment was given to the drought sensible cultivar (BR-16) considering stressed (SS) and its respective control (SC). Tag annotation occurred against transcripts on the soybean Phytozome database. For each library normalized frequencies were considered (tpm: tags per million). Regarding comparisons among different treatments, fold change (FC), regulation (reg) [including upregulated (UR, in red) downregulated (DR, in green) and also non significant differential expression (ns), at 5% level] are given, as well as the tag-mapping position in the respective soybean transcript (Glyma).

Additional file 3: Table S3. Main soybean transcripts similar to known osmoprotectants-genes (I.) from *Arabidopsis thaliana* (used as seed sequences). II. tBLASTn results and sequence evaluation of soybean Osmoprotectants-genes. Information available include number of hits, best match of each gene class and features of hits: size range (maximal and minimal) in nucleotides (n), ORF (Open Reading Frame) size range in amino-acids (aa), e-value range based on SoyBase web resource, as well as number of matching HT-SuperSAGE tags.

List of abbreviations used

ABA: abscisic acid; ALDH: aldehyde dehydrogenase; Arg: arginine; BACs: bacterial artificial chromosome; BADH: betaine aldehyde dehydrogenase; CDS: coding sequence; CMO: choline monooxygenase; DR: down-regulated; FC: fold change value; GB: glycine betaine; GENOSOJA: Brazilian Soybean Genome Consortium; Glu: Glutamic acid; Glucose-6-P: glucose 6-phosphate; HP-SA: high performance sequencing approaches; HSP70: 70 kilodalton heat shock proteins; HT-SuperSAGE: High Throughput Super Serial Analysis of Gene Expression; IMP: L-*Myo*-inositol 1-phosphate phosphatise; MIPS: L-*Myo*-inositol 1-phosphate synthase; NGS: next generation sequencing; OAS: O-acetyl-L-serine; OAS-TL: O-acetyl-L-serine thiol lyase; OAT: ornithine d-aminotransferase; OsTPS: *Oryza sativa* Trehalose-6-phosphate synthase; P5CR: delta(1)-pyrroline-5-carboxylate reductase; P5CS: delta(1)-pyrroline-5-carboxylate synthase; PEG: polyethylene glycol; Pro: Proline; PtdInsP: phosphatidylinositol phosphate; QTLs: quantitative trait loci; RT-qPCR: real-time quantitative PCR; SAL1: *myo*- inositol polyphosphate 1-phosphatase; SAT: serine acetyltransferase; SNP: single nucleotide polymorphism; T6P: trehalose-6-phosphate; TPP: trehalose-6-phosphate phosphatise; *TPPB:* Trehalose 6- phosphate phosphatase B; TPS: Trehalose-6-phosphate synthase; TSM: tolerating a single mismatch; UDP: Glycosyltransferase/trehalose-phosphatase family protein; UR: up-regulated; UTRs: untranslated region.

Acknowledgements

The authors thank Dr Günter Kahl and Dr Björn Rotter for the scientific and technical advices and Dr. Ricardo Vilela Abdelnoor the Brazilian Soybean Genome Consortium coordinator. This work was funded by Brazilian institutions: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE).

Author details

[1]Departament of Genetics/Biological Sciences Center, Federal University of Pernambuco, Recife, Pernambuco, CEP 50.670-420, Brazil. [2]Embrapa Soybean, Brazilian Agricultural Research Corporation, Londrina, PR, CEP 86001-970, Brazil.

Authors' contributions

EAK, VP and ALN generated the HT-SuperSAGE libraries. EAK and JRCFN carried out the identification of tags and differential expression analysis. RLOS, LCB, JPBN, NMSC and MDS generated and curated the seed sequence bank, while JPBN and AMBI generated the data for genome anchoring. AMBI coordinated the research.

Declarations

The publication costs for this article were funded by the first author's institution.

This article has been published as part of *BMC Bioinformatics* Volume 14 Supplement 1, 2013: Computational Intelligence in Bioinformatics and Biostatistics: new trends from the CIBB conference series. The full contents of the supplement are available online at http://www.biomedcentral.com/bmcbioinformatics/supplements/14/S1.

Competing interests

The authors declare that they have no competing interests.

Published: 14 January 2013

References

1. Yancey PH: **Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses.** *J Exp Biol* 2005, **208**:2819-2830.
2. Rontein D, Basset G, Hanson AD: **Metabolic engineering of osmoprotectant accumulation in plants.** *Metab Eng* 2002, **4**:49-56.
3. Serraj R, Sinclair TR: **Osmolyte accumulation: can it really help increase crop yield under drought conditions?** *Plant Cell Environ* 2002, **25**:333-341.
4. McNeil SD, Nuccio ML, Hanson AD: **Betaines and related osmoprotectants: targets for metabolic engineering of stress resistance.** *Plant Physiol* 1999, **120**:945-949.
5. Wang W, Vinocur B, Altman A: **Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance.** *Planta* 2003, **218**:1-14.
6. Nouri M-Z, Toorchi M, Komatsu S: **Proteomics Approach for identifying abiotic stress responsive proteins in soybean.** In *Soybean - Molecular Aspects of Breeding.* Croatia: Intech;Sudaric A 2011:187-214.
7. Chen S, Gollop N, Heuer B: **Proteomic analysis of salt-stressed tomato (*Solanum lycopersicum*) seedlings: effect of genotype and exogenous application of glycinebetaine.** *J Exp Bot* 2009, **60**:2005-2019.
8. Barros OS, Soares-Cavalcanti NM, Vieira-Mello GS, Wanderley-Nogueira AC, Calsa-Junior T, Benko-Iseppon AM: ***In silico* evaluation of osmoprotectants in eucalyptus transcriptome.** *Lect Notes Comput Sci* 2009, **5488**:66-77.
9. Holmström KO, Somersalo S, Mandal A, Palva TE, Welin B: **Improved tolerance to salinity and low temperature in transgenic tobacco producing glycine betaine.** *J Exp Bot* 2000, **51(343)**:177-185.
10. Kathuria H, Giri J, Nataraja KN, Murata N, Udayakumar M, Tyagi AK: **Glycinebetaine-induced water-stress tolerance in codA-expressing transgenic indica rice is associated with up-regulation of several stress responsive genes.** *Plant Biotechnol J* 2009, **7(6)**:512-526.
11. Verbruggen N, Hermans C: **Proline accumulation in plants: a review.** *Amino Acids* 2008, **35(4)**:753-759.
12. Smirnoff N, Cumbes QJ: **Hydroxyl radical scavenging activity of compatible solutes.** *Phytochemistry* 1989, **28**:1057-1060.
13. Bhalu B, Mohanty P: **Molecular mechanisms of quenching of reactive oxygen species by proline under stress in plants.** *Curr Sci* 2002, **82(5)**:525-532.
14. Savouré A, Jaoua S, Hua XJ, Ardiles W, Van Montagu M, Verbruggen N: **Isolation, characterization, and chromosomal location of a gene encoding the delta 1-pyrroline-5-carboxylate synthetase in *Arabidopsis thaliana*.** *FEBS Lett* 1995, **372(1)**:13-19.
15. Parida AK, Dagaonkar VS, Phalak MS, Aurangabadkar LP: **Differential responses of the enzymes involved in proline biosynthesis and degradation in drought tolerant and sensitive cotton genotypes during drought stress and recovery.** *Acta Physiol Plant* 2008, **30(5)**:619-627.
16. Drennan PM: **The occurence of trehalose in the leaves of the desiccation-tolerant angiosperm *Myrothamnus flabellifolius* Welw.** *J Plant Physiol* 1993, **142(4)**:493-496.
17. Liu M-S, Chien C-T, Lin T-P: **Constitutive components and induced gene expression are involved in the desiccation tolerance of *Selaginella tamariscina*.** *Plant Cell Physiol* 2008, **49(4)**:653-663.
18. Fernandez O, Béthencourt L, Quero A, Sangwan RS, Clément C: **Trehalose and plant stress responses: friend or foe?** *Trends Plant Sci* 2010, **15(7)**:409-417.
19. Luo Y, Lib W-M, Wang W: **Trehalose: protector of antioxidant enzymes or reactive oxygen species scavenger under heat stress?** *Environ Exp Bot* 2008, **63(1-3)**:378-384.
20. Wingler A: **The function of trehalose biosynthesis in plants.** *Phytochemistry* 2002, **60(5)**:437-440.
21. Cortina C, Culiáñez-Macià FA: **Tomato abiotic stress enhanced tolerance by trehalose biosynthesis.** *Plant Sci* 2005, **169(1)**:75-82.
22. Chen THH, Murata N: **Glycinebetaine: an effective protectant against abiotic stress in plants.** *Trends Plant Sci* 2008, **13(9)**:499-505.
23. Jagendorf AT, Takabe T: **Inducers of glycinebetaine synthesis in barley.** *Plant Physiol* 2001, **127**:1827-1835.
24. Chen THH, Murata N: **Glycinebetaine protects plants against abiotic stress: mechanisms and biotechnological applications.** *Plant Cell Environ* 2011, **34(1)**:1-20.

25. Weretilnyk EA, Bednarek S, McCue KF, Rhodes D, Hanson AD: **Comparative biochemical and immunological studies of the glycine betaine synthesis pathway in diverse families of dicotyledons.** *Planta* 1989, **178(3)**:342-352.
26. Rathinasabapathi B, Burnet M, Russell BL, Gage DA, Liao PC, Nye GJ, Golbeck JH, Hanson AD: **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* 1997, **94(7)**:3454-3458.
27. Vojtechova M, Hanson AD, Munoz-Clares RA: **Betaine-aldehyde dehydrogenase from amaranth leaves efficiently catalyzes the NAD-dependent oxidation of dimethylsulfoniopropionaldehyde to dimethylsulfoniopropionate.** *Arch Biochem Biophys* 1997, **337(1)**:81-88.
28. Dastidar K, Maitra S, Goswami L, Roy D, Das KP, Majumder AL: **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* 2006, **140**:1279-1296.
29. Karner U, Peterbauer T, Raboy V, Jones DA, Hedley CL, Richter A: ***Myo*-Inositol and sucrose concentrations affect the accumulation of raffinose family oligosaccharides in seeds.** *J Exp Bot* 2004, **55(405)**:1981-1987.
30. Peters S, Mundree SG, Thomson JA, Farrant JM, Keller F: **Protection mechanisms in the resurrection plant *Xerophyta viscosa* (Baker): both sucrose and raffinose family oligosaccharides (RFOs) accumulate in leaves in response to water deficit.** *J Exp Bot* 2007, **58(8)**:1947-1956.
31. Alcazar-Roman AR, Wente SR, Inositol polyphosphates: **A new frontier for regulating gene expression.** *Chromosoma* 2008, **117**:1-13.
32. Majumder AL, Johnson MD, Henry SA: **1L-myoinositol-1-phosphate synthase.** *Biochim Biophys Acta* 1997, **1348(1-2)**:245-256.
33. Parthasarathy L, Vadnal RE, Parthasarathy R, Devi CS: **Biochemical and molecular properties of lithium-sensitive *myo*-inositol monophosphatase.** *Life Sci* 1994, **54(16)**:1127-1142.
34. Matsumura H, Yoshida K, Luo S, Kimura E, Fujibe T, Albertyn Z, Barrero RA, Krüger DH, Kahl G, Schroth GP, Terauchi R: **High-throughput superSAGE for digital gene expression analysis of multiple samples using next generation sequencing.** *PLoS ONE* 2010, **5(8)**:e12010.
35. **GENOSOJA database (Brazilian Soybean Genome Consortium).** [http://bioinfo03.ibi.unicamp.br/soja].
36. Benko-Iseppon AM, Nepomuceno AL, Abdelnoor RV: **GENOSOJA - The Brazilian soybean genome consortium: high throughput omics and beyond.** *Genet Mol Biol* 2012, **35(2)**:i-iv.
37. **Phytozome.** [http://www.phytozome.net/].
38. Kotchoni SO, Jimenez-Lopez JC, Kayodé AP, Gachomo EW, Baba-Moussa L: **The soybean aldehyde dehydrogenase (ALDH) protein superfamily.** *Gene* 2012, **495(2)**:128-33.
39. Ishitani M, Nakamura T, Han SY, Takabe T: **Expression of the betaine aldehyde dehydrogenase gene in barley in response to osmotic stress and abscisic acid.** *Plant Mol Biol* 1995, **27(2)**:307-15.
40. Nakamura T, Nomura M, Mori H, Jagendorf AT, Ueda A, Takabe T: **An isozyme of betaine aldehyde dehydrogenase in barley.** *Plant Cell Physiol* 2001, **42(10)**:1088-1092.
41. Chen JB, Wang SM, Jing RL, Mao XG: **Cloning the PvP5CS gene from common bean (*Phaseolus vulgaris*) and its expression patterns under abiotic stresses.** *J Plant Physiol* 2009, **166(1)**:12-9.
42. Delauney AJ, Verma DP: **A soybean gene encoding delta 1-pyrroline-5-carboxylate reductase was isolated by functional complementation in *Escherichia coli* and is found to be osmoregulated.** *Mol Gen Genet* 1990, **221(3)**:299-305.
43. Williamson CL, Slocum RD: **Molecular cloning and evidence for osmoregulation of the delta 1-pyrroline-5-carboxylate reductase (proC) gene in pea (*Pisum sativum* L.).** *Plant Physiol* 1992, **100**:1464-1470.
44. Simon-Sarkadi L, Kocsy G, Várhegyi A, Galiba G, de Ronde JA: **Genetic manipulation of proline accumulation influences the concentrations of other aminoacids in soybean subjected to simultaneous drought and heat stress.** *J Agric Food Chem* 2005, **53(19)**:7512-7517.
45. Kaur H, Shukla RK, Yadav G, Chattopadhyay D, Majee M: **Two divergent genes encoding L-myo-inositol 1-phosphate synthase1 (CaMIPS1) and 2 (CaMIPS2) are differentially expressed in chickpea.** *Plant Cell Environ* 2008, **31(11)**:1701-1716.
46. Sato Y, Yazawa K, Yoshida S, Tamaoki M, Nakajima N, Iwai H, Ishii T, Satoh S: **Expression and functions of *myo*-inositol monophosphatase family genes in seed development of Arabidopsis.** *J Plant Res* 2011, **124(3)**:385-394.
47. Li HW, Zang BS, Deng XW, Wang XP: **Overexpression of the trehalose-6-phosphate synthase gene OsTPS1 enhances abiotic stress tolerance in rice.** *Planta* 2011, **234(5)**:1007-1018.
48. Ge LF, Chao DY, Shi M, Zhu MZ, Gao JP, Lin HX: **Overexpression of the trehalose-6-phosphate phosphatase gene OsTPP1 confers stress tolerance in rice and results in the activation of stress responsive genes.** *Planta* 2008, **228(1)**:191-201.
49. Oliveira ARS, Brasileiro-Vidal AC, Bortoleti KCA, Bezerra-Neto JP, Abdelnoor RV, Benko-Iseppon AM: **Mining plant genome browsers as a mean for efficient connection of physical, genetic and cytogenetic mapping: an example using soybean.** *Genet Mol Biol* 2012, **35(1)**:335-347, (Suppl 1).
50. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, *et al*: **Genome sequence of the paleopolyploid soybean.** *Nature* 2010, **463**:178-183.
51. Wanderley-Nogueira AC, Soares-Cavalcanti NM, Belarmino LC, Bezerra-Neto JP, Kido EA, Pandolfi V, Abdelnoor R, Bineck E, Carazzole MF, Benko-Iseppon AM: **An overall evaluation of the Resistance (R) and Pathogenesis Related (PR) superfamilies in soybean, as compared with Medicago and Arabidopsis.** *Genet Mol Biol* 2012, **35(1)**:260-271, (Suppl 1).
52. The Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**:796-815.
53. Shultz JL, Kurunam D, Shopinski K, Iqbal MJ, Kazi S, Zobrist K, Bashir R, Yaegashi S, Lavu N, Afzal JA, Yesudas CR, Kassem MA, Wu C, Zhang HB, Town CD, Meksem K, Lightfoot DA: **The soybean genome database (soyGD): a browser for display of duplicated, polyploidy, regions and sequence tagged sites on the integrated physical and genetic maps of *Glycine max*.** *Nucleic Acids Res* 2006, **34(1)**:D758-D756.
54. Liu Q, Wang H, Zhang Z, Wu J, Feng Y, Zhu Z: **Divergence in function and expression of the NOD26-like intrinsic proteins in plants.** *BMC Genomics* 2009, **10**:313.
55. **PubMed Database.** [http://www.ncbi.nlm.nih.gov/pubmed].
56. **National Center for Biotechnology Information (NCBI).** [http://www.ncbi.nlm.nih.gov/].
57. Matsumura H, Kruger DH, Kahl G, Terauchi R: **SuperSAGE: a modern platform for genome-wide quantitative transcript profiling.** *Curr Pharm Biotechnol* 2008, **9(5)**:368-374.
58. Soares-Cavalcanti NM, Belarmino LC, Kido EA, Wanderley-Nogueira AC, Bezerra-Neto JP, Cavalcanti-Lira R, Pandolfi V, Nepomuceno AL, Abdelnoor RV, Nascimento LC, Benko-Iseppon AM: **Overall picture of expressed Heat Shock Factors in *Glycine max, Lotus japonicus* and *Medicago truncatula*.** *Genet Mol Biol* 2012, **35(2)**:247-259.
59. Robertson NM, Oveisi-Fordorei SD, Varhol RJ, Fjell C, Marra M, Jones S, Siddiqui A: **DiscoverySpace: an interactive data analysis application.** *Genome Biol* 2007, **8(1)**:R6.
60. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
61. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19**:1639-1645.

doi:10.1186/1471-2105-14-S1-S7
Cite this article as: Kido *et al.*: Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress. *BMC Bioinformatics* 2013 **14**(Suppl 1):S7.

7. Considerações Finais

Nossos estudos nos permitiram concluir que:

- A construção de bibliotecas DeepSuperSAGE e HT-SuperSAGE se apresenta como uma estratégia robusta e eficiente para análises de transcriptomas;

- O desenho experimental adotado para o presente ensaio, utilizando acessos contrastantes de soja submetidos a déficit hídrico, propiciou a identificação de *unitags* intimamente associadas ao processo de tolerância;

- Entre acessos contrastantes de soja, os transcriptomas produzidos em resposta a déficit hídrico variam tanto quanti quanto qualitativamente;

- A partir de 25 minutos de submissão do estresse, acessos contrastantes de soja já iniciam a reprogramação de seus transcriptomas;

- Transcritos associados a osmoprotetores (P5CS, P5CR, TPS1, TPPB, INPS1, BADH, MIPS) são expressos em ambos acessos em resposta ao déficit hídrico, potencialmente, desempenhando importantes papéis na manutenção da fisiologia dos mesmos.

ANEXO

PRODUÇÃO ACADÊMICA

(2010-2014)

Artigos completos publicados em periódicos

1.

KIDO, EA; **FERREIRA NETO, JRC**; SILVA, RLO; BELARMINO, LC; BEZERRA NETO, JP; SOARES-CAVALCANTI, NM; PANDOLFI, V; SILVA, MD; NEPOMUCENO, AL; BENKO-ISEPPON, AM. Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress. BMC Bioinformatics, v. 14, p. s7, 2013.

2.

BRITO-KIDO, SA; **FERREIRA NETO, JRC**; PANDOLFI, V; MARCELINO, FC; NEPOMUCENO, AL; ABDELNOOR, RV; BENKO-ISEPPON, AM; KIDO, EA. Natural Antisense Transcripts in Plants: A Review and Identification in Soybean Infected with *Phakopsora pachyrhizi* SuperSAGE Library. The Scientific World Journal, v. 2013, p. 1-14, 2013.

3.

KIDO, EA; **FERREIRA NETO, JRC**; SILVA, RLO; PANDOLFI, V; GUIMARÃES, ACR; VEIGA, DT; CHABREGAS, SM ; CROVELLA, S; BENKO-ISEPPON, AM . New Insights in the Sugarcane Transcriptome Responding to Drought Stress as Revealed by Supersage. The Scientific World Journal, v. 2012, p. 1-14, 2012.

4.

EA, KIDO; BARBOSA, PKA; **FERREIRA NETO, JRC**; PANDOLFI, V; HOULLOU-KIDO, LM; CROVELLA, S; MOLINA, C; KAHL, G; BENKO-ISEPPON,

AM. Identification of Plant Protein Kinases in Response to Abiotic and Biotic Stresses Using SuperSAGE. Current Protein and Peptide Science, v. 12, p. 643-656, 2011.

5. FERREIRA NETO, JRC; PANDOLFI, V; MARCELINO-GUIMARAES, FC; BENKO-ISEPPON, AM; ROMERO, C; SILVA, RLO; RODRIGUES, FA; ABDELNOOR, RV; NEPOMUCENO, AL; KIDO, EA. Early Transcriptional Response of Soybean Contrasting Accessions to Root Dehydration. Plos One. v.8, p.e83466, 2013.

6. da SILVA, MD, SILVA, RLO, **FERREIRA NETO, JRC**, GUIMARÃES, ACR; VEIGA, DT; CHABREGAS, SM; BURNQUIST, WL; KAHL, G; BENKO-ISEPPON, AM; KIDO, EA. Expression Analysis of Sugarcane Aquaporin Genes under Water Deficit. Journal of Nucleic Acids. v. 2013, p.1 - 14, 2013.

7. MEDEIROS, CD; **FERREIRA NETO, JRC**, OLIVEIRA, MT; RIVAS, R; PANDOLFI, V; KIDO, EA; BALDANI, JI; SANTOS, MG. Photosynthesis, Antioxidant Activities and Transcriptional Responses in Two Sugarcane (*Saccharum officinarum* L.) Cultivars under Salt Stress. Acta Physiologiae Plantarum. v.1, p.1, 2013.

8. BENKO-ISEPPON, AM; SOARES-CAVALCANTI, NM; BELARMINO, LC; BEZERRA NETO, JP; AMORIM, LLB; **FERREIRA NETO, JRC**; PANDOLFI, VALESCA; AZEVEDO, HMA; SILVA, RLO; SANTOS, MG; ALVES, MVS; KIDO, EA. Prospecção de Genes de Resistência à Seca e à Salinidade em Plantas Nativas e Cultivadas. Revista Brasileira de Geografia Física, v. 6, p. 1112-1134, 2011.

Livros publicados/organizados ou edições

1.

KIDO, EA; **FERREIRA NETO, JRC**; SILVA, RLO; ANDRADE, PP; BENKO-ISEPPON, AM. Gene Discovery and Protection: New Trends and Market Restrictions on Stress Tolerant Transgenic Sugarcane. 1ed. New York: Nova Publishers, 2012.

Capítulos de livros publicados

1.

SILVA, RLO; **FERREIRA NETO, JRC**; PANDOLFI, V; CHABREGAS, SM; BURNQUIST, WL; BENKO-ISEPPON, AM; KIDO, EA. Transcriptomics of Sugarcane Osmoprotectants Under Drought. In: Hemanth KN, Vasanthaiah, Devaiah Kambiranda. (Org.). Plants and Environment. 1ed. InTech, 2011, v. , p. 89-114.

2.

KIDO, EA; **FERREIRA NETO, JRC**; SILVA, RLO; BALDANI, I; BENKO-ISEPPON, A M. Tolerance against Abiotic Stresses in Sugarcane: Can Gene Discovery and Biotechnology Improve the Yield?. In: João F. Goncalves; Kauê D. Correia.. (Org.). Sugarcane: Production, Cultivation and Uses. 1 ed. New York: Nova Publishers, 2011, v., p. 1-.

3.

KIDO, EA; **FERREIRA NETO, JRC**; KIDO, SAB; PANDOLFI, V, BENKO-ISEPPON, AM. DeepSuperSAGE in a Friendly Bioinformatic Approach: Identifying

Molecular Targets Responding to Abiotic Stress in Plants. In: Molecular Approaches in Plant Abiotic Stress. 1 ed. Boca Raton, Florida: CRC Press, 2013, p. 10-.

Resumos expandidos publicados em anais de congressos

1.

FERREIRA NETO, JRC; COSTA, MMR; PANDOLFI, V; SILVA, RLO; DAMASCENO, KJ; NEPOMUCENO, AL; BRONDANI, RP V; BENKO-ISEPPON, AM; KIDO, EA. Modulação transcricional de genes associados à fosforilação oxidativa em genótipos contrastante de feijão-caupi submetidos À desidratação radicular. In: III Congresso Nacional do Feijão-caupi, 2013, Recife. III Congresso Nacional do Feijão-caupi, 2013.

2.

NASCIMENTO, ATB; SILVA, CB; COSTA, MMR; **FERREIRA NETO, JRC**; ROCHA, MM; DAMASCENO, KJ; KIDO, EA. Análise de diversidade em genótipos de feijão-caupi do programa de melhoramento da Embrapa Meio-Norte. In: III Congresso Nacional do Feijão-caupi, 2013, Recife. III Congresso Nacional do Feijão-caupi, 2013.

3.

PANDOLFI, V; AMORIM, LLB; SANTOS, SSR; SANTOS, RF; **FERREIRA NETO, JRC**; GAZZANEO, LRS; SILVA, LCB; SANTANA, KCB; MOURA, RR; KIDO, EA; ANDRADE, GP; PIO-RIBEIRO, G; BENKO-ISEPPON, AM. Seleção de genes de referência e expressão de uma defensina em espécies de *Vigna* sob estresse biótico. In: III Congresso Nacional do Feijão-caupi, 2013, Recife. III Congresso Nacional do Feijão-caupi, 2013.

4.

SILVA, CB; **FERREIRA NETO, JRC**; PANDOLFI, V; DAMASCENO, KJ; ROCHA, MM; NEPOMUCENO, AL; BRONDANI, RPV; BENKO-ISEPPON, AM; KIDO, EA. Transcriptoma SuperSAGE de Feijão-caupi sob desidratação radicular. In: III Congresso Nacional do Feijão-caupi, 2013, Recife. III Congresso Nacional do Feijão-caupi, 2013.

5.

COSTA, MMR; NASCIMENTO, ATB; AQUINO, K; **FERREIRA NETO, JRC**; PANDOLFI, V; DAMASCENO, KJ; NEPOMUCENO, AL; BRONDANI, RPV; BENKO-ISEPPON, AM; KIDO, EA. Mapeamento *in silico* de unitags SuperSAGE de feijão-caupi em genoma tentativo de feijão-comum. In: III Congresso Nacional do Feijão-caupi, 2013, Recife. III Congresso Nacional do Feijão-caupi, 2013.

6.

MOURA, D; **FERREIRA NETO, JRC**; KIDO, EA. Desenvolvimento de marcadores moleculares AFLP para mapeamento genético do feijão-caupi visando tolerância à seca. In: XIX Encontro de Genética do Nordeste, 2012, Petrolina / Juazeiro. XIX Encontro de Genética do Nordeste, 2012.

Resumos publicados em anais de congressos

1.

KIDO, EA; PANDOLFI, V; **FERREIRA NETO, JRC**; BARBOSA, PKA; ANDRADE, PP; SILVA, RLO; MARCELINO, FC; ABDELNOOR, RV; BENKO-ISEPPON, AM. Differential transcriptional profiling of *Phakopsora pachyrhizi*-infected

soybean revealed by high-throughput SuperSage. In: World Soybean Research Conference, 2013, Durban. World Soybean Research Conference, 2013.

2.

BEZERRA NETO, JOÃO P; BELARMINO, LC; PANDOLFI, V; **FERREIRA NETO, JRC**; MARCELINO, FC; ROMERO, C; KIDO, EA; Nepomuceno, AL; BENKO-ISEPPON, AM. Soybean aquaporins: diferentiaal expression, genome distribution and structure. In: World Soybean Research Conference, 2013, Durban. World Soybean Research Conference, 2013.

3.

KIDO, EA; PANDOLFI, V; **FERREIRA NETO, JRC**; SILVA, RLO; BENKO-ISEPPON, AM. Cowpea transcriptomic: looking for genes responding to biotic and abiotic stresses In: Workshop on Biotic and Abiotic Stress Tolerance in Plants: the Challenge for the 21st Century, 2013, Ilhéus. Workshop on Biotic and Abiotic Stress Tolerance in Plants: the Challenge for the 21st Century (book of abstracts), 2013.

4.

KIDO, EA; **FERREIRA NETO, JRC**; SILVA, RLO; SILVA, MD; CORREIA, CN; BALDANI, IJ; CHABREGAS, SM; BURNQUIST, WL. *In silico* expression pattern of sugarcane SuperSAGE tags responsive to Abiotic Stresses. In: Congreso Latinoamericano de Genética, 2010, Viña del Mar - Chile. Congreso Latinoamericano de Genética, 2010.

5.

FERREIRA NETO, JRC; SILVA, RLO; SILVA, MD; CORREIA, CN; ROCHA, MM; FREIRE FILHO, FR; DAMASCENO, KJ; BENKO-ISEPPON, AM; KIDO, EA. Marcadores moleculares no programa brasileiro de melhoramento de feijão-caupi. In: Congreso Latinoamericano de Genética, 2010, Viña del Mar - Chile. Congreso Latinoamericano de Genética, 2010.