



## Expression pattern of drought stress marker genes in soybean roots under two water deficit systems

Anna Cristina Neves-Borges<sup>1\*</sup>, Fábila Guimarães-Dias<sup>2\*</sup>, Fernanda Cruz<sup>2</sup>, Rosilene Oliveira Mesquita<sup>5</sup>, Alexandre Lima Nepomuceno<sup>3</sup>, Eduardo Romano<sup>4</sup>, Marcelo Ehlers Loureiro<sup>5</sup>, Maria de Fátima Grossi-de-Sá<sup>4</sup> and Márcio Alves-Ferreira<sup>2</sup>

<sup>1</sup>*Departamento de Botânica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil.*

<sup>2</sup>*Departamento de Genética, Instituto de Biologia, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brazil.*

<sup>3</sup>*Embrapa Soja, Distrito de Warta, Londrina, PR, Brazil.*

<sup>4</sup>*Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil.*

<sup>5</sup>*Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, MG, Brazil.*

### Abstract

The study of tolerance mechanisms for drought stress in soybean is fundamental to the understanding and development of tolerant varieties. Using *in silico* analysis, four marker genes involved in the classical ABA-dependent and ABA-independent pathways of drought response were identified in the *Glycine max* genome in the present work. The expression profiles of the marker genes *ERD1*-like, *GmaxRD20A*-like, *GmaxRD22*-like and *GmaxRD29B*-like were investigated by qPCR in root samples of drought sensitive and tolerant soybean cultivars (BR 16 and Embrapa 48, respectively), submitted to water deficit conditions in hydroponic and pot-based systems. Among the four putative soybean homologs to *Arabidopsis* genes investigated herein, only *GmaxRD29B*-like was not regulated by water deficit stress. Distinct expression profiles and different induction levels were observed among the genes, as well as between the two drought-inducing systems. Our results showed contrasting gene expression responses for the *GmaxRD20A*-like and *GmaxRD22*-like genes. *GmaxRD20A*-like was highly induced by continuous drought acclimating conditions, whereas *GmaxRD22*-like responses decreased after abrupt water deprivation. *GmaxERD1*-like showed a different expression profile for the cultivars in each system. Conversely, *GmaxRD20A*-like and *GmaxRD22*-like genes exhibited similar expression levels in tolerant plants in both systems.

*Key words:* water deficit, drought adaptation, marker genes, *Glycine max*.

### Introduction

For soybean (*Glycine max* [L.] Merrill), one of the most important agricultural commodities in the world (Clemente and Cahoon, 2009), drought is considered as one of the main causes of yield loss in different countries (Bray *et al.*, 2000). In Brazil, the lack of rainfall in 2009 caused a drop of 4.2% in soybean crop production (Brazilian Institute of Geography and Statistics – IBGE, 2010). The losses in soybean yield in the North and West of the state of Paraná (Brazil) related to drought were 80% in 2008-2009, as compared to the average productivity of the region.

These losses due to drought resulted in a cumulative decline of almost 11 million tons in total production (Franchini *et al.*, 2009). According to Manavalan *et al.* (2009), the drop in production is probably intensified by climate change caused by global warming and the consequent increase in frequency and extension of the water-limited regions. Therefore, comprehension of drought stress response mechanisms is expected to drive the development of drought tolerant soybean cultivars that will be crucial to maintaining soybean yield levels and countering the threat of global warming for this crop (Cutforth *et al.*, 2007).

It has been demonstrated that plants can naturally develop drought tolerance mechanisms, allowing them to prevent or minimize the damaging effects of water deprivation. These response mechanisms involve molecular, cellular and physiological changes, triggered by a molecular signaling cascade (Bray, 1993; Seki *et al.*, 2003; Yamaguchi-Shinozaki and Shinozaki, 2006; Manavalan *et al.*,

Send correspondence to Márcio Alves-Ferreira. Laboratório de Genética Molecular Vegetal, Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Av. Prof. Rodolpho Paulo Rocco s/n, - Prédio do CCS, 2º andar, sala A2-93, 21941-970 Rio de Janeiro, RJ, Brazil. E-mail: marcioaf@ufrj.br.

\*These authors contributed equally to this work.

2009). This complex network of responses to drought stress may involve the abscisic acid (ABA) phytohormone, which orchestrates the production and accumulation of important molecules that trigger and amplify a signaling cascade (Mahajan and Tuteja, 2005; Adie *et al.*, 2007; Urano *et al.*, 2009).

The response to drought stress relies mainly on gene expression regulation, of thousands of genes (Shinozaki and Yamaguchi-Shinozaki, 2007). In *Arabidopsis*, this complex cascade involves the expression of transcription factors that are responsible for regulating downstream genes such as *RD20A*, *RD22* and *RD29B* in the ABA-dependent pathway, as well as *ERD1* in the ABA-independent pathway (Shinozaki and Yamaguchi-Shinozaki, 2007). These four genes, *RD20A*, *RD22*, *RD29B* and *ERD1*, have been widely used as water-deficit markers in *Arabidopsis* and other species, hence, their identification in soybean will be of importance for future research in this crop (Pellegrineschi *et al.*, 2004; Huang *et al.*, 2011; Stolf-Moreira *et al.*, 2011).

Expression profiles of the drought-stress response genes have been investigated under water-deficit stress conditions established in the hydroponic systems (HSys) (Fujita *et al.*, 2004; Ito *et al.*, 2006; Stolf-Moreira *et al.*, 2010a,b, 2011). Notwithstanding, some authors prefer to assess drought responses in pot-based systems (PSys) because they better reproduce field conditions (Casagrande *et al.*, 2001; Qin *et al.*, 2007; Tran *et al.*, 2009). In the PSys condition, drought promotes a slower water deprivation process, allowing the plant to adapt to the water deficit. Conversely, in the HSys condition, the water deficit occurs abruptly by removing the plant from the nutrient solution or by the addition of osmolytes to the solution (Cowan, 1965). The instantaneous water deficit in HSys causes severe consequences to the soybean plants, impeding a continuous acclimation process (Munns *et al.*, 2010). Therefore, the water-deficit stress responses to the PSys and HSys conditions may differ significantly, triggering expression of distinct sets of genes (Martins *et al.*, 2008; Munns *et al.*, 2010).

By means of an *in silico* approach we identified herein the soybean (*G. max*) homologs of the *Arabidopsis* *ERD1*, *RD20A*, *RD22* and *RD29B* genes. These genes are classical markers for the ABA-dependent and ABA-independent pathways of response to drought (Pellegrineschi *et al.*, 2004; Huang *et al.*, 2011). The expression profile for each soybean gene under water deficit stress condition and contrasting the PSys and HSys conditions was assayed by qPCR analysis. The identification and molecular characterization of these drought marker genes (DMGs) under distinct water-deficit stress systems are important not only to reveal differences between the two experimental procedures, but also to provide *bona fide* marker genes for those interested in studying drought stress in soybean.

## Material and Methods

### Identification of DMGs in soybean response to drought

For the identification of DMGs in the response of soybean to drought we employed a search strategy illustrated in Figure S1 (Supplementary Material). DMGs involved in the response to drought in *Arabidopsis* were identified based on published data (Bray, 2002; Kang *et al.*, 2002; Fujita *et al.*, 2005; Shinozaki and Yamaguchi-Shinozaki, 2007) and searches using web tools of the Arabidopsis Information Resource (TAIR). The protein sequences of *Arabidopsis* genes (*ERD1*, *RD20A*, *RD22* and *RD29B*) and their putative paralogs were used to search all possible homologs in the *G. max* and *Oryza sativa* genomes by means of the BLASTP tool. Those meeting the criterion of an E-value  $\leq 10^{-18}$  in the Phytozome and TAIR sites were considered for further investigation. For constructing dendrogram we first performed a multiple alignment of the amino acid sequences for each selected gene using ClustalW2 software (Larkin *et al.*, 2007). These data were then used to build dendrograms using MEGA version 4.0 (Molecular Evolutionary Genetics Analysis) program (Tamura *et al.*, 2007) with the Neighbor-Joining clustering method (Crandall *et al.*, 2008) and data derived from a p-distance matrix (Crandall *et al.*, 2008) in a Poisson model and using the complete deletion option. Tree reconstruction was performed using the Interior Branch Test Phylogeny approach and bootstrapping (1,000 replications) (Sitnikova *et al.*, 1995). For rooting the trees, *O. sativa* sequences were used as the outgroup.

### *In silico* expression analysis

In order to investigate the pattern of induction/repression of the *Arabidopsis*, *ERD1*, *RD20A*, *RD22* and *RD29B* genes, the expression data of the *Arabidopsis* genes during the response to different water privation conditions or ABA stimulus were retrieved from the Genevestigator database (Hruz *et al.*, 2008). Data were presented as absolute expression values or fold change compared with that of the control samples by integrating expression data from thousands of transcriptomic experiments present in the *Genevestigator* database.

### Plant material and drought assays

We used the soybean (*G. max*) cultivars, BR 16 and Embrapa 48, these being sensitive and tolerant to drought, respectively (Casagrande *et al.*, 2001; Texeira *et al.*, 2008). Drought assays were performed in two different water deficit treatments, a pot-based (PSys) and a hydroponic system (HSys) (Martins *et al.*, 2008; Kulcheski *et al.*, 2010).

Plants grown in the PSys condition were maintained at a controlled temperature ( $30 \text{ }^\circ\text{C} \pm 5 \text{ }^\circ\text{C}$ ),  $60\% \pm 20\%$  relative humidity and natural photoperiod. Seeds from both genotypes were germinated in washed sand. After approxi-

mately 10 days, seedlings of each genotype were transplanted to pots containing bovine fertilizer. Plants at the V4 developmental stage (fourth trifoliate leaf fully expanded) (Fehr and Caviness, 1977) were submitted to irrigation (control) or water deficit conditions, by suspension of irrigation for 7-10 days, until reaching a water potential of  $-1.5 \text{ MPa} \pm 0.2 \text{ MPa}$  and  $-3.0 \text{ MPa} \pm 0.2 \text{ MPa}$  (moderate and severe level stress, respectively). The water potential ( $\Psi_w$ ) of each plant was measured at predawn (between 05:00 and 06:00) in the fourth or fifth leaf from the apex using a Scholander-type pressure chamber. Roots were removed from the pot and immediately rinsed with water for 1 min with gentle agitation to remove adhering sand. Biological contaminants were removed by immersion in SDS solution (2%) for 1 min followed by a gentle wash in ultrapure water for 1 min. Finally, the root samples were immediately frozen in liquid nitrogen and stored at  $-80 \text{ }^\circ\text{C}$  for RNA extraction. Two biological replicates for each condition were collected for gene expression studies.

In the HSys condition, the seeds of both cultivars were placed on moist filter paper and pre-germinated in the dark at  $25 \text{ }^\circ\text{C} \pm 1 \text{ }^\circ\text{C}$  and  $65\% \pm 5\%$  relative humidity. Seedling were transferred to polystyrene supports and their roots maintained completely immersed in a nutrient solution (pH balanced at 6.6 and aerated) (Kulcheski *et al.*, 2010), under natural photoperiod daylight (photosynthetic photon flux density (PPFD) =  $1.5 \times 10^3 \text{ } \mu\text{moles m}^{-2} \text{ s}^{-1}$ , equivalent to  $8.93 \times 10^4 \text{ lux}$ ) and a 12 h daylength, at  $25 \text{ }^\circ\text{C} \pm 2 \text{ }^\circ\text{C}$  and  $60\% \pm 5\%$  relative humidity. After 2 weeks, the seedlings at the V4 developmental stage of both genotypes were removed from the HSys condition and kept in the dark without nutrient solution or water for different water deprivation periods: 0 min (T0, control), 50 min (T50), 100 min (T100) and 150 min (T150) of stress. To verify the water deficit, photosynthetic rate, stomatal conductance, intercellular  $\text{CO}_2$  concentration, transpiration rate and leaf temperature were evaluated using a LI-6400 Portable Photosynthesis System (LiCor, Inc.). Measurements were taken on the fully expanded middle leaflet of the basal second leaf node under a photon flux density of  $1,000 \text{ } \mu\text{mol m}^{-2} \text{ s}^{-1}$ . For details, see Figure 2 of Rodrigues *et al.* (2012). Two biological replicates for each condition were used in the expres-

sion studies. The root samples corresponding to a pool composed of 10 plants from each treatment were immediately frozen in liquid nitrogen, followed by storage at  $-80 \text{ }^\circ\text{C}$  for posterior RNA extraction.

### Total RNA isolation

Root samples from the PSys condition were processed for RNA extraction using the Plant RNAeasy kit (Qiagen) following the manufacturer's protocol. RNA extraction of the root samples from the hydroponic experiments was done with Trizol Reagent (Invitrogen). In both cases, the RNA samples were treated with RNase-free DNase I (BioLabs) to eliminate any DNA contamination. RNA integrity was checked by electrophoresis on a 1% agarose gel, and RNA concentration and purity were determined using a NanoDrop spectrophotometer ND-1000 (Thermo Scientific).

### Real-time quantitative polymerase chain reaction (qPCR)

The expression of DMGs in both soybean cultivars submitted to different drought conditions was evaluated by qPCR analysis. Primers pairs of 20 bp and  $T_m$  of  $60 \text{ }^\circ\text{C} \pm 1 \text{ }^\circ\text{C}$  designed with Primer 3 plus software (Untergasser *et al.*, 2007) were used to amplify a region of 80-200 bp of the respective target gene. For normalization of target gene expression, *ACT11* (cytoskeletal structural protein) and *FBOX* (F-Box protein family) reference genes (RGs) were used (Kulcheski *et al.*, 2010). The stability of expression of each of these RGs under the experimental conditions was checked by means of the NormFinder program (Andersen *et al.*, 2004) (data not shown). All primer sequences and amplicon lengths are listed in Table 1.

qPCR assays were carried out in a Realplex 4 Eppendorf Mastercycler Eppgradient (Eppendorf) sequence detection system using a Power SYBR® Green RNA-to-Ct One-Step Kit (Applied Biosystems) following the manufacturer's protocol. For each sample, 25 ng of RNA was used in the reaction mixtures in a final volume of 20  $\mu\text{L}$ . For each primer combination, all samples were evaluated in technical triplicates and including a no-template control. Reaction mixtures were incubated for 30 min at  $48 \text{ }^\circ\text{C}$ , followed by 10 min at  $95 \text{ }^\circ\text{C}$  and 40 amplification cycles of

**Table 1** - Primer sequences used in the qPCR protocols and amplicon lengths.

Gene model	Soybean gene	Forward primer sequence [5'-3']	Reverse primer sequence [5'-3']	Amplicon length (bp)
<i>RD20A</i>	<i>Glyma03g41030</i> ( <i>GmaxRD20A</i> -like)	GTGGCACATGACTGAAGGAA	ATCTTCCAGCAGCACCTCT	195
<i>RD22</i>	<i>Glyma14g20450</i> ( <i>GmaxRD22</i> -like)	AATGCCGAAAGCCATTACAG	GCTTTGTTTTCCCTGCGTTA	110
<i>RD29B</i>	<i>Glyma16g31330</i> ( <i>GmaxRD29B</i> -like)	AGCTGACAAAGCCATCACTG	CTCTGTCAAGGACTGAGCAA	88
<i>ERD1</i>	<i>Glyma04g38050</i> ( <i>GmaxERD1</i> -like)	CGTCCAGAATTGCTCAACAG	TGGGGTTATAGCCTTGTGG	184
<i>ACT11</i>	<i>Glyma18g52780</i>	CGGTGGTTCTATCTTGGCATC	GTCTTTCGCTTCAATAACCCTA	142
<i>FBOX</i>	<i>Glyma12g05510</i>	CTAATGGCAATTGCAGCTCTC	AGATAGGGAAATGGTGCAGGT	93

15 s at 95 °C and 1 min at 60 °C (fluorescence measurement step). At the end of 40 cycles, a melting-curve analysis was run (15 s at 95 °C, 15 s at 60 °C – fluorescence measurement step, and 15 s at 95 °C). Melting curve and 1% gel electrophoresis analysis of the amplification products were employed to confirm the presence of only a single amplified product of expected size. Primer set efficiencies were estimated for each experimental set by using the Miner software (Zhao and Fernald, 2005) through a nonlinear regression algorithm without the need for a standard curve, and the values were used in all subsequent analyses. In addition, the values of the threshold cycle (quantification cycle value – C<sub>q</sub>) were converted by the program QBASE v1.3.5 (Hellemans *et al.*, 2007), into normalized relative quantities (NRQ).

### Promoter analysis

The promoter sequences of the DMG (1000 bp length upstream from the start codon) were obtained from the genome browser tool in the Phytozome database. *Cis*-regulatory elements responsive to drought stress, salinity stress, osmotic stress and ABA (Table S1) were identified through a web tool in the database of Plant *Cis* Program-acting Regulatory DNA Elements – PLACE (Higo *et al.*, 1999) and also from published data (Busk and Pages, 1998; Li and Chen, 1999; Choi *et al.*, 2000; Simpson *et al.*, 2003; Nakashima *et al.*, 2006; Lenka *et al.*, 2009; Mochida *et al.*, 2009; Umezawa *et al.*, 2010).

The frequency of the respective *cis*-regulatory elements in the promoter region of each gene of interest was compared to the expected frequency in the genes of the *G. max* genome. The statistical analysis of *cis*-elements of the gene of interest promoters was performed by the POBO web tool (Kankainen and Holm, 2004).

## Results

### Identification and *in silico* characterization of drought marker genes from ABA-dependent and ABA-independent pathways involved in the drought-stress response in soybean

In order to identify and characterize Drought Marker Gene (DMG) homologs for *ERD1*, *RD20A*, *RD22* and *RD29B* in soybean we initially used an *in silico* approach, followed by qPCR validation. We also evaluated the promoter region of these genes for the presence and frequency of *cis*-elements related to drought stress. The diagram of the search strategy employed is illustrated in Figure S1 (Supplementary Material).

To assess the expression pattern of the *ERD1*, *RD20A*, *RD22* and *RD29B* genes under different water-deficit stress conditions and ABA stimulus in *Arabidopsis*, we used the Genevestigator web tool (Hruz *et al.*, 2008). The digital expression analyses confirmed previous results showing that *RD20A*, *RD22* and *RD29B* are induced by

drought stress and ABA, whereas *ERD1* is induced predominantly by drought stress (Figure S2).

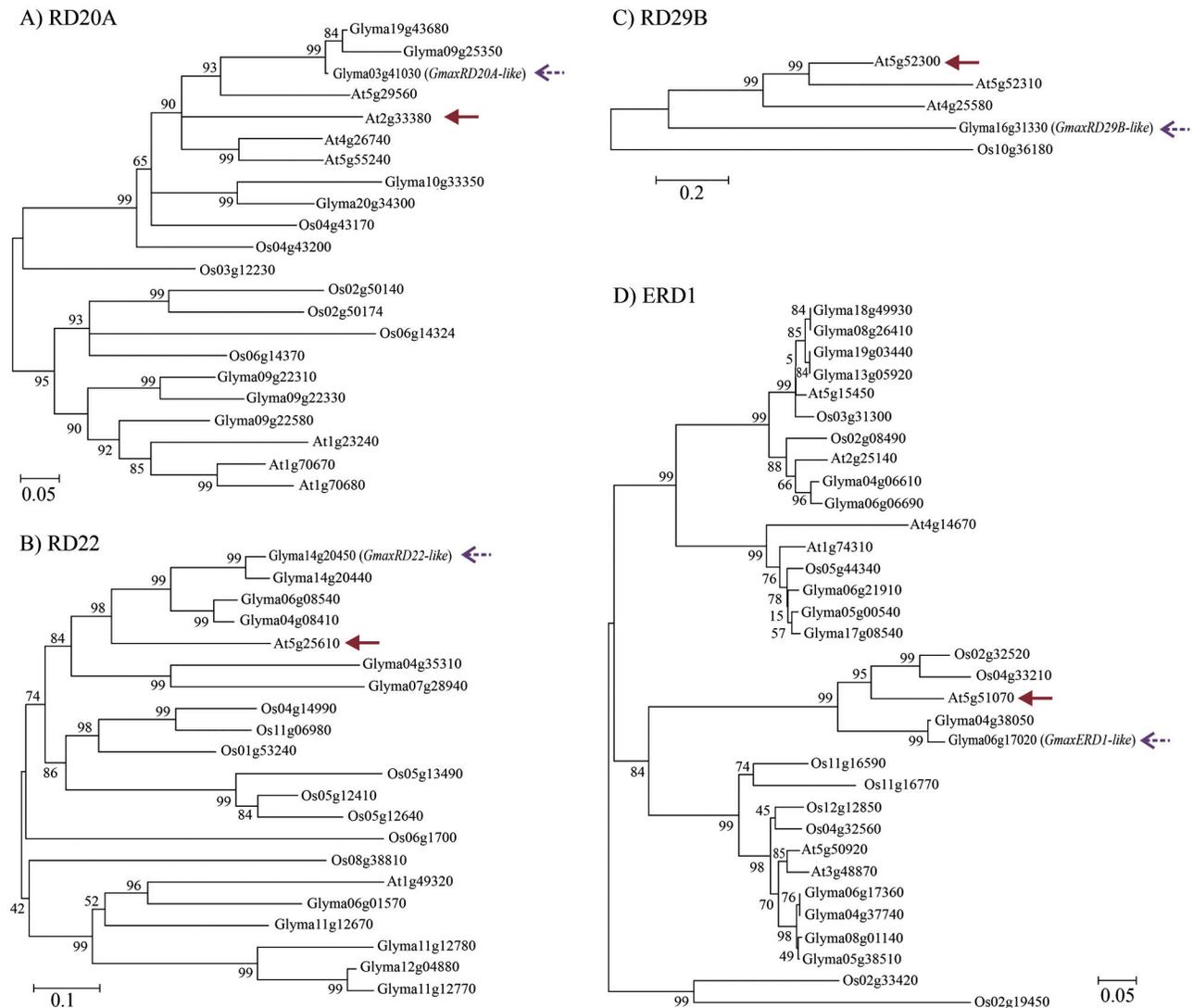
The *Arabidopsis* gene models, as well as their respective amino acid sequences and functions, were crucial for the search for putative homologs in the soybean genome. The putative homologs for each *Arabidopsis* gene in the soybean genome were identified through a BLASTP search in the Phytozome database combined with a Neighbor-joining analysis. For each *Arabidopsis* gene under consideration we identified the putative homologs in the *G. max* and *O. sativa* genomes. The threshold *e*-value used for the identification of the putative homologs and their use in a dendrogram analysis was determined according to the size of the gene families evaluated. For instance, *ERD1* belongs to a large gene family, indicating the use of an *e*-value threshold of  $\leq 10^{-50}$ . In contrast, for gene families with only few members, such as the *RD20A* or *RD22* protein families, the *e*-value threshold was set at  $\leq 10^{-30}$ . Finally, the *RD29B* gene presents only a few putative homologs with very low similarity, hence an *e*-value threshold of  $\leq 10^{-18}$  was indicated.

The dendrogram analysis was performed on the selected protein sequences to uncover putative soybean homologs for the *Arabidopsis* DMGs (Figure 1). When the dendrogram analysis could not pinpoint a single putative homolog, the top *e*-value obtained in the BLASTP analysis was considered as selection criteria. This strategy allowed us to identify the four putative soybean homolog genes, *GmaxERD1*-like (Glyma04g38050), *GmaxRD20A*-like (Glyma03g41030), *GmaxRD22*-like (Glyma14g20450) and *GmaxRD29B*-like (Glyma16g31330), for the *Arabidopsis* DMGs *ERD1* (AT5G51070), *RD20A* (AT2G33380), *RD22* (AT5G25610) and *RD29B* (AT5G52300), respectively (Figure 1). As previously mentioned, the *RD20A*, *RD22* and *RD29B* genes belong to the ABA-dependent pathways, while the *ERD1* gene belongs to the ABA-independent pathways of drought stress response.

### Expression profile analysis of putative soybean DMGs by real-time quantitative polymerase chain reaction (qPCR)

The expression pattern of each soybean DMG involved in the drought response was investigated by qPCR. The analysis was performed on RNA samples from roots of two soybean cultivars that presented contrasting responses to drought, the cultivar BR 16 being highly sensitive and the cultivar Embrapa 48 moderately tolerant to drought (Casagrande *et al.*, 2001; Texeira *et al.*, 2008). These plants had been submitted to different water deficit conditions (PSys and HSys) and also several stress levels.

Among the four DMGs identified by our analysis, *GmaxRD29B*-like was not induced under any of the test conditions (data not shown), whereas *Gmaxerd1*-like, *GmaxRD20A*-like and *GmaxRD22*-like were induced in both cultivars in the two cultivation conditions, PSys and



**Figure 1** - Dendrograms of the *Arabidopsis thaliana*, *Glycine max* and *Oryza sativa* key genes responsive to drought, based on the amino acid sequences. The multiple alignment was made using ClustalW2 and the dendrograms were built using MEGA4.0 software, with the Neighbor-Joining and pair-wise deletion options as a consensus of 1,000 bootstrap replicates. *O. sativa* was used as outgroup. The solid brown arrows indicate the *Arabidopsis* reference key genes and the dotted purple arrows indicate the respective soybean homologs selected for qPCR validation of expression. A) Dendrogram of the RD20A family of *A. thaliana*, *G. max* and *O. sativa*; B) Dendrogram of the RD22 family of *A. thaliana*, *G. max* and *O. sativa*; C) Dendrogram of the RD29B family of *A. thaliana*, *G. max* and *O. sativa*; D) Dendrogram of the ERD1 family of *A. thaliana*, *G. max* and *O. sativa*.

HSys. Nevertheless, these genes showed distinct expression dynamics and induction levels depending on drought conditions assessed.

The *GmaxRD20A*-like gene was induced progressively in both soybean cultivars under P Sys. It reached the highest expression level in the sensitive cultivar under the most severe stress condition ( $\Psi_w$  -3.0 MPa) in the P Sys (Figure 2A). In contrast, the gene expression level under HSys was very low for all tested conditions (Figure 2A).

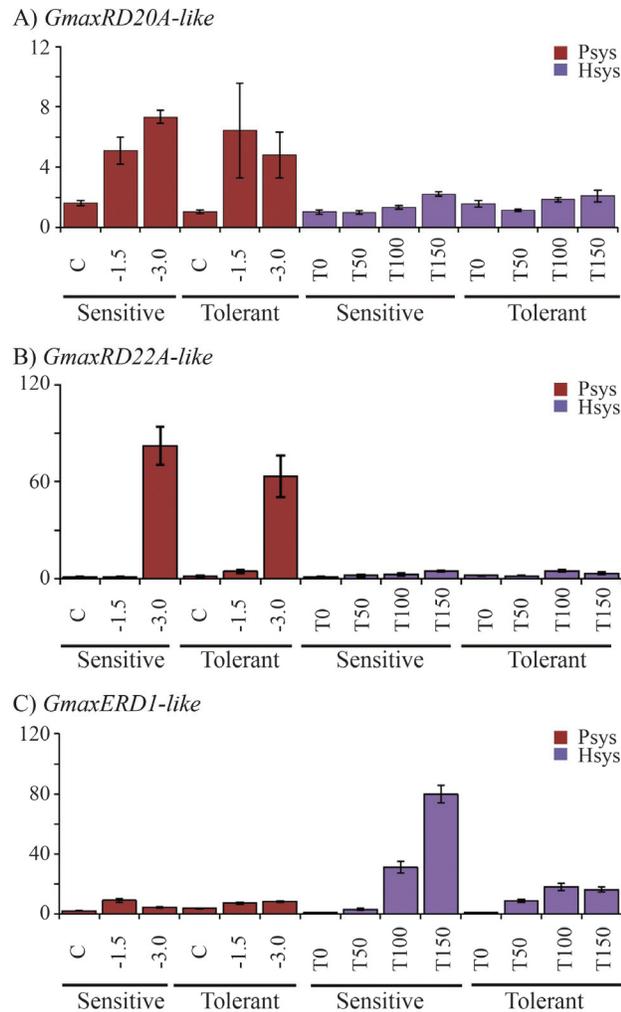
The *GmaxRD22*-like gene showed quite similar expression levels and dynamics for the two cultivars in the P Sys condition, with high expression restricted to the severe water deficit condition,  $\Psi_w$  -3.0 MPa (Figure 2B). In HSys, *GmaxRD22*-like presented very low expression lev-

els in both tolerant and sensitive cultivars. Interestingly, the two genes that belong to the ABA-dependent pathway, *GmaxRD22*-like and *GmaxRD20A*-like showed similar expression profiles in the test conditions.

In contrast to the other genes evaluated in this work, the *GmaxERD1*-like, which belongs to the ABA-independent pathway, was highly expressed in the HSys condition. It is worthy of note that this gene was about five times more expressed in the sensitive cultivar than in the tolerant one under the most severe stress condition, T150 (Figure 2C).

#### Promoter analysis

We employed the POBO program (Kankainen and Holm, 2004) to evaluate the enrichment of 17 *cis*-elements



**Figure 2** - Expression profile analyses for the soybean genes *GmaxRD20A*-like (A), *GmaxRD22A*-like (B) and *GmaxERD1*-like (C) in the roots of sensitive and tolerant cultivars during water deficit stress in pot-based (Psys) and hydroponic (Hsys) systems. Marker genes responsive to soybean genotypes are differentially regulated in roots, during water deficit stress conditions established in Psys (C – control no stress; Ψw -1.5 MPa and Ψw -3.0 MPa) or in Hsys (T0, T50, T100; and T150 min after water privation, respectively). The relative expression values, represented on the Y-axis, were obtained by qPCR experiments and calculated using the  $2^{-Ct}$  method with *ACT* and *FBOX* as endogenous control genes for normalization. The qPCR assay for each gene was performed in triplicate for each of the two independent biological replicates used in the validation. Means and Standard errors of three technical replications are shown. The brown and purple bars indicate the pot-based and hydroponic systems, respectively.

related to drought, ABA, or osmotic stress, such as ABRE, DRE, MYB, MYC or NAC *cis*-elements (Table S1), that we had found previously in the PLACE data bank (Higo *et al.*, 1999). The statistical significance of their enrichment in the promoters of *GmaxRD20A*-like, *GmaxRD22*-like and *GmaxERD1*-like was also evaluated. *GmaxRD29B*-like was excluded from this analysis because it was not responsive to the drought conditions evaluated in this work. To do so, we compared the frequency of the 17 motifs in a group

comprising these genes and in a background set (comprising all promoters regions of the *G. max* genome – BG model). This analysis revealed that three *cis*-elements (ACGT, WAACCA and CANNTG) are significantly enriched in the promoter region of the three genes analyzed when compared to the background set, while the ACCGAC motif is significantly enriched in the promoter region of the *GmaxRD20A*-like and *GmaxRD22*-like genes, also compared to the background set (Table S2). The most striking result was found for the ACGT motif, an element functionally that is important in a variety of promoters responding to ABA, among other stimuli (Guiltinan *et al.*, 1990). The analysis in POBO indicated that the ACGT motif is present up to 16 times in the promoter regions of the DMG genes, with an average of 5.33 compared to an average of 3.02 for the genome background (Table S2).

## Discussion

We herein identified and characterized the *G. max* putative homologs of *Arabidopsis* DMGs. The putative soybean homologs of the *Arabidopsis* genes *RD20A*, *RD22* and *ERD1* were induced during drought stress in the two experimental systems tested: the Psys and the Hsys conditions. These four genes have been employed as markers of water deficit in several works in *Arabidopsis* and other species (Pellegrineschi *et al.*, 2004; Huang *et al.*, 2011; Stolf-Moreira *et al.*, 2011).

In order to identify the soybean homologs for each *Arabidopsis* gene previously characterized we performed a dendrogram analysis for each gene family, using the genome information for *Arabidopsis*, rice and soybean. We could not find a *bona fide* soybean homolog for the *Arabidopsis RD29B* gene. Based on the dendrogram analysis we selected the soybean genes that showed the highest degree of similarity for each *Arabidopsis* DMG (Figure 1), aiming to uncover functional similarities. The selected genes had their expression evaluated by qPCR assays in two cultivars grown under two different water-deficit stress systems.

In Psys, the drought conditions should occur in a mode similar to field conditions, triggering genes involved in a continuous acclimation process. Since Psys can entail problems of heterogeneity and inconstant water potential, among other interferences, this has led researchers to use also Hsys in water-deficit stress experiments (Martins *et al.*, 2008; Munns *et al.*, 2010). Hsys allows for quick drought induction and straightforward sample collection (Martins *et al.*, 2008; Munns *et al.*, 2010). However, in Hsys, the water deficit stress is promoted suddenly by removing the plant from the nutrient solution, and it completely differs from field conditions, causing shock and injuries that do not allow a continuous acclimation process. After 100 min in Hsys without the nutrient solution, the water deficit is so severe that it leads to wilting in soybean plants (Martins *et al.*, 2008). Since there is no consensus

about the best system to promote water deficit, we decided to investigate the expression profiles of genes responsive to drought in both water-deficit stress induction systems. Therefore, besides providing the identification of soybean DMGs, our work also revealed discrepancies in gene expression patterns between the two systems.

The expression profiles of the putative soybean homologs *GmaxRD20A*-like, *GmaxRD22*-like and *GmaxERD1*-like were notably different in HSys when compared to PSys (Figure 2). *GmaxRD20A*-like and *GmaxRD22*-like putative homologs of *Arabidopsis* genes from the ABA-dependent pathway, were predominantly expressed in PSys, whereas *GmaxERD1*-like, a putative homolog of an *Arabidopsis* gene from the ABA-independent pathway, was mainly expressed in HSys (Figure 2C). The expression levels and profiles also varied among genes. For instance, *GmaxRD22*-like was highly expressed in both cultivars in PSys, and its expression levels in severe stress ( $\Psi_w$  -3.0 MPa) were nine to eleven times higher than in moderate stress ( $\Psi_w$  -1.5 MPa). In contrast, *GmaxRD20A*-like showed lower expression levels and only a minor difference in gene expression between severe and moderate stress. Considerable differences in DMG expression between the sensitive and tolerant soybean cultivars were exclusively found for *GmaxERD1*-like in plants grown under HSys. The difference was limited to the expression levels, but this may have biological significance. These results suggest that drought stress levels and pathway activation may vary considerably between the two systems investigated in this work. Martins *et al.* (2008) also observed differences in the expression profiles of *Axi 1*, *PITP* and *bHLH* soybean genes between the two systems and reported significant differences among time points in stress recognition and the subsequent adaptive response between mild or more severe stress levels. *GmaxERD1*-like showed expression levels of about four times higher in the sensitive cultivar in the T150 condition of HSys (Figure 2C), suggesting that the sensitive cultivar may have a stronger response to water deficit stress at the molecular level due to the lack of traits that are important for plant adaptation to drought. Another explanation could be the occurrence of damage-related responses that may not have been induced in the tolerant cultivar under the same stress level. *GmaxRD20A*-like and *GmaxRD22*-like did not show such responses, which may indicate that *GmaxERD1*-like belongs to an alternative response pathway. In fact, *GmaxERD1*-like should be part of the ABA-dependent and *GmaxRD20A*-like and *GmaxRD22*-like of the ABA-independent pathway. However, more genes from each pathway are needed to confirm this hypothesis. Stolf-Moreira *et al.* (2011) identified other putative *Arabidopsis ERD1* homolog genes in the *G.max* genome, named *GmERD1*, which showed low expression in the tolerant cultivar in all drought conditions tested in the HSys condition. *GmERD1*

was not included in our analysis because it has a very low similarity to *Arabidopsis ERD1* (data not shown).

*GmaxRD22*-like gene expression was highly induced in plants grown under severe stress in PSys. When compared to other genes evaluated herein, *GmaxRD22*-like could be ranked as the best DMG for PSys. However, no significant differences were observed between sensitive and tolerant cultivars (Figure 2B). Despite the fact that a gradual increase in expression of *GmaxRD22*-like was observed in HSys, the discrepancy seen in expression levels highlights the differences of the two experimental procedures employed in this work.

*GmaxRD20A*-like showed a similar expression profile in the sensitive and tolerant cultivars. However, *GmaxRD20A*-like was more expressed in the tolerant cultivar under severe stress ( $\Psi_w$  -3.0 MPa) in PSys (Figure 2A). In *Arabidopsis*, the RD20A gene is regulated by the AREB1/ABF2 transcription factor, which is also involved in the regulation of the RD29B gene (Fujita *et al.*, 2005; Shinozaki and Yamaguchi-Shinozaki, 2007). Even after exhaustive searches in the soybean genome, the most similar gene we found to *Arabidopsis RD29B*, *Glyma16g31330*, showed low similarity (e-value  $10^{-18}$ ). Furthermore, this putative *RD29B* soybean homolog is not regulated by the water deficit stress according to our experiments (data not shown). These results indicate that *RD29B* homologs are absent in the soybean genome. Recently, it was shown that *Arabidopsis RD29B* is unlikely to perform any protective effects during drought stress (Msanne *et al.*, 2011). Therefore, the lack of *RD29B* homologs in the soybean genome should not interfere in the tolerance of this species to drought.

The higher frequency of the *cis*-elements ACGT, ACCGAC, WAACCA and CANNTG in the promoter regions of DMGs suggests that these elements may be important in expression activity during water deficit. In addition, this result also indicates that these genes probably share *cis*-regulatory elements related to drought stress responses. These motifs are important molecular keys involved in the transcriptional regulation of a dynamic network of gene activities (Simpson *et al.*, 2003; Li *et al.*, 2006; Lenka *et al.*, 2009; Mochida *et al.*, 2009; Umezawa *et al.*, 2010). Additionally, these motifs have a well-documented role in regulating the expression of tolerance in response during drought stress (Busk and Pages, 1998; Simpson *et al.*, 2003; Nakashima *et al.*, 2006).

The soybean DMGs have in their promoter region the WAACCA and CANNTG motifs, which are MYB and MYC recognition sites, respectively. The MYC transcription factor, MYC2, and a MYB transcription factor, MYB2, have been shown to bind to these motifs in the *Arabidopsis RD22* promoter and cooperatively activate the RD22 gene (Abe *et al.*, 1997). Moreover, analysis of the *Arabidopsis ERD1* promoter, a gene that participates in ABA-independent pathways, showed that a MYC recogni-

tion motif is necessary for the induction of the *ERD1* gene by dehydration stress (Simpson *et al.*, 2003). Thus, the promoters of the DMGs investigated in this work share some *cis*-elements with their *Arabidopsis* homologs.

In conclusion, we identified three genes as homologs of traditional *Arabidopsis* DMGs, which were induced by water deficit in soybean. The *GmaxRD20A*-like and *GmaxRD22*-like genes, homologs of *Arabidopsis* genes of the ABA-independent pathway, are highly induced by water deficit in the PSys condition, whereas *GmaxERD1*-like, a homolog of an *Arabidopsis* gene of the ABA-dependent pathway, was highly induced by water deficit in HSys. Hence, these three genes can be very useful as stress markers for studying water deficit in soybean. Moreover, the differences in gene expression in the two systems revealed by our work emphasize that both systems need to be employed to get a *bona fide* picture of the pathways activated by water deficit in soybean. Finally, the differences in gene expression of the *GmaxERD1*-like in the HSys condition between sensitive and tolerant cultivars are robust and may provide a means to better understand the drought tolerance phenotype in soybean cultivars.

## Acknowledgments

We thank Camila Maistro Patreze and Sarah Muniz Nardeli for comments on a previous version manuscript and Durvalina Felix for technical support. This work was part of Guimarães-Dias' PhD research in Genetics, at the Department of Genetics of the Universidade Federal do Rio de Janeiro (UFRJ), and was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; M. Alves-Ferreira: # 306025/2010-8) and Fundação de Amparo à Pesquisa do Rio de Janeiro (FAPERJ).

## References

- Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D and Shinozaki K (1997) Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* 9:1859-1868.
- Adie BAT, Pérez-Pérez J, Pérez-Pérez MM, Godoy M, Sánchez-Serrano J-J, Schmelz EA and Solano R (2007) ABA is an essential signal for plant resistance to pathogens affecting JA biosynthesis and the activation of defenses in *Arabidopsis*. *Plant Cell* 19:1665-1681.
- Andersen CL, Jensen JL and Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 64:5245-5250.
- Bray EA (1993) Molecular responses to water deficit. *Plant Physiol* 103:1035-1040.
- Bray EA (2002) Classification of genes differentially expressed during water-deficit stress in *Arabidopsis thaliana*: An analysis using microarray and differential expression data. *Ann Bot* 89:803-811.
- Bray EA, Bailey-Serres J and Weretilnyk E (2000) Responses to abiotic stresses. In: Buchanan BB, Gruissem W and Jones RL (eds) *Biochemistry and Molecular Biology of Plants*. American Society of Plant Physiologists, Rockville, pp 1158-1203.
- Busk PK and Pages M (1998) Regulation of abscisic acid-induced transcription. *Plant Mol Biol* 37:425-435.
- Casagrande EC, Farias JRB, Neumaier N, Oya T, Pedroso J, Martins PK, Breton MC and Nepomuceno AL (2001) Expressão gênica diferencial durante déficit hídrico em soja. *Rev Bras Fisiol Veg* 13:168-184.
- Choi H-I, Hong J-H, Ha J-o, Kang J-Y and Kim SY (2000) ABFs, a family of ABA-responsive element binding factors. *J Biol Chem* 275:1723-1730.
- Clemente TE and Cahoon EB (2009) Soybean oil: Genetic approaches for modification of functionality and total content. *Plant Physiol* 151:1030-1040.
- Cowan IR (1965) Transport of water in the soil-plant-atmosphere system. *J Appl Ecol* 2:221-239.
- Crandall K, Lagergren J, Simonsen M, Mailund T and Pedersen C (2008) Rapid neighbour-joining. In: Crandall K and Lagergren J (eds) *Algorithms in Bioinformatics*, vol. 5251. Springer Berlin, Heidelberg, pp 113-122.
- Cutforth HW, McGinn SM, McPhee KE and Miller PR (2007) Adaptation of pulse crops to the changing climate of the northern Great Plains. *Agron J* 99:1684-1699.
- Fehr W and Caviness C (1977) Stages of Soybean Development. Special Report n. 80. Iowa State University, Ames, 11 pp.
- Franchini JC, DeBias H, Sacoman A, Nepomuceno AL and Farias JRB (2009) Manejo do Solo para Redução das Perdas de Produtividade pela Seca. Embrapa Soja, Londrina, 39 pp.
- Fujita M, Fujita Y, Maruyama K, Seki M, Hiratsu K, Ohme-Takagi M, Tran L-SP, Yamaguchi-Shinozaki K and Shinozaki K (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant J* 39:863-876.
- Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, Hiratsu K, Ohme-Takagi M, Shinozaki K and Yamaguchi-Shinozaki K (2005) AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in *Arabidopsis*. *Plant Cell* 17:3470-3488.
- Guiltinan M, Marcotte W and Quatrano R (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. *Science* 250:267-271.
- Hellemans J, Mortier G, De Paepe A, Speleman F and Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8:R19.
- Higo K, Ugawa Y, Iwamoto M and Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27:297-300.
- Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W and Zimmermann P (2008) Genevestigator V3: A reference expression database for the meta-analysis of transcriptomes. *Adv Bioinform* 2008:e420747.
- Huang G-T, Ma S-L, Bai L-P, Zhang L, Ma H, Jia P, Liu J, Zhong M and Guo Z-F (2011) Signal transduction during cold, salt, and drought stresses in plants. *Mol Biol Rep* 39:969-987.

- Ito Y, Katsura K, Maruyama K, Taji T, Kobayashi M, Seki M, Shinozaki K and Yamaguchi-Shinozaki K (2006) Functional analysis of rice DREB1/CBF-type transcription factors Involved in cold-responsive gene expression in transgenic rice. *Plant Cell Physiol* 47:141-153.
- Kang JY, Choi HI, Im MY and Kim SY (2002) *Arabidopsis* basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. *Plant Cell* 14:343-357.
- Kankainen M and Holm L (2004) POBO, transcription factor binding site verification with bootstrapping. *Nucleic Acids Res* 32:W222-W229.
- Kulcheski FR, Marcelino-Guimaraes FC, Nepomuceno AL, Abdelnoor RV and Margis R (2010) The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. *Anal Biochem* 406:185-192.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, *et al.* (2007) Clustal W and clustal X ver. 2.0. *Bioinformatics* 23:2947-2948.
- Lenka S, Lohia B, Kumar A, Chinnusamy V and Bansal K (2009) Genome-wide targeted prediction of ABA responsive genes in rice based on over-represented cis-motif in co-expressed genes. *Plant Mol Biol* 69:261-271.
- Li YH, Lee KK, Walsh S, Smith C, Hadingham S, Sorefan K, Cawley G and Bevan MW (2006) Establishing glucose- and ABA-regulated transcription networks in *Arabidopsis* by microarray analysis and promoter classification using a Relevance Vector Machine. *Genome Res* 16:414-427.
- Li Z-Y and Chen S-Y (1999) Inducible expression of translation elongation factor 1A gene in rice seedlings in response to environmental stresses. *Acta Bot Sin* 41:800-806.
- Mahajan S and Tuteja N (2005) Cold, salinity and drought stresses: An overview. *Arch Biochem Biophys* 444:139-158.
- Manavalan LP, Guttikonda SK, Tran LSP and Nguyen HT (2009) Physiological and molecular approaches to improve drought resistance in soybean. *Plant Cell Physiol* 50:1260-1276.
- Martins PK, Jordao 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* 31:512-521.
- Mochida K, Yoshida T, Sakurai T, Yamaguchi-Shinozaki K, Shinozaki K and Tran L-SP (2009) LegumeTFDB: An integrative database of *Glycine max*, *Lotus japonicus* and *Medicago truncatula* transcription factors. *Bioinformatics* 26:290-291.
- Msanne J, Lin J, Stone J and Awada T (2011) Characterization of abiotic stress-responsive *Arabidopsis thaliana* RD29A and RD29B genes and evaluation of transgenes. *Planta* 234:97-107.
- Munns R, James RA, Sirault XRR, Furbank RT and Jones HG (2010) New phenotyping methods for screening wheat and barley for beneficial responses to water deficit. *J Exp Bot* 61:3499-3507.
- Nakashima K, Fujita Y, Katsura K, Maruyama K, Narusaka Y, Seki M, Shinozaki K and Yamaguchi-Shinozaki K (2006) Transcriptional regulation of ABI3-and ABA-responsive genes including RD29B and RD29A in seeds, germinating embryos, and seedlings of *Arabidopsis*. *Plant Mol Biol* 60:51-68.
- Pellegrineschi A, Reynolds M, Pacheco M, Brito RM, Almeraya R, Yamaguchi-Shinozaki K and Hoisington D (2004) Stress-induced expression in wheat of the *Arabidopsis thaliana* DREB1A gene delays water stress symptoms under greenhouse conditions. *Genome* 47:493-500.
- Qin F, Kakimoto M, Sakuma Y, Maruyama K, Osakabe Y, Tran L-SP, Shinozaki K and Yamaguchi-Shinozaki K (2007) Regulation and functional analysis of ZmDREB2A in response to drought and heat stresses in *Zea mays* L. *Plant J* 50:54-69.
- Rodrigues FA, Marcolino J, Carvalho JdFC, Nascimento LCd, Neumaier N, Farias JRB, Carazzolle MF, Marcelino FC and Nepomuceno AL (2012) Using subtractive libraries to prospect differentially expressed genes in soybean plants submitted to water deficit. *Genet Mol Biol* 35(suppl 1): 304-314.
- Seki M, Kamei A, Yamaguchi-Shinozaki K and Shinozaki K (2003) Molecular responses to drought, salinity and frost: Common and different paths for plant protection. *Curr Opin Biotechnol* 14:194-199.
- Shinozaki K and Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. *J Exp Bot* 58:221-227.
- Simpson SD, Nakashima K, Narusaka Y, Seki M, Shinozaki K and Yamaguchi-Shinozaki K (2003) Two different novel cis-acting elements of erd1, a clpA homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. *Plant J* 33:259-270.
- Sitnikova T, Rzhetsky A and Nei M (1995) Interior-branched and bootstrap tests of phylogenetic trees. *Mol Biol Evol* 12:319-333.
- Stolf-Moreira R, Medri M, Neumaier N, Lemos N, Brogin R, Marcelino F, de Oliveira M, Farias J, Abdelnoor R and Nepomuceno A (2010a) Cloning and quantitative expression analysis of drought-induced genes in soybean. *Genet Mol Res* 11:858-867.
- Stolf-Moreira R, Medri M, Neumaier N, Lemos N, Pimenta J, Tobita S, Brogin R, Marcelino-Guimarães F, Oliveira M, Farias J, *et al.* (2010b) Soybean physiology and gene expression during drought. *Genet Mol Res* 5:1946-1956.
- Stolf-Moreira R, Lemos E, Carareto-Alves L, Marcondes J, Pereira S, Rolla A, Pereira R, Neumaier N, Binneck E, Abdelnoor R, *et al.* (2011) Transcriptional profiles of roots of different soybean genotypes subjected to drought stress. *Plant Mol Biol Rep* 29:19-34.
- Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software ver. 4.0. *Mol Biol Evol* 24:1596-1599.
- Texeira LR, Braccini AdLe, Sperandio D, Scapim CA, Schuster I and Viganó J (2008) Avaliação de cultivares de soja quanto à tolerância ao estresse hídrico em substrato contendo polietileno glicol. *Acta Sci Agron* 30:217-223.
- Tran L-S, Quach T, Guttikonda S, Aldrich D, Kumar R, Neelakandan A, Valliyodan B and Nguyen H (2009) Molecular characterization of stress-inducible GmNAC genes in soybean. *Mol Genet Genomics* 281:647-664.
- Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K and Yamaguchi-Shinozaki K (2010) Molecular basis of the core regulatory network in ABA responses:

- Sensing, signaling and transport. *Plant Cell Physiol* 51:1821-1839.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R and Leunissen JAM (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35:W71-W74.
- Urano K, Maruyama K, Ogata Y, Morishita Y, Takeda M, Sakurai N, Suzuki H, Saito K, Shibata D, Kobayashi M, *et al.* (2009) Characterization of the ABA-regulated global responses to dehydration in *Arabidopsis* by metabolomics. *Plant J* 57:1065-1078.
- Xu H, Li Y, Yan Y, Wang K, Gao Y and Hu Y (2010) Genome-scale identification of Soybean BURP domain-containing genes and their expression under stress treatments. *BMC Plant Biol* 10:e197.
- Yamaguchi-Shinozaki K and Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu Rev Plant Biol* 57:781-803.
- Zhao S and Fernald RD (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J Comput Biol* 12:1047-1064.

### Internet Resources

- ClustalW2 site, <http://www.ebi.ac.uk/Tools/clustalw2/index.html> (May 11, 2011).
- Genevestigator web tool, <https://www.genevestigator.com/gv/index.jsp> (July 11, 2011).
- Phytozome, <http://www.phytozome.net/soybean> v5.0 (December 15, 2010).

- Plant Cis program-acting Regulatory DNA Elements, PLACE, <http://www.dna.affrc.go.jp/PLACE> (September 1, 2010).
- Pobo web tool, <http://ekhidna.biocenter.-helsinki.fi:9801/pobo> (May 15, 2011).
- The *Arabidopsis* Information Resource, TAIR, <http://www.arabidopsis.org> (December 15, 2010).
- Brazilian Institute of Geography and Statistics (IBGE), <http://www.ibge.gov.br> (October 20, 2010).

### Supplementary Material

The following online material is available for this article:

Table S1 - Cis-regulatory element responsive to drought stress, salinity stress, osmotic stress and ABA.

Table S2 - Frequency of the *cis*-elements present in promoters of the DMGs.

Figure S1 - Search strategy for DMGs involved in the soybean response to water deficit.

Figure S2 - Digital expression pattern of *Arabidopsis* DMGs involved in the response to different water privation conditions and ABA stimulus

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

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

**Table S1** - Cis-regulatory elements responsive to drought stress, salinity stress, osmotic stress and ABA.

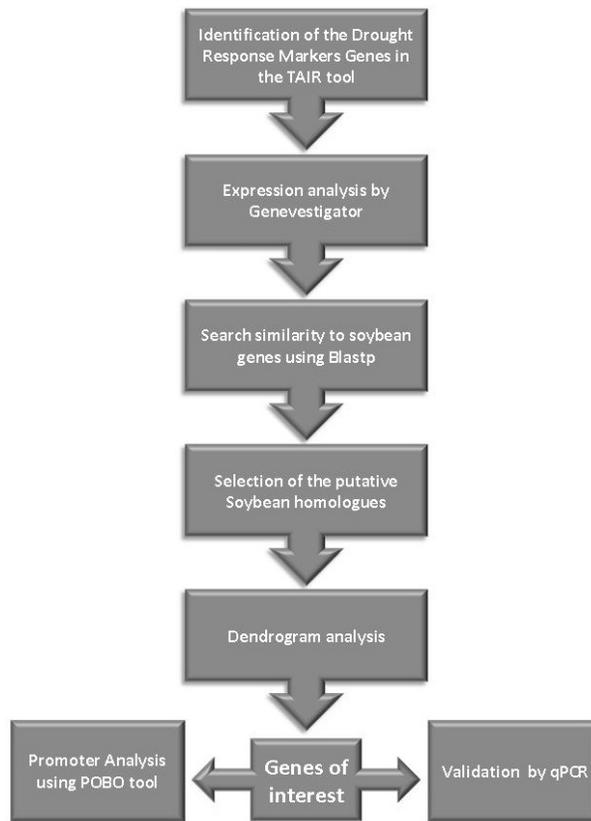
<b>Pattern</b>	<b>Description pattern</b>
ACGT	This sequence required for etiolation-induced expression of erd1.
ACCGAC	DBF2" bound to "DRE2"; rab17 is expressed during late embryogenesis, and is induced by ABA.
WAACCA	MYB recognition site found in the promoters of the dehydration-responsive gene rd22.
CANNTG	MYC recognition site found in the promoters of the dehydration-responsive gene rd22 and many other genes in Arabidopsis; Binding site of ATMYC2.
ACGTGKC	DRE and ABRE are interdependent in the ABA-responsive expression of the rd29A.
YAACKG	MYB recognition site found in the promoters of the dehydration-responsive gene rd22.
RYACGTGGYR	ABRE in Arabidopsis dehydration-responsive gene rd22.
CCACGTGG	ABRE; ABA and water-stress responses.
YACGTGGC	ABA responsive element found in the promoter of stress regulated.
CACATG	MYC binding site in rd22 gene of Arabidopsis thaliana; ABA-induction.
CATGTG	MYC recognition sequence necessary for expression of erd1 in dehydrated Arabidopsis.
CNGTTR	ATMYB2 is involved in regulation of genes that are responsive to water stress in Arabidopsis.
CCGAC	Core of low temperature responsive element (LTRE) of cor15a gene in Arabidopsis.
TACCGACAT	Related to responsiveness to drought, low-temperature or high-salt stress.
TAACTG	ATMYB2 is involved in regulation of genes that are responsive to water stress in Arabidopsis.
CTAACCA	Binding site for MYB in dehydration-responsive gene, rd22.
WK[TAGC]CGTR	NAC binding site (NACBS).

Note: The symbol W was used in addition to A or T; the symbol R was used in addition to A or G; the symbol Y was used in addition to C or T; the symbol K was used in addition to G or T; The symbol W was used in addition to T or A; and the symbol N was used in addition to A,C,G or T.

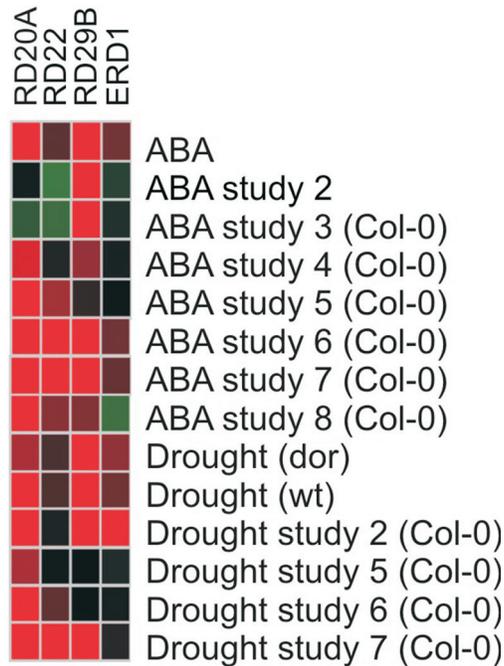
Table S2 – Frequency of cis elements present in promoters of the DMG.

Motif	Data set	Number of promoters in a each dataset	Number of promoters containing the pattern	Total number of patterns in each dataset	Promoter mean
ACGT	BG	77222	55801 (72.26%)	233388	3.02
	Cluster 1	3	3 (100%)	16	5.33
ACCGAC	BG	77222	8254 (10.7%)	10276	0.13
	Cluster 1	3	2 (66.7%)	2	0.66
WAACCA	BG	77222	58808 (76.2%)	121330	1.56
	Cluster 1	3	3 (100%)	6	2
CANNTG	BG	77222	73865 (95.7%)	540426	7.07
	Cluster 1	3	3 (100%)	24	7.99

Note: All analyzed promoter sequences presented 1,000 bp and POBO was run with the following parameters: Number pseudoclusters 50 and length of the background promoter 1,000 bp, bootstrap 1000. Calculated t-test using the linked on line GrapPad web site: <http://www.graphpad.com/quickcalcs/DistMenu.cfm>. P<0.0001. The symbol W was used in addition to A or T and the symbol N was used in additional to A,C,G or T. BG - background set that comprise all promoters regions of the *G. max* genome. Cluster 1 – promoters set that comprise the promoters regions of the GOI genes.



**Figure S1** - Search strategy for DMGs involved in the soybean response to water deficit.



**Figure S2** - Digital expression pattern of *Arabidopsis* DMGs involved in the response to different water privation conditions and ABA stimulus, evaluated by the Genevestigator web tool. The intensity of the red color is equivalent to increase in the gene induction in response to drought stress or ABA. The increase in the intensity of the green color is equivalent to increase in the gene repression in response to drought stress or ABA. The black color indicates that it is not induced or repressed under drought stress conditions. Description of experiments: ABA: wild type (Col-0) seedlings, treated with 10  $\mu$ M ABA for 1 h; ABA study 2: Leaf samples from wild type (Col-0) plants, not watered for 7 days; ABA study 3 (Col-0): After stratification at 4  $^{\circ}$ C for 4 days, wild type seeds were allowed to germinate on the MS-2% sucrose plate containing 0.5  $\mu$ M ABA for two days; ABA study 4 (Col-0): wild type leaves of 4 week-old plants, sprayed with 50  $\mu$ M ABA. After 4 h, samples were collected; ABA study 5 (Col-0): Plant samples of Col-0 grown for 2 weeks on MS agar medium with 3% sucrose and then treated with 100  $\mu$ M ABA for 1 h; ABA study 6 (Col-0): Plant samples of wild type (Col-0) grown for 2 weeks on MS agar medium with 3% sucrose and then exposed to dehydration stress for 1 h; ABA study 7 (Col-0): Isolated guard cell samples of 5-week-old wild-type plants, treated with 50 mM ABA for 3 h; ABA study 8 (Col-0): Excised leaf samples of 5-week-old wild-type plants, treated with 50 mM ABA for 3 h. Drought (dor): Rosette leaf samples of dor plants, grown under normal watering conditions for 24 days and then stressed by completely depriving them of irrigation for 10 days; Drought (wt): Rosette leaf samples of wild type (Col-0) plants, grown under normal watering conditions for 24 days and then stressed by completely depriving them of irrigation for 10 days; Drought study 2 (Col-0): Leaf samples from wild-type (Col 0) plants, not watered for 7 days; Drought study 5 (Col-0): Rosette leaf samples of dor plants, grown under normal watering conditions for 24 days and then stressed by completely depriving them of irrigation for 10 days; Drought study 6 (Col-0): Plant samples of Col-0 grown for 2 weeks on MS agar medium with 3% sucrose and then exposed to dehydration stress for 1 h; Drought study 7 (Col-0): Plant samples of Col-0 grown for 2 weeks on MS agar medium with 3% sucrose and then exposed to dehydration stress for 4 h.