In silico identification of known osmotic stress responsive genes from Arabidopsis in soybean and Medicago

Nina M. Soares-Cavalcanti¹, Luis C. Belarmino¹, Ederson A. Kido¹, Ana C. Wanderley-Nogueira¹, João P. Bezerra-Neto¹, Rafaela Cavalcanti-Lira¹, Valesca Pandolfi¹, Alexandre L. Nepomuceno², Ricardo V. Abdelnoor², Leandro C. Nascimento³ and Ana M. Benko-Iseppon¹

¹Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal de Pernambuco, Recife, PE, Brazil.
²Embrapa Soja, Londrina, PR, Brazil.
³Laboratório de Genômica e Expressão, Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.

Abstract

Plants experience various environmental stresses, but tolerance to these adverse conditions is a very complex phenomenon. The present research aimed to evaluate a set of genes involved in osmotic response, comparing soybean and medicago with the well-described Arabidopsis thaliana model plant. Based on 103 Arabidopsis proteins from 27 categories of osmotic stress response, comparative analyses against Genosoja and Medicago truncatula databases allowed the identification of 1,088 soybean and 1,210 Medicago sequences. The analysis showed a high number of sequences and high diversity, comprising genes from all categories in both organisms. Genes with unknown function were among the most representative, followed by transcription factors, ion transport proteins, water channel, plant defense, protein degradation, cellular structure, organization & biogenesis and senescence. An analysis of sequences with unknown function allowed the annotation of 174 soybean and 217 Medicago sequences, most of them concerning transcription factors. However, for about 30% of the sequences no function could be attributed using in silico procedures. The establishment of a gene set involved in osmotic stress responses in soybean and barrel medic will help to better understand the survival mechanisms for this type of stress condition in legumes.

Key words: osmotic stress, stress-responsive genes, Glycine max, Medicago truncatula.

Introduction

In the course of evolution, plants have acquired a myriad of developmental and metabolic strategies to cope with the adverse effects of environmental stresses during vegetative growth and reproduction (Parry et al., 2005), making stress tolerance a complex phenomenon.

Stress perception and the immediate induction of signals that culminate in adaptive responses are key steps leading to plant stress tolerance. Tolerance stress differences between genotypes or different developmental stages of a single genotype may arise from peculiarities in signal perception and transduction mechanisms (Chinnusamy et al., 2004). Under osmotic stress conditions diverse sets of physiological responses are activated, including metabolic and defense systems used to sustain growth and for survival.

The stress-inducible genes are classified into two major groups: one of them protects the plant directly against stresses, whereas the other regulates gene expression and signal transduction (Valliyodan and Nguyen, 2006).

Because plant tolerance against osmotic stress is a complex multigenic trait, a demand exists for genome wide analysis, including ‘omics’ approaches suitable for uncovering important gene sets involved in this important process (Hirayama and Shinozaki, 2010).

After the ‘sequencing era’, genetic information was then available for several non-model plants, including some legume species, a group that exhibits unique features, such as the ability to carry the nodulation process. Nitrogen fixation mediated by nodule activities abolishes the need for external nitrogen sources from fertilizers, while providing the so-called ‘green manuring’ that enriches the soil. Moreover, some legumes, such as soybean, barrel medic and cowpea, are important economic crops that provide humans with food, livestock for feeding purposes, and industry with raw materials (Graham and Vance, 2003).
Soybean is an example of a non-model plant with plentiful transcriptome information available. Among available databases, the Genosoja platform connects public and restricted data, providing 60,747 unigenes (Nascimento et al., 2012, this issue).

The identification of candidate genes in soybean and barrel medic will provide additional evidence of the response mechanisms for osmotic stresses in Fabaceae, yielding useful information for crop improvement. As osmotic stress cannot be solved solely via remedial land management, tolerant crops - able to maintain cellular turgor and osmotic balance - may contribute significantly to reduce this economic burden. The key to plant engineering for osmotic tolerance lies in the knowledge of the underlying mechanisms of plant adaptive responses (Hariadi et al., 2011).

In the present work the main categories of osmotic stress genes known from *A. thaliana* were identified in the soybean (Genosoja Project) and barrel medic (*M. truncatula* database) transcriptomes through an *in silico* approach, in order to contribute to a better understanding of the early molecular adaptation to osmotic (drought and salinity) stress in both leguminous plants.

**Materials and Methods**

In a previous study based on 7,000 Arabidopsis genes, Seki et al. (2002) identified 103 coding genes distributed over 27 functional categories (Table 1) whose expression increased more than five times in response to osmotic stress. The protein sequences of these stress-inducible genes were obtained at the RIKEN Arabidopsis Full-Length Clone Database, and used as query sequences.

After this step, a local bank with the retrieved sequences was generated in order to make searches for similar sequences against the Genosoja platform (Nascimento et al., 2012) and the *M. truncatula* database (Quackenbush et al., 2000) using the tBLASTn algorithm (Altschul et al., 1990) with a cut-off of 1e-05. The results were annotated in other local database for further analyses and for comparisons among studied organisms and literature information. In view of the different number of seed sequences per category, the results obtained from each category and organism were normalized. The soybean and *Medicago* genes with unknown function were submitted to the AutoFACT program (Koski et al., 2005), and annotated according to the data available in the largest functional annotation databases (KEGG, COG, PFAM, SMART, nr). This step was performed in order to categorize these sequences and assign function to them, based on a comparative analysis.

**Results and Discussion**

The stress-inducible gene products were classified into two main groups: (I) those that are at the front line of defense, protecting the plant against adverse conditions and (II) those that regulate genic expression and signal transduction in the stress response (Seki et al., 2003). The first group included proteins that probably act in the protection of plant cells from dehydration, such as the enzymes required for the biosynthesis of various osmoprotectants, LEA proteins, antifreeze proteins, chaperones and detoxification enzymes. The second group included signaling mol-

| Functional category          | # Seed sequence | Functional category          | # Seed sequence |
|-----------------------------|-----------------|-----------------------------|-----------------|
| bZIP TF                     | 1               | WRKY TF                     | 2               |
| Photosynthesis              | 1               | Osmoprotectant              | 3               |
| Signaling                   | 1               | ZincFinger TF               | 3               |
| Reproductive development    | 1               | Detoxification enzyme       | 2               |
| Respiration                 | 1               | Cellular metabolism         | 3               |
| DNA nucleus                 | 1               | DREB and ERF TF             | 2               |
| Ferritin                    | 1               | Ethylene biosynthesis       | 2               |
| LEA protein                 | 1               | Cytochrome P450              | 2               |
| MYB TF                      | 1               | Fatty Acid metabolism       | 4               |
| Homeodomain TF              | 1               | Heat Shock protein          | 2               |
| Membrane protein            | 2               | Kinase protein              | 2               |
| Senescence-related          | 1               | Carbohydrate metabolism     | 6               |
| Degradation protein         | 1               | Plant defense                | 4               |
| Secondary metabolism        | 1               | Transport protein ion channel carrier | 4 |
| Water channel protein       | 1               | Cellular struct. organiz. and biogenesis | 5 |
| NAC TF                      | 2               | Unknown protein             | 37              |
| Protein phosphatase         | 2               |                              |                 |
| Total                       | 103             |                              |                 |
molecules such as transcription factors and protein kinases, among others (Seki et al., 2003). Twenty-seven categories of these two groups classified according to Seki et al. (2002) were analyzed, resulting in 1,088 (soybean) and 1,210 (Medicago) sequences (Table S1, supplementary material). In both genomes the ‘unknown protein’ category was the most representative (Figure 1), with 268 candidates for soybean and 331 for Medicago, followed by ‘cellular structure organization and biogenesis’, ‘plant defense’ and ‘transport protein ion channel carrier’ categories (Figure 1).

The highest number of sequences for genes with ‘unknown function’ - a very common category in expression essays regarding osmotic stress response in plants – attracting great interest from researchers, since those genes represent a clear source of new candidates for breeding purposes. Previous studies highlighted the importance of analyzing the role of stress-induced genes, not only for a further understanding of the molecular mechanisms of stress tolerance in higher plants, but also for improving crop performance using gene manipulation (Seki et al., 2002).

Osmotic stress greatly affects cells both at the micro (i.e., membrane structure), and at the macro level (i.e. the physiology of the whole plant), with results that reflect the variety of responses involved in the acquisition of tolerance. At the microcellular level, the activation of genes in the categories ‘cellular structure, organization and biogenesis’ (soybean: 62; Medicago: 66) and ‘transport protein ion channel carrier’ (soybean: 64; Medicago: 60) was observed, showing the importance of the maintenance of cellular structures and of the control of ion exchange with the environment.

Furthermore, we observed the activation of genes in the category ‘plant defense’ (soybean: 66; Medicago: 60), indicating the presence of a cross-talk process between pathways, a common mechanism in plants under stressful conditions. In addition to stress-specific adaptive responses, plants also share responses that protect them from more than one type of stress (Seki et al., 2002; DeFalco et al., 2010; Nuruzzaman et al., 2010), a response also observed in cowpea, another Fabaceae member (Kido et al., 2011).

Amongst the candidates of the second group of responses, composed of genes involved in signal transduction and regulation of expression (203 in soybean and 190 in Medicago; Figure 2), the category transcription factor (TF) was the most prevalent, representing up to 80% in soybean and 82% in Medicago (Figure 2). The high number of transcription factors suggests that transcriptional regulation is an important mechanism in the signal transduction triggered by osmotic stresses in both legumes.

A surprising result was the absence of a bZIP representative in the soybean database, while in Medicago this category was represented by three candidates (Figure 3). This transcription factor has been identified in many plants and is known to participate in various responsive pathways, including abiotic stress response.

Among the transcription factors, the DREB/ERF and Zinc-finger families had the highest number of sequences (Figure 3). This result was expected, since from more than 1,600 transcription factors encoded by A. thaliana, 9% are members of the DREB/ERF-like family (Dietz et al., 2010). Due to the versatility of functions that the zinc finger family may have, as well as the variety of their structural proteins, the obtained result was expected. According to

---

**Figure 1** - Main categories of Group I stress-inducible genes (protective molecules), indicating the number of orthologs identified in Glycine max and Medicago truncatula.

**Figure 2** - Percentage of stress-inducible genes (Group II), including cell signaling factors identified in Glycine max and Medicago truncatula.
Takatsuji (1998), plants seem to have adopted preexisting prototype zinc-finger motifs, generating new zinc-finger domains to adapt them to various regulatory processes. The zinc finger domain can be present in a number of transcription factors and play critical roles in interactions with other molecules. Mutations in some of the genes coding for zinc-finger proteins have been found to cause profound developmental aberrations or defective responses to environmental cues (Takatsuji, 1998). Zinc finger proteins are required for key cellular processes including transcriptional regulation, development, pathogen defense, and stress responses (Ciftci-Yilmaz and Mittler, 2008). A recent study of rice showed that the C2H2-type zinc finger family alone was represented by 189 members and demonstrated that at least 26 of them respond to different environmental stresses (Agarwal et al., 2007). Moreover, Gong et al. (2010), in a study on transcriptional regulation in drought-tolerant tomato genotypes, also identified and characterized the zinc-finger family as the main activated group during the drought response.

It is important to note that the number of seed-sequences used in the search was different for each category; the ‘unknown protein’ category, for example, was represented by 37 sequences, while the ‘bZIP transcription factor’ category comprised a single sequence. Thus, it was expected that the more abundant orthologous categories would be those obtained through comparative searches with the categories composed of more query sequences.

As for the remainder, after normalizing the results, proportionally the most representative categories (7% each) were: ‘water channel proteins’, ‘protein degradation’ and ‘senescence-related’ (Figure 4). Without doubt, all categories analyzed may contribute to an improvement in osmotic tolerance, although some functions are more relevant than others. Proteins associated with ion channels and water channels are essential in the acquisition of resistance in the presence of soluble salts and water shortages, the former controlling the entry and exit of ions such as Na+, which are toxic in high concentrations, and the latter controlling water loss to the environment. Besides these proteins, those falling into the category ‘protein degradation’ are required for protein turnover and recycling of essential amino acids, while ‘senescence-related’ genes are key components in the abiotic stress response, with genes controlling subcellular changes that lead to tolerance (Seki et al., 2002).
While the normalized results evidenced similar amounts of data in the most representative categories for both organisms, in some categories there were significant variations in the number of sequences between both leguminous species (Figure 4); this difference was even greater than 50% for the categories ‘Reproductive development’ (soybean: 1,395; Medicago: 465), ‘Ferritin’ (soybean: 651; Medicago: 1,392), ‘Respiration’ (soybean: 186; Medicago: 1,302) and ‘Ethylene biosynthesis’ (soybean: 791; Medicago: 1,721). Nevertheless, this variation may be related to the conditions under which the data were generated and deposited, as well as to the number of sequences available in the respective databases. Additionally, species-specific features could be responsible for these variations, to a lesser extent.

Regarding the category ‘Unknown Protein’, screened candidates from soybean (268) and Medicago (331) were subjected to the AutoFACT program in order to assign function to these sequences, allowing the recognition of the function of 174 and 217 sequences, respectively.

As a result, 42 and 57 G. max and M. truncatula were categorized according to the COG (Cluster of Orthologous Groups) functional database in five categories (Table 2; Figure 5). Within each category, the annotation revealed that they present the same description as the matched sequences deposited in the databank. For example, the ‘Amino acid transport and metabolism’ functional category was represented just by ‘Amino Acid Permease’ sequences (Table 2).

### Table 2 - Sequence description annotated according to the COG (Cluster of Orthologous Groups) functional category in Glycine max and Medicago truncatula.

| COG functional category                              | Sequence description                  | G. max | M. truncatula |
|------------------------------------------------------|---------------------------------------|--------|---------------|
| Amino acid transport and metabolism                  | Amino acid permease                   | 9      | 8             |
| Carbohydrate transport and metabolism                | Beta-galactosidase                    | 0      | 2             |
| General function prediction only                      | Patatin                              | 4      | 17            |
| Posttranslational modification, protein turnover, chaperones | DnaJ-like protein                | 14     | 13            |
| Signal transduction mechanisms                       | Universal Stress Protein (USP) family protein | 15    | 17            |
| **Total**                                            |                                       | 42     | 57            |

![Figure 5 - Categorization of soybean and Medicago ‘unknown category’ candidates based on COG (Cluster of Orthologous Groups) functional database.](image)

#### Table 3 - Description of sequences with unknown function after AutoFACT analysis.

| Description                                      | G. max | M. truncatula |
|--------------------------------------------------|--------|---------------|
| Amino acid permease                              | 7      | 4             |
| ATP binding / kinase / protein serine/threonine kinase | 0      | 3             |
| Auxin-responsive GH3 product [Glycine max]       | 2      | 8             |
| BTB/POZ domain-containing protein                | 0      | 2             |
| Calcium ion binding                              | 2      | 4             |
| Calmodulin binding                               | 10     | 14            |
| CCT_2 domain containing protein                  | 4      | 5             |
| Copper ion binding / electron transporter         | 4      | 1             |
| Cu-binding-like domain containing protein         | 4      | 10            |
| Dev_Cell_Death domain containing protein         | 9      | 17            |
| DFL1 (DWARF IN LIGHT 1)                          | 1      | 0             |
| Dnal-like protein [Phaseolus vulgaris]            | 3      | 2             |
| F-box family protein                              | 5      | 4             |
| Heat shock protein binding                        | 3      | 4             |
| Herpes_BLLF1 domain containing protein           | 1      | 0             |
| Hydroxyproline-rich glycoprotein family protein   | 0      | 1             |
| IFRD1; interferon-related developmental regulator 1 | 8      | 0             |
| Indole-3-acetic acid-amido synthetase GH3.17, putative | 3    | 5             |
| NAC Transcription Factor                         | 4      | 10            |
| Nucleic acid binding / transcription factor       | 18     | 14            |
| Patain B2 precursor, putative                     | 1      | 0             |
| PHI-1 (PHOSPHATE-INDUCED 1)                       | 19     | 20            |
| Plastocyanin-like domain-containing protein       | 0      | 1             |
| RC12A (RARE-COLD-INDUCIBLE 2A)                    | 0      | 2             |
| SMC_N multi-domain protein                       | 1      | 3             |
| SPX domain-containing protein                     | 2      | 0             |
| Stress-inducible protein                         | 0      | 2             |
| Tify domain containing protein                    | 8      | 12            |
| Triacylglycerol lipase                            | 5      | 5             |
| Uncharacterized protein family/Unassigned protein/Protein of unknown function | 94 | 114 |
| Universal stress protein (USP) family protein     | 1      | 3             |
| Zinc finger family protein                        | 7      | 4             |
ble 2). Two candidates of Medicago, which were functionally classified into the ‘Carbohydrate transport and metabolism’ category, were also annotated on the KEGG database as involved in the beta-galactosidase pathway (Galactose Metabolism Glycan Structure – degradation), (Table 2).

The remaining previously ‘unknown’ sequences were annotated as shown in Table 3. The analysis through AutoFACT allowed a function assignment to 132 and 160 soybean and Medicago sequences, respectively. In general, the highest number of sequences was categorized as transcription factors, essential genes participating in the transcriptional regulation of plants. Although it was possible to record more than 65% of the sequences, 35% of ‘unknown’ soybean and 34% of ‘unknown’ Medicago sequences remained without their putative function identified. These are relevant data to be worked out in future functional studies, since they may represent new genes not yet described and unique to legumes.

In conclusion, even in the absence of libraries restricted to osmotic stress in the Genosoja databank, this study indicated that most of the genes involved in the osmotic stress pathways were expressed by the non-stressed soybean and Medicago libraries at least in a baseline way. The data also revealed that soybean and Medicago are a rich source of stress-responsive candidates, which can be also applied to improve soybean and other legumes. It also highlights the existence of significant diversity for most genes, useful for comparative physiological essays. The obtained data are available for gene-targeted functional evaluation using qRT-PCR, as well as other biotechnological approaches. The molecular differences detected between the compared libraries will permit the identification of important candidates by additional approaches including PCR walking, as previously done for other crops (e.g. Coemans et al., 2005).

The identified candidates are also being monitored in further expression assays carried out in the Genosoja project (considering contrasting combinations of tolerant and susceptible plants under drought stress as compared with their negative control in a time frame) providing a more complete picture of genes involved in osmotic stress response and useful for breeding and biotechnological purposes.

Acknowledgments

The authors would like to thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FACEPE (Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco), and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for their financial support.

References

Agarwal P, Arora R, Ray S, Singh AK, Singh VP, Takatsuji H, Kapoor S and Tyagi AK (2007) Genome-wide identification of C(2)H(2) zinc-finger gene family in rice and their phylogeny and expression analysis. Plant Mol Biol 65:467-485.

Autschluf SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410.

Chinnusamy V, Schumaker K and Zhu J-K (2004) Molecular genetics perspectives on cross-talk and specificity in abiotic stress signaling in plants. J Exp Bot 55:225-236.

Ciftci-Yilmaz S and Mittler R (2008) The zinc finger network of plants. Cell Mol Life Sci 65:1150-1160.

Coemans B, Matsumura H, Terauchi R, Remy S, Swennen R and Sági L (2005) SuperSAGE combined with PCR walking allows global gene expression profiling of banana (Musa acuminate), a non-model organism. Theor Appl Genet 111:1118-1126.

DeFalco TA, Bender KW and Snedden WA (2010) Breaking the code: Ca+27 sensors in plant signalling. Biochem J 425:27-40.

Dietz K-J, Vogel MO and Viehhauser A (2010) AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. Protoplasma 245:3-14.

Gong P, Zhang J, Li H, Yang C, Zhang C, Zhang X, Khurram Z, Zhang Y, Wang T, Fei Z, et al. (2010) Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. J Exp Bot 61:3563-3575.

Graham PH and Vance CP (2003) Legumes: Importance and constraints to greater use. Plant Physiol 131:872-877.

Hariadi Y, Marandon K, Tian Y, Jacobsen S-E and Shabala S (2011) Ionic and osmotic relations in quinoa (Chenopodium quinoa Willd.) plants grown at various salinity levels. J Exp Bot 62:185-193.

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.

Kido EA, Barbosa PK, Ferreira Neto JCR, PandolﬁV, Houllou-Kido LM, Crovella S and Benko-Issepon AM (2011) Identiﬁcation of plant protein kinases in response to abiotic and biotic stresses using SuperSAGE. Curr Prot Pept Sci 12:643-656.

Koski LB, Gray MW, Lang BE and Burger G (2005) AutoFACT: An Automatic Functional Annotation and Classiﬁcation Tool. BMC Bioinformatics 16:151-161.

Nascimento LC, Costa GGL, Binneck E, Pereira GAG and Carazzolle MF (2012) A web-based bioinformatics interface applied to Genosoja Project: Databases and pipelines. Genet Mol Biol 35(suppl 1): 203-211.

Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H and Kikuchi S (2010) Genome-wide analysis of NAC transcription factor family in rice. Gene 465:30-44.

Parry MAJ, Flexas J and Medrano H (2005) Prospects for crop production under drought: Research priorities and future directions. Ann Appl Biol 147:211-226.

Quackenbush J, Liang F, Holt I, Pertea G and Upton J (2000) The TIGR Gene Indices: Reconstruction and representation of expressed gene sequences. Nucleic Acids Res 28:141-145.

Sekif M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, et al. (2002) Monitoring the expression proﬁles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J 31:279-292.

Sekif 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 Plant Biol 14:194-199.
Takatsuji H (1998) Zinc-finger transcription factors in plants. Cell Mol Life Sci 54:582-596.
Valliyodan B and Nguyen HT (2006) Understanding regulatory networks and engineering for enhanced drought tolerance in plants. Curr Opin Plant Biol 9:189-195.

Internet Resources
RIKEN Arabidopsis Full-Length Clone Database, http://www.brc.riken.go.jp/lab/epd/catalog/cdnaclone.html (May, 2011)
Genosoja platform, http://bioinfo03.ibi.unicamp.br/soja/ (May, 2011)
Medicago truncatula database, http://www.medicago.org/ (May, 2011)

Supplementary Material
The following online material is available for this article:
Table S1 - Identified candidates among abiotic stress responsive gene categories in soybean and Medicago genomes.
This material is available as part of the online article from http://www.scielo.br/gmb.
Table S1 - Identified candidates among abiotic stress responsive gene categories in soybean and *Medicago* genomes based on selected arabidopsis seed sequences, as well as number of other hits, e-value and score, against the respective database of *Medicago truncatula* (Mt) and *Glycine max* (Gm).

| Arabidopsis Information | Query Sequence | Blast Result |
|-------------------------|---------------|--------------|
| **bZIP Transcription Factor** | At1g42990 | Mt_bZIP_1 | *M. truncatula* | 2 | 7.00 e-12 | 67.4 |
| | At3g10740 | Mt_Carb_Met_1 | *M. truncatula* | 7 | 0.0 | 959 |
| | Gm_Carb_Met_1 | G. max | 8 | 0.0 | 947 |
| | Gm_Carb_Met_2 | G. max | 7 | 6.00 e-121 | 430 |
| | At3g06500 | Mt_Carb_Met_2 | *M. truncatula* | 6 | 0.0 | 788 |
| | Gm_Carb_Met_3 | G. max | 2 | 8.00 e-76 | 280 |
| | At3g60130 | Mt_Carb_Met_3 | *M. truncatula* | 5 | 1.00 e-107 | 386 |
| | Gm_Carb_Met_4 | G. max | 3 | 1.00 e-166 | 582 |
| | At5g18670 | Mt_Carb_Met_4 | *M. truncatula* | 3 | 8.00 e-141 | 496 |
| | Gm_Carb_Met_5 | G. max | 19 | 0.0 | 919 |
| | At3g04240 | Mt_Carb_Met_5 | *M. truncatula* | 17 | 0.0 | 1166 |
| | Gm_Carb_Met_6 | G. max | 5 | 5.00 e-98 | 353 |
| | At2g43820 | Mt_Carb_Met_6 | *M. truncatula* | 10 | 2.00 e-106 | 382 |
| | Gm_Cell_Met_1 | G. max | 1 | 8.00 e-83 | 304 |
| | At3g53180 | Mt_Cell_Met_1 | *M. truncatula* | 7 | 9.00 e-175 | 610 |
| | Gm_Cell_Met_2 | G. max | 4 | 2.00 e-60 | 229 |
| | At3g45300 | Mt_Cell_Met_2 | *M. truncatula* | 7 | 6.00 e-153 | 536 |
| | Gm_Cell_Met_3 | G. max | 14 | 2.00 e-156 | 352 |
| | At2g39210 | Mt_Cell_Met_3 | *M. truncatula* | 8 | 1.00 e-91 | 333 |
| | Gm_Cell_Met_4 | G. max | 2 | 3.00 e-91 | 331 |
| | At2g42970 | Mt_Cell_Met_4 | *M. truncatula* | 11 | 7.00 e-145 | 510 |
| | Gm_Cell_Met_5 | G. max | 19 | 1.00 e-38 | 155 |
| | At1g68620 | Mt_Cell_Met_5 | *M. truncatula* | 19 | 7.00 e-40 | 159 |
| | Gm_Cell_Stru_Org_Biog_1 | G. max | 16 | 4.00 e-110 | 394 |
| | At1g03220 | Mt_Cell_Stru_Org_Biog_1 | *M. truncatula* | 19 | 4.00 e-145 | 510 |
| | Gm_Cell_Stru_Org_Biog_2 | G. max | 7 | 1.00 e-123 | 437 |
| | At3g10720 | Mt_Cell_Stru_Org_Biog_2 | *M. truncatula* | 7 | 1.00 e-87 | 318 |
| | Gm_Cell_Stru_Org_Biog_3 | G. max | 16 | 1.00 e-47 | 184 |
| | At5g62350 | Mt_Cell_Stru_Org_Biog_3 | *M. truncatula* | 18 | 9.00 e-47 | 182 |
| | Gm_Cell_Stru_Org_Biog_4 | G. max | 19 | 7.00 e-21 | 95.9 |
| | At5g20230 | Mt_Cell_Stru_Org_Biog_4 | *M. truncatula* | 18 | 1.00 e-22 | 102 |
| | Gm_Cytoch_P450_1 | G. max | 11 | 2.00 e-101 | 365 |
| | Cytochrome P450 | At2g34500 | Mt_Cytoch_P450_1 | *M. truncatula* | 12 | 0.0 | 684 |
| | Gm_Cytoch_P450_2 | G. max | 4 | 2.00 e-85 | 311 |
| | At3g26220 | Mt_Cytoch_P450_2 | *M. truncatula* | 10 | 8.00 e-97 | 350 |
| | Gm_Detox_Enz_1 | G. max | 14 | 4.00 e-43 | 169 |
| | Detoxification Enzyme | At2g31570 | Mt_Detox_Enz_1 | *M. truncatula* | 17 | 9.00 e-40 | 157 |
| | Gm_Detox_Enz_2 | G. max | 19 | 4.00 e-51 | 196 |
| | At2g29450 | Mt_Detox_Enz_2 | *M. truncatula* | 19 | 2.00 e-50 | 194 |
| Module                        | Gene Name          | Species | Score  | Identity  |
|------------------------------|--------------------|---------|--------|-----------|
| DNA Nucleus                  | Gm_Detox_Enz_3     | G. max  | 4.00 e-107 | 374       |
|                              | Mt_Detox_Enz_3     | M. truncatula | 1.00 e-95 | 346       |
|                              | Gm_DNA_Nuc_1       | G. max  | 3.00 e-11  | 63.2      |
|                              | Mt_DNA_Nuc_1       | M. truncatula | 1.00 e-11  | 65.1      |
| DREB/ERF Transcription Factor | Gm_DREB_ERF_TF_1   | G. max  | 2.00 e-41  | 164       |
|                              | Mt_DREB_ERF_TF_1   | M. truncatula | 2.00 e-37  | 152       |
| Ethylene Biosynthesis        | Gm_DREB_ERF_TF_2   | G. max  | 6.00 e-33  | 136       |
|                              | Mt_DREB_ERF_TF_2   | M. truncatula | 2.00 e-52  | 202       |
| Fatty Acid Metabolism        | Gm_Ethyl_Bios_1    | G. max  | 1.00 e-112 | 402       |
|                              | Mt_Ethyl_Bios_1    | M. truncatula | 4.00 e-106 | 380       |
|                              | Gm_Ethyl_Bios_2    | G. max  | 5.00 e-61  | 230       |
|                              | Mt_Ethyl_Bios_2    | M. truncatula | 7.00 e-101 | 363       |
| Fatty Acid Metabolism        | Gm_Fatty_Acid_Met_1| G. max  | 6.00 e-96  | 347       |
|                              | Mt_Fatty_Acid_Met_1| M. truncatula | 5.00 e-26  | 115       |
|                              | Gm_Fatty_Acid_Met_2| G. max  | 3.00 e-94  | 340       |
|                              | Mt_Fatty_Acid_Met_2| M. truncatula | 3.00 e-124 | 441       |
|                              | Gm_Fatty_Acid_Met_3| G. max  | 2.00 e-58  | 222       |
|                              | Mt_Fatty_Acid_Met_3| M. truncatula | 3.00 e-134 | 446       |
|                              | Gm_Fatty_Acid_Met_4| G. max  | 2.00 e-92  | 335       |
| Ferritin                     | Gm_Ferritin_1      | G. max  | 1.00 e-85  | 311       |
|                              | Mt_Ferritin_1      | M. truncatula | 1.00 e-87  | 318       |
|                              | Gm_HSF_1           | G. max  | 2.00 e-51  | 197       |
|                              | Mt_HSF_1           | M. truncatula | 1.00 e-51  | 198       |
| Heat Shock Protein           | Gm_HSF_2           | G. max  | 0.0        | 996       |
|                              | Mt_HSF_2           | M. truncatula | 0.0        | 994       |
| Homeodomain Transcription Factor | Gm_Homeodom_TF_1    | G. max  | 8.00 e-88  | 320       |
|                              | Mt_Homeodom_TF_1   | M. truncatula | 4.00 e-116 | 415       |
| LEA Protein                  | Gm_LEA_1           | G. max  | 6.00 e-09  | 55.5      |
|                              | Mt_LEA_1           | M. truncatula | 2.00 e-07  | 50.4      |
| Membrane Protein             | Gm_Memb_Prot_1     | G. max  | 1.00 e-116 | 415       |
|                              | Mt_Memb_Prot_1     | M. truncatula | 3.00 e-113 | 332       |
|                              | Gm_Memb_Prot_2     | G. max  | 6.00 e-44  | 174       |
| MYB Transcription Factor     | Gm_MYB_TF_1        | G. max  | 8.00 e-57  | 217       |
|                              | Mt_MYB_TF_1        | M. truncatula | 4.00 e-25  | 112       |
| NAC Transcription Factor     | Gm_NAC_TF_1        | G. max  | 1.00 e-89  | 325       |
|                              | Mt_NAC_TF_1        | M. truncatula | 4.00 e-91  | 330       |
|                              | Gm_NAC_TF_2        | G. max  | 2.00 e-95  | 344       |
|                              | Mt_NAC_TF_2        | M. truncatula | 2.00 e-94  | 341       |
| Osmoprotectant               | Gm_Osmoprot_1      | G. max  | 2.00 e-142 | 500       |
|                              | Mt_Osmoprot_1      | M. truncatula | 3.00 e-157 | 550       |
|                              | Gm_Osmoprot_2      | G. max  | 2.00 e-11  | 65.5      |
|                              | Mt_Osmoprot_3      | G. max  | 5.00 e-09  | 57.8      |
|                              | Mt_Osmoprot_2      | M. truncatula | 9.00 e-16  | 80.9      |
| Function                | Gene                  | Species | Gene ID  | Function          | Gene                  | Species | Gene ID  |
|-------------------------|-----------------------|---------|----------|-------------------|-----------------------|---------|----------|
| Photosynthesis          | Gm_Osmoprot_4         | G. max  | At3g57520| Mt_Osmoprot_3     | M. truncatula         | M. truncatula | At5g20830 |
|                         | Gm_Osmoprot_5         | G. max  |          |                   |                       |          |          |
|                         | Gm_Photosyn_1         | G. max  | At4g15530| Mt_Photosyn_1     | M. truncatula         |          |          |
|                         |                       |         |          |                   |                       |          |          |
| Plant Defense           | Gm_Osmoprot_4         | G. max  | At3g55430| Mt_Osmoprot_4     | M. truncatula         |          |          |
|                         |                       |         |          |                   |                       |          |          |
|                         | Gm_Photosyn_1         | G. max  | At4g13580| Mt_Photosyn_1     | M. truncatula         |          |          |
| Protein degradation    | Gm_Prot_Degrad_1      | G. max  | At2g40000| Mt_Prot_Degrad_1  | M. truncatula         |          |          |
| Protein Kinase          | Gm_Prot_Kinase_1      | G. max  | At5g25110| Mt_Prot_Kinase_1  | M. truncatula         |          |          |
| Protein Phosphatase     | Gm_Prot_Phosphat_1    | G. max  | At4g26080| Mt_Prot_Phosphat_1| M. truncatula         |          |          |
| Reproductive Development| Gm_Reprod_Develop_1   | G. max  | At5g56750| Mt_Reprod_Develop_1| M. truncatula        |          |          |
| Secondary Metabolism    | Gm_Second_Metabol_1   | G. max  | At2g38240| Mt_Second_Metabol_1| M. truncatula          |          |          |
| Senescence-Related      | Gm_Senesc_Relat_1     | G. max  | At5g13170| Mt_Senesc_Relat_1 | M. truncatula         |          |          |
| Signalling              | Gm_Sigalnling_1       | G. max  | At5g33380| Mt_Sigalnling_1   | M. truncatula         |          |          |
| Transport Protein       | Gm_Transp_Prot_Ion_1  | G. max  | At1g58360| Mt_Transp_Prot_Ion_1| M. truncatula        |          |          |
| Ion Channel Carrier     | Gm_Transp_Prot_Ion_2  | G. max  | At1g08930| Mt_Transp_Prot_Ion_2| M. truncatula          |          |          |
|                         | Gm_Transp_Prot_Ion_3  | G. max  | At5g20380| Mt_Transp_Prot_Ion_3| M. truncatula          |          |          |
|                         | Gm_Transp_Prot_Ion_4  | G. max  | At2g22500| Mt_Transp_Prot_Ion_4| M. truncatula          |          |          |
| Unknown Protein         | Gm_Unknown_1          | G. max  | At5g22290| Mt_Unknown_1      | M. truncatula         |          |          |
|                         | Gm_Unknown_2          | G. max  | At1g11210| Mt_Unknown_2      | M. truncatula         |          |          |
|                         |                       |         | At1g15430| Gm_Unknown_2      | G. max                |          |          |

**Notes:**
- **Function** indicates the biological function of the gene.
- **Gene** and **Species** represent the gene name and species, respectively.
- **Gene ID** indicates the gene identification number.
- **G. max** and **M. truncatula** refer to the species of Glycine max and Medicago truncatula, respectively.
- **Signalling** and **Transport Protein Ion Channel Carrier** are categories for specific functions.
- **Unknown Protein** category indicates genes with unknown functions.
| Gene Symbol | Species        | Log2 Fold Change | P-Value   | Window Size |
|-------------|----------------|-----------------|-----------|-------------|
| MtUnknown_3 | M. truncatula  | 7               | 1.00e-31  | 132         |
| GmUnknown_3 | G. max         | -               | 9.00e-34  | 140         |
| MtUnknown_4 | M. truncatula  | 5               | 2.00e-20  | 95.5        |
| GmUnknown_4 | G. max         | 1               | 2.00e-28  | 122         |
| MtUnknown_5 | M. truncatula  | 7               | 1.00e-76  | 281         |
| GmUnknown_5 | G. max         | 2               | 2.00e-66  | 248         |
| MtUnknown_6 | M. truncatula  | 5               | 1.00e-16  | 80.5        |
| GmUnknown_6 | G. max         | 17              | 3.00e-124 | 441         |
| MtUnknown_7 | M. truncatula  | 11              | 3.00e-81  | 298         |
| GmUnknown_7 | G. max         | 2               | 2.00e-67  | 254         |
| MtUnknown_8 | M. truncatula  | 13              | 1.00e-39  | 160         |
| GmUnknown_8 | G. max         | 12              | 2.00e-56  | 214         |
| MtUnknown_9 | M. truncatula  | 11              | 1.00e-81  | 320         |
| GmUnknown_9 | G. max         | 13              | 3.00e-106 | 380         |
| MtUnknown_10| M. truncatula  | 15              | 1.00e-61  | 233         |
| GmUnknown_10| G. max         | 4               | 3.00e-58  | 202         |
| MtUnknown_11| M. truncatula  | 11              | 6.00e-50  | 154         |
| GmUnknown_11| G. max         | 12              | 6.00e-79  | 290         |
| MtUnknown_12| M. truncatula  | 9               | 1.00e-20  | 94.0        |
| GmUnknown_12| G. max         | 13              | 3.00e-106 | 380         |
| MtUnknown_13| M. truncatula  | 4               | 4.00e-58  | 202         |
| GmUnknown_13| G. max         | 3               | 5.00e-29  | 123         |
| MtUnknown_14| M. truncatula  | 1               | 1.00e-25  | 112         |
| GmUnknown_14| G. max         | 7               | 7.00e-20  | 94.0        |
| MtUnknown_15| M. truncatula  | 12              | 3.00e-30  | 129         |
| GmUnknown_15| G. max         | 6               | 2.00e-13  | 70.1        |
| MtUnknown_16| M. truncatula  | 6               | 4.00e-18  | 86.3        |
| GmUnknown_16| G. max         | -               | 9.00e-18  | 84.7        |
| MtUnknown_17| M. truncatula  | 3               | 2.00e-15  | 77.8        |
| GmUnknown_17| G. max         | 1               | 8.00e-10  | 58.2        |
| MtUnknown_18| M. truncatula  | 3               | 2.00e-07  | 50.8        |
| GmUnknown_18| G. max         | 8               | 8.00e-66  | 246         |
| MtUnknown_19| M. truncatula  | 17              | 2.00e-70  | 262         |
| GmUnknown_19| G. max         | 1               | 3.00e-60  | 197         |
| MtUnknown_20| M. truncatula  | 1               | 2.00e-46  | 181         |
| GmUnknown_20| G. max         | 19              | 5.00e-102 | 366         |
| MtUnknown_21| M. truncatula  | 19              | 2.00e-101 | 365         |
| GmUnknown_21| G. max         | 5               | 0.0        | 510         |
| MtUnknown_22| M. truncatula  | 12              | 0.0        | 754         |
| GmUnknown_22| G. max         | 3               | 3.00e-20  | 95.1        |
| MtUnknown_23| M. truncatula  | 4               | 5.00e-32  | 134         |
| GmUnknown_23| G. max         | 1               | 3.00e-36  | 146         |
| MtUnknown_24| M. truncatula  | 5               | 4.00e-11  | 63.5        |
| MtUnknown_25| M. truncatula  | 5               | 6.00e-25  | 109         |
| MtUnknown_26| M. truncatula  | -               | 2.00e-40  | 161         |
| Accession | Organism | Symbol | Gene | Log2 Fold Change | p-value | FPKM |
|-----------|----------|--------|------|------------------|---------|------|
| At2g40140 | G. max  | Gm_Unknown_24 | 19 | 1,00 e-161 | 565 |
| MtUnknown_27 | M. truncatula | MtUnknown_27 | 19 | 5,00 e-153 | 537 |
| GmUnknown_25 | G. max | GmUnknown_25 | 20 | 1,00 e-28 | 121 |
| At4g36040 | M. truncatula | MtUnknown_28 | 19 | 2,00 e-26 | 114 |
| GmUnknown_26 | G. max | GmUnknown_26 | 9 | 1,00 e-59 | 225 |
| At4g33050 | M. truncatula | MtUnknown_29 | 14 | 7,00 e-74 | 273 |
| GmUnknown_27 | G. max | GmUnknown_27 | 18 | 3,00 e-87 | 317 |
| At5g09440 | M. truncatula | MtUnknown_30 | 19 | 2,00 e-79 | 291 |
| GmUnknown_28 | G. max | GmUnknown_28 | 9 | 9,00 e-38 | 152 |
| At1g19180 | M. truncatula | MtUnknown_31 | 14 | 3,00 e-23 | 105 |
| GmUnknown_29 | G. max | GmUnknown_29 | 1 | 4,00 e-11 | 64.3 |
| At1g17380 | M. truncatula | MtUnknown_32 | 1 | 7,00 e-14 | 74.3 |
| GmUnknown_30 | G. max | GmUnknown_30 | 4 | 4,00 e-59 | 224 |
| At1g02660 | M. truncatula | MtUnknown_33 | 5 | 7,00 e-168 | 587 |
| GmUnknown_31 | G. max | GmUnknown_31 | 2 | 6,00 e-66 | 245 |
| At2g21620 | M. truncatula | MtUnknown_34 | 1 | 2,00 e-65 | 244 |
| GmUnknown_32 | G. max | GmUnknown_32 | 7 | 2,00 e-126 | 448 |
| At1g27760 | M. truncatula | MtUnknown_35 | 1 | 6,00 e-107 | 384 |
| GmUnknown_33 | G. max | GmUnknown_33 | 4 | 3,00 e-77 | 285 |
| At1g63010 | M. truncatula | MtUnknown_36 | 3 | 8,00 e-89 | 324 |
| GmUnknown_34 | G. max | GmUnknown_34 | 1 | 1,00 e-56 | 216 |
| At2g41640 | M. truncatula | MtUnknown_37 | 2 | 9,00 e-160 | 559 |
| GmUnknown_35 | G. max | GmUnknown_35 | 14 | 5,00 e-67 | 249 |
| At1g11360 | M. truncatula | MtUnknown_38 | 19 | 1,00 e-59 | 225 |
| GmWater_Chan_1 | G. max | GmWater_Chan_1 | 19 | 2,00 e-116 | 414 |
| MtWater_Chan_1 | M. truncatula | MtWater_Chan_1 | 19 | 3,00 e-115 | 410 |
| GmWRKY_TF_1 | G. max | GmWRKY_TF_1 | 10 | 4,00 e-65 | 244 |
| MtWRKY_TF_1 | M. truncatula | MtWRKY_TF_1 | 15 | 3,00 e-62 | 235 |
| GmWRKY_TF_2 | G. max | GmWRKY_TF_2 | 12 | 9,00 e-43 | 167 |
| MtWRKY_TF_2 | M. truncatula | MtWRKY_TF_2 | 8 | 1,00 e-40 | 161 |
| GmZF_TF_1 | G. max | GmZF_TF_1 | 7 | 9,00 e-31 | 129 |
| MtZF_TF_1 | M. truncatula | MtZF_TF_1 | 15 | 4,00 e-32 | 134 |
| GmZF_TF_2 | G. max | GmZF_TF_2 | 11 | 1,00 e-24 | 107 |
| MtZF_TF_3 | M. truncatula | MtZF_TF_3 | 17 | 7,00 e-68 | 252 |
| GmZF_TF_3 | G. max | GmZF_TF_3 | 11 | 1,00 e-24 | 107 |
| MtZF_TF_2 | M. truncatula | MtZF_TF_2 | 16 | 4,00 e-67 | 250 |