*Article*

# Proteomic and Metabolomic Analysis of a Drought Tolerant Soybean Cultivar from Brazilian Savanna

**Lucas Leal Lima [1], Bruno Pereira Balbi [1], Rosilene Oliveira Mesquita [2], José Cleydson Ferreira da Silva [1], Flaviane Silva Coutinho [1], Flavia Maria Silva Carmo [3], Camilo Elber Vital [1,4], Angela Mehta [5], Marcelo Elher Loureiro [6], Elizabeth Pacheco Batista Fontes [1], Everaldo Gonçalves Barros [7], Humberto Josué Oliveira Ramos [1,3,*]**

[1] Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Viçosa, Viçosa, Minas Gerais 36570-000, Brazil
[2] Departamento de Fitotecnia, Universidade Federal do Ceará, Fortaleza, Ceará 60020-181, Brazil
[3] Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, Minas Gerais 36570-000, Brazil
[4] Núcleo de Analise de Biomoléculas (NuBioMol), Centro de Ciências Biológicas, Universidade Federal de Viçosa, Viçosa, Minas Gerais 36570-000, Brazil
[5] Embrapa Recursos Genéticos e Biotecnologia, Cenargen, Brasília, DF 770770-917, Brazil
[6] Departamento de Biologia Vegetal, Universidade de Viçosa, Viçosa, Minas Gerais 36570-000, Brazil
[7] Universidade Católica de Brasília, Brasília, DF 71966-700, Brazil
* Correspondence: Humberto Josué Oliveira Ramos, Email: humramos.ufv.br@gmail.com; Tel.: +55-31-3899-3437.

## ABSTRACT

Drought is the major abiotic stress which contributes to the reduction in the productivity of soybeans in Brazil, especially in the savanna regions. Two parental genotypes contrasting in drought tolerance (Embrapa 48–tolerant and BR 16–sensitive) were used to study the molecular mechanism underlying this process in soybeans. The hydric potential of the BR 16 reached values of −1.0 and −1.5 MPa, three days before the Embrapa 48, confirming the contrasting drought tolerance. The proteomic, phosphoproteomic and metabolomic profiles were evaluated to detect the metabolic pathways, which were affected by the drought stress. An integrative overview showed that the tolerant plants maintain cell homeostasis and the photosynthetic metabolism was unchanged under the stress condition in contrast to the sensitive genotype that showed several dysregulated pathways. These findings were confirmed by the protein expression and protein regulation by phosphorylation. Furthermore, just small deviations in the metabolic pathways were observed for drought-tolerant plants in comparison to the sensitive genotype. Complex post-translational modification patterns by phosphorylation were detected in some key enzymes, such as carbonic

**Open Access**

Received: 19 June 2019
Accepted: 17 October 2019
Published: 22 October 2019

anhydrase, rubisco activase, transketolase and RNA binding proteins during water deficiency. Osmoprotection does not appear to be the major mechanism for tolerance, as indicated by the accumulation of the metabolite and the phytohormone profiles from tolerant and sensitive soybean plants. Thus, regulatory cascades of the metabolic activities, mediated by protein phosphorylation, may cause a higher water use efficiency in the leaves as well as the translocation from the root to the shoot system.

**KEYWORDS:** soybean; drought tolerance; proteome; metabolome; phosphoproteome

## INTRODUCTION

Water scarcity results in huge reductions in crop yield and is one of the greatest limitations on the expansion of agriculture areas. Although the total annual precipitation in Brazil is sufficient for soybean cultivation, water deficiency could be caused by a dry spell of many weeks, without any rain in the growth period in a particular location. As a result, soybean yields could be affected by a water deficit, particularly during flowering and early pod expansion [1]. Drought stress induces several changes at morphological, physiological and biochemical levels in all plant organs. On the other hand, plants have developed several strategies in dealing with drought stress, including a short life cycle, enhanced water uptake and reduced water loss, as well as osmotic adjustment, antioxidant capacity, and desiccation tolerance [2].

Metabolic adjustments in response to the environmental conditions may alter pools of metabolites that play important roles in the physiological response. The study of these metabolites may indicate which pathways have been disturbed by the stress [3–5] and could be used to indicate the cell adjustment during drought [6]. Some of the most important responses of plants against drought stress are related to the accumulation of osmoprotectant metabolites, which contribute to water retention in stressed cells [7]. These osmoprotectants maintain relatively low cellular osmotic potential and may greatly differ in their chemical features and concentrations among plant species.

Several genes respond to water stresses at pre- and post-transcriptional and translational levels [8,9]. Understanding the role of stress-induced proteins is important in explaining the plant tolerance processes. Therefore, the study of proteomics is an approach that could be used to discover proteins and pathways associated with crop physiological and stress responses [10]. Analyses of the soybean leaf proteome have been used to identify the differential expression of various abiotic stress-responsive proteins that regulate many molecular processes and signaling cascades [9,11,12]. On the other hand, the metabolomic profiling shows the metabolites involved in the tolerance to drought and heat stress [5,13].

Thus, studying metabolome and proteome of the plants could help in understanding the pathways involved in water stress tolerance.

Brazil is the second leading producer of soybean in the world and has the potential to become the largest due to increases in cultivated areas in recent years (OCDE-FAO, 2017). However, the effects of drought which is the major challenge for soybean production in Brazilian savannas must be reduced. Therefore, the development of soybean cultivars which are drought-tolerant is required in order to improve their productivity and meet the ever-increasing demand for soybean grains worldwide. Therefore, it is important to understand the mechanism of drought tolerance in soybean plants.

It has already been reported that soybean plants can adapt to these conditions by triggering protective mechanisms that enable them to survive drought conditions [14,15]. The development of two genotypes; Brazilian soybeans (*Glycine max* L. Merrill), Embrapa 48 and BR 16, were evaluated under drought conditions in a greenhouse [16] and in the field [17]. The results showed that the Embrapa 48 cultivar may be characterized as drought-tolerant and the BR 16 cultivar as drought-sensitive. Physiological responses and gene expression profiles were also evaluated to understand the mechanism of tolerance for these cultivars [17,18]. Until now, no studies have investigated proteome and metabolome, so this research can lead to a new understanding of the contrasting molecular mechanisms for drought tolerance between these genotypes that share a common ancestor.

Therefore, in this work, we performed a combination of biochemical, proteomics and metabolomics approaches to investigate the molecular components that have shown greater variations when compared to the responsive profiles of the Embrapa 48 and BR 16 cultivars. To achieve this goal, we focused on metabolites and proteins responsive to stress conditions that directly affect the drought tolerance between genotypes. Our results showed an agreement between the metabolic response and the physiological status, that allowed for higher rates of leaf growth, which was not caused by osmoprotection by amino acids biosynthesis or by ABA-dependent cascades. Proteomic and metabolomics data indicate that the metabolism of the Embrapa 48 tolerant genotype is not impaired, when compared to the sensitive BR 16. Thus, regulatory cascades of the metabolic activities, mediated by protein phosphorylation, may be operating for a more efficient water use.

## MATERIALS AND METHODS

### Plant Growth and Drought Stress Treatments

The soybean genotypes BR 16 and Embrapa 48 which are sensitive and tolerant to drought, respectively, were selected for a study based on previous experiments [16,17]. We prepared 6 pots containing 10 L of a mixture of soil, sand, and dung (3:1:1). 3 pots were used to sow each of the

soybean cultivars BR 16 and Embrapa 48, using 3 seeds in each pot. The pots were taken to a greenhouse and kept under natural light and photoperiod, with relative humidity varying daily between 65% and 85% and a temperature varying between 15 °C and 35 °C. The seedlings of both cultivars were watered once a day with 100 mL of water, for 43 days, until the pre-flowering period (stage V4), when the water supply was interrupted to set the experimental treatments.

The experimental design was completely randomized, in a 3 × 2 factorial design. The first factor was the plants' water potential (−1.5 MPa, −1.0 MPa and control) and the second was the two different cultivars. The water regimes were defined as irrigated (IR), related to the control treatment, and the non-irrigated treatment (NI Ψ1.0 and NI Ψ1.5). We used a Scholander pressure pump to measure the water potential of the plants and set the different treatments. In the morning (Ψam) (4:30–5:30 h) we monitored the water pressure in the fourth leaf from the apical meristem of each plant, according to DaMatta [19]. We cut off the individual leaf when the water potential of the plants reached −1 MPa and −1,5 MPa, and then stored them in liquid nitrogen at −80 °C. For each treatment 3 pots were used and each pot contained 3 plants. A trifoliate leaf from each plant from one pot was collected together (3 plants by replicate). Thus, the biochemical analyses were performed using 3 distinct pools, resulting in 3 biological replicates. These procedures were performed on both plant cultivars.

**Protein Extraction and Fractionation by PEG 15%**

The protein extraction of the soybean leaves was performed on the plants under drought stress (−1 MPa) and on the irrigated controls, according to the method for fractionation/precipitation of abundant proteins, described by Aryal [20] with some modifications. All the washing and centrifugation steps were performed using 50 mL centrifuge tubes or 15 mL Falcon tubes. The soybean leaves were grounded with liquid nitrogen in pre-cooled porcelain mortars. The powdered tissue (approximately 3 g) was suspended in a 10 mL extraction buffer [0.5 M Tris-HCl (pH 8.0), 20 mM $MgCl_2$, 2% Triton X-100, 2% β-mercaptoethanol, 0.2% protease inhibitor cocktail (P-9599, Sigma Aldrich, San Luiz, USA), 1% phosphatase inhibitor cocktail (P-0044, Sigma Aldrich, San Luiz, USA), 25 mM Sodium Fluoride (NaF), 1 mM Sodium Molybdate ($Na_2MoO_4$) and 0.5% polyvinylpolypyrrolidone (PVPP)]. After homogenization in a vortex for 30 s, the solution was filtered using four layers of gauze and the filtrate was transferred into a new 50 mL tube. A 10 mL extraction buffer containing 30% polyethylene glycol 4000 (PEG 4000, Sigma-Aldrich, San Luiz, USA) was added to the filtrate. The solution was homogenized again using a vortex for 30 s and kept on ice for 30 min, followed by centrifugation at 3200× *g* for 20 min at 4 °C. After the two extraction procedures, the supernatants and the pellet (designed as PEG-pellet) were collected for further analysis.

The supernatant, containing low abundance proteins, was transferred into a new 50 mL tube. The protein precipitation was carried out by adding 20 mL of 20% trichloroacetic acid (20% TCA) to the tube, and kept overnight at −20 °C before centrifuging the solution at 6000× *g* for 30 min at 4 °C. The supernatant was discarded, and the pellet was then collected, dried at room temperature, followed by a solubilization in 10 mL of dense SDS-buffer (30% sucrose, 2% SDS, 0.1 M Tris-HCl pH 8.0 and 2% β-mercaptoethanol). The next step was to add 5 mL of buffered phenol to the solution, homogenized using a vortex for 30 s and transferred into a 15 mL tube. The solution was kept on ice for 30 min, while it was homogenized using a vortex 4 times, and then centrifuged at 6000× *g* for 30 min at 4 °C. The phenolic phase was collected and transferred into a 50 mL tube, followed by adding 5 volumes of 0.1 M ammonium acetate in methanol in order to precipitate the proteins. The tube was kept overnight at −20 °C. After that, the solution was centrifuged at 6000× *g* for 30 min at 4 °C. The supernatant was discarded and the pellet was washed performing 3 cycles of resuspension and centrifugation at 6000× *g* for 10 min at 4 °C using 80% acetone (ice cold) and 1 cycle using 70% ethanol. After being washed, the pellet was dried at room temperature. The protein extract was solubilized in a solution containing 7 M urea, 2 M thiourea, and 2.0% CHAPS. The resulting solution was sonicated using an UltraSonic Processor (Model GE 50) for cycles of 5 to 10 s until complete solubilization of the pellet. The sample was then stored at −80 ºC.

The pellet-PEG, containing high abundance of proteins, was solubilized in SDS-dense buffer (30% sucrose, 2% SDS, 0.1M Tris-HCl pH 8.0 and 2% β-mercaptoethanol) at room temperature. After the solubilization, 5 mL of buffered phenol was added, and the solution was homogenized using a vortex for 30 s and transferred to a 15 mL tube. The 15 mL tube was kept on ice for 30 min, using a vortex every 4 min and centrifuged at 6000× *g* for 30 min at 4 °C. The phenolic phase obtained after the centrifugation was collected and transferred into a 50 mL tube, and 5 volumes of 0.1 M ammonium acetate in methanol were added to precipitate the proteins. The tube was kept overnight at −20 °C, before centrifuging the solution at 6000× *g* for 30 min at 4 °C. The supernatant was then discarded and the pellet obtained was collected, washed (3 cycles of resuspension and centrifugation at 6000× *g* for 10 min at 4 °C) with 80% acetone (ice cold) and 70% ethanol before being dried at room temperature. The protein extracted was solubilized with 7 M urea, 2 M thiourea, and 2.0% CHAPS. The solution was sonicated using an UltraSonic Processor (Model GE 50, Cole-Parmer, Vernon Hills, USA) for cycles of 5 to 10 seconds until complete solubilization of the pellet. The sample was then stored at −80 °C.

### Separation of Proteins by Two-Dimensional Gel Electrophoresis (2DE)

The isoelectric focusing (IEF) was performed in a 13 cm IPG strip pH 3–10 (GE Healthcare, Chicago, USA) as described by Coutinho [21]. SDS-

PAGE electrophoresis was carried out at 20 °C in the SE 600 Ruby system (GE Healthcare) at 40 mA/gel for 5 h. The gels were stained with Coomassie blue G-250 solution (8% ammonium sulfate (v/v), 0.8% phosphoric acid (v/v), 0.08% Coomassie Blue G-250 (v/v) and 20% methanol (v/v)), scanned using an ImageScanner III (GE Healthcare, Chicago, USA), and calibrated using the Labscan software (GE Healthcare, Chicago, USA). To correctly determine the protein spots with differential abundance, a comparison of the protein profiles between the genotypes submitted to irrigated and non-irrigated treatments were performed using the ImageMaster2D Platinum 7 software (GE Healthcare, Chicago, USA). For the protein spot quantitative analysis, only the spots present in three gels were considered as differentially expressed. The spots greater than 1.5, indicated differences between the protein expressions of the classes. One-way ANOVA was performed considering a *p*-value less than 0.05. The gels were performed in triplicates.

The abundances (%V) of the protein spots expressed differentially using the criteria above were converted to fold-change and classified as up- or down-regulated (fold-change higher than 1.3 were plotted).

### Detection and Differential Quantification of Phosphoprotein in 2DE Gels

The detection of the phosphorylated proteins was carried out by staining the 2DE gels using a Pro-Q® Diamond Phosphoprotein Gel Stain kit (Invitrogen, Calrsbad, USA), according to the protocol described by Agrawal and Thelen [22]. The gels were then scanned with a FLA 5100 laser scanner and the images were analyzed as described by Vital [23].

The spots stained with Pro-Q Diamond with ratio values greater than 1.5 were considered as differentially expressed ($p < 0.05$) phosphorylated proteins. The three replicates of phosphoprotein spots were selected to be identified using mass spectrometry.

### In-Gel Digestion and Mass Spectrometry for the Protein Identification

Protein spots were excised and destained in 50 mM ammonium bicarbonate and 50% methanol (v/v), followed by acetonitrile addition for dehydration. The proteins were then reduced by adding 200 mM DTT solution in 100 mM of ammonium bicarbonate for 30 min of incubation at 56 °C, followed by alkylation with 200 mM of iodoacetamide solution containing 100 mM of ammonium bicarbonate for 30 min at room temperature. The spots were washed with 100 mM of ammonium bicarbonate solution, dehydrated with acetonitrile and dried using vacuum centrifugation. the spots were then rehydrated using trypsin digestion solution overnight (20 h) at 37 °C. The digested peptides were extracted using an extraction buffer (50% acetonitrile, 5% formic acids) and dried using vacuum centrifugation. The peptides were dissolved in 0.1% formic acid and analyzed online using LC-MS coupled to a UPLC

system (nanoAcquity, Waters, Milford, USA), equipped with a column and a capillary trap C18 BEH130 1.7 uM 100 nm × 100 mm operating at a flow rate of 0.5 μL/min. The peptides eluted were automatically injected into a mass spectrometer Ion Trap (Amazon ETD, Bruker, Billerica, USA), using a nanoESI ionization needle. The scanning of the ions from the mass spectrometer was carried out between 300 and 1500 *m*/*z* in positive mode and the data were acquired for 70 min through LC-MS/MS analysis. The mass spectrometer was operated in the auto-MSn mode. The acquisition of data from the LC-MS instrument was controlled using Hystar software (Bruker, Billerica, USA) and the spectra were processed with the Data Analysis software (Bruker, Billerica, USA) using the default settings for proteomics. A peak list was converted to the mascot generic format (.mgf). Peak lists were used for protein identification MASCOT algorithm with a local Client license connected to a remote server. The parameters used in the program were: protein list obtained from Phytozome containing all the proteins described for soybeans in the database, methionine oxidation as a variable modification of cysteine, cysteine carbamidomethylation as a fixed modification, a missed cleavage, peptide charge state 2+, 3+, 4+, trypsin-like cleavage enzyme and a mass error of 0.1 Da. SCAFFOLD 4.2.1 version was used to validate MS/MS based peptide and the protein identifications. Peptide identification was accepted with a probability greater than 95% using three unique peptides.

### Prediction of the Putative Phosphorylated Sites and the Presence in the Public Database

The FASTA sequences of the proteins stained as phosphorylated were analyzed for possible phosphorylation sites. The prediction of the number of phosphorylation sites was performed online through the MUSITE website (http://www.musite.net), designed for plant protein, allowing for variation in the probability of occurrence of phosphorylation sites. For this approach, the general phosphorylation on serine and threonine were chosen, and all the green plants were considered for group analysis. To identify the phosphorylation sites, the algorithm takes the downstream and upstream amino acid sequence into the serine and threonine residues to determine if, and in which position, these amino acids will be phosphorylated. As there are not a large number of phosphorylated tyrosine residues in plants, the site does not identify the phosphorylation sites of this amino acid residue. In order to have differentiated levels of requirement, the number of phosphorylation sites was determined to be 70% (low requirement) and 95%.

The presence of homologous proteins or phosphorylated sites were confirmed using the P3DB site (Plant Protein Phosphorylation DataBase: http://www.p3db.org). In this site, the input of the search is the name of the proteins and the output is the details of the results as phosphoproteins, such as the phosphorylated amino acid sites, the position, the protein ontology, the plant species and the description of the experimental

conditions. Therefore, it was possible to verify the existence or not of the phosphorylation sites from the identified proteins in our assays. It is worth noting that the non-identification of phosphorylation sites of a protein by the P3DB site does not necessarily imply that this protein is not phosphorylated because the information on this protein is not found on the database.

### Metabolite Extractions, Chemical Derivatization, Metabolic Profile Analysis by GC/MS

The metabolite extractions and derivatization were performed as described by Coutinho [21]. The metabolic extract was analyzed using an Agilent 7890A GC System coupled with a Mass Spectrometric TruTOF® HT TOFMS (Leco, Setor Joseph, USA) equipped with a capillary column of 30 mm (MDN-35) operating according to Lisec [24]. The samples were injected in splitless mode at 230 °C using a gas flow (continuous flow of helium) of 2 mL/min. The oven temperature was initially maintained at 80 °C and then increased to 15 °C/min, until reaching 330 °C, before being kept at this temperature for 5 min. Mass spectrum was obtained by the full-scan method with a range from 33 to 600 *m*/*z*. A series of *n*-alkanes was used along with samples to calculate the retention indices.

### Data Processing and Metabolite Identification

The raw GC/MS data were processed and converted to the CDF format (NetCDF) using the ChromaTOF package and analyzed using TargetSearch algorithms [25]. A script was used to identify and quantify the metabolites designed to run on the R package [25]. The processing parameters and alignment used was optimized for our GC/TOF platform. The parameters used were: mass range of 85–500 Da; threshold 50; TopMasses 10; r threshold of 0.05; quality index threshold 600. The identification of the compounds was determined by database searches based on the combination of the mass spectrum and the chromatographic retention indices according to the TAG-based method [25]. In this study, a metabolites fragmentation library was characterized by electron impact (EI) and retention index (RI) GMDB_FAMELib_TS_20110228_IS.txt, generated by the GMD Mass Spectrum Reference Library (http://gmd.mpimp-golm.mpg.de/download/). It was generated using the same configuration of our GC/MS platform. A filtered identification table containing the identified compounds and their intensities were normalized by the fresh mass from the leaves, and used as input data for processing and statistical analysis using the MetaboAnalyst platform (http://www.metaboanalyst.ca/).

### Metabolite Data Analysis

Tables containing the identified metabolites by GC/MS and phytohormonal concentrations were used as input data for the

characterization of metabolites with differential relative abundances. A display of the metabolic pathway maps significantly disturbed using the MetaboAnalyst 3.0 package was generated. The Quality Filters based on the standard deviation methods were used to automatically remove the low-quality data. The intensity values were then normalized. The data were analyzed using the Partial Least Squares Discriminant Analysis (PLS-DA) for grouping the treatments, and to arrange the metabolites (VIP score) in order of their importance in the grouping of the samples. The significantly different metabolites (One-way ANOVA, $p < 0.05$) were analyzed using the module Pathway Analysis search tool present in the MetaboAnalyst (v 3.0) platform. To do this, the intensity of each identified metabolite showing differences in relative abundance was converted to fold-change and classified as up- and down dysregulated (fold-changes higher than 2.0 were used as input). A pathway library of *Arabidopsis thaliana* was used.

### Phytohormone Analysis

Ten milligrams of powdered tissue (~110 mg fresh weight) was mixed with 400 µL of methanol, isopropanol and acetic acid (20:79:1). The phytohormones were extracted and analyzed as described by Coutinho [21]. The sample was scanned by MRM (multiple reaction monitoring) to detect each hormone as shown in Supplementary Table S1. Three biological replications were performed. The generated mass spectra were processed using the MassHunter software to obtain the extracted chromatograms (XIC) of each transition and the areas to quantify each hormone. A standard curve of each of the hormones varying concentrations from 0.1 to 300 ng/mL was used to convert the area values from XIC in ng/g of plant tissue. The synthetic plant hormone naphthalene acetic acid (NAA) was added into the extraction buffer and used as an internal normalizer in the quantitative analysis.

## RESULTS

### Drought Stress Assays

In the control plants of both soybean cultivars (BR 16–sensitive and Embrapa 48–tolerant) the water potential was higher than −0.25 MPa. The tolerant cultivar had lower potentials (Figure 1) throughout the experimental period in the irrigated condition. After the irrigation was interrupted, an increasing reduction of $\psi_{am}$ was observed. After three days of stress conditions, the decrease of $\psi_{am}$ was more noticeable in the sensitive BR 16 plants (Figure 1). The BR 16 plants reached $\psi_{am}$ values of −1.0 and −1.5 MPa, on the seventh and ninth day respectively, after the irrigation suspension. The Embrapa 48 plants reached the same $\psi_{am}$ levels, on the ninth and eleventh days respectively, after the irrigation suspension. These results are consistent with a more efficient water use by the cultivar Embrapa 48 and confirm this cultivar as drought-tolerant.

As the severe water potential of −1.5 MPa caused considerable damage to the BR 16 plants (Supplementary Figure S1), only the irrigated control plants and those that showed water potential of −1.0 MPa were used for the proteomic analyses.

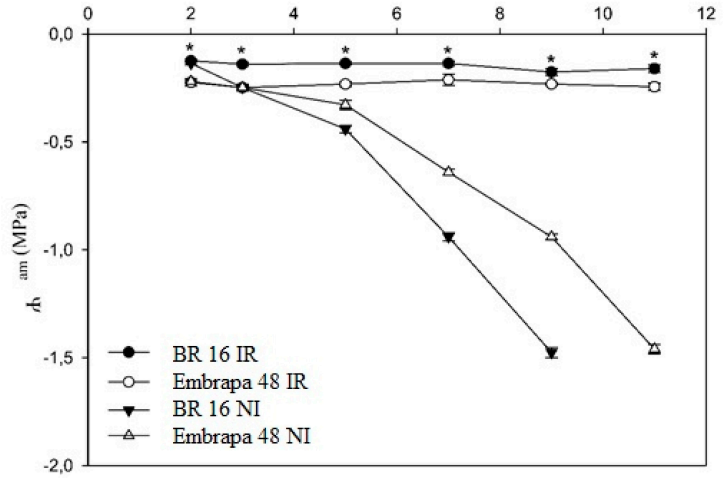


**Figure 1.** Time course (days) of the water potentials ($\psi_{am}$ in MPa) of soybean leaves after the suspension of the irrigation. Irrigated (IR) and non-irrigated (NI) treatments, respectively, for the tolerant (Embrapa 48) and the sensitive (BR 16) plants.

### Protein Fractionation by PEG and Phosphoprotein Detection

A significant portion of the cellular proteins are post-translationally modified by phosphorylation. However, phosphoproteins were observed at the low stoichiometric ratio and at low levels. 2DE gel staining tests for phosphoproteins showed a few spots stained with Pro Q Diamond (data not shown). Therefore, a strategy was applied for differential precipitation of high abundance proteins based on the addition of polyethylene glycol (PEG) to the aqueous protein extract, under non-denaturing conditions. The efficiency of the fractionation by PEG to remove higher abundance protein was evaluated in one-dimensional (1-DE) and 2DE gel electrophoresis (Supplementary Figure S2) by comparing the profile of the PEG-fractionated and non-fractionated samples. In all of the non-fractionated samples, the presence of the large and small rubisco subunits was observed, as indicated in Supplementary Figure S2. In contrast, the protein profile of PEG-fractionation on the supernatant samples did not show the protein bands corresponding to the rubisco subunits (Supplementary Figure S2). This method made it possible to see a larger number of stained spots using Pro Q Diamond.

### Differential Proteome Analysis under Drought Stress

The supernatant of the PEG-fractionated protein extracts, from the sensitive and the drought-tolerant soybean plants, was separated by 2DE. As the genotypes have different genetic backgrounds, contrasts were performed for each cultivar in irrigated (IR) and non-irrigated (NI)

conditions. Afterwards, the drought-responsive protein profiles were analyzed to characterize the molecular and physiological responses. For the Embrapa 48, 643 well-separated spots were used to compare plants with −1.0 MPa water potential (NI) with the corresponding irrigated control (IR). 72 out of 643 were identified as differentially expressed spots, of which 60 were identified by mass spectrometry (Supplementary Table S2). These 60 spots corresponded to 45 different proteins as putative isoforms or modifications of the same protein. Among the 45 unique spots, 38 showed differential intensities and 7 were present only in plants under drought stress. In contrast, 4 spots were detected only in the control NI and 11 spots showed a reduction in relative abundance when comparing IR × NI. For the sensitive cultivar BR 16, 574 spots were used to compare IR *vs* NI plants. Only 58 were classified as differentially expressed spots, of which 42 were identified by mass spectrometry (Supplementary Table S3). These 42 spots corresponded to 35 different proteins, because of some possible isoforms or modifications. Among the 42 spots identified, 25 spots had an increase in relative abundance in NI treatment compared to the IR, 11 spots were present only in the IR treatment. The abundances (%V) of the responsive proteins to drought of each genotypes were converted to fold-change (Figures 2–4), and classified as up- or down-regulated, when an increase or reduction of the protein abundance was observed in the NI treatment, respectively. The protein abundance data were also separated by the response pattern to each genotype, when expressed in both (Figure 2) and when it was genotype-specific (Figures 3 and 4). For a better visualization of the general response differences between genotypes, the proteins were classified into functional categories and frequent biological processes, as the function of the interruption of water supply. The GO classification of the responsive proteins for both genotypes was in general similar, (Supplementary Figure S3) however, the specific expression profile was distinct (Supplementary Figure S4). The protein lists were also used for functional enrichment analyses using a ClueGO plugin to detect the significantly enriched GO terms when the proteome was up or down-regulated. A distinct response pattern was observed for the genotypes. Two main clusters were seen when comparing up-regulated proteins between genotypes (Supplementary Figure S5). The up-regulated proteins of the sensitive genotype BR16 share clusters related to biological processes, such as response to oxygen radical and regulation of translational termination (Supplementary Figure S5). However, for the tolerant Embrapa 48 (Supplementary Figure S5), the up-regulated proteins showed a distinct cluster related to positive regulation of catalytic activity, photosynthesis, and photosystem II assembly. The analysis applied for down-regulated proteins also showed distinct results for the genotypes, notably a cluster related to plastid translation and transcription, and chlorophyll biosynthesis for the drought sensitive BR 16 (Supplementary Figures S4 and S5). The individual analyses of the identified proteins showed a similar up-fold change for both genotypes such as the

glutathione *S*-transferase TAU 19, glutathione peroxidase 6, transcriptional coactivator/pterin dehydratase and chloroplast drought-induced stress protein of 32 KD (Figure 2). This is an indicator of the water potential of the leaves and the general response of these soybean genotypes. However, some proteins classified as up-regulated were only identified in the tolerant Embrapa 48, such as rubisco activase and chloroplast RNA-binding protein 31B (Figure 3). Some proteins were only identified in the sensitive BR 16, such as the eukaryotic elongation factor 5A-1, transcriptional coactivator/pterin dehydratase and ascorbate peroxidase (Figure 4). These results are in accordance with the enriched GO terms (Supplementary Figure S5). Complex expression profiles were observed in the 2DE gels for some identified proteins because different spots were detected coding isoform for the same protein. For example, ten spots were detected for carbonic anhydrase (CA): spots 872, 881 and 69 as down-regulated in the Embrapa 48; 436 and 997 as down-regulated in the BR 16. One spot (93) was up-regulated in the Embrapa 48. Two spots were observed for gamma CA 1, as up-regulated in both genotypes and one spot for CA 2 as up-regulated in the BR 16.

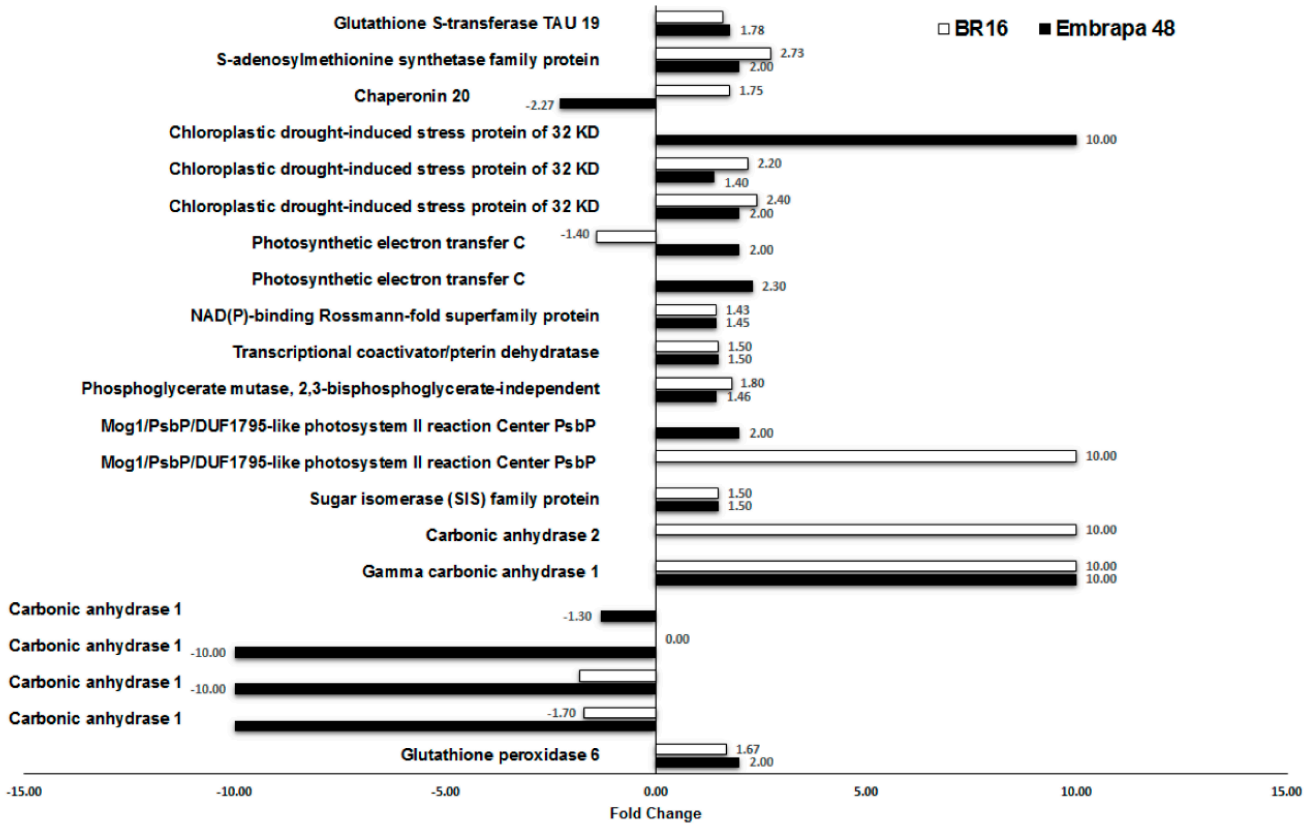


**Figure 2.** Up- and down-regulated proteins of the soybean leaves from the drought sensitive BR 16 and the tolerant Embrapa 48 genotypes. Fold change was calculated between non-irrigated (NI) and irrigated (IR) plants from the same genotype. Different values of protein fold changes are shown with the same name because spots with different PIs were identified for the same Glyma ID by LC/MS. Values of fold change for protein spots when present only in the non-irrigated (NI) or irrigated (IR) treatments are shown as 10.00 or −10.00, respectively.

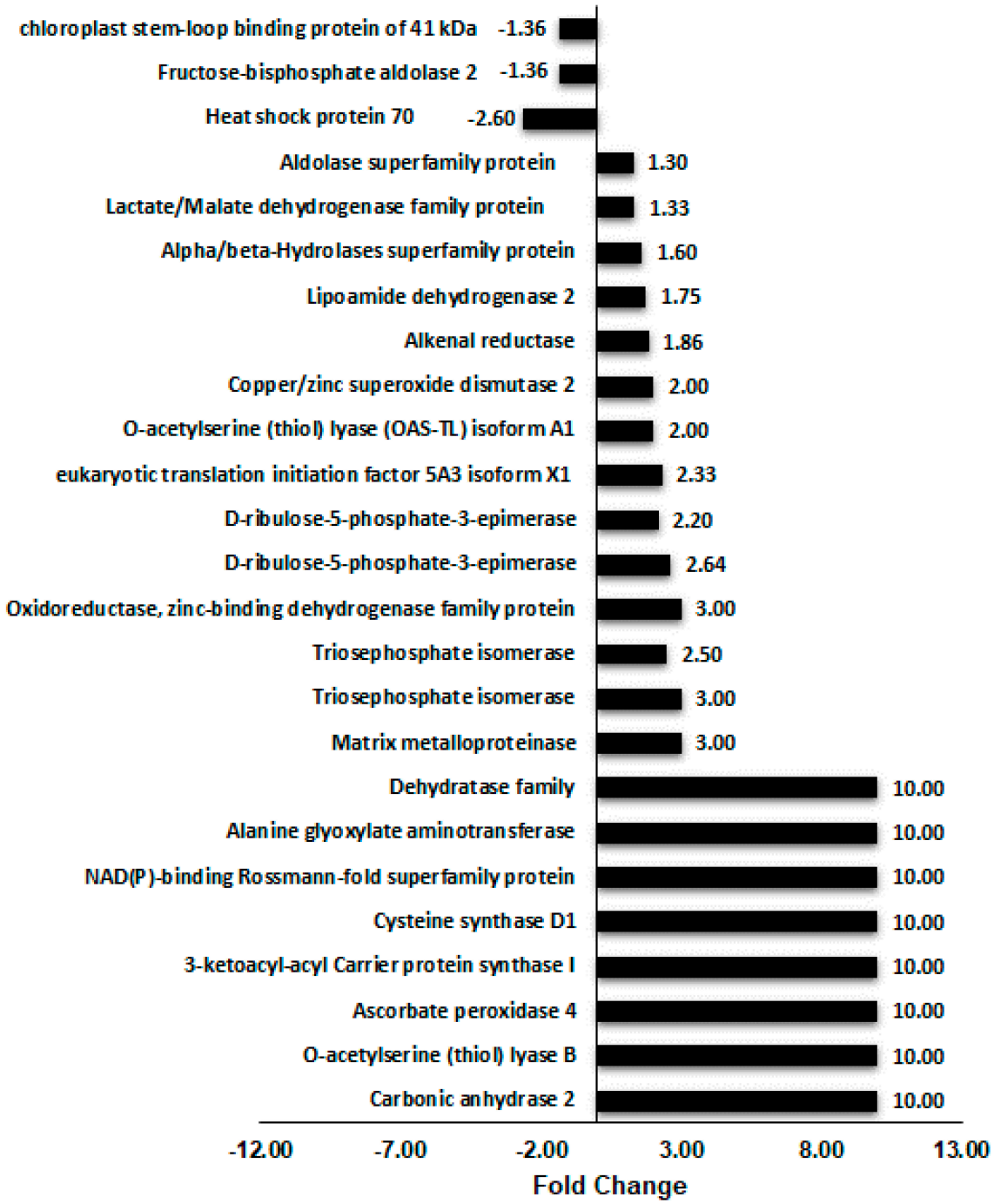


**Figure 3.** Up- and down-regulated proteins of the soybean leaves of drought sensitive BR 16. Fold change was calculated between non-irrigated (NI) and irrigated (IR) plants. Different values of protein fold changes are shown with the same name because spots with different PIs were identified for the same Glyma ID by LC/MS. Values of fold change for protein spots present only in the non-irrigated (NI) and irrigated (IR) treatments are shown as 10.00 or −10.00, respectively.

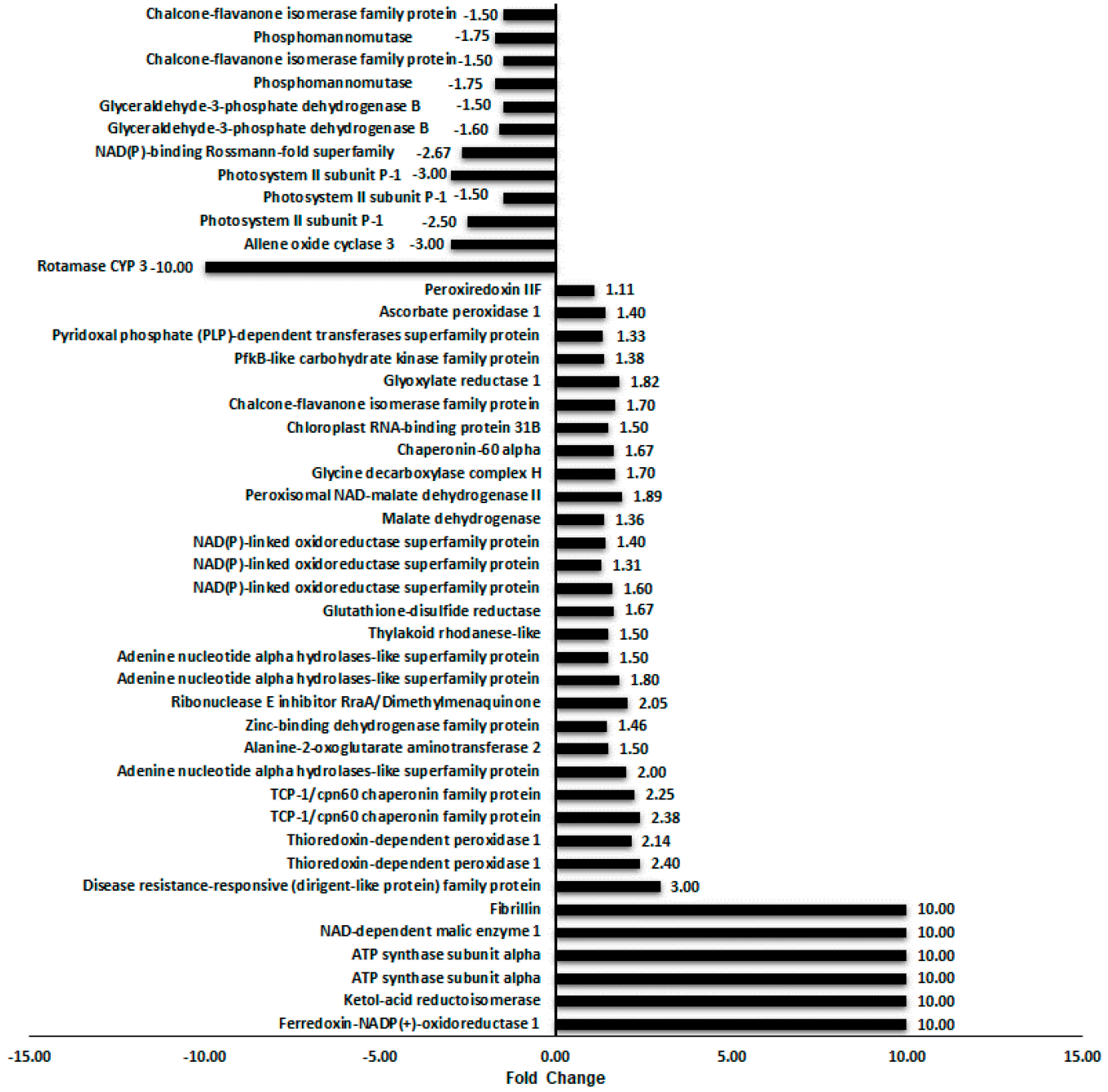


**Figure 4.** Up- and down-regulated proteins of the soybean leaves of drought sensitive Embrapa 48. Fold change was calculated between non-irrigated (NI) and irrigated (IR) plants. Different values of protein fold changes are shown with the same name because spots with different PIs were identified for the same Glyma ID by LC/MS. Values of fold change for protein spot present only in the non-irrigated (NI) and irrigated (IR) treatments are shown as 10.00 or −10.00, respectively.

### Differential Phosphoproteome

The spot proteins stained with Pro Q Diamond were identified by mass spectrometry (Supplementary Tables S5 and S6) and were analyzed for differential expression, classified as up- and down-regulated for both genotypes (Figure 5) and genotype-specific (Figures 6 and 7). The phosphoprotein lists were also classified for GO functional categories and

functional enrichment analyses by ClueGO plugin to detect the significantly enriched GO terms when the proteome was up-regulated (Supplementary Figure S6). A distinct profile was also observed for the drought contrasting soybean genotypes when the phosphoproteome was analyzed. Significant functional enrichment results were observed for the overexpressed phosphoprotein in the tolerant Embrapa 48 (in red), in which sharing clusters related to biological processes, such as oligosaccharides biosynthesis process, *S*-glycoside biosynthesis process and response to disaccharide stimulus (Supplementary Figure S6). On the other hand, the sensitive BR 16 enrichments were more evident for the down-regulated phosphoproteins mainly for biological processes related to response to disaccharide stimulus and glycolysis (Supplementary Figure S6). Stress-responsive proteins, which were detected as phosphorylated, showed a similar profile for both genotypes, such as the stress-responsive alpha-beta barrel domain protein and ascorbate peroxidase 1 (Figure 5). Six spots were identified as phosphorylated isoforms of the transketolase as up-regulated in the tolerant Embrapa 48. An enzyme involved in the biosynthesis of saccharides were also identified in the Embrapa 48, while four phosphorylated isoforms were down-regulated in the sensitive BR 16 (Figure 6). Further proteins were identified including; phosphoprotein and glycine decarboxylase P which showed expression up-regulated for Embrapa 48 and down-regulated for BR 16. This is a multi-protein complex, which plays a major role in photorespiration in plants, and is involved and related to the tolerance of the oxidative stress [26]. Spots were also identified as regulated phosphoproteins involved in the photosynthetic apparatus, such as photosystem II subunit O-2 and photosystem II subunit P-1 as up-regulated for tolerant Embrapa 48 and down-regulated for sensitive BR 16 (Figures 5 and 6). Other phosphoproteins related to carbon fixation processes such as carbonic anhydrase and rubisco activase were also observed as different protein spots. Spots for isoforms of the cobalamin-independent synthase family protein were found up-regulated in the Embrapa 48 (Figure 7). Protein homologous related to methionine biosynthesis is important in oxidative stress [27]. The protein IDs, for each phosphoprotein detected in the 2DE profiles, were confirmed for the presence of phosphoprotein in the public plant. Databases were used to verify those already described as phosphorylated proteins (Supplementary Table S4). In addition, the FASTA amino acid sequences were submitted to *in silico* analysis to predict the presence of conserved phosphorylated sites in these proteins (Supplementary Table S4). Most of the proteins staining in the gels as phosphoproteins, were validated as phosphoproteins present in the databases, or as showing a high score for the presence of the phosphorylated sites from the prediction results.

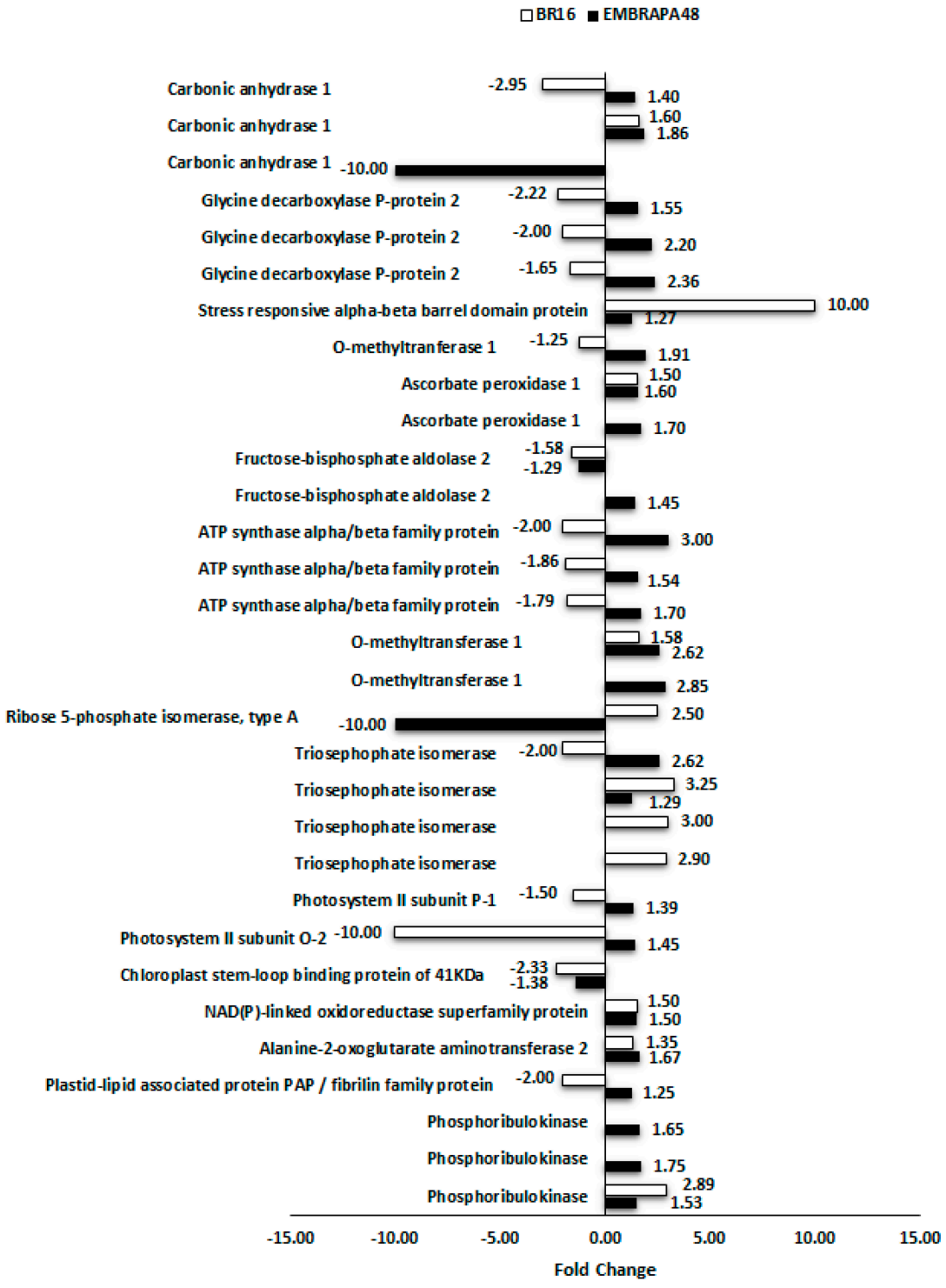


**Figure 5.** Up- and down-regulated phosphoproteins of the soybean leaves drought sensitive BR 16 and tolerant Embrapa 48 genotypes. Fold change was calculated between non-irrigated (NI) and irrigated (IR) plants from the same genotype. Different values of protein fold changes are shown with the same name because spots with different PIs were identified for the same Glyma ID by LC/MS. Values of fold change for phosphoprotein spot present only in the non-irrigated non-irrigated (NI) and irrigated (IR) treatments are shown as 10.00 or −10.00, respectively.

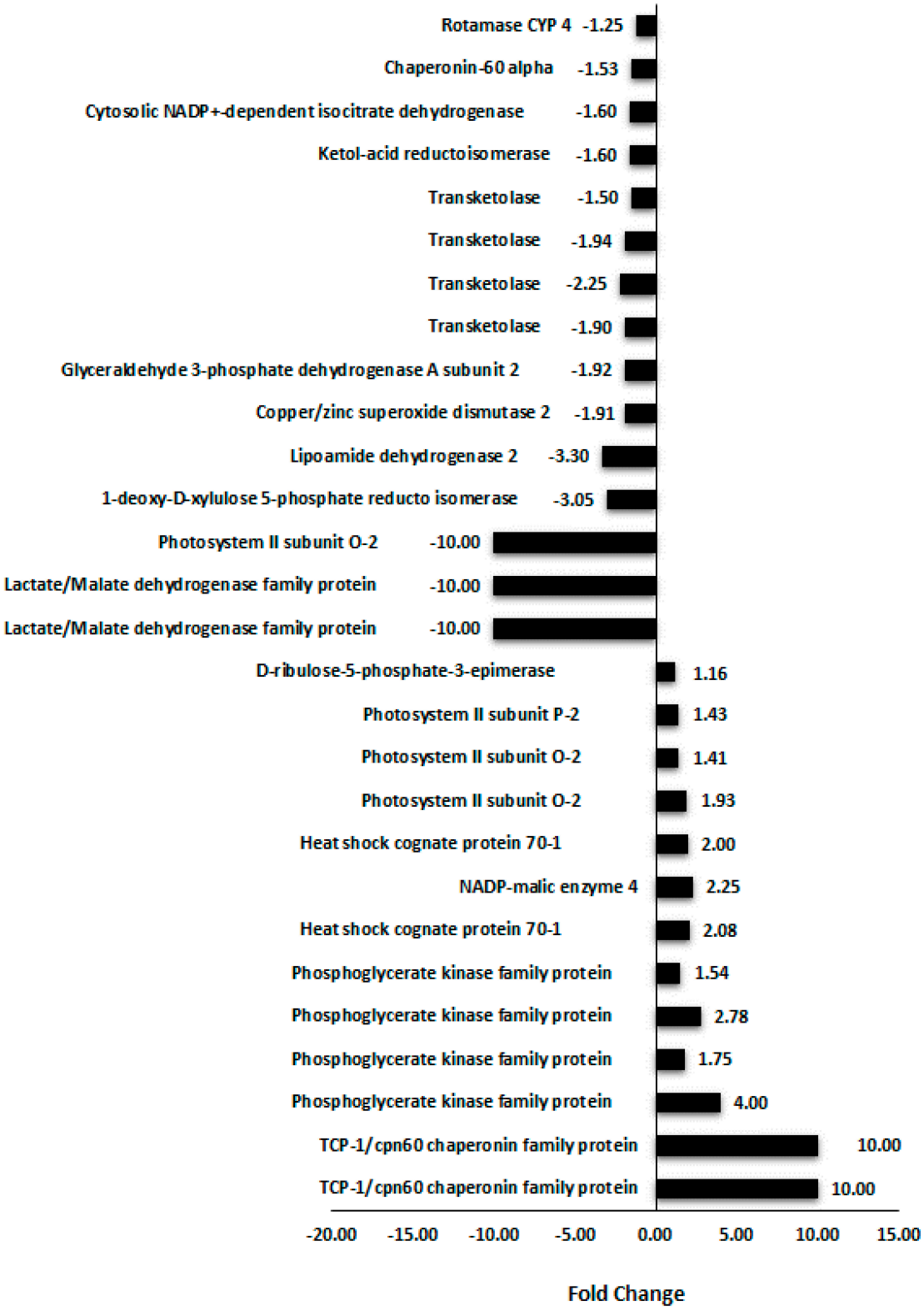


**Figure 6.** Up- and down-regulated phosphoproteins of the soybean leaves drought sensitive BR 16 genotype. Fold change was calculated between non-irrigated (NI) and irrigated (IR) plants. Different values of protein fold changes are shown with the same name because spots with different PIs were identified for the same Glyma ID by LC/MS. Values of fold change for phosphoprotein spot present only in the non-irrigated (NI) and irrigated (IR) treatments are shown as 10.00 or −10.00, respectively.


**Figure 7.** Up- and down-regulated phosphoproteins of the soybean leaves drought sensitive Embrapa 48 genotype. Fold change was calculated between non-irrigated (NI) and irrigated (IR) plants. Different values of protein fold changes are shown with the same name because spots with different PIs were identified for the same Glyma ID by LC/MS. Values of fold change for phosphoprotein spot present only in the non-irrigated (NI) and irrigated (IR) treatments are shown as 10.00 or −10.00, respectively.

Determining the impact of differentially phosphorylated proteins under water stress deficit in molecular and physiological processes may be challenging, as phosphoproteins are shown as many isoforms, and such spots could be interconverted by the addition of a phosphate group to the protein, therefore, changing the isoelectric points (PI). In this case, a spot will show an increase and the other one will show a reduction in volume.

This could be more complex when multiples phosphorylated isoforms are present in the profile. We detected 13 spot clusters for BR 16 and 14 for Embrapa 48, containing different phosphorylated isoforms in the 2DE profiles (Supplementary Tables S6–S8). This phosphorylation behavior could be important in regulating the photosynthesis in the soybean genotypes.

**Metabolic Profile**

The analysis of the metabolome was performed by GC/MS and the metabolic profiles were compared between the contrasting genotypes to verify the fluctuation of the metabolites during drought treatments related to the tolerance mechanism. This approach allows us to identify metabolites that were quantified and used for data analysis by MetaboAnalyst platform. The PLSA-DA method was used to analyse the most significant variables and observed differential grouping of the samples, in order to compare the genotypes and water regimes (Supplementary Figure S10). The BR 16 showed a better separation of the treatments, whereas, in Embrapa 48, the treatments remained more grouped. This indicates a lower fluctuation of the metabolites in this genotype, especially under moderate stress (Supplementary Figure S9).

The PLS-DA method was also used to identify the variables that had the biggest effect in grouping the samples (Supplementary Figure S11) and indicate the responsive amino acids in both genotypes. The metabolites that showed significant abundance variations as a function of the applied treatments are more responsive to drought (Figure 8). For both genotypes, the proline was the metabolite that showed higher variations, presenting a higher value VIP score when compared to the other fourteen most important metabolites. Proline is the most responsive amino acid to drought. The main responsive metabolites belong to the classes of organic acids, amino acids, and carbohydrates. When the metabolites of these classes were evaluated individually for the response to drought between genotypes, a pattern of behavior with a trend of greater abundance and response in the sensitive genotype BR 16 (Figure 8), such as the amino acids alanine, asparagine, aspartic acid, cysteine, isoleucine, lysine, methionine, tyrosine, phenylalanine, tryptophan and ornithine was observed. The amino acid proline showed a higher abundance in the sensitive BR 16, especially in severe drought stress. Likewise, some carbohydrates, such as cellobiose, xylulose, sucrose, xylose, melibiose, and trehalose, had the same effect of higher abundance in the sensitive BR 16 (Figure 8). Other metabolites that were responsive to drought treatment, such as putrescine, spermidine and ornithine, also presented the same behavior. The identified metabolites and the fold-change were used to map the pathway affected during stress by using the MetaboAnalyst platform. The number and the impact in the pathways were higher for the sensitive BR 16 (Tables 1 and 2). In these analyses, the pathway for amino acids metabolism showed higher impact indexes.

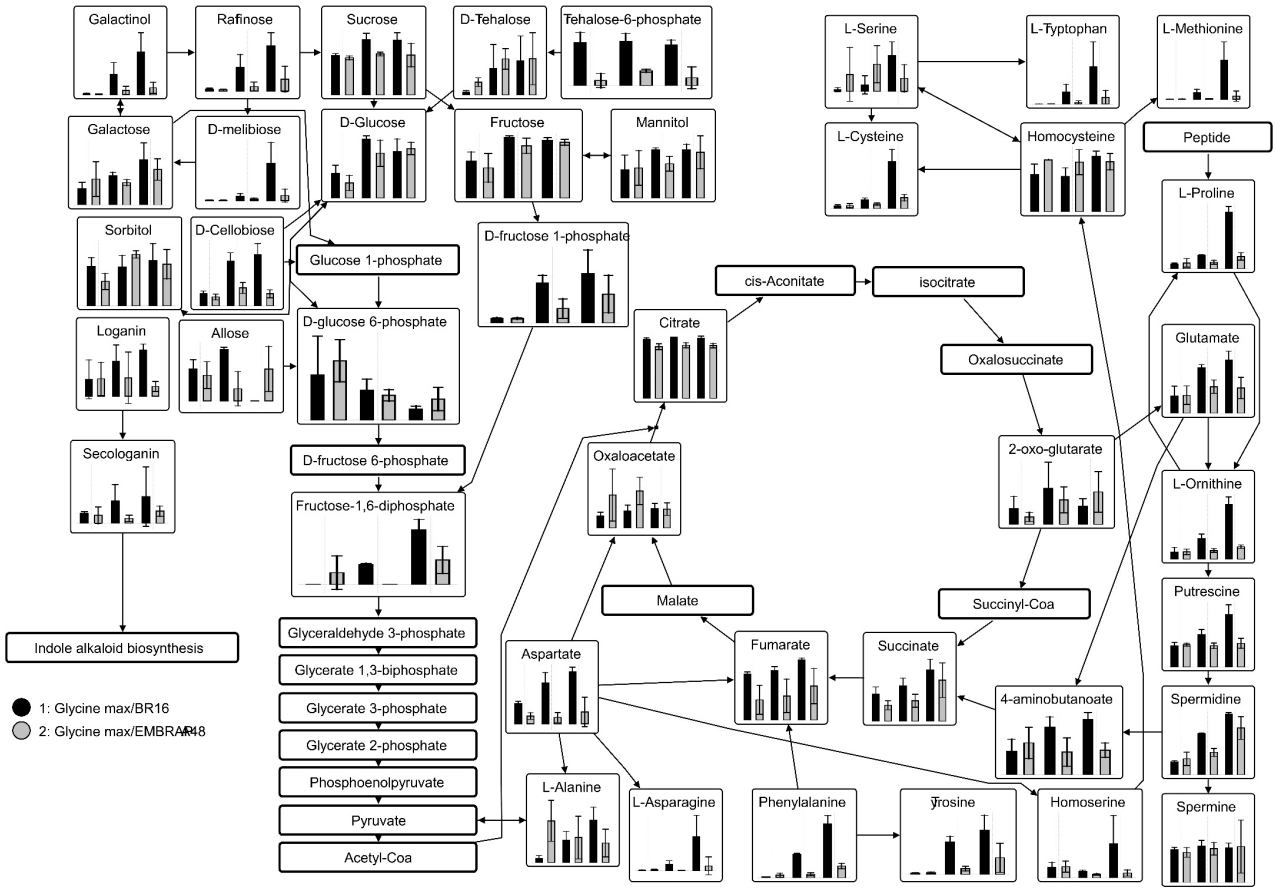


**Figure 8.** Relative abundances of some of the significantly responsive metabolites to non-irrigated treatments of the soybean leaves. Black bars indicate the metabolite intensities for the tolerant genotype Embrapa 48 and grey bars for the sensitive BR 16, irrigated and non-irrigated in the leaf potentials of −0.2 MPa (IR), −1 MPa (NI) and 1.5 MPa (NI), respectively.

**Table 1.** Metabolic pathways significantly affected by drought stress for the tolerant Embrapa 48. Fold change was calculated between non-irrigated (NI) and irrigated (IR) plants from the same genotype and was used as an input for the pathway analysis.

| Metabolic pathway | Total Cmpd | Hits | –log(p) | FDR | Impact |
|---|---|---|---|---|---|
| Propanoate metabolism | 15 | 5 | 3.69 | 0.14762 | 0.45455 |
| Galactose metabolism | 26 | 6 | 3.6219 | 0.14762 | 0.30001 |
| Citrate cycle (TCA cycle) | 20 | 4 | 3.7236 | 0.14762 | 0.21428 |
| Glycerophospholipid metabolism | 25 | 2 | 4.2749 | 0.14762 | 0.17242 |
| Glucosinolate biosynthesis | 54 | 6 | 4.1493 | 0.14762 | 0.15384 |
| Starch and sucrose metabolism | 30 | 4 | 3.4193 | 0.1488 | 0.14285 |
| Arginine and proline metabolism | 38 | 6 | 3.585 | 0.14762 | 0.13159 |
| Phenylalanine, tyrosine and tryptophan biosynthesis | 21 | 3 | 4.2819 | 0.14762 | 0.05263 |
| Purine metabolism | 61 | 3 | 4.5759 | 0.14762 | 0.02899 |
| Aminoacyl-tRNA biosynthesis | 67 | 12 | 3.5225 | 0.14762 | 0.2708 |
| Glyoxylate and dicarboxylate metabolism | 17 | 3 | 3.7434 | 0.14762 | 0.125 |
| Pentose phosphate pathway | 18 | 2 | 3.1036 | 0.18702 | 0.04 |

Total Cmpd: the number of metabolites from the pathway; Hits: the number of the identified dysregulated metabolites; −log(p): Holm *p*-value; FDR: the False Discovery Rate; Impact: the index of impact in the pathway, relative to the fold-change magnitude.

**Table 2.** Metabolic pathways significantly affected by drought stress for the tolerant BR 16. Fold change was calculated between non-irrigated (NI) and irrigated (IR) plants from the same genotype and was used as an input for the pathway analysis.

| Metabolic pathway | Total Cmpd | Hits | −log(p) | FDR | Impact |
|---|---|---|---|---|---|
| Propanoate metabolism | 15 | 5 | 4.9669 | 0.021481 | 0.45455 |
| beta-Alanine metabolism | 12 | 6 | 4.8294 | 0.021481 | 0.44444 |
| Phenylalanine metabolism | 8 | 1 | 4.6112 | 0.021481 | 0.42857 |
| Glycine, serine and threonine metabolism | 30 | 5 | 5.2874 | 0.021481 | 0.33334 |
| Galactose metabolism | 26 | 6 | 4.7511 | 0.021481 | 0.30001 |
| Alanine, aspartate and glutamate metabolism | 22 | 6 | 5.1002 | 0.021481 | 0.28572 |
| Citrate cycle (TCA cycle) | 20 | 4 | 4.7508 | 0.021481 | 0.21428 |
| Glycerophospholipid metabolism | 25 | 2 | 4.4402 | 0.021481 | 0.17242 |
| Starch and sucrose metabolism | 30 | 5 | 4.5093 | 0.021481 | 0.17142 |
| Cyanoamino acid metabolism | 11 | 4 | 4.802 | 0.021481 | 0.16667 |
| Ubiquinone and other terpenoid-quinone biosynthesis | 23 | 2 | 4.4539 | 0.021481 | 0.15789 |
| Cysteine and methionine metabolism | 34 | 5 | 4.7607 | 0.021481 | 0.15625 |
| Glucosinolate biosynthesis | 54 | 6 | 3.8465 | 0.029434 | 0.15384 |
| Fructose and mannose metabolism | 16 | 1 | 3.6275 | 0.033893 | 0.11765 |
| Lysine biosynthesis | 10 | 1 | 4.6736 | 0.021481 | 0.11111 |
| Nicotinate and nicotinamide metabolism | 12 | 2 | 4.6736 | 0.021481 | 0.09091 |
| Sulfur metabolism | 12 | 1 | 4.0258 | 0.026146 | 0.09091 |
| Glycerolipid metabolism | 13 | 1 | 3.6555 | 0.033893 | 0.07692 |
| Nitrogen metabolism | 15 | 2 | 4.7005 | 0.021481 | 0.07143 |
| Valine, leucine and isoleucine degradation | 34 | 4 | 3.2373 | 0.045517 | 0.06978 |
| Pyrimidine metabolism | 38 | 4 | 3.65 | 0.033893 | 0.06383 |
| Phenylalanine, tyrosine and tryptophan biosynthesis | 21 | 3 | 4.5303 | 0.021481 | 0.05263 |
| Phenylpropanoid biosynthesis | 45 | 2 | 4.6851 | 0.021481 | 0.04348 |
| Pyruvate metabolism | 21 | 1 | 3.4413 | 0.037981 | 0.03846 |
| Glycolysis or Gluconeogenesis | 25 | 1 | 3.4413 | 0.037981 | 0.03125 |
| Purine metabolism | 61 | 3 | 4.4703 | 0.021481 | 0.02899 |
| Aminoacyl-tRNA biosynthesis | 67 | 12 | 4.3671 | 0.022313 | 0.2708 |
| Amino sugar and nucleotide sugar metabolism | 41 | 1 | 3.0825 | 0.051956 | 0.02703 |
| Carbon fixation in photosynthetic organisms | 21 | 2 | 4.6736 | 0.021481 | 0.125 |
| Glyoxylate and dicarboxylate metabolism | 17 | 3 | 4.5996 | 0.021481 | 0.125 |
| Sphingolipid metabolism | 13 | 1 | 4.0258 | 0.026146 | 0.125 |

**Table 2.** *Cont.*

| Metabolic pathway | Total Cmpd | Hits | −log(p) | FDR | Impact |
|---|---|---|---|---|---|
| Zeatin biosynthesis | 19 | 1 | 4.0205 | 0.026146 | 0.125 |
| Methane metabolism | 11 | 2 | 3.989 | 0.026233 | 0.125 |
| Tyrosine metabolism | 18 | 3 | 4.6706 | 0.021481 | 0.25 |
| Tryptophan metabolism | 27 | 3 | 4.6264 | 0.021481 | 0.25 |
| Tropane, piperidine and pyridine alkaloid biosynthesis | 8 | 1 | 4.6112 | 0.021481 | 0.25 |
| Isoquinoline alkaloid biosynthesis | 6 | 1 | 4.1563 | 0.026146 | 0.5 |
| Pentose phosphate pathway | 18 | 2 | 4.5036 | 0.021481 | 0.04 |
| Indole alkaloid biosynthesis | 7 | 2 | 4.0864 | 0.026146 | 0.4 |
| Pentose and glucuronate interconversions | 12 | 2 | 4.5825 | 0.021481 | 0.2 |
| Pantothenate and CoA biosynthesis | 14 | 2 | 4.5318 | 0.021481 | 0.2 |
| Biosynthesis of unsaturated fatty acids | 42 | 2 | 4.6593 | 0.021481 | 0 |
| Fatty acid biosynthesis | 49 | 2 | 4.6589 | 0.021481 | 0 |
| Butanoate metabolism | 18 | 2 | 4.0843 | 0.026146 | 0 |
| C5-Branched dibasic acid metabolism | 4 | 1 | 3.4654 | 0.037981 | 0 |

Total Cmpd: the number of metabolites from the pathway; Hits: the number of the identified dysregulated metabolites; −log(p): Holm *p*-value; FDR: the False Discovery Rate; Impact: the index of impact in the pathway, relative to the fold-change magnitude.

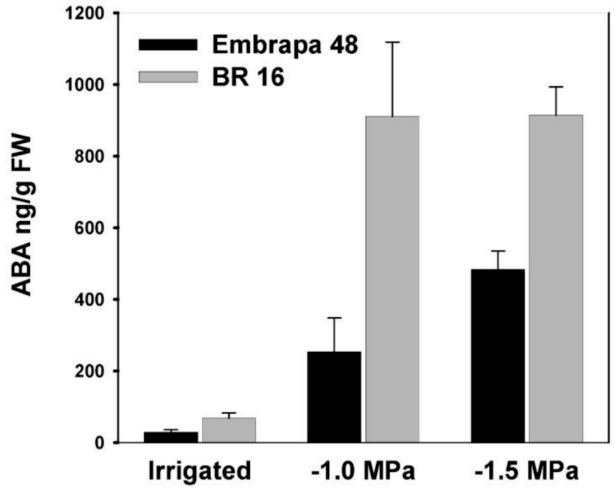


**Figure 9.** Absolute abundance of the phytohormone Abscisic Acid (ABA) of the soybean leaves quantified by LC MS. The treatments were irrigated (T1), under moderate drought stress (T2, −1.0 MPa) and under severe drought stress (T3, −1.5 MPa).

### Phytohormones

The absolute concentrations of nine phytohormones were determined by LC/MS and were used to identify which one had an effect on a water deficit. In Embrapa 48, only abscisic acid (ABA) showed significant variations in the relative concentrations between the analyzed treatments (Figure 9 and Supplementary Figure S12). In BR 16 genotype, ABA,

carboxylic-1-aminocyclopropane (ACC, Ethylene precursor) and auxin (AIA) were statistically significant (Supplementary Figure S12). For both cultivars, the phytohormones response to drought treatment showed an increase in the relative concentration with the severity of the water deficit. As expected, ABA showed big variations in the absolute concentration in response to drought treatment, it was however, more noticeable for the sensitive genotype BR 16 (Figure 9).

## DISCUSSION

Soybeans have been reported to have a wide variation in drought tolerance and these genetic resources have been used to develop new genotypes. Furthermore, genetic crosses were obtained to combine agronomic and tolerance characteristics [2,15,16]. The productivity analysis aims at distinguishing the different levels of tolerance among the cultivars, this can be carried out by many different complex physiological and molecular mechanisms. Trying to determine those contributing to the phenotype requires a combination of different experimental data. The cultivars BR 16 and Embrapa 48, share a common ancestor, and were studied by Oya [16] under drought conditions. They found that at the vegetative stage, the tolerant cultivar (Embrapa 48) showed a higher growth rate and a larger leaf area when compared to the sensitive BR 16 plants.

Initially the tolerance drought phenotype of the Embrapa 48 was confirmed by monitoring water potential under a reduction of the water supply (Figure 1), which was postponed by seven days to reach to −1.0 Mpa. Afterwards, the proteome and metabolomic profiles were compared. The proteome of the plant under drought stress was extensively studied and characterized, with a list of the responsive proteins [28]. However, in this study, we focused on the protein expression patterns that were distinct between the genotypes, aiming at identifying the physiological behavior and specific candidates that explain the drought tolerance phenotype observed. The global analysis of the protein abundance of both genotypes (Figure 2) indicated that some molecular and physiological processes for the Embrapa 48 were less affected by the reduction in the water supply. This can be explained by the presence of proteins related to functional categories that are indicative of the oxidative and water stress status, such as glutathione *S*-transferase TAU 19, transcriptional coactivator/pterin dehydratase and chloroplast drought-induced stress protein of 32 KD (CDSP32). CDSP32 is composed of two thioredoxin modules that have a critical role in plastid defense against oxidative damage. Furthermore, this is related to its role as a physiological electron donor to the peroxiredoxin [29]. An isoform of the Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP protein was up-regulated only in the sensitive BR 16, which required photosystem II core assembly and stability [30].

The proteome response profiles were different between the genotypes, and indicated that the physiological and biochemical processes, such as photosynthesis, were less affected in the tolerant genotype. Under osmotic stress, the plant cells respond to limiting cellular damages and restores a new homeostasis, which includes a coordination of biochemical and physiological changes, including the photosynthesis inhibition [8]. In this context, it was observed that the photosynthetic electron transfer C protein showed a distinct expression between the tolerant and the sensitive genotypes (Figure 2), up-regulated in Embrapa 48 and down-regulated in BR 16. This encodes the rieske FeS center of the cytochrome b6f complex and the *hcf* mutant, in which the biosynthesis was inhibited, reducing the electron transport at saturating light intensities. Generally, drought reduces transpiration and at the same time photosynthesis. However, we observed that in the tolerant genotype the regulation of the protein expression could maintain higher photosynthetic activities in comparison to the sensitive one. This is in accordance with the higher growth rate that was observed for Embrapa 48 [16]. We identified nine spots for carbonic anhydrase 1 (CA1) and one for carbonic anhydrase 2 (CA2). However, six protein isoforms were down-regulated (Figure 2). CA1 and CA2 are involved in photosynthesis, and are usually depressed under drought stress in plants [10], and have a great effect on photosynthesis and water use efficiency [31]. Overexpression of CA1 or CA4 decreased stomatal density and improved water use efficiency [31]. CA is a metalloenzyme located in the chloroplast stroma, very close to rubisco, that catalyzes the reversible reaction of bicarbonate to carbon dioxide and maintains the supply of $CO_2$ for rubisco activity [32]. These mechanisms could be important in differentiating the responses between the genotypes and improving the water-use efficiency, and the leaf growth of the Embrapa 48, as reported by [16]. Other proteins related to photosynthesis were also stained as phosphorylated in 2DE profiles (Figure 7). A phosphorylated isoform for the rubisco activase and carbonic anhydrase was observed as up-regulated only in the tolerant genotype, which might reduce the damage on rubisco by drought stress (Figures 7 and 8). Hierarchical clustering by ClueGO revealed phosphoproteins related to some biological processes that were specific for the genotypes (Supplementary Figures S4 and S6). As reported by Rocha [33], phosphorylation could regulate the behavior of the soybean's response to tolerance. The chloroplastic drought-induced stress protein was observed in different phosphorylated isoforms stained in the gel (32 KD). (Supplementary Table S8). This protein was also detected in spots showing reduced molecular weight, which is an indication that it is also regulated by proteolytic cleavage of the polypeptide chain.

A key role in the regulation of the response to drought is related to the hormone abscisic acid (ABA). The higher levels of this phytohormone were observed in the sensitive BR 16. Additionally, it was observed that this sensitive genotype had a greater number of metabolites showing

significantly altered abundances in drought conditions and therefore, more noticeable perturbations in the metabolic pathways (Supplementary Figure S11). Proline is widely regarded as the main osmoprotectant in drought stress tolerance in plants. In our study on plants at the vegetative stage, we found that drought triggers an increase in the proline synthesis or accumulation on both genotypes (Supplementary Figure S11). However, higher levels were found for the sensitive BR 16. The role of proline in stress tolerance remains controversial as some authors have reported high proline levels in the susceptible cultivars subjected to stress conditions [34,35], while others have highlighted the opposite trend [5]. It has been suggested that proline functions are related to the plant water status but not as a measure of the level of tolerance [36], as also observed in soybean by Silvente [5]. In our study, the higher levels of ABA and proline identified in the sensitive BR 16 is in agreement with these findings. In addition, the metabolic profiles indicated that in the sensitive BR 16 genotype, the leaf cellular system is suffering greater damage when exposed to the same water potential of −1.0 Mpa. Thus, the signal for drought by ABA and proline were more noticeable in the sensitive BR 16. In the same way, amino acids and sugar were more abundant during drought in the sensitive genotype, which suggests that these compounds were not involved in osmoprotection of the tolerant genotype. This indicates that BR 16 has more evident cell damage and probably presents a higher protein degradation [37]. The proteome analysis is in accordance with the hierarchical clustering analysis by ClueGO, which found clusters of proteins related to the regulation of translational termination in the sensitive BR 16. The eukaryotic translation initiation factor 5A3 isoform X1 was found up-regulated in the BR 16 (Figure 2) and could be involved as an attempt to balance protein status. Wang [38] showed that this protein is a stress-responsive gene involved in the ABA signal transduction pathway, which facilitates protein synthesis and regulates several physiological pathways to improve stress tolerance. Another up-regulated protein only found in the intolerant Embrapa 48 was the chloroplast RNA-binding protein 31B (RBP). The importance of the mRNA in stabilizing mechanisms against oxidative or enzymatic degradation was observed in plants tolerant to salt and drought stress. Recently, it was shown that the RNA-binding protein (RBP) participates in the selective mRNA translation mechanism during hypoxia and drought by sequestrating mRNAs in stress granules. RBP was found up-regulated in both cells and plants upon long-term exposure to PEG and ABA [39]. Interestingly, when exposed to drought stress, RBP-overexpressing plants also accumulate less proline, suggesting that the higher tolerance observed is independent of the proline accumulation [39]. For sensitive BR 16, a differential expression of these RNA binding proteins was not found (Figure 2). However, the phosphorylated isoform for the chloroplast stem-loop binding protein of 41 kDa, that also binds and cleaves RNA, particularly in stem-loops and participates in chloroplast ribosomal RNA metabolism, was down-

regulated for both genotypes (Figure 2). This protein was also reported as down-regulated in tomato plants [40] under stress drought.

A notable distinct protein expression profile between the genotypes was also observed for the protein Transketolase (Figures 7 and 8). We found nine phosphorylated isoforms in the tolerant Embrapa 48, and eight in the sensitive BR 16. Six were up-regulated (spots 322, 323, 320, 318, 921 and 316) and one was down-regulated (spot 319) for tolerant. However, for the sensitive genotype four were down-regulated (spots 1092, 1082, 607, 1021). Transketolases are key enzymes of the reductive and oxidative pentose phosphate pathways, and are responsible for the synthesis of sugar phosphate intermediates. They are expressed in three enzymatically active isoforms and their accumulation is associated with the rehydration process [41]. This observation suggests a possible role for these transketolase isoforms in the conversion of sugars during drought stress response. The complex phosphorylation behavior of this protein was confirmed by *in silico* prediction analysis of the phosphorylated sites (Supplementary Tables S6 and S7) and experimentally by Rocha [33].

## CONCLUDING REMARKS

Drought tolerance in plants is due to many complex mechanisms and to evaluate which is the most significant for this phenotype is challenging because an extensive gene reprogramming is activated. Thus, a physiological and molecular analysis of the contrasting genotypes, showing a similar genetic background, could be an efficient approach. In this study, two Brazilian soybean genotypes were confirmed as drought-sensitive and tolerant by dissecting curves, protein expression and metabolites profiles, showing distinct responses. Our results indicate that the metabolic response observed is in accordance with a physiological study and explains the higher leaf growth rates observed for the tolerant genotype Embrapa 48. However, the tolerance appears not to because of osmoprotection by amino acids biosynthesis or by ABA-dependent cascades, as determined by metabolic profiles. In contrast, it is indicated that the tolerant genotypes maintain a photosynthetic steady-state even during drought conditions. A lower perturbation level of the metabolic pathways for the Embrapa 48 was observed. This behavior was also found in the gene expression profiles and suggests a lack of adaptive traits, which are important for the plants resistance to drought in the sensitive genotype [18]. 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. In physiological assays [17] lower rates of chlorophyll degradation were found for Embrapa 48, which could have allowed the recovery of the photosynthesis after re-irrigation and higher stomatal conductance values after re-irrigation [17]. In addition, the transpiration rate of both cultivars decreased with the development of water stress. However, transpiration was higher for BR 16 plants and reduced the soil moisture more rapidly than the Embrapa 48 [17]. Our

proteomic and metabolomic data is in accordance with physiological and expression gene results and indicate that the metabolism of the tolerant Embrapa 48 is not impaired, or relative to the sensitive BR 16. Thus, regulatory cascades of the metabolic activities, mediated by protein phosphorylation, may be operating as a result of water saving mechanisms and of a more efficient translocation of the root-shoot system.

## SUPPLEMENTARY MATERIALS

The following supplementary materials are available online at https://doi.org/10.20900/cbgg20190022:

Supplementary Figure S1: Leaf turgor of the soybean leaves under drought and water supply.

Supplementary Figure S2: SDS-PAGE 1D profiles of the protein extracts fractioned with PEG and stained by Coomassie blue.

Supplementary Figure S3: Gene Ontology (GO) classification of the differentially expressed proteins identified from the leaves of soybean Embrapa 48 (A) and BR 16 (B) genotypes under drought stress.

Supplementary Figure S4: Biological processes enriched by ClueGO analysis of the proteins that were up-regulated in the Embrapa 48 (A) and BR16 (B) genotypes under drought stress.

Supplementary Figure S5: Clustering analysis generated by ClueGO plugin showing functionally grouped networks of the enriched GO terms using a kappa score level ≥ 0.3.

Supplementary Figure S6: Clustering analysis generated by *ClueGO* showing functionally grouped networks of the enriched GO terms using a kappa score level ≥ 0.3.

Supplementary Figure S7: Biological Processes enriched generated ClueGO analysis from the proteins that were down-regulated of the Embrapa 48 (A) and BR 16 (B) genotypes under drought stress.

Supplementary Figure S8: Biological Process enriched generated ClueGO analysis from the phosphoproteins that were UP-regulated of the Embrapa 48 (A) and BR 16 (B) genotypes under drought stress.

Supplementary Figure S9: Clustering analysis by tridimensional score plot *PCA* for the metabolite abundances of the soybean leaves identified by GC/MS.

Supplementary Figure S10: Clustering analysis by *bidimensional score plot* PLS-DA for the metabolite abundances of the soybean leaves identified by GC/MS.

Supplementary Figure S11: Determination of the metabolites that contributed for sample distributions on the PLS-DA plot (Figure 10).

Supplementary Figure S12: Absolute abundance of the phytohormones Abscisic Acid (ABA), ACC and AIA of the soybean leaves quantified by LC/MS.

Supplementary Table S1: Transition list of the ions monitored for quantification analysis of the phytohormones from soybean leaves by MRM.

Supplementary Table S2: Differentially expressed proteins of the soybean leaves from the Embrapa 48 identified by mass spectrometry.

Supplementary Table S3: Differentially expressed proteins of the soybean leaves from BR 16 identified by mass spectrometry.

Supplementary Table S4: Differentially expressed phosphoproteins of the soybean leaves from Embrapa 48 identified by mass spectrometry.

Supplementary Table S5: Differentially expressed phosphoproteins of the soybean leaves from BR 16 identified by mass spectrometry.

Supplementary Table S6: Phosphorylated sites predicted in silico to the protein sequences of BR 16, stained as phosphoprotein in 2DE gels, and present in plant phosphoprotein database P3DB.

Supplementary Table S7: Phosphorylated sites predicted in silico to the protein sequences of Embrapa 48, stained as phosphoprotein in 2DE gels, and present in plant phosphoprotein database P3DB

Supplementary Table S8: Protein isoforms from BR 16 detected in the 2DE gels and identified by mass spectrometry.

Supplementary Table S9: Protein isoforms from Embrapa 48 detected in the 2DE gels and identified by mass spectrometry.

## DATA AVAILABILITY

All data generated from the study are available in the manuscript or supplementary files.

## AUTHOR CONTRIBUTIONS

HJOR, EGB, MEL, AM and EPBF designed the study; BPB, ROM, CEV and FSC performed greenhouse experiments; BPB, HJOR performed proteomic assays; LLL, FSC, CEV and HJOR performed metabolomic assays; HJOR, LLL, JCFS, FSC, FMSC analyzed the data; LLL, FSC, and HJOR wrote the paper with input from all authors; LLL, FSC, CEV and HJOR performed data integration.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## FUNDING

This research was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq) and Instituto Nacional de Ciência e Tecnologia (INCT).

**ACKNOWLEDGMENTS**

The authors would like to thank the Núcleo de Análise de Biomoléculas (NuBioMol), the Universidade Federal de Viçosa, MG, Brazil, the Instituto de Biotecnologia Aplicada a Agropecuária (BIOAGRO), the Instituto Nacional de Ciência e Tecnologia em Interações Planta-Praga (INCTIPP) and the Brazilian Soybean Genome Consortium (GENOSOJA). This work was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq) and Instituto Nacional de Ciência e Tecnologia (INCT).

**REFERENCES**

1. Pedersen P. Soybean growth and development. Ames (US): Iowa State University, University Extension; 2004.
2. Fang Y, Xiong L. General mechanisms of drought response and their application in drought resistance improvement in plants. Cell Mol Life Sci. 2015;72(4):673-89. doi: 10.1007/s00018-014-1767-0
3. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry–based metabolite profiling in plants. Nat Protoc. 2006;1(1):387-96. doi: 10.1038/nprot.2006.59
4. Obata T, Fernie AR. The use of metabolomics to dissect plant responses to abiotic stresses. Cell Mol Life Sci. 2012;69(19):3225-43. doi: 10.1007/s00018-012-1091-5
5. Silvente S, Sobolev AP, Lara M. Metabolite adjustments in drought tolerant and sensitive soybean genotypes in response to water stress. PLoS One. 2012;7(6):e38554. doi: 10.1371/journal.pone.0038554
6. Bowne JB, Erwin TA, Juttner J, Schnurbusch T, Langridge P, Bacic A, et al. Drought responses of leaf tissues from wheat cultivars of differing drought tolerance at the metabolite. level. Mol Plant. 2012;5(2):418-29. doi: 10.1093/mp/ssr114
7. Hare PD, Cress WA, Van Staden J. Dissecting the roles of osmolyte accumulation during stress. Plant Cell Environ. 1998;21(6):535-53. doi: 10.1046/j.1365-3040.1998.00309.x
8. Zhu JK. Salt and Drought Stress Signal Transduction in Plants. Annu Rev Plant Biol. 2002;53:247-73. doi: 10.1146/annurev.arplant.53.091401.143329
9. Wang X, Cai X, Xu C, Wang Q, Dai S. Drought-responsive mechanisms in plant leaves revealed by proteomics. Int J Mol Sci. 2016;17(10):1706. doi: 10.3390/ijms17101706
10. Ahmad P, Latef A, Arafat AH, Rasool S, Akram NA, Ashraf M, et al. Role of proteomics in crop stress tolerance. Front Plant Sci. 2016;7:1336. doi: 10.3389/fpls.2016.01336
11. Das A, Eldakak M, Paudel B, Kim DW, Hemmati H, Basu C, et al. Leaf proteome analysis reveals prospective drought and heat stress response mechanisms in soybean. Biomed Res Int. 2016;2016:6021047. doi: 10.1155/2016/6021047

12. Wang X, Komatsu S. Proteomic approaches to uncover the flooding and drought stress response mechanisms in soybean. J Proteomics. 2018;172:201-15. doi: 10.1016/j.jprot.2017.11.006
13. Das A, Rushton PJ, Rohila JS. Metabolomic profiling of soybeans (*Glycine max* L.) reveals the importance of sugar and nitrogen metabolism under drought and heat stress. Plants. 2017;6(2):21. doi: 10.3390/plants6020021
14. Bray EA. Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. J Exp Bot. 2004;55(407):2331-41. doi: 10.1093/jxb/erh270
15. Ku YS, Au-Yeung WK, Yung YL, Li M-W, Wen C-Q, Liu X, et al. Drought stress and tolerance in soybean. In: Board, J, editor. A comprehensive survey of international soybean research-genetics, physiology, agronomy and nitrogen relationships. London (UK): InTech; 2013. doi: 10.5772/52945
16. Oya T, Nepomuceno AL, Neumaier N, Farias JR, Tobita S, Ito O. Drought tolerance characteristics of Brazilian soybean cultivars. Plant Prod Sci. 2004;7(2):129-37. doi: 10.1626/pps.7.129
17. Carvalho JDFC, Crusiol LGT, Perini LJ, Sibaldelli RN, Ferreira LC, Marcelino-Guimarães FC, et al. Phenotyping soybeans for drought responses using remote sensing techniques and non-destructive physiological analysis. Global Sci Technol. 2015;8(2). doi: 10.14688/1984-3801/gst.v8n2p1-16
18. Neves-Borges AC, Guimarães-Dias F, Cruz F, Mesquita RO, Nepomuceno AL, Romano E, et al. Expression pattern of drought stress marker genes in soybean roots under two water deficit systems. Genet Mol Biol. 2012;35(1):212-21. doi: 10.1590/S1415-47572012000200003
19. Da Matta FM, Maestri M, Mosquim PR, Barros RS. Photosynthesis in coffee (*Coffea arabica* and *C. canephora*) as affected by winter and summer conditions. Plant Sci. 1997;128(1):43-50. doi: 10.1016/s0168-9452(97)00142-8
20. Aryal UK, Krochko JE, Ross AR. Identification of phosphoproteins in *Arabidopsis thaliana* leaves using polyethylene glycol fractionation, immobilized metal-ion affinity chromatography, two-dimensional gel electrophoresis and mass spectrometry. J Proteome Res. 2012;11(1):425-37. doi: 10.1021/pr200917t
21. Coutinho FS, Santos DS, Lima LL, Vital CE, Santos LA, Pimenta MR, et al. Mechanism of the drought tolerance of a transgenic soybean overexpressing the molecular chaperone BiP. Physiol Mol Biol Plants. 2019;25(2):457-72. doi: 10.1007/s12298-019-00643-x
22. Agrawal GK, Thelen JJ. A high-resolution two dimensional Gel-and Pro-Q DPS-based proteomics workflow for phosphoprotein identification and quantitative profiling. Methods Mol Biol. 2009;527:3-19. doi: 10.1007/978-1-60327-834-8_1
23. Vital CE, Giordano A, Soares EA, Williams TCR, Mesquita RO, Vidigal PMP, et al. An integrative overview of the molecular and physiological responses of sugarcane under drought conditions. Plant Mol Biol. 2017;94:577-94. doi: 10.1007/s11103-017-0611-y
24. Lisec J, Steinfath M, Meyer RC, Selbig J, Melchinger AE, Willmitzer L, et al. Identification of heterotic metabolite QTL in Arabidopsis thaliana RIL and IL populations. Plant J. 2009;59(5):777-88. doi: 10.1111/j.1365-313X.2009.03910.x

25. Cuadros-Inostroza Á, Caldana C, Redestig H, Kusano M, Lisec J, Peña-Cortés H, et al. TargetSearch—a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. BMC Bioinformatics. 2009;10(1):428. doi: 10.1186/1471-2105-10-428
26. Palmieri MC, Lindermayr C, Bauwe H, Steinhauser C, Durner J. Regulation of plant glycine decarboxylase by *S*-nitrosylation and glutathionylation. Plant Physiol. 2010;152(3):1514-28. doi: 10.1104/pp.109.152579
27. Dixon DP, Skipsey M, Grundy NM, Edwards R. Stress-induced protein *S*-glutathionylation in *Arabidopsis*. Plant Physiol. 2005;138(4):2233-44. doi: 10.1104/pp.104.058917
28. Wang X, Oh M, Sakata K, Komatsu S. Gel-free/label-free proteomic analysis of root tip of soybean over time under flooding and drought stresses. J Proteomics. 2016;130:42-55. doi: 10.1016/j.jprot.2015.09.007
29. Broin M, Cuiné S, Eymery F, Rey P. The plastidic 2-cysteine peroxiredoxin is a target for a thioredoxin involved in the protection of the photosynthetic apparatus against oxidative damage. Plant Cell. 2002;14(6):1417-32. doi: 10.1105/tpc.001644
30. Yi XP, Zhang YL, Yao HS, Han JM, Chow WS, Fan DY, et al. Changes in activities of both photosystems and the regulatory effect of cyclic electron flow in field-grown cotton (*Gossypium hirsutum* L) under water deficit. J Plant Physiol. 2018;220:74-82. doi: 10.1016/j.jplph.2017.10.011
31. Hu L, Wang Z, Huang B. Diffusion limitations and metabolic factors associated with inhibition and recovery of photosynthesis from drought stress in a $C_3$ perennial grass species. Physiolog Plant. 2010;139(1):93-106. doi: 10.1111/j.1399-3054.2010.01350.x
32. Igamberdiev AU. Control of Rubisco function via homeostatic equilibration of $CO_2$ supply. Front Plant Sci. 2015;6:106. doi: 10.3389/fpls.2015.00106
33. Rocha AG, Mehlmer N, Stael S, Mair A, Parvin N, Chigri F, et al. Phosphorylation of *Arabidopsis* transketolase at $Ser^{428}$ provides a potential paradigm for the metabolic control of chloroplast carbon metabolism. Biochem J. 2014;458(2):313-22. doi: 10.1042/BJ20130631
34. Sundaresan S, Sudhakaran PR. Water stress-induced alterations in the proline metabolism of drought-susceptible and-tolerant cassava (*Manihot esculenta*) cultivars. Physiolog Plant. 1995;94(4):635-42.
35. Hieu DT, Jacobs M, Angenon G, Hermans C, Thu TT, Son LV, et al. Proline Accumulation and Delta(1)-pyrroline-5-carboxylate synthetase gene properties in three rice cultivars differing in salinity and drought tolerance. Plant Sci. 2003;165:1059-68.
36. Lazcano-Ferrat I, Lovatt CJ. Relationship between relative water content, nitrogen pools, and growth of *Phaseolus vulgaris* L. and *P. acutifolius* A. Gray during water deficit. Crop Sci. 1999;39(2):467-75. doi: 10.2135/cropsci1999.0011183X0039000200028x
37. Huang T, Jander G. Abscisic acid-regulated protein degradation causes osmotic stress-induced accumulation of branched-chain amino acids in *Arabidopsis thaliana*. Planta. 2017;246(4):737-47. doi: 10.1007/s00425-017-2727-3

38. Wang L, Xu C, Wang C, Wang Y. Characterization of a eukaryotic translation initiation factor 5A homolog from *Tamarix androssowii* involved in plant abiotic stress tolerance. BMC Plant Biol. 2012;12(1):118. doi: 10.1186/1471-2229-12-118
39. Ambrosone A, Batelli G, Nurcato R, Aurilia V, Punzo P, Bangarusamy DK, et al. The Arabidopsis AtRGGA RNA binding protein regulates tolerance to salt and drought stress. Plant Physiol. 2015;168(1):292-306. doi: 10.1104/pp.114.255802
40. Tamburino R, Vitale M, Ruggiero A, Sassi M, Sannino L, Arena S, et al. Chloroplast proteome response to drought stress and recovery in tomato (*Solanum lycopersicum* L.). BMC Plant Biol. 2017;17(1):40. doi: 10.1186/s12870-017-0971-0
41. Bernacchia G, Schwall G, Lottspeich F, Salamini F, Bartels D. The transketolase gene family of the resurrection plant *Craterostigma plantagineum*: differential expression during the rehydration phase. EMBO J. 1995;14(3): 610-8.

How to cite this article:
Lima LL, Balbi BP, Mesquita RO, da Silva JCF, Coutinho FS, Carmo FMS, et al. Proteomic and Metabolomic Analysis of a Drought Tolerant Soybean Cultivar from Brazilian Savanna. Crop Breed Genet Genom. 2019;1:e190022. https://doi.org/10.20900/cbgg20190022