



Subtractive libraries for prospecting differentially expressed genes in the soybean under water deficit

Fabiana Aparecida Rodrigues¹, Juliana Marcolino-Gomes¹, Josirlei de Fátima Corrêa Carvalho¹, Leandro Costa do Nascimento², Norman Neumaier¹, José Renato Bouças Farias¹, Marcelo Falsarella Carazzolle², Francismar Corrêa Marcelino¹ and Alexandre Lima Nepomuceno¹

¹*Embrapa Soja, Laboratório de Biotecnologia Vegetal, Londrina, PR, Brazil.*

²*Departamento de Genética, Evolução e Bioagentes, Universidade Estadual de Campinas, Campinas, SP, Brazil.*

Abstract

Soybean has a wide range of applications in the industry and, due to its crop potential, its improvement is widely desirable. During drought conditions, soybean crops suffer significant losses in productivity. Therefore, understanding the responses of the soybean under this stress is an effective way of targeting crop improvement techniques. In this study, we employed the Suppressive Subtractive Hybridization (SSH) technique to investigate differentially expressed genes under water deficit conditions. Embrapa 48 and BR 16 soybean lines, known as drought-tolerant and -sensitive, respectively, were grown hydroponically and subjected to different short-term periods of stress by withholding the nutrient solution. Using this approach, we have identified genes expressed during the early response to water deficit in roots and leaves. These genes were compared among the lines to assess probable differences in the plant transcriptomes. In general, similar biochemical processes were predominant in both cultivars; however, there were more considerable differences between roots and leaves of Embrapa 48. Moreover, we present here a fast, clean and straightforward method to obtain drought-stressed root tissues and a large enriched collection of transcripts expressed by soybean plants under water deficit that can be useful for further studies towards the understanding of plant responses to stress.

Key words: gene expression, drought, transcriptome.

Introduction

Soybean is one of the most important crops in the world, and Brazil is its second largest producer. Due to its use in animal feed, vegetable oil and biodiesel, there is a sustained demand for soybeans. Nevertheless, abiotic stresses, such as water deficit, cold, wind, heat or waterlogging, reduce productivity and present a major challenge in the quest for sustainable soybean production. In order to meet the ever-increasing demand, studies to improve productivity are very important. Crop yield is highly affected by water deficit stress, particularly when it occurs during flowering and early pod expansion (Pedersen *et al.*, 2005). However, soybean plants can acclimate to such conditions by triggering protective mechanisms that enable these plants to survive subsequent events of drought (Bray, 2004). While studying soybean plants exposed to stress at different developmental stages, Kron *et al.* (2008) observed

that plant responses were more uniform in the V4 stage (Fehr *et al.*, 1971). The genetic background of the cultivars BR 16 (drought sensitive) and Embrapa 48 (drought tolerant) under drought conditions was studied by Oya *et al.* (2004), and they found that at the vegetative stage, tolerant cultivars present higher growth rate and a larger leaf area compared to sensitive plants. However, to better understand plant physiology under stress, productivity analyses are hardly enough to distinguish significant tolerance differences among cultivars. The data indicate how cultivars behave under stress and, therefore, suggest which ones are suitable for further biochemical and molecular studies.

A stress response is initiated when plants recognize the stress at the cellular level, it triggers a complex signal transduction network which is modulated by molecular events. Gene products involved in water-deficit responses can be classified into two groups: functional or regulatory genes. Functional genes include the following: proteins, osmolytes, chaperones, heat-shock proteins (HSPs), water-channel proteins (multifunctional proteins involved in transport facilitation of solutes and water in leaves and roots), and other compounds probably conferring direct tol-

Send correspondence to Alexandre Lima Nepomuceno. Embrapa Soja, Laboratório de Biotecnologia Vegetal, Rodovia Carlos João Strass s/n, 86001-970 Londrina, PR, Brazil. E-mail: nepo@cnpso.embrapa.br.

erance to abiotic stresses. Under water stress conditions, a chemical imbalance in the cell leads to production of reactive oxygen species (ROS), which are damaging to lipids and proteins (Carvalho, 2008; Molina *et al.*, 2008). Enzymes involved in detoxification metabolism are found to protect against the stress in the tolerant common bean cultivar (*Phaseolus acutifolius*) (Türkan *et al.*, 2005). In drought-stressed tissues, the water movement within and between cells can be facilitated by aquaporins (Cocozza *et al.*, 2010). The second group of genes involved in the response to water deficit includes molecules involved in further regulation of signal transduction and stress-responsive gene expression. These molecules are regulatory proteins, represented by various transcription factors such as DREB1, ABA-responsive element, nitrogen assimilation control protein, ethylene-response factors, kinases, phosphatases, enzymes involved in phospholipid metabolism, and other signaling molecules, such as calmodulin-binding protein (Shinozaki and Yamaguchi-Shinozaki, 2007). In the peanut, a predicted drought tolerant legume, a set of functional and regulatory genes were detected in enriched subtractive libraries probably acting in a gradual process of drought acclimation (Govind *et al.*, 2009).

As discussed, the ability to tolerate water deficit is a complex trait controlled by many genes (Molina *et al.*, 2008; Ergen and Budak, 2009). Collections of cDNAs expressed under stressed conditions have been a useful resource to study gene expression and regulation. The cDNA libraries were associated to normalization steps and PCR amplification to obtain just the transcripts presented in a target situation (Diatchenko *et al.*, 1996). A suppressive subtractive library approach makes it possible to search for genes expressed in treated samples but not in the control samples, if the forward subtraction is done. It can be used to assess changes in metabolism and in quantitative gene expression studies, such as those based on arrays. The method has been employed successfully in different research areas, such as the monitoring of citrus responses to fungus infection (González-Candelas *et al.*, 2010) and in fish infected by iridovirus (Xu *et al.*, 2010). Using the subtracted library of pigeonpea, Priyanka *et al.* (2010) isolated a gene encoding a hybrid-rich-proline protein, which affords tolerance to water deficit and other abiotic stresses such as salt, heat, cold and abscisic acid treatment. In addition, cDNA subtractive libraries have also allowed the identification of genes involved in the flower development of orange plants (Zhang *et al.*, 2008) and of genes involved in the *Varroa destructor* mite tolerance to honey bees (Zhang *et al.*, 2010).

In recent years, drought has resulted in severe losses in soybean productivity worldwide. In order to expand previous knowledge about the water stress-sensitive BR 16 and water stress-tolerant Embrapa 48 soybean cultivars, the present work aimed to study early responses in gene expression of leaves and roots under water stress.

Material and Methods

Plant growth and water stress treatment

Seeds of the soybean cultivars Embrapa 48 and BR 16, classified as tolerant and sensitive to water deficit, respectively, were germinated on filter paper during four days in a growth chamber at 25 ± 1 °C of temperature and 100% relative humidity. Seedlings were placed in 36 L boxes containing 50% Hoagland's solution (Hoagland and Arnon, 1950), which was continuously aerated and replaced on a weekly basis. These boxes were then transferred to a greenhouse under a natural photoperiod of approximately 12/12 h light/dark cycle, temperature of 30 ± 5 °C and $60 \pm 10\%$ relative humidity (RH), where the plants were allowed to grow until the V4 stage (Fehr *et al.*, 1971). The experimental plan was a randomized complete block 2x7 factorial design with three repetitions. The factors were two cultivars (BR 16 and Embrapa 48) and seven times of water deficit (0, 25, 50, 75, 100, 125 and 150 min). The stress was applied by removing the plants out of the hydroponic solution and leaving them in boxes without nutrient solution for up to 150 min under ambient-air exposure. For each time of water deficit, roots from 10 plants were collected, pooled and frozen in liquid nitrogen before storage at - 80 °C. The same procedure was performed for the leaf samples.

Physiological parameters measurements

Photosynthetic rate (A), stomatal conductance (g_s), intercellular CO₂ concentration (c_i), transpiration rate (E) and leaf temperature (T_l) were evaluated using a LI-6400 Portable Photosynthesis System (LiCor, Inc.). Measurements were taken on the middle leaflet of the youngest soybean leaf, fully expanded, under photon flux density of $1.000 \mu\text{mol m}^{-2} \text{s}^{-1}$. The water use efficiency (WUE) of plants was determined through the ratio between photosynthetic (A) and transpiration (E) rate. Data was submitted to ANOVA analysis through the programs Sanest and SAS (Statistical Analysis System *version 8.0*), and the treatment means were compared by the Tukey test ($p < 0.05$).

cDNA subtracted libraries synthesis

Total RNA was extracted from roots and leaves using Trizol reagent (Invitrogen) according to the manufacturer's instructions. Samples from each biological replicate were extracted separately, and equimolar amounts of total RNA were pooled before the mRNA purification. The FastTrack MAG mRNA Isolation (Invitrogen) was applied to isolate the mRNA Poly A⁺ molecules by using oligo-dT conjugated magnetic beads. The mRNA samples were pooled as illustrated in Figure 1 to perform the cDNA reverse transcription. Libraries of roots and leaves from Embrapa 48 and BR 16 plants were constructed based on the method of Diatchenko *et al.* (1996) with the PCR Select cDNA Subtraction (Clontech) Kit. First strand cDNA was synthesized

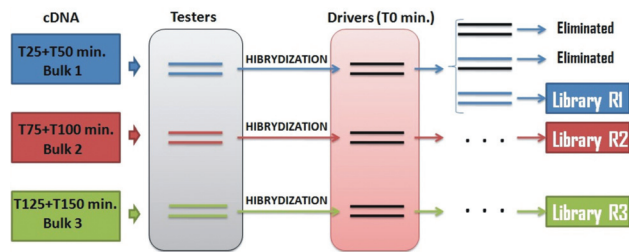


Figure 1 - Illustration of the subtracted libraries synthesis. cDNAs extracted from drought-treated tissues were applied as tester and normalized using the control sample as driver. The subtractive effect occurred in the second hybridization cycle when an excess of driver was added to obtain transcripts expressed only in tester samples (forward subtraction). Libraries were named as R/L1, R/L2 and R/L3 according to bulks 1 (25-50 min), 2 (75-100 min) and 3 (125-150 min), respectively.

from 2 μg of mRNA Poly A⁺ pooled samples (tester and driver). After the second strand synthesis, cDNAs (tester and driver) were digested with the restriction enzyme *Rsa I* to produce blunt end fragments. Once digested, the cDNAs (only tester samples) were divided into two samples and each sample was ligated with two different adapters (adaptor 1 and 2R, respectively). As our aim was to detect transcripts expressed only in water-deficit stressed plants, the sample T0 (control plant) was used as the driver in the hybridization step. The cDNA driver (with no adaptor) was added to tester samples (1 and 2R) and they were submitted to the first hybridization step to eliminate sequences common in both samples or present in the control (driver). For the second hybridization, tester samples 1 and 2R from the first hybridization were mixed together, and the denatured cDNA driver was added again to enrich the expressed sequences in the tester. The PCR amplification was performed using the double-strand cDNAs with different adapters as template. Sequences were exponentially amplified due to the specificity of primers to adapters. The second PCR was carried out using the nested primers to enrich the differentially expressed sequences. The PCR products were directly applied to next generation sequencing.

Data analysis

The PCR products were sequenced using the Sequencing by Synthesis Technology (Illumina). The six libraries from BR 16 cultivar (three from roots and three from leaves) were sequenced through 36 bp (base pairs) single reads in a Genome Analyzer *Ile* machine. The Embrapa 48 libraries were sequenced in a Genome Analyzer *Ilz* using single reads of 76 bp. Libraries tagged with a barcode produced thousands of reads from each library. After base-calling, reads were mapped on a reference genome (Schmutz *et al.*, 2010) using the SOAP (Li *et al.*, 2008), and non-mapped reads were assembled into contigs through the Edena program following the default parameters (Hernandez *et al.*, 2008).

Contig sequences were submitted to the non-redundant protein database NCBI through Blast X (Altschul *et*

al., 1997), to search for similarity to known proteins. In addition, sequences were analyzed by the software AutoFact (Koski *et al.*, 2005), which is an automated annotation tool that assigns biological information for a given sequence by comparing different databases. We used the UniRef90 e UniRef100, KEGG (Kanehisa and Goto, 2000), Pfam (Finn *et al.*, 2010), Smart (Schultz *et al.*, 1998) databases. In order to establish the GO (Gene Ontology) terms (Ashburner *et al.*, 2000) we employed the Blast2Go program to classify the sequences according to the molecular function and biological process described to the respective protein (Götz *et al.*, 2008; Carbon *et al.*, 2009).

Results

Plants grown hydroponically were exposed to a short-term water deficit and the physiological parameters A , g_s , c_i , E , T_i , Troom and WUE are presented in Figure 2.

We employed the Subtractive Suppressive Hybridization (SSH) technique to obtain an enriched collection of ESTs expressed early in response to water deficit stress. cDNA libraries were constructed to detect genes up-regulated when compared to control plants. After sequencing and base calling, reads were mapped on the soybean genome and assembled into contigs. Approximately 65% of the total reads produced for each library was mapped on the genome. Contigs assembled were searched in the NCBI database in order to access the information about proteins. The data from each sequencing run are shown in Table 1. After the AutoFact analysis, the unknown/hypothetical genes decreased significantly. All genes identified by using the subtractive libraries are stored in the Soybean Database.

Biochemical roles of the proteins were determined by using the Biological Process level 3 from Gene Ontology (Figure 3). Approximately 71% of the sequences in BR 16 and 61% of Embrapa 48 showed an associated putative biological or molecular function. Consequently, the number of no match and unknown genes was larger in Embrapa 48 than in BR 16 plants. In BR 16, we observed 121 and 452 different roles for the genes expressed in the roots and leaves, respectively. In Embrapa 48, there were 251 and 128 biological processes, respectively. Categories employed in functional classification include the following: Biological regulation (gene expression regulation, protein modification regulation); Signaling (kinases, phosphatases, secondary messengers); Response to stress (water deficit, osmotic and salt stress, heat, cold); Response to stimulus (chemical stimulus, endogenous stimulus); Response to biotic stimulus (pathogen, fungus, insect); Cell organization (cytoskeleton components, cell wall organization, cellular component assembly); Cell cycle (actin filament-based process, cell division); Senescence (aging, cell death); Other biological process (cellular component organization, cell recognition, developmental process); Biosynthetic process (carbohydrate, fatty acid, hormone, small and macromolecules, pigment, nucleotide); Catabolic process (pro-

tein, fatty acid, cell wall, polysaccharide); Nitrogen compound metabolism (glutamate and glutamine biosynthetic process, cellular nitrogen compound metabolic process); Oxidation reduction (electron transport, lipid oxidation) and Secondary metabolism (terpenoid, phytochelatin, phytoalexin and phenylpropanoid metabolic process). Each category was normalized by the total number of sequences observed in the respective library, which enables comparison of the results during the time course of the water deficit (Figure 3).

Applying the Venn-Master Diagram, we also searched for genes expressed in more than one library as described in Figure 4. BR 16 root tissue genes were categorized as involved in the electron transport process (R1 and R2 libraries) or as encoding proteins related to the stress response class (R2-R3 library) (Figure 4A). In BR 16 leaves (L1-L2 libraries), we also detected genes involved in signaling reactions or involved in the transport of metabolites and electrons (oxidation reduction process) (Figure 4B). In addition, the genes in the L2-L3 libraries were also related to stress response (Figure 4B). A general view of L1-L2-L3

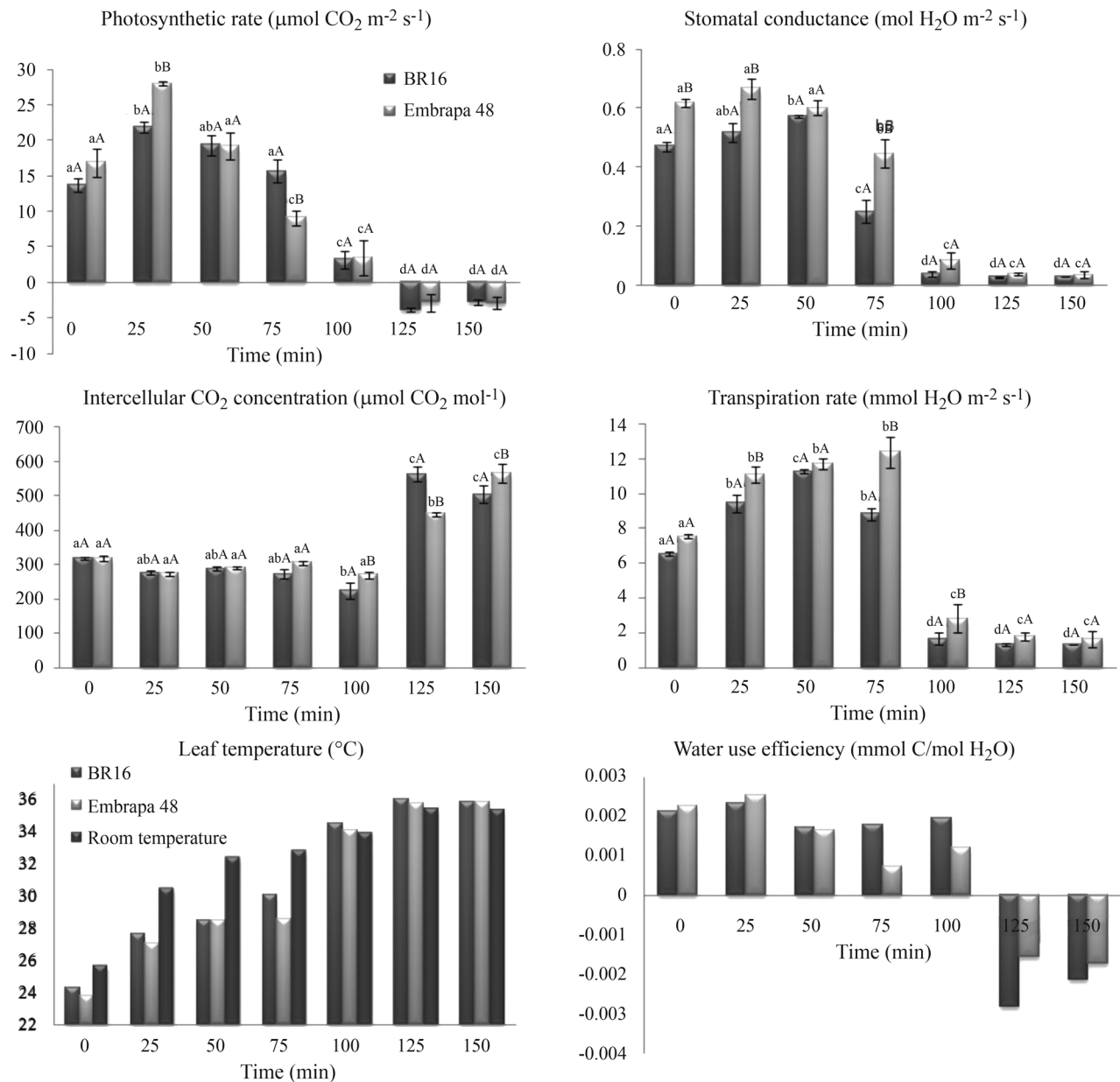


Figure 2 - Physiological parameters measured on drought-stressed plants. Photosynthetic rate ($\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$), stomatal conductance ($\text{mol H}_2\text{O m}^{-2} \text{ s}^{-1}$), internal CO_2 concentration ($\mu\text{mol CO}_2 \text{ mol}^{-1}$), transpiration rate ($\text{mmol H}_2\text{O m}^{-2} \text{ s}^{-1}$), and leaf temperature ($^{\circ}\text{C}$) were measured with a LiCor Portable Photosynthesis System. The water use efficiency ($\text{mmol C/mol H}_2\text{O}$) was determined based on transpiration and photosynthetic rate. From each stress treatment, two plants were evaluated in each block. Results represent the mean value of all plants analyzed. The (\pm) standard error bars are shown.

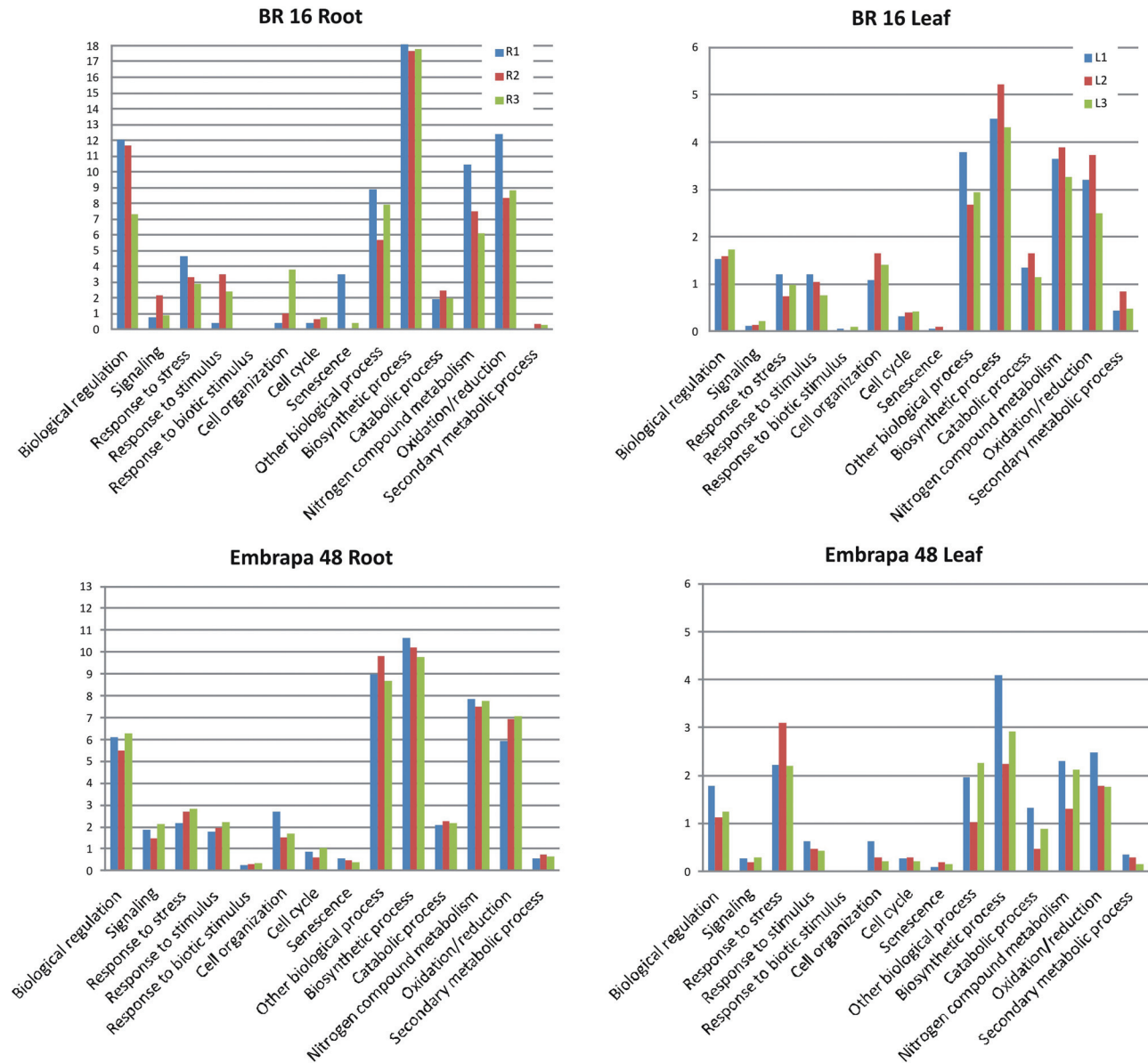


Figure 3 - Functional classification. The expressed genes were distributed in categories according to the Biological Process (level 3) of the Gene Ontology database. Categories were normalized through the ratio between the number of genes present in each category by the number of genes identified in the library. The Y axis represents the percentage (%) of genes identified.

libraries confirmed that genes functionally related to transport and stress responses were also expressed in leaves over the water deficit period (Figure 4B). It is important to emphasize that transcripts observed in R1-R2, for example, were different from the transcripts found in R1-R2-R3, although similar biological processes could be detected in both libraries. Interestingly, in the Embrapa 48 cultivar, transcription factors and genes involved in signal transduction were the most abundant transcripts in both R1 and R2 root tissues (Figure 4C). Genes encoding transcription factors were also predominant in the R2-R3 libraries in addition to the ones related to transport process (electrons and molecules). The same pattern was found in R1-R2-R3 li-

braries (Figure 4C). Processes related to protein metabolism (assembly, folding and modification) were very active in L2-L3 libraries as well as genes involved in transport process, response to stress and signaling. Stress response genes stand for 22% of the genes identified in L1-L2-L3 libraries (Figure 4D).

We performed a comparison between genes expressed in roots and leaves in either cultivar using the Venn-Master Diagram tool and looking for the biological processes observed in each group. When the same contig or glyma was detected in more than one library it was counted as one transcript. Then, considering only the non-repeated genes from each group (tissue or cultivar), we were able to

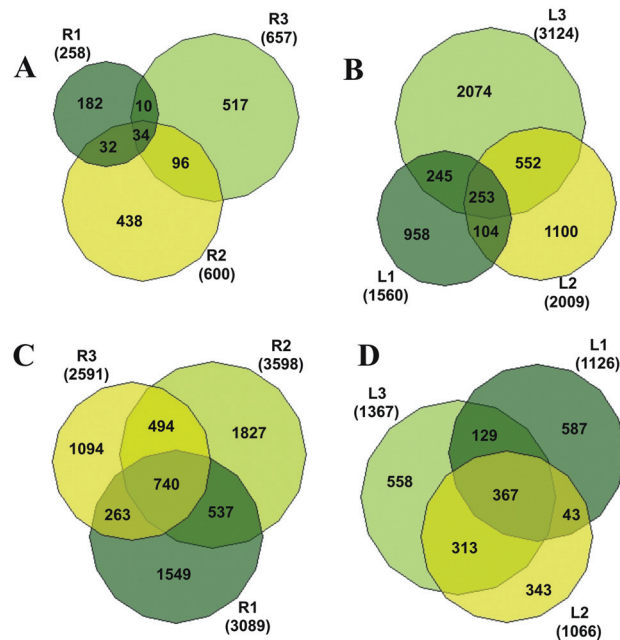


Figure 4 - Venn-Master diagram. Libraries were compared to each other to show the number of genes expressed only in one library and the ones common to different libraries. R(L)1, R(L)2 and R(L)3 represent genes expressed in 25-50, 75-100 and 125-150 min under stress. (A) BR 16 root; (B) BR 16 leaf; (C) Embrapa 48 root; (D) Embrapa 48 leaf.

observe genes expressed in root and leaves as well as genes expressed in Embrapa 48 and BR 16 plants (Figure 5A). To assess biological processes in roots and leaves of both cultivars, we highlighted the more abundant putative proteins and cellular components in each water-deficit stressed tissue (Figure 5B). According to our data, transcripts expressed by BR 16 plants in roots and leaves showed that the most frequent biological processes were similar in both tissues (Figure 5B). However, in Embrapa 48, the process of response to stress was predominant among transcripts from leaves and less common in roots (Figure 5B).

Among genes expressed by plants in response to water deficit, we observed that transcripts classified in the response to stress category presented similarity to proteins with responsiveness to biotic as well as to abiotic stress. These transcripts represent proteins with antimicrobial activity or proteins involved in defense and immune responses. Genes related to the response to virus, pathogenic bacterium, nematodes and fungi were also detected. Under abiotic stress, we identified transcripts in response to osmotic and oxidative stress, desiccation, heat, cold, water deprivation, hyperosmotic salinity and high light intensity. To assess which transcripts are expressed in response to other adverse conditions, the genes placed in the response to stress category ($n = 458$) were compared to transcripts from plants infected by *Phakopsora pachyrhizi*, *Uromyces appendiculatus*, virus and treated with *Bradirhizobium japonicum* (Figure 6).

Discussion

Plant water deficit responses and tolerance are complex biological processes that need to be analyzed at system-level using genomics and physiological approaches (Harb *et al.*, 2010). Studies aiming to detect differences among plants in response to stress using contrasting cultivars have been performed in some legumes (Türkan *et al.*, 2005). Empiric observations during growing seasons in the field are the first step to explore plant variability based on its yield under water deficit conditions. Oya *et al.* (2004) examined the physiological characteristics of the Embrapa 48 and BR 16 cultivars and determined that they were moderately tolerant and highly sensitive to drought, respectively. Here, we studied these cultivars at the molecular level, and our data demonstrated that the water deficit applied was effective as it triggered changes in the plant physiological behavior (Figure 2). As the stress increased in the plants, photosynthesis of both cultivars was affected and was strongly inhibited after 100 min. This inhibition was related to a drastic decrease in the stomatal conductance (g_s) and transpiration rate (E) of both cultivars. We also observed an increase in the leaf temperature (from 100 min) and internal concentration of CO_2 (from 125 min). The water deficit did not produce significant physiological differences between Embrapa 48 and BR 16 cultivars. This was expected as the short-term procedure to induce stress used here is usually not enough to produce effects on plant morphological and physiological characteristics needed to distinguish the two cultivars. The experimental treatment described here aimed to study the effects of the stress on the early plant responses at a molecular level. Additionally, the hydroponic system allowed the obtainment of a clean root sample, as demonstrated by Martins *et al.* (2008).

As described in the literature, subtractive libraries have been employed in a wide range of studies such as gene expression studies (Kojima *et al.*, 2009; González-Candelas *et al.*, 2010; Jain and Chattopadhyay, 2010; Soria-Guerra *et al.*, 2010). Most of these studies were based on the construction of a physical cDNA collection by cloning fragments into a vector and on the traditional big dye sequencing technology. However, *in vivo* cloning can be considered a laborious and expensive procedure and also has technical limitations. In order to overcome these limitations, both cultivar cDNAs were sequenced using two different chemistries based on Sequencing by Synthesis Technology (Illumina). As shown in Table 1, we identified many genes being regulated under water deficit, more than is often detected through subtractive libraries. Larger sets of genes were also detected in the characterization of stress-responsive genes to heat in wheat (Chauhan *et al.*, 2011) and by studying salt-induced genes of grapevine (Daldoul *et al.*, 2010). Both suppressive subtractive libraries were done through *in vivo* cloning and by the capillary sequencing method. Interestingly, while investigating invasive species, Prentis *et al.* (2010) found 12442 contigs in

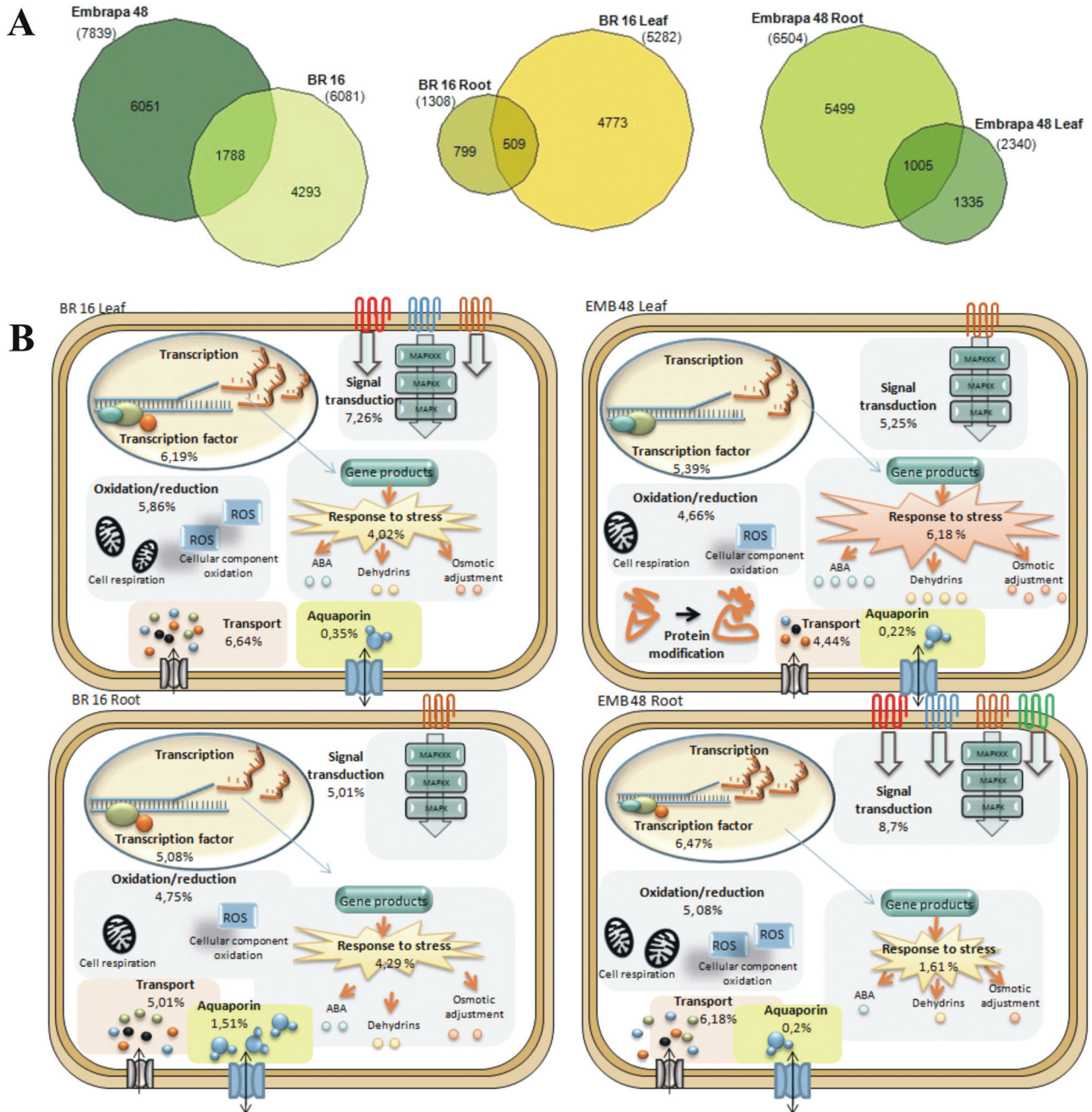


Figure 5 - Genes expressed in BR 16 and Embrapa 48 cultivars. The number of non-redundant genes identified for the cultivar/tissues is shown for each set. Subsets of genes that are shared between the sets under comparison are presented in the intersection (A). The most frequent biological processes observed for each plant and tissue are shown (B). The percentage (%) represents a set of transcripts normalized against the total number of genes identified in the respective tissue.

the transcriptome by combining subtractive libraries and massively parallel sequencing (pyrosequencing). One of the advantages of the next generation sequencing (NGS) over the traditional Sanger methodology is a more precise measurement of the transcript levels and isoforms, in addition to other revolutionary benefits (Marioni *et al.*, 2009; Wang *et al.*, 2009). The advantages and limitations found in the NGS and capillary sequencing have led to comparative studies of both methods (Cheung *et al.*, 2008; Hert *et al.*,

2008; Tedersoo *et al.*, 2010). Lower costs and quickness have made the NGS a powerful tool for functional genomics studies (Shendure and Ji, 2008; Harismendy *et al.*, 2009).

Under water deficit, plants can up- or down-regulate expression of responsive genes. Different biochemical roles have been suggested for these responsive genes, and their expression patterns make an intricate network that can be related to either water deficit acclimation or the cel-

Table 1 - Summary of the total reads sequenced and number of sequences assembled and mapped in each library. (A) BR 16 was sequenced by using a 36 bp single read in Genome Analyzer *Ile* whereas the run of (B) Embrapa 48 cultivar was carried out through a 76 bp single read in the sequencer Genome Analyzer *Ilz*.

Library	Reads	Contigs	No hits	Matched sequences	Known sequences	Unknown/hypothetical	GO annotation (%)
BR 16							
L1	1854641	1560	82	1479	1308	170	71
L2	519031	2009	51	1958	1677	281	76
L3	2035320	3124	132	2993	2616	376	71
R1	2486569	258	6	252	212	40	71
R2	2458847	600	34	566	499	67	68
R3	2428923	657	53	604	525	79	68
Embrapa 48							
L1	5731156	1126	66	1060	510	550	65
L2	5644473	1066	89	977	427	550	58
L3	5545375	1367	108	1259	562	697	59
R1	5144645	3089	198	2891	1365	1526	61
R2	5359395	3598	194	3404	1585	1819	62
R3	3095694	2591	161	2430	1174	1256	62

lular damage response (Shinozaki and Yamaguchi-Shinozaki, 2007). The expressed genes have been shown to belong to many classes of genes (Figure 3). However, the largest class of genes expressed under stress remains unknown, partly because there is no similarity in databases available to date, and partly due to the gene functions are not understood yet (Reddy *et al.*, 2008; Nelson and Shoemaker, 2006). Many genes expressed under water deficit conditions are present in the contrasting cultivars Embrapa 48 and BR 16 (Figure 5A). Among others, common processes affected by the stress are those related to transport, protein folding and assembling as well as protein degradation and ubiquitination. Some proteins are synthesized in requirement of defense pathways triggered by the stress whereas other proteins are degraded because its aggregation or denaturation events are predominant in water stress conditions. The transport category involves the translocation of ions, lipids, sugar and others molecules. Also included in this category are the genes for aquaporins, which are proteins that might be directly involved in water deficit acclimation by facilitating water movement across the membranes (Zhao *et al.*, 2008). We have identified several transcripts encoding the tonoplast and plasma membrane intrinsic proteins that were more frequently found in roots rather than in leaves, especially in the BR 16 plants (Figure 5B). Roles played by aquaporins in plants can be regulated in different complex ways, but the expression of aquaporins throughout the plant would most likely be regulated by water potential gradient (Zhao *et al.*, 2008). However, there are differences between the mRNA changes and its respective protein level. The changes expressed here by soybean plants in response to water stress are only at the transcriptional

level. BR 16 plants did not show an expressive quantitative difference in the biological processes affected by water deficit between roots and leaves (Figure 5B). Nonetheless, Embrapa 48 did show a significant contrast between both tissues (Figure 5B) regarding the genes included in the response to stress class (water deficit, osmotic and salt stress, heat and cold).

In abiotic stress responses, genes are induced to protect the plants against a disturbance in cell homeostasis. Under water stress, plants can become susceptible to other abiotic or biotic stresses because their normal metabolism is directed towards the acclimation to the first stress. Several studies have tried to make a connection between plant responses to water deficit and heat (Sakuma *et al.*, 2006; Montero-Barrientos *et al.*, 2010), cold (Li *et al.*, 2010), or reactive oxygen species enhanced due to drought (Kocsy *et al.*, 2005; Khanna-Chopra and Selote, 2007). In the response to stress category, we observed that various transcripts were similarly involved in several abiotic stresses such as osmotic stress, cold, heat and salinity as well as those involved in response to biotic stimulus (bacterium, fungus, nematode, virus) (Figure 6). From subtracted libraries of pigeonpea, Priyanka *et al.* (2010) identified a hybrid-proline-rich encoding gene whose multiple stress-responsive characteristics have been demonstrated under PEG, mannitol, salt and heat stress, cold and ABA-treatment. Another gene from pigeonpea, a cyclophilin gene, presented a stress-responsive nature in *Arabidopsis thaliana* (Sekhar *et al.*, 2010). A bZIP transcription factor gene found in soybean plants was also associated with multiple stress responses, being its expression induced by ABA-treatment, drought, high salt and low temperature (Gao *et al.* 2011).

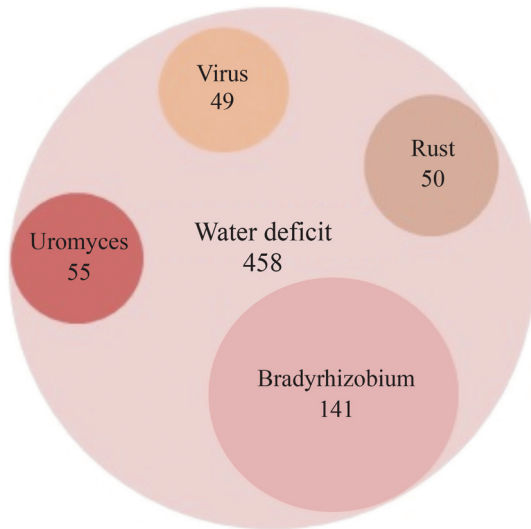


Figure 6 - Multiple stress genes. Genes identified in drought-stressed plants were compared to transcripts expressed when soybean plants were infected by *Phakopsora pachyrhizi*, *Uromyces appendiculatus*, virus and treated with *Bradyrhizobium japonicum*. Contig sequences were aligned and sequences common among experiments are indicated.

In conclusion, a hydroponic system was applied successfully to grow soybean plants and to produce clean roots in a short-term water deficit experiment. We described a large collection of cDNAs expressed by soybean plants under water deficit and their respective biochemical processes. Although the analyses presented here represent a first step in gene expression investigation, the results indicate changes which are triggered in roots and leaves of tolerant and sensitive soybean plants.

Acknowledgments

This work was sponsored by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Project # 552735/2007-8).

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25:25-29.

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

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, the AmiGO Hub and the Web Presence Working Group (2009) AmiGO: Online access to ontology and annotation data. *Bioinformatics* 25:288-289.

Carvalho MHCd (2008) Drought stress and reactive oxygen species. *Plant Signal Behav* 3:156-165.

Chauhan H, Khurana N, Tyagi AK, Khurana JP and Khurana P (2011) Identification and characterization of high temperature stress responsive genes in bread wheat (*Triticum aestivum* L.) and their regulation at various stages of development. *Plant Mol Biol* 75:35-51.

Cheung F, Win J, Lang JM, Hamilton J, Vuong H, Leach JE, Kamoun S, Lévesque CA, Tisserat N and Buell CR (2008) Analysis of the *Pythium ultimum* transcriptome using Sanger and Pyrosequencing approaches. *BMC Genomics* 9:e542.

Cocozza C, Cherubini P, Regier N, Saurer M, Frey B and Tognetti R (2010) Early effects of water deficit on two parental clones of *Populus nigra* grown under different environmental conditions. *Funct Plant Biol* 37:244-254.

Daldoul S, Guillaumie S, Reustle GM, Krczal G, Ghorbel A, Delrot S, Mliki A and Hofer MU (2010) Isolation and expression analysis of salt induced genes from contrasting grapevine (*Vitis vinifera* L.) cultivars. *Plant Sci* 179:489-498.

Diatchenko L, Lau Y-FC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, *et al.* (1996) Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* 93:6025-6030.

Ergen NZ and Budak H (2009) Sequencing over 13,000 expressed sequence tags from six subtractive cDNA libraries of wild and modern wheats following slow drought stress. *Plant Cell Environ* 32:220-236.

Fehr WR, Caviness CE, Burmood DT and Pennington JS (1971) Stage of development description for soybeans [*Glycine max* (L.) Merrill]. *Crop Sci* 11:929-931.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res* 38:D211-D222.

Gao S-Q, Chen M, Xu Z-S, Zhao C-P, Li L, Xu H-j, Tang Y-m, Zhao X and Ma Y-Z (2011) The soybean GmbZIP1 transcription factor enhances multiple abiotic stress tolerances in transgenic plants. *Plant Mol Biol* 75:537-553.

González-Candelas L, Alamar S, Sánchez-Torres P, Zacarías L and Marcos JF (2010) A transcriptomic approach highlights induction of secondary metabolism in citrus fruit in response to *Penicillium digitatum* infection. *BMC Plant Biol* 10:e194.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J and Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420-3435.

Govind G, Thammegowda HV, Kalaiarasi PJ, Iyer DR, Muthappa SK, Nese S and Makarla UK (2009) Identification and functional validation of a unique set of drought induced genes preferentially expressed in response to gradual water stress in peanut. *Mol Genet Genomics* 281:591-605.

Harb A, Krishnan A, Ambavaram MMR and Pereira A (2010) Molecular and physiological analysis of drought stress in *Arabidopsis* reveals early responses leading to acclimation in plant growth. *Plant Physiol* 154:1254-1271.

Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, *et al.* (2009) Evaluation of next generation sequencing platforms

- for population targeted sequencing studies. *Genome Biol* 10:R32.
- Hernandez D, François P, Farinelli L, Østerås M and Schrenzel J (2008) *De novo* bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Res* 18:802-809.
- Hert DG, Fredlake CP and Barron AE (2008) Advantages and limitations of next generation sequencing technologies: A comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* 29:4618-4626.
- Hoagland DR and Arnon DI (1950) The water-culture method for growing for plants without soil. *Agric Exp Stn Circ* 347:1-32.
- Jain D and Chattopadhyay D (2010) Analysis of gene expression in response to water deficit of chickpea (*Cicer arietinum* L.) varieties differing in drought tolerance. *BMC Plant Biol* 10:e24.
- Kanehisa M and Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27-30.
- Khanna-Chopra R and Selote DS (2007) Acclimation to drought stress generates oxidative stress tolerance in drought-resistant than -susceptible wheat cultivar under field conditions. *Environ Exp Bot* 60:276-283.
- Kocsy G, Laurie R, Szalai G, Szilágyi V, Simon-Sarkadi L, Galiba G and de Ronde JA (2005) Genetic manipulation of proline levels affects antioxidants in soybean subjected to simultaneous drought and heat stresses. *Physiol Plant* 124:227-235.
- Kojima H, Hitomi Y, Numata T, Tanaka C, Imai K and Omokawa H (2009) Analysis of gene expression in rice root tips treated with R-1-a-methylbenzyl-3-p-tolylurea using PCR-based suppression subtractive hybridization. *Pestic Biochem Physiol* 93:58-64.
- Koski LB, Gray MW, Lang BF and Burger G (2005) AutoFACT: An automatic functional annotation and classification tool. *BMC Bioinformatics* 6:e151.
- Kron AP, Souza GM and Ribeiro RV (2008) Water deficiency at different developmental stages of *Glycine max* can improve drought tolerance. *Bragantia* 67:43-49.
- Li R, Li Y, Kristiansen K and Wang J (2008) SOAP: Short oligonucleotide alignment program. *Bioinformatics* 24:713-714.
- Li X-W, Feng Z-G, Yang H-M, Zhu X-P, Liu J and Yuan H-Y (2010) A novel cold-regulated gene from *Camellia sinensis*, *CsCOR1*, enhances salt- and dehydration-tolerance in tobacco. *Biochem Biophys Res Commun* 394:354-359.
- Marioni JC, Mason CE, Mane SM, Stephens M and Gilad Y (2009) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18:1509-1517.
- Martins PK, Jordão BQ, Yamanaka N, Farias JRB, Beneventi MA, Binneck E, Fuganti R, Stolf R and Nepomuceno 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* 32:512-521.
- Molina C, Rotter B, Horres R, Udupa SM, Besser B, Bellarmino L, Baum M, Matsumura H, Terauchi R, Kahl G, *et al.* (2008) SuperSAGE: The drought stress-responsive transcriptome of chickpea roots. *BMC Genomics* 9:e553.
- Montero-Barrientos M, Hermosa R, Cardoza RE, Gutiérrez S, Nicolás C and Monte E (2010) Transgenic expression of the *Trichoderma harzianum* hsp70 gene increases *Arabidopsis* resistance to heat and other abiotic stresses. *J Plant Physiol* 167:659-665.
- Nelson RT and Shoemaker R (2006) Identification and analysis of gene families from the duplicated genome of soybean using EST sequences. *BMC Genomics* 7:e204.
- Oya T, Nepomuceno AL, Neumaier N, Farias JRB, Tobita S and Ito O (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.
- Pedersen P, Kumudini S, Board J and Conley S (2005) Soybean growth and development. In: Dorrance AE, Draper MA and Hershman DE (eds) *Using Foliar Fungicides to Manage Soybean Rust*. Ohio State University, Columbus, pp 41-47.
- Prentis PJ, Woolfit M, Thomas-Hall SR, Ortiz-Barrientos D, Pavašovic A, Lowe AJ and Schenck PM (2010) Massively parallel sequencing and analysis of expressed sequence tags in a successful invasive plant. *Ann Bot* 106:1009-1017.
- Priyanka B, Sekhar K, Reddy VD and Rao KV (2010) Expression of pigeonpea hybrid-proline-rich protein encoding gene (CcHyPRP) in yeast and *Arabidopsis* affords multiple abiotic stress tolerance. *Plant Biotechnol J* 8:76-87.
- Reddy PCO, Sairanganayakulu G, Thippeswamy M, Reddy PS, Reddy MK and Sudhakar C (2008) Identification of stress-induced genes from the drought tolerant semi-arid legume crop horsegram (*Macrotyloma uniflorum* (Lam.) Verdc.) through analysis of subtracted expressed sequence tags. *Plant Sci* 175:372-384.
- Sakuma Y, Maruyama K, Qin F, Osakabe Y, Shinozaki K and Yamaguchi-Shinozaki K (2006) Dual function of an *Arabidopsis* transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc Natl Acad Sci USA* 103:18822-18827.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178-183.
- Schultz J, Milpetz F, Bork P and Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci USA* 95:5857-5864.
- Sekhar K, Priyanka B, Reddy VD and Rao KV (2010) Isolation and characterization of a pigeonpea cyclophilin (CcCYP) gene, and its over-expression in *Arabidopsis* confers multiple abiotic stress tolerance. *Plant Cell Environ* 33:1324-1338.
- Shendure J and Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135-1145.
- Shinozaki K and Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. *J Exp Bot* 58:221-227.
- Soria-Guerra RE, Rosales-Mendoza S, Chang S, Haudenschild JS, Zheng D, Rao SS, Hartman GL, Ghabrial SA and Korban SS (2010) Identifying differentially expressed genes in leaves of *Glycine tomentella* in the presence of the fungal pathogen *Phakopsora pachyrhizi*. *Planta* 232:1181-1189.
- Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G and Kõljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical

- mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytol* 188:291-301.
- Türkan I, Bor M, Özdemir F and Koca H (2005) Differential responses of lipid peroxidation and antioxidants in the leaves of drought-tolerant *P. acutifolius* Gray and drought-sensitive *P. vulgaris* L. subjected to polyethylene glycol mediated water stress. *Plant Sci* 168:223-231.
- Wang Z, Gerstein M and Snyder M (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57-63.
- Xu D, Wei J, Cui H, Gong J, Yan Y, Lai R and Qin Q (2010) Differential profiles of gene expression in grouper *Epinephelus coioides*, infected with Singapore grouper iridovirus, revealed by suppression subtractive hybridization and DNA microarray. *J Fish Biol* 77:341-360.
- Zhang JZ, Li ZM, Liu L, Mei L, Yao JL and Hu CG (2008) Identification of early-flower-related ESTs in an early-flowering mutant of trifoliolate orange (*Poncirus trifoliata*) by suppression subtractive hybridization and macroarray analysis. *Tree Physiol* 28:1449-1457.
- Zhang Y, Liu X, Zhang W and Han R (2010) Differential gene expression of the honey bees *Apis mellifera* and *A. cerana* induced by *Varroa destructor* infection. *J Insect Physiol* 56:1207-1218.
- Zhao C-X, Shao H-B and Chu L-Y (2008) Aquaporin structure-function relationships: Water flow through plant living cells. *Colloids and Surf B Biointerfaces* 62:163-172.

Internet Resources

- Blast2Go, http://www.blast2go.org/start_blast2go (October 11, 2011).
- Exact DE Novo Assembler, <http://www.genomic.ch/edena.php> (October 11, 2011).
- Gene Ontology database, <http://www.geneontology.org/> (October 11, 2011).
- Kegg database, <http://www.genome.jp/kegg/> (October 11, 2011).
- National Center Biotechnology Information, <http://www.ncbi.nlm.nih.gov> (October 11, 2011).
- Pfam database, <http://pfam.sanger.ac.uk/> (October 11, 2011).
- Smart database, <http://smart.embl-heidelberg.de/> (October 11, 2011).
- Short Oligonucleotide Alignment Program, <http://soap.genomics.org.cn/> (October 11, 2011).
- Soybean database, <http://bioinfo03.ibi.unicamp.br/soja/> (October 11, 2011).
- UniRef databases, <http://www.ebi.ac.uk/uniref/> (October 11, 2011).
- Venn Diagram, <http://bioinfogp.cnb.csic.es/tools/venny/index.html> (October 11, 2011).

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