Research Article

# Expression analysis in response to drought stress in soybean: Shedding light on the regulation of metabolic pathway genes

Fábia Guimarães-Dias[1*], Anna Cristina Neves-Borges[2*], Antonio Americo Barbosa Viana[3], Rosilene Oliveira Mesquita[5], Eduardo Romano[3], Maria de Fátima Grossi-de-Sá[3], Alexandre Lima Nepomuceno[4], Marcelo Ehlers Loureiro[5] and Márcio Alves-Ferreira[*1]

[1]*Laboratório de Genética Molecular Vegetal, Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil.*
[2]*Laboratório Integrado de Biologia Vegetal II, Departamento de Botânica, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brazil.*
[3]*Embrapa Recursos Genéticos e Biotecnologia, Parque Estação Biológica, Brasília, DF, Brazil.*
[4]*Embrapa Soja, Londrina, PR, Brazil.*
[5]*Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, MG, Brazil.*

## Abstract

Metabolomics analysis of wild type *Arabidopsis thaliana* plants, under control and drought stress conditions revealed several metabolic pathways that are induced under water deficit. The metabolic response to drought stress is also associated with ABA dependent and independent pathways, allowing a better understanding of the molecular mechanisms in this model plant. Through combining an *in silico* approach and gene expression analysis by quantitative real-time PCR, the present work aims at identifying genes of soybean metabolic pathways potentially associated with water deficit. Digital expression patterns of *Arabidopsis* genes, which were selected based on the basis of literature reports, were evaluated under drought stress condition by Genevestigator. Genes that showed strong induction under drought stress were selected and used as bait to identify orthologs in the soybean genome. This allowed us to select 354 genes of putative soybean orthologs of 79 *Arabidopsis* genes belonging to 38 distinct metabolic pathways. The expression pattern of the selected genes was verified in the subtractive libraries available in the GENOSOJA project. Subsequently, 13 genes from different metabolic pathways were selected for validation by qPCR experiments. The expression of six genes was validated in plants undergoing drought stress in both pot-based and hydroponic cultivation systems. The results suggest that the metabolic response to drought stress is conserved in *Arabidopsis* and soybean plants.

*Key words:* *Glycine max*, drought resistance, qPCR, metabolic pathway, bioinformatics.

## Introduction

Crop plants are often exposed to various biotic (viruses, bacteria and fungi) and abiotic stress factors (such as water deficit and salinity) that may impair their growth, development and ultimately affect productivity (Kang *et al.*, 2002; Mahajan and Tuteja, 2005). Damage caused by these stresses represents a major concern for producers, consumers and governments, especially in relation to crops of great economic importance, such as wheat, corn and soybean, whose losses may range between 78%-87% of maximum yield under ideal conditions (Bray *et al.*, 2000).

Soybean [*Glycine max* (L.) Merr.], the most important legume grown worldwide, is an essential source of oil, protein, macronutrients and minerals (Clemente and Cahoon, 2009). Despite increased global demand, the current losses in soybean production are estimated to be over one fifth of the crop worldwide. Most of these losses are attributed to abiotic factors, responsible for a decrease of 69% in comparison to the record yield capacity (Bray *et al.*, 2000). In Brazil, the occurrence of prolonged drought during summer has become increasingly common in recent years (Brando *et al.*, 2010). In the state of Paraná, Brazil, soybean yields have fallen due to drought resulting in a cumulative decline of almost 11 million tons in total production (Franchini *et al.*, 2009). In 2008-2009, losses due to

drought in the north and west of the state of Paraná, were 80% (Franchini *et al.*, 2009). This situation may become even more dramatic in light of current environmental predictions, which point to global warming and subsequent occurrence of drought in water-stressed regions, which represent one-third of the world's culturable land (Manavalan *et al.*, 2009).

In order to better cope with drought stress, plants possess a large repertoire of morphological, biochemical, physiological and molecular adaptations and responses (Bray, 1993; Seki *et al.*, 2003; Yamaguchi-Shinozaki and Shinozaki, 2006). Recent functional genomics studies using combined strategies of transcriptomics, proteomics, and metabolomics revealed a wide range of important genes involved in the synthesis of metabolites in response to drought, such as osmoprotectants, osmolytes, compatible solutes, or signaling molecules (Shinozaki and Yamaguchi-Shinozaki, 2007; Verbruggen and Hermans, 2008; Urano *et al.*, 2010).

The accumulation of osmolytes in plant cells results in a decrease in osmotic potential, water absorption and cell turgor pressure, which contribute to the maintenance of physiological processes such as stomata opening, photosynthesis and plant growth (Hsiao, 1973; Shinozaki and Yamaguchi-Shinozaki, 2000; Baxter *et al.*, 2007). Solute accumulation under stress is probably the most distinctive feature of an adaptive response to stresses that involve a component of water deficit, such as drought, freezing and salinity (Hsiao, 1973; Thomashow, 1999; Zhu, 2002). A specific physiological response to drought represents combinations of events that are activated and turned off by the perception of stress. An understanding of how these events interact is an important step towards the development of crops with greater tolerance to drought.

Two experimental procedures are usually applied to assess a gene expression profile during drought stress conditions in soybean: the **p**ot-based **sys**tem (**PSys**) (Casagrande *et al.*, 2001; Qin *et al.*, 2007; Martins *et al.*, 2008; Tran *et al.*, 2009) and the **h**ydroponic **sys**tem (**HSys**) (Martins *et al.*, 2008; Kulcheski *et al.*, 2010). Drought stress in plants cultured in PSys is more similar to field conditions, where the rate of water loss is slower, allowing acclimation to the drought condition (Cowan, 1965). In the HSys, the plants are placed in containers where a nutrient solution composed of water and nutrients circulates, without the presence of soil as a substrate. In this system, the simulation of drought is carried out by removing the plants from the nutrient medium, so water loss is more rapid, causing a shock in the plant, and within minutes it is possible to observe the physical effects caused by the stress. HSys does not allow plant acclimation (Munns *et al.*, 2010).

In this work, we investigated several metabolic pathways potentially associated with water deficit in soybean (*G. max*). For this purpose, we employed different strategies, combining an *in silico* approach and gene expression analysis by qPCR. The gene expression analysis was performed with plants cultivated under HSys and PSys, which allowed us to compare the effects and responses to differences in acclimation. The identification of such genes is the first step to better understand the effects of water deficit on the regulation of expression of metabolic pathway genes in soybean. This knowledge should also be helpful in the identification of drought tolerant soybean cultivars and provide better tools to develop water-stress tolerant crops.

## Material and Methods

### Plant material, growth conditions and treatments

The *Glycine max* L. Merrill cultivars BR 16 and Embrapa 48 have been shown to have contrasting responses to water deficit; BR 16 is very sensitive to drought, and Embrapa 48 shows a high tolerance to this stress (Casagrande *et al.*, 2001; Texeira *et al.*, 2008).

We used two different water deficit treatments, a pot-based system (PSys) in which plant were grown in sand and a hydroponics system (HSys) in which plants were grown in a nutrient solution (Martins *et al.*, 2008; Kulcheski *et al.*, 2010).

Plants grown in the PSys were maintained in a greenhouse at 30 °C ± 5 °C temperature and 60% ± 20% relative humidity. The cultivars BR16 and Embrapa 48 were germinated in washed sand where they remained for about 10 days. After this period, seedlings were transplanted to pots. Seedlings at the V4 development stage (fourth trifoliate fully expanded) (Fehr *et al.*, 1971) were watered on a daily basis in the control pots, whereas watering was suspended in the pots of plants under drought stress. The water potential ($\Psi$w) was measured daily (always between 05:00 and 06:00) after the second day of the interruption of watering. The $\Psi$w for each plant was measured by the Scholander-type pressure chamber. Seven days after the interruption of watering the $\Psi$w was -1.5 ± 0.2 MPa (moderate stress level) and after ten days -3.0 ± 0.2 MPa (severe stress level). The roots with sand were removed from their pots and then immediately and gently rinsed with water for 1 min, in order to remove all the sand. To remove biological contaminants, the roots were carefully immersed in 2% SDS solution for 1 min, and washed gently with ultrapure water for 1 min. After this process, the root samples for one plant from each treatment, in total, two plants (two biological replicates), were immediately frozen in liquid nitrogen and stored at -80 °C for RNA extraction.

For cultivation in the hydroponic system (HSys), seeds were pre-germinated on moist filter paper in dark conditions at 25 °C ± 1 °C and 65% ± 5% relative humidity. Plantlets were then placed in polystyrene supports in such a way that the roots of the seedlings were completely immersed in the solution. Each tray containing seedlings was maintained in a greenhouse at 25 °C ± 2 °C and 60% ± 5%

relative humidity, under natural daylight (photosynthetic photon flux density (PPFD) = 1.5 x $10^3$ μmoles $m^{-2}$ $s^{-1}$, equivalent to 8.93 x $10^4$ lux) and a 12 h day length. After 15 days, seedlings at V4 development stage were submitted to different treatments in which they were removed from the hydroponic solution and kept in a tray in the dark without nutrient solution or water for 0 min (T0, or unstressed), 50 min (T50), 100 min (T100) and 150 min (T150). Two biological replicates of root samples from both cultivars were collected at these time points and immediately frozen in liquid nitrogen followed by storage at -80 °C for posterior RNA extraction.

## Total RNA isolation

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

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

Primers were designed using the Primer 3 plus software (Untergasser *et al.*, 2007) using as criteria the generation of amplicons ranging from 80 to 200 bp with a Tm of 60 °C ± 1 °C (primer sequences are listed in Table S1). Both candidate and housekeeping genes were amplified in a one step protocol. As housekeeping genes, *ACT11* (cytoskeleton structural protein) and *FBOX* (F-Box protein family) (Kulcheski *et al.*, 2010) were used for normalization of target gene expression. Melting curve and gel electrophoresis analysis of the amplification products confirmed that the primers amplified only a single product of expected size (data not shown).

PCRs were carried out in an optical 96-well plate with a Realplex 4 Eppendorf Masterclycler® Ep gradient sequence detection system (Eppendorf) Power SYBR® Green RNA-to-Ct TM 1-Step Kit (Applied Biosystems) was used as recommended by the manufacturer. For each sample, 25 ng of RNA was used in the reaction mixture in a final volume of 20 μL. Reaction mixtures were incubated for 30 min at 48 °C and 10 min at 95 °C, followed by 40 amplification cycles of 15 s at 95 °C, and 1 min at 60 °C. Primer set efficiencies were estimated for each experimental set by Miner software (Zhao and Fernald, 2005) and these values were used in all subsequent analyses. Miner software was used to determine the starting and ending points of the exponential phase of PCR from raw fluorescence data. It also estimated primer set amplification efficiencies through a nonlinear regression algorithm without the need for a standard curve. In addition, the values of the threshold cycle (quantification cycle value – Cq) were converted by the program QBASE v1.3.5 (Hellemans *et al.*, 2007) into relative amounts normalized (NRQ). All references and samples for each experimental condition were evaluated in technical triplicates.

## Bioinformatic tools

### Identification of metabolic pathway genes in soybean

*Arabidopsis* genes associated with response to drought in different pathways were selected based on information from the literature (Sanchez *et al.*, 2008; Bundy *et al.*, 2009; Urano *et al.*, 2009; Hey *et al.*, 2010). Gene models for the metabolic pathway genes were obtained using the tools AraCyc metabolic pathway from the TAIR (The *Arabidopsis* Information Resource) and KEGG pathways websites. The digital expression pattern of these genes under drought conditions in *Arabidopsis* was evaluated by using the Genevestigator web tool (Hruz *et al.*, 2008). Subsequently, the protein sequences of possible orthologs in soybean were used to conduct Blastp searches in Phytozome. All sequences with an *e-value* = 0, or, in the absence of sequences with *e-value* = 0, the first five with *e-value* lower than $10^{-30}$ were analyzed for their presence in subtractive libraries available in the GENOSOJA LGE (Laboratory of Genomic and Expression: Project GENOSOJA) database (Rodrigues *et al.*, 2012). These subtractive libraries are composed of samples from leaves and roots in three separate bulks with regard to the dehydration period: bulk 1 (T25-50 min); bulk 2 (T75-100 min) and bulk 3 (T125-150 min), for both cultivars (Rodrigues *et al.*, 2012). The presence of a given gene in these libraries is indicative of the induction of its expression during water deficit. The selected genes represented in the libraries were also submitted to a dendrogram analysis, as well as a validation of their expression pattern through qPCR.

### Generation of dendrograms

The protein sequences of *A. thaliana* were used to search for all aligned genes in *G. max* and *Oryza sativa* (out group) genomes, as well as in *Arabidopsis*. The alignment of amino acid sequences was done using the ClustalW2 software (Larkin *et al.*, 2007). The software MEGA v.4 was used to construct dendrograms by means of the Neighbor-Joining algorithm (Tamura *et al.*, 2007), under a Poisson model, complete deletion, and bootstrapping with 1,000 replications (Sitnikova *et al.*, 1995). *G. max, O. sativa* and *A. thaliana* genes were selected considering *e-values* smaller than $10^{-15}$ in the Phytozome and TAIR databases.

### Promoter analysis

Sequences of 1,000 bp upstream to the start codon of the genes of the soybean genome were obtained by using

the genome browse tool in the Phytozome database. *Cis*-regulatory elements related to drought stress, salinity stress and ABA were identified in the database of Plant Cis program-acting Regulatory DNA Elements –(PLACE) by a keyword search (Higo *et al.*, 1999). The POBO tool (Kankainen and Holm, 2004) was used for comparison of motif occurrences in promoters of putative orthologous genes by using the whole genome of *G. max* as background information.

## Results

### *In silico* identification and characterization of soybean genes involved in different pathways in response to dehydration

The metabolic pathways of *Arabidopsis* involved the synthesis and degradation of metabolites during drought stress were selected based on information from the literature (Sanchez *et al.*, 2008; Bundy *et al.*, 2009; Urano *et al.*, 2009; Hey *et al.*, 2010). Each step of the metabolic pathways was investigated in the AraCyc metabolic pathway (Zhang *et al.*, 2005) and KEGG pathway tools (Zhang and Wiemann, 2009). The digital expression profile for each gene under water deficit was evaluated through clustering analysis by the Genevestigator web tools (Hruz *et al.*, 2008). This procedure allowed us to select 80 genes from *Arabidopsis* belonging to 39 different metabolic pathways that are regulated during water deficit (Table S2). For simplicity, this group was named "Arabidopsis Genes of the Metabolic Pathways" (AGMPs). The diagram of the search strategy employed is illustrated in Figure 1.

The 354 putative soybean orthologs of the 80 *Arabidopsis* genes were identified by Blastp searches on the Phytozome website. The putative soybean ortholog genes had their expression pattern evaluated by subtractive library tools of the GENOSOJA LGE (Laboratory of Genomic and Expression: Project GENOSOJA) (Rodrigues *et al.*, 2012). This step allowed us to check whether the expression of these genes is induced during drought stress. The selection criteria were the presence of the gene in at least two subtractive libraries related to drought stress. This strategy allowed us to identify 13 putative soybean ortholog genes belonging to seven different metabolic pathways (data not shown). We herein focus on the description of three pathways: lysine degradation, putrescine biosynthesis and stachyose biosynthesis.

In order to identify the best candidates in the soybean genome for the AGMPs, we performed dendrogram analyses. These included the genes *GmaxLKR/SDH*-like1, *GmaxLKR/SDH*-like2 and *GmaxADC2*-like1 (Figure 3) and also *GmaxGOLS2*-like1, *GmaxGOLS2*-like2, and *GmaxGOLS2*-like3 (Figure 4). These genes are part of the metabolic pathways of lysine degradation II, putrescine biosynthesis I and stachyose biosynthesis, respectively (Figure 2). For those soybean genes where the neighbor-
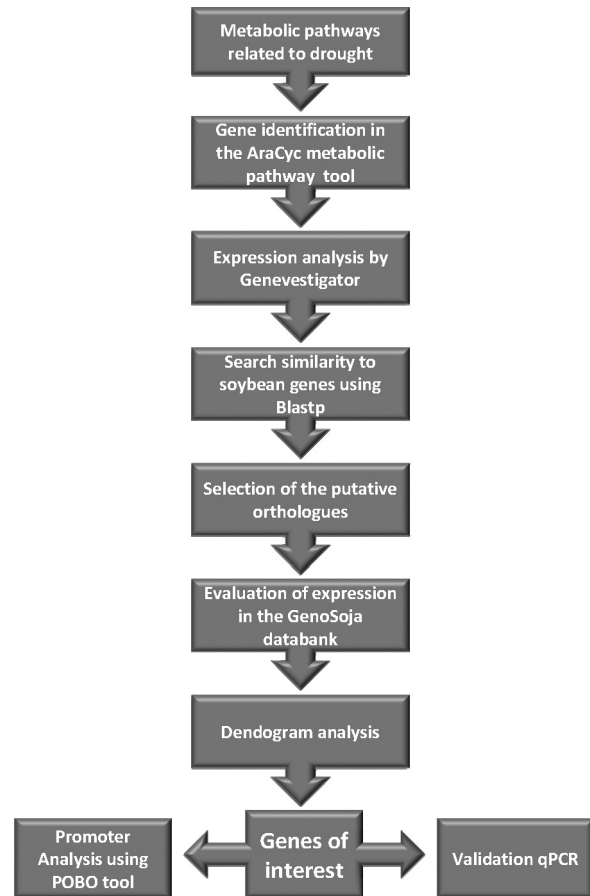


**Figure 1** - Strategy of ortholog gene search in soybean subjected to drought stress.

joining analysis was not able to determine the closest *Arabidopsis* ortholog, the selection of the soybean gene(s) for posterior analysis was based on their expression frequency in the drought induced subtractive library of the GENOSOJA LGE database (Table S2). The putative soybean orthologs of AGMPs were identified through Blastp searches in the soybean genome on the Phytozome website, followed by dendrogram analysis. For each AGMP, we identified a putative ortholog in the *G. max* and *O. sativa* genomes. The dendrogram analysis indicated that the
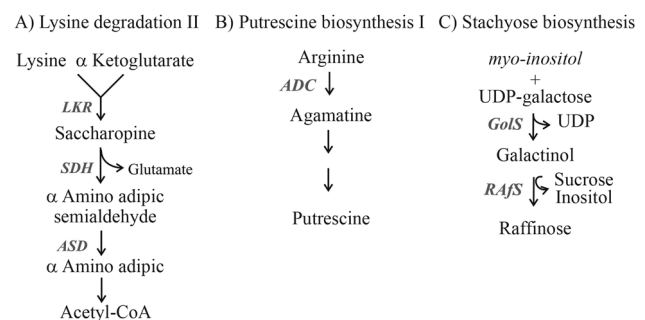


**Figure 2** - Schematic diagram of pathways for (A) Lysine degradation II, (B) Putrescine biosynthesis I, and (C) Stachyose biosynthesis. Enzyme names are in green letters and italics.
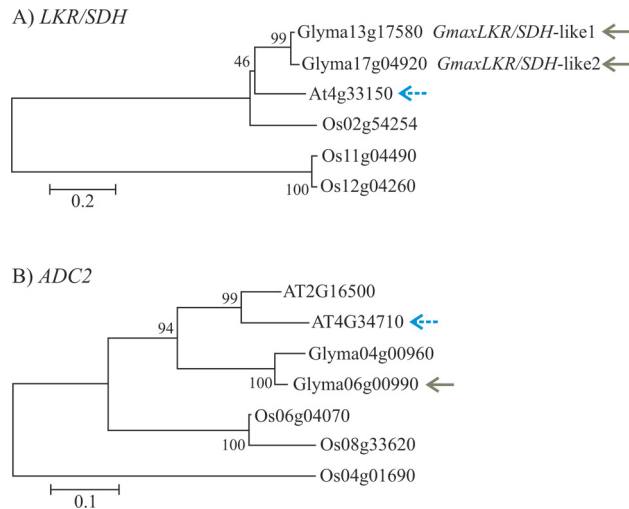
**Figure 3** - Dendrogram using a gene model of drought responsive genes in *Arabidopsis thaliana*, *Oryza sativa* and *Glycine max* based on the amino acid sequences. (A) Dendrogram of *LKH/SDH*-like1 and *LKH/SDH*-like2, and (B) of *ADC2*-like 1. The green solid arrows indicate the soybean candidates and the blue dotted arrows the respective *Arabidopsis* reference genes. Bootstrap values (1,000 replications) are indicated at the base of each branch.

*Arabidopsis* genes *AtLKR/SDH* (At4g33150) and *AtGOLS2* (At1g56600) have two putative orthologs in the soybean genome. For the gene *GmaxLKR/SDH* the putative orthologs are Glyma13g17580 and Glyma17g0492, while for the gene *GmaxGOLS2* the putative orthologs are Glyma20g22700, Glyma03g38080 and Glyma19g40680 (Figures 3A and 4). The dendrogram analysis of *ADC2* pointed to Glyma04g00960 as being the closest gene to *AGMP*. However, *Glyma04g00960* was present only in a single subtractive library whereas Glyma06g00990 was represented in four. Therefore, *Glyma06g00990* was also selected to be validated by qPCR (Figure 3B).

## RT-qPCR

Through *in silico* analysis we selected six genes for validation by qPCR of root samples of the sensitive (BR16) and tolerant (Embrapa 48) cultivars submitted to water deficit in PSys and HSys.

The genes *GmaxLKR/SDH*-like1 and *GmaxLKR/SDH*-like2 showed higher expression in PSys compared to HSys (Figure 5A, B). The expression profile in the sensitive cultivars showed a gradual increase in all conditions tested. Interestingly, the expression of *GmaxLKR/SDH*-like1 and *GmaxLKR/SDH*-like2 in the tolerant cultivar was down-regulated in the PSys when exposed to drought. In the HSys condition, these genes showed a higher increase in expression at a later time (T100 min and T150 min) in both cultivars.

The *GmaxADC2*-like1 gene showed similar expression dynamics for both cultivars in the two systems studied, with a peak of relative expression under moderate stress in
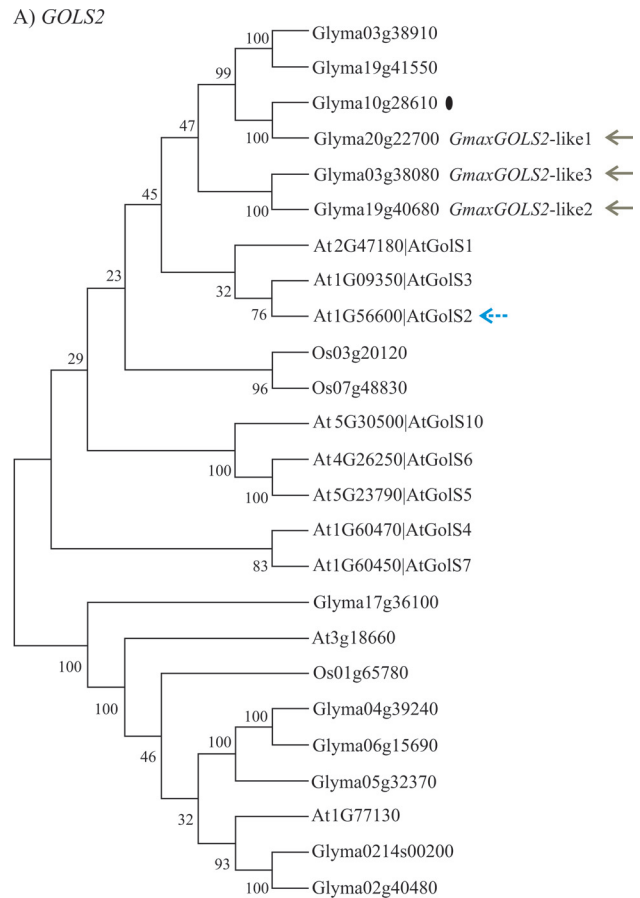


**Figure 4** - Dendrogram of the *GOLS* gene using a gene model based on amino acid sequences for drought of responsive genes in *Arabidopsis thaliana*, *Oryza sativa* and *Glycine max*,. The green solid arrow indicates soybean candidates, while the blue dotted arrow point to *Arabidopsis* reference gene, and the black dot indicates another *GmGOLS* candidate gene. Bootstrap values (1,000 replications) are indicated at the base of each branch.

PSys (Ψw -1.5 MPa) at 100 min (T100) in the HSys culture condition. Furthermore, expression levels were significantly higher in the HSys condition (Figure 5C).

The *GmaxGOLS2*-like1 gene presented a quite different expression profile during drought stress in the two tested systems when compared with the other two *GmaxGOLS2* soybean orthologs, *GmaxGOLS2*-like2 and *GmaxGOLS2*-like3. It is worthy of note that the level of expression of *GmaxGOLS2*-like1 is eight times higher in the tolerant cultivar at an early time point (T50 min) in HSys compared to the non-stress sample, while the sensitive cultivar showed a level of expression four times higher for the same time point (T50 min) compared to the control sample. In PSys, the tolerant cultivar showed a subtle increase in the *GmaxGOLS2*-like1 expression level under moderate stress (-1.5 MPa) compared to the control, while the sensitive cultivar exhibited mild repression under the same stress level (Figure 5D)

The *GmaxGOLS2*-like2 and *GmaxGOLS2*-like3 showed fairly similar gene expression profiles for both
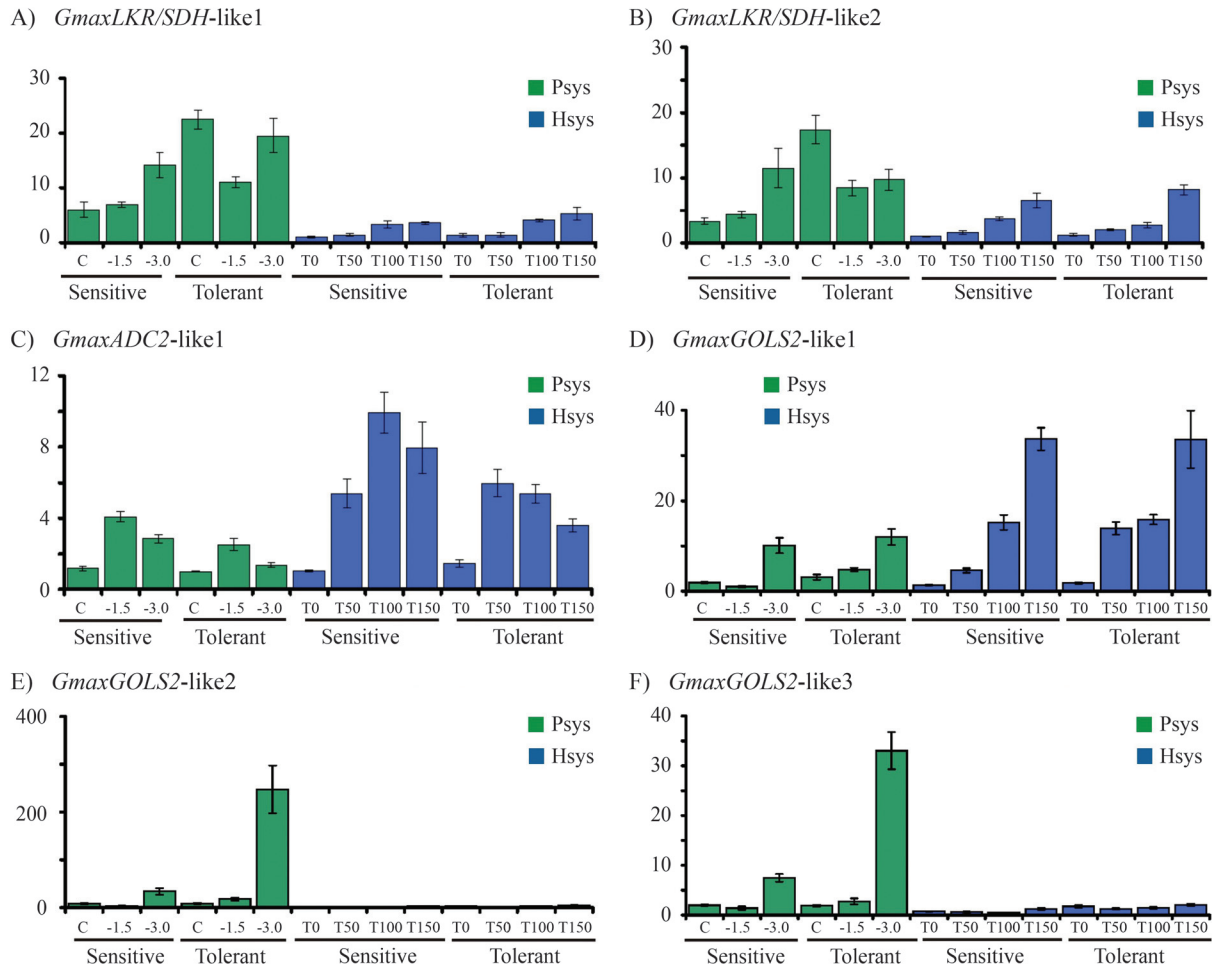
**Figure 5** - Expression profile analyses of drought stress-related genes in pot-based (PSys) and hydroponic (HSys) cultivation conditions. A) *GmaxLKR/SDH*-like1, B) *GmaxLKR/SDH*-like2, C) *GmaxADC2*-like1, D) *GmaxGOLS2*-like1, E) *GmaxGOLS2*-like2 and F) *GmaxGOLS2*-like3. The PSys condition is represented by green solid bars and HSys by blue solid gray bars. The sensitive (BR16) and tolerant (Embrapa 48) cultivars are indicated at the bottom of the bars. Relative expression levels of these genes are represented on the Y-axis, relation to the reference genes ACT and FBOX in both cultivars and systems. The C, -1.5 and -3.0 represent control and the water potentials of soybean plants (Ψw) measured after the second day of the interruption of watering. Seven days after the interruption of watering the Ψw was -1.5 ± 0.2 MPa (moderate stress level) and after ten days -3.0 ± 0.2 MPa (severe stress level). The T0 (control), T50 (50 min), T100 (100 min) and T150 (150 min) indicate the different times under drought stress which the soybean seedlings were submitted after removed of the hydroponic solution.

cultivars in the two systems studied. These genes reached the highest level of relative expression under the most severe stress (Ψw -3.0 MPa) in the PSys condition. Notwithstanding, it is important to note that the expression level of *GmaxGOLS2*-like2 was about ten times higher than that of *GmaxGOLS2*-like3. In the HSys conditions, expression levels were very low for both cultivars which indicates that these genes are not regulated during water deficit stress in this system (Figure 5E,F).

In addition to the gene expression studies we investigated the presence of *cis*-regulatory elements in soybean drought-response genes selected for *in silico* analysis. By means of the Place tool, 17 candidate motifs related to drought were identified (data not show) and the statistical significance of their enrichment was assessed using the POBO tool, which compares motif abundance in the given promoter set relative to *G. max* background (BG) frequen-

cies. The analysis revealed that two ABA responsive binding elements, named AREBs, (ACGTG and ACGTGKC) and one motif for the early response to dehydration, named ERD (ACGT) are enriched in the promoter of the selected genes when compared to the background genome. The analysis in POBO also indicated that the ACGTG motif was present in 54.5% of the promoters of all genes of interest. The average number of promoters that presented this motif was 2.55 compared to an average of 0.88 for all *G. max* promoters (BG) (t-test; p > 0.0001). The ACGTGKC motif was present in 54.5% of the promoters of all genes of interest. The average number of promoters that showed this motif in the selected gene set was 1.46 compared to an average of 0.13 for all *G. max* promoters (t-test; p > 0.0001). The ACGT motif is the most representative one within the set of target genes, being present in 81.8% of the promoters. The average number of promoters harboring this motif was

5.96 compared to an average of 3.03 in the promoter regions of the *G. max* genome (Table S3).

## Discussion

Herein we identified several soybean genes that are responsive to drought stress. These belong to different metabolic pathways based on previous information of the model plant *Arabidopsis* (Taji *et al.*, 2002; Sanchez *et al.*, 2008; Urano *et al.*, 2009, 2010). We identified 354 putative orthologs in the soybean genome within 39 metabolic pathways. We used the subtractive libraries performed on soybean root tissues obtained from the GENOSOJA database to direct us in the selection of the key genes. Through *in silico* analysis, we selected six soybean genes from three metabolic pathways for qPCR validation. The expression was assayed in roots of plants under water deficit in two ways: (i) PSys, in which the rate of water loss is slower, and allows the plant to adapt to the unfavorable environmental conditions, and (ii) HSys, in which the rate of water loss is very rapid, not giving the plant time to adapt to the stress conditions (Bray, 1993). Employing these alternative systems helped us to understand the control of gene expression involved in drought-induced metabolism.

Drought in plants starts as a complex set of responses, beginning with the perception of stress, which triggers a cascade of molecular events that comprise various levels of physiological, metabolic and developmental responses (Mahajan and Tuteja, 2005). Previous studies indicate that PSys and HSys physiological responses were observed at a stress level of -3.0 MPa and T100 min, respectively (Martins *et al.*, 2008). At this point, soybean plants begin a process of wilting, where the rate of photosynthesis decreases, leading to stomata closure and increased leaf temperature. Our expression analysis allowed to characterize the two systems, revealing a distinct perception of stress in the plants kept under PSys and HSys in cultivars that are tolerant and sensitive to drought, respectively.

In previous studies carried out with different soybean cultivars, the Embrapa 48 cultivar showed a reduced response to the evaluated characteristics, such as lower rates of reduction in germination rate, lower percentage of reduction in primary root length, and lower photosynthetic rate under moderate and severe water deficit, compared to other cultivars, including BR16 (Casagrande *et al.*, 2001; Texeira *et al.*, 2008). Hence, the Embrapa 48 cultivar is considered more tolerant to water deficit because it reacts more rapidly to the adverse situation. In our analysis, *GmaxGOLS2*-like2 and *GmaxGOLS2*-like3, for instance, were expressed in both cultivars in the Psys condition, but expression levels were significantly higher in Embrapa 48 (Figure 5E,F). Differences in the regulation of gene expression between cultivars were also noted when the expression of *GmaxLKR/SDH*-like1 and *GmaxLKR/SDH*-like-2 were evaluated in the Psys condition. Both presented high expression levels under this control condition, which may in-

dicate that the Embrapa 48 cultivar presents naturally higher levels of protective compounds and can better cope with a water deficit. These conclusions do not apply to the HSys experiment, where practically no differences were observed between the cultivars. These results strongly suggest that a water deficit in the sensitive and tolerant cultivar activates distinct molecular switches depending on the cultivation system.

The adaptive response to stress at cellular and molecular levels involves the accumulation of osmolytes and proteins related to stress tolerance (Kishor *et al.*, 1995; Kiyosue *et al.*, 1996; Zhu, 2002; Mahajan and Tuteja, 2005; Fujita *et al.*, 2006; Hummel *et al.*, 2010; Ashraf *et al.*, 2011). In *Arabidopsis*, drought stress responses are perceived by the biosynthetic genes *BCAT2*, *LKR/SDH*, *P5CS1* and *ADC2* pertaining to the ABA-dependent pathway, while the raffinose (RFO) and galactinol (*GOLS2*) genes are not regulated by ABA during dehydration stress (Taji *et al.*, 2002; Sanchez *et al.*, 2008; Hirayama and Shinozaki, 2010). If the *GOLS2* ABA independent response is conserved in the three putative soybean homologues, our results suggest that an ABA independent response is activated in both systems tested (PSys and HSys).

Among the genes expected to participate in the ABA-dependent pathway in soybean*, GmaxLKR/SDH*-like1, *GmaxLKR/SDH*-like2 and *GmaxADC2*-like1 showed different expression dynamics during water deprivation. The putative paralogs *GmaxLKR/SDH*-like1 and *GmaxLKR/SDH*-like2 displayed a quite similar expression pattern (Figure 5A,B). Moreover, the gene *GmaxADC2*-like1 showed higher levels of expression in the HSys condition (Figure 5C). On the other hand, genes belonging to the ABA-independent pathway presented distinct patterns of gene expression, such as those displayed by *GmaxGOLS2*-like1, *GmaxGOLS2*-like2 and *GmaxGOLS2*-like3 (Figure 5D-F)

Lysine is catabolized in plants from saccharopine to glutamic acid and acetyl-CoA. Lysine catabolism is largely regulated by two enzymes, lysine-ketoglutarate reductase (LKR) and saccharopine dehydrogenase (SDH). These are linked to each other by a single bi-functional protein encoded by a single *LKR/SDH* gene (Arruda *et al.*, 2000; Galili *et al.*, 2001; Anderson *et al.*, 2010) (Figure 2A). The response of *LKR/SDH* gene expression to ABA as well as to biotic and abiotic stresses (Moulin *et al.*, 2000) implies that the Lys catabolism pathway participates in a metabolic networks that helps plants withstand such stresses. A dendrogram analysis allowed us to identify the putative soybean orthologs of *LKR/SDH* (Figure 3A). The analysis also suggests that duplication events occurred in the soybean *LKR/SDH* genes, generating the two genes found in the soybean genome, *GmaxLKR/SDH*-like1 and *GmaxLKR/SDH*-like2 (Figure 3A). This event has already been described in other crop species, such as sugarcane, coffee, cotton, maize and tobacco, and generated a large

number of paralogous genes for *LKR/SDH* (Soltis and Soltis, 1999; Schmutz *et al.*, 2010). This is in accordance with previous studies that indicated two major duplication events in the soybean genome, resulting in a current conformation with almost 75% of the genes represented in multiple copies that were maintained over time (Schmutz *et al.*, 2010). In the gene expression analysis, the *GmaxLKR/SDH*-like1 and *GmaxLKR/SDH*-like2 soybean genes presented quite similar expression regulation indicating that the respective promoter regions may not have diverged among the duplicated genes. However, these genes showed a rather distinct gene expression profile between sensitive and tolerant cultivars in the Psys condition (Figure 5A, B).

Arginine decarboxylase (ADC) is a key plant enzyme that converts arginine into putrescine, an important mediator of abiotic stress tolerance (Figure 2B) (Peremarti *et al.*, 2010). The over-expression of *ADC2* in transgenic *Arabidopsis* showed that higher levels of putrescine increased drought tolerance (Alcazar *et al.*, 2006, 2010). Dendrogram analysis allowed us to identify two paralogs, *GmaxADC2*-like1 (Glyma06g00990) and *GmaxADC2*-like2 (Glyma04g0960) (Figure 3B). An analysis by qPCR was not done for *GmaxADC2*-like2 because previous information from subtractive library data did not indicate its expression during water deficit. The *GmaxADC2*-like1 reached peak expression at a water deficit of -1.5 MPa in the PSys and at the T100 time point in the HSys condition in both cultivars. Interestingly, unlike the *GmaxLKR/SDH*-like1 and *GmaxLKR/SDH*-like2 genes, the expression levels of *GmaxADC2*-like1 were lower in the PSys when compared to the HSys condition (Figure 5C). This indicates that the regulation of *GmaxADC2*-like1 expression may be early and transient after the onset of a water deficit sensitivity.

The conversion of myo-inositol to galactinol or to other raffinose series oligosaccharides (Figure 2C) under abiotic stress was studied in *Arabidopsis* (Seki *et al.*, 2002; Taji *et al.*, 2002; Shinozaki and Yamaguchi-Shinozaki, 2007; Urano *et al.*, 2009,2010). Among the key genes of this pathway, *AtGOLS1* and *AtGOLS2* are the best studied. Their expression patterns are tightly regulated by drought stress and the over-expression of *AtGOLS2* in *Arabidopsis* increases dehydration tolerance (Taji *et al.*, 2002). The neighbor joining analysis suggests that there are six genes in the soybean genome related to *AtGOLS1, AtGOLS2 and AtGOLS3: GmGOLS* (Glyma10g28610), *GmaxGOLS2*-like1 (Glyma20g22700) *GmaxGOLS2*-like2 (Glyma19g40680), *GmaxGOLS2*-like3, *GmaxGOLS2*-like4 (Glyma03g33910) and *GmaxGOLS2*-like5 (Glyma19g41550) (Figure 4). The genes *GmaxGOLS2*-like4 and *GmaxGOLS2*-like5 were not selected for validation by qPCR because they were absent in the subtractive libraries (Table S2). Our analysis in HSys revealed that *GmaxGOLS2*-like1 shows higher levels of gene expression at earlier stages (T50 min) in the tolerant cultivar (Embrapa

48), while the sensitive cultivar (BR16) shows a slower response to water deficit (Figure 5D). A similar expression profile was also observed in PSys, but expression levels were significantly lower when compared with HSys (Figure 5D). In contrast, *GmaxGOLS2*-like2 and *GmaxGOLS2*-like3 were induced exclusively in the PSys condition (Figure 5E, F). Moreover, the expression levels in the tolerant cultivars were dramatically higher under severe stress (Figure 5E, F). This result indicates that the expression of *GmaxGOLS2*-like2 and *GmaxGOLS2*-like3 is not regulated during the sudden water deficit promoted by the HSys treatment, but may be fundamental during the slow adaptation to drought in a PSys condition. The disparity observed in the regulation of gene expression between *GmaxGOLS2*-like1 and the two paralogs *GmaxGOLS2*-like2 and *GmaxGOLS2*-like3 fits with the well-accepted model according to which changes in the transcriptional regulation of duplicated genes play an important role for their fixation in the genome (Carroll, 2000). The distinct regulation of expression of the *GmaxGOLS2* genes may be important to soybean plants to promote tight control of GOLS2 expression under a multitude of environmental conditions.

The analysis of soybean gene promoters, using the POBO tool, revealed a cluster composed of up-regulated genes in PSys or HSys, where the frequency of the ACGT (ERD1), ACGTG (ABRE) and ACGTGKC (ABRE) *cis*-elements is higher. The high frequency of these *cis*-elements suggests that they may function as important regulatory players in genes that participate in different metabolic pathways during drought stress.

The results presented here indicate that several genes of different metabolic pathways have their expression tightly regulated by drought stress in soybean. Moreover, the data show that the dynamics and the expression level can change drastically depending on the drought stress system and also among closely related orthologs. Our work has shed light on the gene expression response of key genes involved in soybean metabolism during drought stress. The information provided here is important to better understand the molecular mechanisms involved in water deficit tolerance in soybean and may contribute to the development of soybean varieties that are more apt to cope with water stress.

## References

Alcazar R, Marco F, Cuevas JC, Patron M, Ferrando A, Carrasco P, Tiburcio AF and Altabella T (2006) Involvement of polyamines in plant response to abiotic stress. Biotechnol Lett 28:1867-1876.

Alcazar R, Planas J, Saxena T, Zarza X, Bortolotti C, Cuevas J, Bitrian M, Tiburcio AF and Altabella T (2010) Putrescine accumulation confers drought tolerance in transgenic *Arabidopsis* plants over-expressing the homologous Arginine decarboxylase 2 gene. Plant Physiol Biochem 48:547-552.

Anderson OD, Coleman-Derr D, Gu YQ and Heath S (2010) Structural and transcriptional analysis of plant genes encoding the bifunctional lysine ketoglutarate reductase saccharopine dehydrogenase enzyme. BMC Plant Biol 10:e113.

Arruda P, Kemper EL, Papes F and Leite A (2000) Regulation of lysine catabolism in higher plants. Trends Plant Sci 5:324-330.

Ashraf M, Akram NA, Al-Qurainy F and Foolad MR (2011) Drought tolerance: Roles of organic osmolytes, growth regulators, and mineral nutrients. Adv Agron 111:249-296.

Baxter CJ, Redestig H, Schauer N, Repsilber D, Patil KR, Nielsen J, Selbig J, Liu JL, Fernie AR and Sweetlove LJ (2007) The metabolic response of heterotrophic *Arabidopsis* cells to oxidative stress. Plant Physiol 143:312-325.

Brando PM, Goetz SJ, Baccini A, Nepstad DC, Beck PSA and Christman MC (2010) Seasonal and interannual variability of climate and vegetation indices across the Amazon. Proc Natl Acad Sci USA 107:14685-14690.

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

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.

Bundy JG, Davey MP and Viant MR (2009) Environmental metabolomics: A critical review and future perspectives. Metabolomics 5:3-21.

Busk PK and Pages M (1998) Regulation of abscisic acid-induced transcription. Plant Mol Biol 37:425-435.

Carroll SB (2000) Endless forms: The evolution of gene regulation and morphological diversity. Cell 101:577-580.

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.

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.

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.

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, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K and Shinozaki K (2006) Crosstalk between abiotic and biotic stress responses: A current view from the points of convergence in the stress signaling networks. Curr Opin Plant Biol 9:436-442.

Galili G, Tang GL, Zhu XH and Gakiere B (2001) Lysine catabolism: A stress and development super-regulated metabolic pathway. Curr Opin Plant Biol 4:261-266.

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.

Hey SJ, Byrne E and Halford NG (2010) The interface between metabolic and stress signalling. Ann Bot 105:197-203.

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.

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

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.

Hsiao TC (1973) Plant responses to water stress. Annu Rev Plant Physiol 24:519-570.

Hummel I, Pantin F, Sulpice R, Piques M, Rolland G, Dauzat M, Christophe A, Pervent M, Bouteille M, Stitt M, *et al.* (2010) *Arabidopsis* plants acclimate to water deficit at low cost through changes of carbon usage: An integrated perspective using growth, metabolite, enzyme, and gene expression analysis. Plant Physiol 154:357-372.

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.

Kishor PBK, Hong ZL, Miao GH, Hu CAA and Verma DPS (1995) Overexpression of delta-pyrroline-5-carboxylate synthetase increases proline production and confers osmotolerance in transgenic plants. Plant Physiol 108:1387-1394.

Kiyosue T, Yoshiba Y, YamaguchiShinozaki K and Shinozaki K (1996) A nuclear gene encoding mitochondrial proline dehydrogenase, an enzyme involved in proline metabolism, is upregulated by proline but downregulated by dehydration in Arabidopsis. Plant Cell 8:1323-1335.

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.

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.

Moulin M, Deleu C and Larher F (2000) L-Lysine catabolism is osmo-regulated at the level of lysine-ketoglutarate reductase and saccharopine dehydrogenase in rapeseed leaf discs. Plant Physiol Bioch 38:577-585.

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.

Peremarti A, Bassie L, Zhu CF, Christou P and Capell T (2010) Molecular characterization of the Arginine decarboxylase gene family in rice. Transgenic Res 19:785-797.

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.

Sanchez DH, Siahpoosh MR, Roessner U, Udvardi M and Kopka J (2008) Plant metabolomics reveals conserved and divergent metabolic responses to salinity. Physiol Plantarum 132:209-219.

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, *et al.* (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Seki M, Ishida J, Narusaka M, Fujita M, Nanjo T, Umezawa T, Kamiya A, Nakajima M, Enju A, Sakurai T, *et al.* (2002) Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. Funct Integr Genomics 2:282-291.

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 Biotech 14:194-199.

Shinozaki K and Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: Differences and cross-talk between two stress signaling pathways. Curr Opin Plant Biol 3:217-223.

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.

Soltis DE and Soltis PS (1999) Polyploidy: Recurrent formation and genome evolution. Trends Ecol Evol 14:348-352.

Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K and Shinozaki K (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. Plant J 29:417-426.

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.

Thomashow MF (1999) Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol 50:571-599.

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.

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, Kurihara Y, Seki M and Shinozaki K (2010) 'Omics' analyses of regulatory networks in plant abiotic stress responses. Curr Opin Plant Biol 13:132-138.

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.

Verbruggen N and Hermans C (2008) Proline accumulation in plants: A review. Amino Acids 35:753-759.

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.

Zhang JD and Wiemann S (2009) KEGGgraph: A graph approach to Kegg pathway in R and bioconductor. Bioinformatics 25:1470-1471.

Zhang PF, Foerster H, Tissier CP, Mueller L, Paley S, Karp PD and Rhee SY (2005) MetaCyc and AraCyc. Metabolic pathway databases for plant research. Plant Physiol 138:27-37.

Zhao S and Fernald RD (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. J Comput Biol 12:1047-1064.

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

## Internet Resources

The Arabidopsis Information Resource, TAIR site, http://www.Arabidopsis.org (August 1, 2010).

Genevestigator shaping biological discovery, http://www.genevestigator.com/gv/index.jsp (August 15, 2010).

Soybean Genome Project GENOSOJA LGE,
    http://bioinfo03.ibi.unicamp.br/soja (August 25, 2010).

Phytozome, http://www.phytozome.net/soybean v6.0 (August
    20, 2010).

ClustalW2-Multiple Sequence Alignment,
    http://www.ebi.ac.uk/Tools/clustalw2/index.html (September 10, 2010).

Plant Cis program-acting Regulatory DNA Elements, PLACE,
    http://www.dna.affrc.go.jp/PLACE/ (September 10, 2010).

POBO tool, http://ekhidna.biocenter.-helsinki.fi:9801/pobo
    (February 2, 2011).

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

## Supplementary Material

The following online material is available for this article:

Table S1 - Sequences and features of primers used in this study.

Table S2 - Prevalence of soybean matches in different metabolic pathways responsive to drought in the subtractive libraries.

Table S3 - Transcription factor binding site verification performed with the POBO tool.

**Table S1** - Sequences and features of primers used in this study.

| Gene model | Forward primer sequence [5'3'] | Reverse primer sequence [5'3'] | Amplicon length (pb) |
|---|---|---|---|
| *GmaxACT11* | CGGTGGTTCTATCTTGGCATC | GTCTTTCGCTTCAATAACCCTA | 142 |
| *GmaxFBOX* | AGATAGGGAAATGGTGCAGGT | CTAATGGCAATTGCAGCTCTC | 93 |
| *GmaxLKR/SDH1* | ATCCTGCCACCTACAAATGG | ACGGAAAATGGTTGATGCTT | 182 |
| *GmaxLKR/SDH 2* | GGGGAATGGTGTGATATGCT | ATTGGCTATGCAAGCTCTCC | 166 |
| *GmaxADC2* | CAGGAGTATGTCAGCCACGA | CAGATCTTGAGCAGCAGGAA | 144 |
| *GmaxGOLS2 like-1* | CCTGAGAACGTTGAGCTTGA | CCACCACTTCTTCACCAACA | 132 |
| *GmaxGOLS2 Like-2* | AGTCACCACTCCCACTTCGT | CCCGTATATCTCCACGGTTT | 192 |
| *GmaxGOLS2 Like-3* | TTGCCATGGCTTATTACGTC | TACCTCAATGTCTCCGTCCA | 98 |

**Table S2** - Prevalence of soybean matches regarding different metabolic pathways respor

| Metabolic Pathways | *Arabidopsis thaliana* Gene | Soybean Matches | L1_T25-50min (BR16) | L2_T75-100min (BR16) | L3_T125-150 min (BR16) | L4-T25-T50min (EMB48) | L5-T75-T100min (EMB48) | L6-T125-T150min (EMB48) | R1_T25-T50min (BR16) | R2_T75-T100min (BR16) | R3_T125-T150 min (BR16) | R4_T25-50min (EMB48) | R5_T75-100min (EMB48) | R6T125-150min (EMB48) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Glutamate degradation IV | At5g17330 | Glyma02g40840.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39170.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma18g04940.1 | **Yes** | No | No | No | No | No | No | No | No | No | **Yes** | **Yes** |
| | | Glyma09g29900.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g34450.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma08g09670.1 | **Yes** | **Yes** | **Yes** | No | No | No | No | No | No | **Yes** | **Yes** | **Yes** |
| | | Glyma08g09660.1 | No | No | No | No | No | No | No | No | No | **Yes** | No | No |
| | | Glyma05g26660.1 | **Yes** | No | **Yes** | No | No | No | No | No | No | **Yes** | **Yes** | **Yes** |
| | | Glyma08g09650.1 | **Yes** | **Yes** | **Yes** | No | No | No | No | No | No | No | No | No |
| | | Glyma11g33280.1 | No | **Yes** | **Yes** | No | No | No | No | No | No | No | No | **Yes** |
| | At3g22200 | Glyma12g02510.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g02510.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At3g17720 | Glyma09g29900.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g34450.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma18g04940.1 | **Yes** | No | No | No | No | No | No | No | No | No | **Yes** | **Yes** |
| | | Glyma08g09660.1 | No | No | No | No | No | No | No | No | No | **Yes** | No | No |
| | | Glyma08g09670.1 | **Yes** | **Yes** | **Yes** | No | No | No | No | No | No | **Yes** | **Yes** | **Yes** |
| | At1g79440 | Glyma15g41690.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma08g17450.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Methionine biosynthesis II | At4g23590 | Glyma13g37080.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g33350.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g26170.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g35630.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At4g23600 | Glyma06g35580.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma13g37080.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g33350.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g26170.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g35630.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g35580.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At3g22740 | Glyma20g28720.5 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma20g28720.4 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma20g28720.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma20g28720.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g34120.1 | No | No | No | No | No | No | No | No | No | No | No | No |

| Pathway | At gene | Glyma gene | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Stachyose biosynthesis | At1g56600 | Glyma03g38080.1 | No | No | No | No | **Yes** | **Yes** | No | No | No | No | No | No |
| | | Glyma03g38910.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g40680.1 | No | No | No | **Yes** | **Yes** | **Yes** | No | No | No | No | No | No |
| | | Glyma19g41550.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma20g22700.1 | No | No | No | No | **Yes** | **Yes** | No | No | No | **Yes** | **Yes** | **Yes** |
| | At3g57520 | Glyma03g29440.1 | No | No | No | **Yes** | **Yes** | **Yes** | No | No | No | No | No | **Yes** |
| | | Glyma14g01430.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma02g47330.1 | No | No | No | **Yes** | No | No | No | No | No | No | No | No |
| | | Glyma17g11970.1 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g11970.3 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g11970.4 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma13g22890.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma04g36410.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g01430.2 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g11970.2 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| | | Glyma09g01940.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| TCA cycle variation III (eukaryotic) | At2g18450 | Glyma11g07250.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma05g08550.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g38200.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g38200.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At5g66760 | Glyma11g07250.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g07250.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma02g06400.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma01g38200.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g38200.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma05g08550.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At3g47520 | Glyma17g10880.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g10880.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g10880.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g34190.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma05g01010.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At1g53240 | Glyma17g10880.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g19520.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g04720.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma07g30430.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g34190.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Ascorbate biosynthesis I (L-galactose pathway) | At1g70820 | Glyma20g02220.1 | No | No | No | No | No | **Yes** | No | No | No | No | No | No |
| | At2g45790 | Glyma18g46390.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma09g39800.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma15g12230.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At3g02870 | Glyma09g01380.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma07g39620.1 | No | No | No | No | No | No | No | No | No | No | No | No |

| Category | At ID | Gene | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Glyma11g10320.1 | No | **Yes** | No | No | No | **Yes** | No | No | No | **Yes** | No | **Yes** |
| | | Glyma12g02610.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g02140.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g02240.1 | No | No | **Yes** | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma14g39880.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g38120.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At4g38420 | Glyma14g39880.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g31920.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g46350.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g01580.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g10420.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma07g39160.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g21490.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma01g38980.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g06290.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At5g66920 | Glyma11g06290.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g06290.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g21530.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g10320.1 | No | **Yes** | No | No | No | **Yes** | No | No | No | **Yes** | No | **Yes** |
| | | Glyma04g02140.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g38120.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At1g21850 | Glyma12g02610.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g46350.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g31920.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g01580.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g10420.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g07780.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At1g07890 | Glyma12g07780.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g07780.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g33700.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma18g04510.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At5g16710 | Glyma11g33700.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma20g38440.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma10g43730.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g01580.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Ascorbate glutathione | | Glyma07g39160.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g38120.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.1 | No | No | No | No | No | No | No | No | No | No | No | No |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| glutathione cycle | At4g22010 | Glyma04g02140.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g02610.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g02240.1 | No | No | **Yes** | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma12g31920.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g46350.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g10420.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma07g39160.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g38980.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g06290.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g06290.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g06290.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g21490.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| | At1g76160 | Glyma04g02140.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g02240.1 | No | No | **Yes** | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma17g38120.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g02610.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g39880.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g46350.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g31920.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g10420.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g01580.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma07g39160.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g46350.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g38980.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g21490.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma11g06290.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g06290.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g06290.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma07g39160.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At1g75790 | Glyma11g36390.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g10420.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g46350.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g02140.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g02240.1 | No | No | **Yes** | No | No | No | No | No | No | No | **Yes** | No |
| | At5g03630 | Glyma10g07820.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma0169s00210.1 | No | No | No | No | No | **Yes** | No | No | No | **Yes** | **Yes** | **Yes** |
| | | Glyma16g07970.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g14500.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma08g02100.1 | No | No | No | No | No | No | No | No | No | No | No | No |

| Category | At-gene | Glyma-gene | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ascorbate biosynthesis I (L-galactose pathway) | At2g39770 | Glyma10g07820.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma0169s00210.1 | No | No | No | No | No | Yes | No | No | No | Yes | Yes | Yes |
| | | Glyma16g07970.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g14500.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma08g02100.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At4g30570 | Glyma18g03840.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g34550.1 | No | No | No | No | No | No | No | No | No | Yes | No | No |
| | | Glyma14g07150.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma02g41820.1 | No | No | No | No | No | No | No | No | No | Yes | Yes | Yes |
| | At4g39120 | Glyma04g01170.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g01170.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g01210.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g01170.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| Ascorbate and aldarate metabolism | At4g09010 | Glyma02g37160.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g35440.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At4g35970 | Glyma11g08320.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g03610.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g11460.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Glutamate biosynthesis IV, ammonia assimilation cycle II | At5g04140 | Glyma03g28410.2 | No | No | No | No | No | No | Yes | No | No | No | No | No |
| | | Glyma03g28410.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g31120.2 | No | No | No | No | No | No | Yes | No | No | No | No | No |
| | | Glyma19g31120.1 | No | No | No | No | No | No | Yes | No | No | No | No | No |
| | | Glyma14g32500.1 | No | Yes | Yes | No | No | No | No | No | No | No | No | Yes |
| | | Glyma06g13280.1 | Yes | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g13280.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At2g41220 | Glyma04g41540.1 | No | No | No | Yes | No | No | No | No | No | No | Yes | No |
| | | Glyma19g16450.1 | No | No | No | No | No | No | No | No | No | No | Yes | Yes |
| | At5g53460 | Glyma06g13280.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g41540.1 | No | No | No | Yes | No | No | No | No | No | No | Yes | No |
| | | Glyma14g32500.1 | No | Yes | Yes | No | No | No | No | No | No | No | No | Yes |
| | | Glyma19g16450.1 | No | No | No | No | No | No | No | No | No | No | Yes | Yes |
| | | Glyma06g13280.1 | Yes | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma03g28410.2 | No | No | No | No | No | No | Yes | No | No | No | No | No |
| | | Glyma03g28410.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g31120.2 | No | No | No | No | No | No | Yes | No | No | No | No | No |
| | | Glyma19g31120.1 | No | No | No | No | No | No | Yes | No | No | No | No | No |
| | At2g41220 | Glyma03g28410.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma03g28410.2 | No | No | No | No | No | No | Yes | No | No | No | No | No |
| | | Glyma19g31120.2 | No | No | No | No | No | No | Yes | No | No | No | No | No |
| | | Glyma19g31120.1 | No | No | No | No | No | No | Yes | No | No | No | No | No |
| | | Glyma06g13280.1 | Yes | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g13280.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g41540.1 | No | No | No | Yes | No | No | No | No | No | No | Yes | No |
| | | Glyma01g41310.1 | No | No | No | No | No | No | No | No | No | No | No | No |

| Pathway | At gene | Glyma gene | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Glutamate degradation I | At5g07440 | Glyma05g05460.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g15740.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g04560.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g04560.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g04560.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g28770.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g28770.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g26940.1 | No | **Yes** | **Yes** | No | No | No | No | No | No | No | **Yes** | **Yes** |
| | | Glyma02g07940.1 | No | **Yes** | No | No | No | No | No | No | No | **Yes** | **Yes** | **Yes** |
| Phosphatidyleth anolamine biosynthesis I, II | At1g15110 | Glyma10g41430.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| | At2g26830 | Glyma04g18940.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Starch degradation | At3g23920 | Glyma09g29840.1 | No | No | No | **Yes** | **Yes** | **Yes** | No | No | No | **Yes** | **Yes** | **Yes** |
| | | Glyma16g34360.1 | No | No | No | No | **Yes** | **Yes** | No | No | No | No | No | **Yes** |
| Glycine biosynthesis | At1g08630 | Glyma20g36590.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma20g36590.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma10g30880.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma10g30880.2 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| | | Glyma10g30880.1 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| | | Glyma03g41120.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| Valine biosynthesis | At1g10070 | Glyma06g05280.4 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g05280.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g05190.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g05190.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Phenylalanine biosynthesis II | At5g22630 | Glyma12g30660.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g05290.1 | No | No | No | No | No | **Yes** | No | No | No | **Yes** | No | No |
| | | Glyma11g19430.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g09050.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g01610.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Phenylalanine degradation III | At1g62960 | Glyma10g40140.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma07g13320.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma11g03070.1 | No | No | No | No | No | No | No | No | No | **Yes** | No | No |
| | At5g53970 | Glyma06g35580.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At5g11520 | Glyma01g42290.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g08670.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g08560.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Tryptophan biosynthesis | At2g28880 | Glyma20g31970.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma10g35580.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At3g55870 | Glyma18g03270.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma20g23680.1 | No | No | **Yes** | No | No | No | No | No | No | No | No | No |
| | | Glyma18g03260.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g05810.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma02g42680.1 | No | No | No | No | No | No | No | No | No | No | No | No |

| Pathway | At gene | Glyma gene | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | At5g48220 | Glyma14g05810.4 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g05810.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g05810.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| UDP-L-arabinose biosynthesis I (from UDP-xylose) | At2g34850 | Glyma05g30410.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At4g20460 | Glyma09g03490.1 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| | | Glyma09g03490.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma08g13540.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma09g03490.2 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| | At1g30620 | Glyma05g30410.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma08g13540.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma09g03490.1 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| | | Glyma09g03490.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma09g03490.2 | No | **Yes** | No | No | No | No | No | No | No | No | No | No |
| Glutamine biosynthesis I | At5g35630 | Glyma13g28180.4 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma13g28180.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma13g28180.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma13g28180.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma15g10890.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma15g10890.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma15g10890.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At3g17820 | Glyma09g30370.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Histidine biosynthesis | At5g10330 | Glyma16g27220.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g27220.1 | No | No | **Yes** | No | No | No | No | No | No | No | No | No |
| | At1g71920 | Glyma16g27220.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g27220.1 | No | No | **Yes** | No | No | No | No | No | No | No | No | No |
| | At5g63890 | Glyma15g13910.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma09g02960.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma08g08630.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At1g09795 | Glyma19g36070.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma03g33360.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Putrescine biosynthesis I, II,IV | At4g08870 | Glyma17g14040.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma03g03270.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g33750.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g33640.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At4g08900 | Glyma17g14040.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma03g03270.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g33750.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g33640.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At4g34710 | Glyma06g00990.1 | No | No | No | No | No | **Yes** | No | No | No | **Yes** | **Yes** | **Yes** |
| | | Glyma04g00960.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| Proline degradation II | At5g38710/At3g30775 | Glyma18g51400.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma13g07110.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma08g28460.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g05570.1 | No | No | No | No | No | No | No | No | No | **Yes** | **Yes** | **Yes** |

| Pathway | At gene | Glyma gene | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Glyma19g05580.1 | No | No | No | No | No | No | No | No | No | **Yes** | No | No |
| Proline biosynthesis II (from arginine) | At5g46180 | Glyma05g27360.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma08g10340.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma05g27360.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| IAA biosynthesis I | At1g70560 | Glyma01g03340.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma02g04270.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g36040.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g18880.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g09400.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Galactose degradation III | At1g22430 | Glyma14g04610.1 | No | No | No | No | No | No | No | No | No | **Yes** | No | No |
| | | Glyma20g10240.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma02g44170.1 | No | No | No | No | **Yes** | No | No | No | No | No | No | No |
| | | Glyma02g44160.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma20g10240.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At1g26570 | Glyma02g12870.1 | No | No | No | No | No | No | No | No | No | No | No | **Yes** |
| | | Glyma01g06970.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma08g26520.1 | **Yes** | **Yes** | **Yes** | No | No | No | No | No | No | **Yes** | **Yes** | **Yes** |
| | | Glyma18g50000.1 | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | No | No | No | **Yes** | **Yes** | **Yes** |
| | | Glyma19g03500.1 | No | No | **Yes** | No | No | No | No | No | No | **Yes** | **Yes** | No |
| | | Glyma13g06050.1 | No | No | **Yes** | No | No | No | No | No | No | **Yes** | No | No |
| | | Glyma05g00590.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Xylose degradation I | At5g57655 | Glyma17g07380.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At5g49650 | Glyma04g09340.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g09490.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| sine degradation | At4g33150 | Glyma17g04920.1 | No | No | No | **Yes** | No | **Yes** | No | No | No | **Yes** | **Yes** | No |
| | | Glyma13g17580.1 | No | No | No | **Yes** | No | **Yes** | No | No | No | No | No | No |
| Chorismate biosynthesis | At2g21940 | Glyma08g14980.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Tyrosine degradation I | At1g12050 | Glyma15g12100.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma09g01270.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At1g06570 | Glyma14g03410.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | At5g54080 | Glyma12g20220.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma06g34940.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| phenylethanol biosynthesis | At5g19440 | Glyma18g10270.1 | No | No | No | **Yes** | **Yes** | **Yes** | No | No | No | No | No | **Yes** |
| | | Glyma18g10260.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g02240.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g02250.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma12g02230.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Glucosinolate biosynthesis from tryptophan | At5g57220 | Glyma16g26520.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Galactose | | Glyma04g39190.1 | No | No | No | No | No | No | No | No | No | No | No | No |

| Pathway | At gene | Glyma | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Galactose degradation III | At1g77120 | Glyma06g12780.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma04g41990.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Spermidine biosynthesis and spermine biosyntesis | At5g15950 | Glyma02g14180.1 | **Yes** | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma02g14180.2 | **Yes** | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g10080.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Tyrosine biosynthesis II | At5g34930 | Glyma17g13150.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma05g07870.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma13g06340.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Glutamate degradation I | At5g18170 | Glyma16g04560.3 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g04560.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g04560.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g28770.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma19g28770.2 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma16g26940.1 | No | **Yes** | **Yes** | No | No | No | No | No | No | No | **Yes** | **Yes** |
| | | Glyma02g07940.1 | No | **Yes** | No | No | No | No | No | No | No | **Yes** | **Yes** | **Yes** |
| | | Glyma05g05460.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma17g15740.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma01g41310.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Ascorbate glutathione cycle | At1g55570 | Glyma11g36390.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma07g35180.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma20g03030.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma07g35170.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| Choline biosynthesis III | At3g25585 | Glyma12g08720.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma02g14210.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| Jasmonic acid biosynthesis | At4g16760 | Glyma05g04940.1 | No | No | No | No | No | No | No | No | No | No | **Yes** | No |
| | | Glyma11g03800.1 | No | No | No | **Yes** | **Yes** | **Yes** | No | No | No | **Yes** | No | **Yes** |
| | | Glyma01g41600.1 | No | No | No | No | No | **Yes** | No | No | No | No | No | No |
| | | Glyma17g15320.1 | No | No | No | No | No | No | No | No | No | No | No | No |
| | | Glyma14g14990.1 | No | No | No | No | No | No | No | No | No | No | No | No |

Genes present in the subtractive libraries are represented by "Yes" and absent genes are represented by "No".

nt times. Sensitive (BR16) and tolerant (Embrapa48) cultivars are indicated in relation to the different times and tissues evaluated.

**Table S3 -** Results obtained after transcription factor binding site verification performed with POBO tool.

| Motif | Data set | Number of promoters in each dataset | Number of promoters containing the pattern | | Total number of patterns in each dataset | Promoter mean |
|-------|----------|------|--------|-----------|-------|-------|
| ACGT | BG | 77222 | 55801 | ( 72.3%) | 233388 | 3.03 |
|  | Cluster 1 | 11 | 9 | (81.8%) | 66 | 5.96 |
| ACGTGKC | BG | 77222 | 8532 | (11.0%) | 10189 | 0.13 |
|  | Cluster 1 | 11 | 6 | (54.5%) | 16 | 1.46 |
| ACGTG | BG | 77222 | 37257 | (48.2%) | 68328 | 0.88 |
|  | Cluster 1 | 11 | 6 | (54.5%) | 28 | 2.55 |

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 1,000. The symbol K was used in addition to A or T. Calculated t-test using the linked on line GrapPad web site: ttp://www.graphpad.com/quickcalcs/DistMenu.cfm; $p < 0.0001$.