VERSITA

CELLULAR & MOLECULAR BIOLOGY LETTERS
http://www.cmbl.org.pl

Received: 19 May 2014
Final form accepted: 01 December 2014

DOI: 10.2478/s11658-014-0223-3

Short communication

IDENTIFICATION OF DROUGHT-INDUCED TRANSCRIPTION FACTORS IN *Sorghum bicolor* USING GO TERM SEMANTIC SIMILARITY

MANOJ KUMAR SEKHWAL[1], AJIT KUMAR SWAMI[2], VINAY SHARMA[1,*] and RENU SARIN[2]

[1]Department of Bioscience & Biotechnology, Banasthali University, P.O. Banasthali Vidyapith 304022 Vanasthali, Rajasthan, India, [2]Department of Botany and Biotechnology, University of Rajasthan, JLN Marg, Jaipur 302055, Rajasthan, India

Abstract: Stress tolerance in plants is a coordinated action of multiple stress response genes that also cross talk with other components of the stress signal transduction pathways. The expression and regulation of stress-induced genes are largely regulated by specific transcription factors, families of which have been reported in several plant species, such as *Arabidopsis*, rice and *Populus*. In sorghum, the majority of such factors remain unexplored. We used 2DE refined with MALDI-TOF techniques to analyze drought stress-induced proteins in sorghum. A total of 176 transcription factors from the MYB, AUX_ARF, bZIP, AP2 and WRKY families of drought-induced proteins were identified. We developed a method based on semantic similarity of gene ontology terms (GO terms) to identify the transcription factors. A threshold value (≥ 90%) was applied to retrieve total 1,493 transcription factors with high semantic similarity from selected plant species. It could be concluded that the identified transcription factors regulate their target proteins with endogenous signals and environmental cues, such as light, temperature and drought stress. The regulatory network and *cis*-acting elements of the identified transcription factors in distinct families are involved in responsiveness to auxin, abscisic acid, defense, stress and light. These responses may be highly important in the modulation of plant growth and development.

* Author for correspondence. Email: vinaysharma30@yahoo.co.uk; phone: +91 1438 228302; fax: +91 1438 228365

Abbreviations used: BP – biological process; CC – cellular component; GO – gene ontology; MF – molecular function; SS – semantic similarity

Keywords: 2D Gel electrophoresis, Drought-stress, *Sorghum bicolor*, Semantic similarity, Transcription factor, GO terms, Clusters, MALDI-TOF, Functional annotation, Regulatory network, Gene ontology, Protein family

INTRODUCTION

Environmental stresses such as drought and salinity are major causes of crop loss worldwide [1]. It has been estimated that approximately 80% of sorghum production in the world occurs in dry land conditions [2]. To cope with these abiotic stresses, numerous defense response genes are transcriptionally activated, eventually leading to physiological and metabolic changes that increase the chance of plant survival [3]. Stress tolerance in plants is a coordinated action of multiple stress-induced genes that cross talk with other components of the stress signal transduction pathways [1].

The expression and regulation of stress-induced genes are largely controlled by specific transcription factors [4]. Transcription factors are important regulators of gene transcription that generally consist of at least two domains, such as the DNA-binding and activation/repression domains [5]. The transcription factors often function in networks, in which a regulatory protein controls the expression of another protein. Transcription factors are often expressed in tissue- and developmental stage-specific and stimulus-dependent pathways [6]. Several studies have demonstrated that a single transcription factor may function in several stress signaling pathways [7]. The transcription factors regulate the expression of their target genes by physically binding to the promoter regions. The binding site and common pattern (motif) of each transcription factor naturally share similarity with each other [8]. The transcription factors interact with *cis*-acting elements present in the promoter region of various stress-induced genes and activate cascades or networks of genes that act together to enhance tolerance towards multiple stresses simultaneously [9]. This property of transcription factors makes them an attractive category of genes for the manipulation of abiotic stress tolerance. Thus, stress responsive transcription factors are highly important for genetic engineering, which may lead to upregulation or downregulation of a whole array of genes under their control. Several transcription factors are involved in plant stress tolerance, regulating plant responses to different stresses [9]. Most of the stress-related transcription factors are grouped into several large families, such as AP2/ERF, bZIP, NAC, MYB, MYC, Cys2His2 zinc finger and WRKY [10]. Various transcription factors, such as DREB2, AREB1, RD22BP1 and MYC/MYB, are known to regulate ABA-responsive gene expression via interaction with their corresponding *cis*-acting elements, such as DRE/CRT, ABRE and MYCRS/MYBRS [1]. Understanding these mechanisms is important to improve the stress tolerance of crop plants.

Computational identification and annotation of transcription factors at the genomic scale are the first steps to understanding the mechanism of gene expression and regulation [11]. Using microarray data enables researchers to identify genes that have similar expression profiles and that may be involved in similar biological processes [12]. Analyzing the promoters of such genes may identify common *cis*-elements that are responsible for their co-expression [13]. System biology approaches facilitate the identification of regulatory hubs in complex networks [14], where the interaction of proteins determines the outcome of most cellular processes. The protein interaction network during stress tolerance was recently elucidated at the transcript level in *Arabidopsis thaliana* and *Oryza sativa* [15]. The gene network in sorghum that is responsive to water-limiting environments has also been reported on [16].

Sorghum bicolor is one of the most tolerant grass species to abiotic stresses, including drought, salt and heat stress. The recently completed genome sequence for sorghum [17] and other online resources provide a unique opportunity to obtain genes and gene networks [16]. The availability of the complete genome sequence has facilitated access to essential information for genes, proteins and their function, transcript level, putative *cis*-regulatory elements and alternative splicing patterns [18].

Although the number of sequenced genomes continues to grow, the functional annotation of whole genomes remains unreliable. However, several attempts have been made to annotate the function of genes. Various computational approaches to identify proteins are available, such as sequence similarity [19], phylogenetic profiles [20], protein–protein interaction (PPI) [21] and gene expression [22]. Sequence similarity-based approaches are widely used for function prediction, but they are often insufficient if the similarity is not statistically sound [23]. GO term semantic similarity provides a functional relationship between biological process (BP), molecular function (MF) and cellular component (CC). The semantic similarity between two genes is usually calculated on the basis of their GO term similarity [24]. The GO terms in the Gene Ontology database are organized as directed acyclic graphs (DAG) in those three aspects of ontologies (BP, MF, CC) [25].

Several methods exist to determine semantic similarity [26, 27]. Semantic similarities of GO terms have been used to identify clusters and functional annotation of genes and proteins [28]. We also recently reported on GO term semantic similarity-based methods to predict the function of proteins [29–31]. In this paper, we report on a more dedicated method to identify the transcription factors of drought-induced proteins. The method is based on the concept that highly semantically similar proteins are involved in same pathway. These highly semantically similar transcription factors are involved in the pathway to regulate abiotic stress responsive proteins in sorghum.

MATERIALS AND METHODS

Plant material and growth conditions

Seeds of *Sorghum bicolor* (L.) genotype csv-17 were obtained from MPUAT. The seeds were washed twice with tap water and then with distilled water. Subsequently, the seeds were sterilized with 0.1% $HgCl_2$ (Merck Ltd.) for 10 min and washed again thoroughly with distilled water. These sterilized seeds were grown hydroponically in plastic containers with 2 l distilled water in Hoagland's solution and maintained in a culture chamber as described previously [32].

Drought stress condition

After 7 days of germination, the germinated seeds were subjected to drought stress (no water supply) and their leaves were harvested after 24, 48 and 96 h. Hydroponic assemblies for control and stress were used for further analysis. Leaves from 10 to 15 seedlings from each assembly were pooled for the extraction of proteins.

Protein extraction and determination

The proteins were extracted from plant samples as per the method reported in [33]. Fresh leaves from control and stressed plants were ground to fine powder in liquid nitrogen using a mortar and pestle. About 1 g (1 part) of the powder was mixed with 10 ml (10 parts) of precipitation solution containing 10% TCA (w/v), 0.07% 2-mercaptoethanol (w/v) in 20 ml acetone (Sigma-Aldrich). The suspension was incubated at −20ºC for 60 min with intermittent mixing (every 10 min) using a cyclomixer (Bangalore Genie). The precipitated material was collected by centrifugation (25,000 × g, 4ºC, 15 min). The pellet was washed twice with a washing solution containing 20 ml acetone and 14 µl 2-mercaptoethanol, and the precipitate was air dried for 20 min. The pellet was stored at −80ºC until further use. Proteins were dissolved from the dried precipitate into lysis-buffer (8 M urea, 2% CHAPS; Sigma-Aldrich) by repeated pipetting. Insoluble material was removed by centrifugation (20,000 × g, 20ºC, 20 min) and the supernatant was clarified by passing through a 0.22 µm syringe filter (Millipore). Protease activity was kept low by maintaining the cell material at 4ºC during centrifugation and adding protease inhibitor (protease inhibitor cocktail, Sigma-Aldrich) to the lysis buffer. The total protein concentration was determined using a Quick Start Bradford Protein Assay kit (Bio-Rad) according to the manufacturer's instructions. The protein concentration was determined according to [34] using bovine serum albumin (BSA; Sigma-Aldrich) as standard.

2D gel electrophoresis

Proteins samples were purified using a 2D-cleanup kit (Bio-Rad) and the protein pellet was finally re-suspended in sample rehydration buffer consisting of 8 M urea, 2% CHAPS, 15 mM DTT and 0.5% IPG buffer (pH 4–7; Sigma-Aldrich). The isoelectric focusing was performed using IPG strips (Bio-Rad). IPG strips with a pH range from 4 to 7 were used to determine the distribution of

differentially expressed spots. For the first dimension 250 and 500 µg of protein samples in 150 and 300 µl of rehydration solution were used to rehydrate IPG strips 7 cm respectively. The IPG strips were rehydrated over night and then the proteins were focused for 10,000 Vh at 20ºC under mineral oil. After focusing, the strips were incubated for 10 min in 1ml (for 7 cm strip) of equilibration buffer I, consisting of 6 M urea, 30% glycerol, 2% SDS, 1% DTT, 50 mM Tris-HCl buffer (pH 8.8) followed by equilibration buffer II, consisting of 6 M urea, 30% glycerol, 2% SDS and 4% iodo-acetamide in 50 mM Tris/HCl buffer (pH 8.8). After equilibration steps, the strips were transferred to 12% SDS-PAGE for the second dimension using a previously described method [35]. Protein spots were visualized by staining with coomassie brilliant blue G-250. Gel images were captured using a GS800 densitometer (Bio-Rad). The relative abundance of the spots and the differential protein expression were determined using PD Quest software (Bio-Rad).

Protein spot identification using MALDI-TOF

The drought-induced protein spots were excised using thin-walled PCR tubes (200 µl) cut at the bottom with the help of a new surgical scalpel blade. The gel spots were washed with proteomic grade de-ionized water and the proteins were identified using a MALDI-TOF mass spectrometer (Ultraflex III, Bruker Daltonics). The gel pieces containing proteins were de-stained and trypsin digested using the Montage In-Gel Digest Kit (Millipore). For MALDI-TOF, 1 µl of the digest was mixed with 2 µl of the matrix solution (5 mg alpha-cyano-4-hydroxycinnamic acid, 80% acetonitrile, 0.1% trifluoroacetic acid; Sigma-Aldrich) and 1 µl of this mixture was deposited onto the MALDI target. The spectrum was obtained in the mass range of 500–4000 Da and calibrated using a calibration mixture consisting of angiotensin I, substance P, ACTH (1–17), ACTH (18–39) and somatostatin (28). The expressed proteins were analyzed using a mascot sequence-matching server (http://www.matrixscience.com) using Matrix Science Database (MSDB) in the taxonomy group of green plants. All of the expressed proteins were identified from closely related homolog proteins in *Oryza sativa*. While performing the mascot, the search parameters included a maximum of one missed cleavage by trypsin, fixed modification of oxidation, charged state of +1, peptide mass tolerance of 50 ppm, and fragment mass tolerance of ± 1.0 Da.

Bioinformatics analysis

Protein retrieval and analysis

The transcription factors of each of the expressed proteins were identified using the Stress Responsive Transcription Factor Database (STIFDBV2.0; http://caps.ncbs.res.in/stifdb/help.html#intro) [36] with closely related *A. thaliana*. The families, transcript and protein sequence of the transcription factors were noted at plant transcription factors database (plnTFDB) version 3.0 (http://plntfdb.bio.uni-potsdam.de/v3.0/) [37]. A total of 176 transcription

factors were manually identified using the PSI-BLAST database (http://www.ebi.ac.uk/Tools/sss/psiblast/) [38]. Manually, we searched new transcription factors for identified families, namely MYB, AUX_ARF, bZIP, AP2 and WRKY, and removed duplicates if they had been reported previously at the plnTFDB database. To find similar sequence proteins using PSI-BLAST, we applied the following parameters: protein database-Uniprot; *E*-value-1.0e - 3; metrix-blosum62; gap opening-11; gap extend-1; scores and alignments-1000; dropoff-15 default; final dropoff-25 default; and alignment view-pairwise with active filter.

GO terms and families were identified using the UniProt database (http://www.uniprot.org/) [39], which is cross-linked with other databases, such as Gene Ontology [24] and Pfam [40]. The *cis* regulatory elements and motifs in the transcription factors were noted using the PlantCare database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [41]. The orthologous groups for distinct families of transcription factors in sorghum were identified using blast at InParanoid7 (http://inparanoid.sbc.su.se/cgi-bin/index.cgi) [42]. While finding orthologous groups, we applied the following parameters: maximum number of hits to show-50, *E*-value cutoff-0.01 and Score cutoff-50. The gene regulatory network information was retrieved from the *Arabidopsis thaliana* Transcription Factor Database (AtTFDB; http://arabidopsis.med.ohio-state.edu/AtTFDB/) and *A. thaliana cis*-Regulatory Databases (ATcisDB; http://arabidopsis.med.ohio-state.edu/AtcisDB/) of the *Arabidopsis* Gene Regulatory Information Server (AGRIS; http://arabidopsis.med.ohio-state.edu/) [43]. The visual access to regulatory networks was via the Grassius Regulatory Grid eXplorer (GRG-X; http://arabidopsis.med.ohio-state.edu/grgx/) at the AGRIS database.

The gene ontology and domain information of similar sequence proteins was retrieved using the *Arabidopsis* Information Resource (TAIR) Database [44], which is cross-linked with several databases, such as AGRIS, InterPro (http://www.ebi.ac.uk/interpro/) [45] and InParanoid [41].

Semantic similarity conception

A method was developed to calculate GO term semantic similarity in transcription factors from *S. bicolor*, *O. sativa indica*, *O. sativa japonica* and *A. thaliana* (Fig. 1). The G-SESAME tool (http://bioinformatics.clemson.edu/G-SESAME/) with Wang's method [27] was used to calculate semantic similarities of GO terms. Semantic similarity analysis was performed individually with three GO terms (BP, MF, CC).

Several methods, including Resnik's [26], Wang's [27] and Jiang–Conrath's [46], are used to calculate semantic similarity. Previously, it was reported that highly (≥ 90%) semantically similar proteins in the distinct families of major facilitator superfamily (MFS) and glycoside hydrolase (GH) in sorghum had complex interactions [30, 31]. Kinase–protein interactions in rice also show high (> 70%) semantic similarity [47]. Clustering allowed researchers to identify

protein complexes, detect functional modules and predict protein functions [28]. Hence, semantic similarity conception attempted to identify clusters and the functional annotation of transcription factors in sorghum.

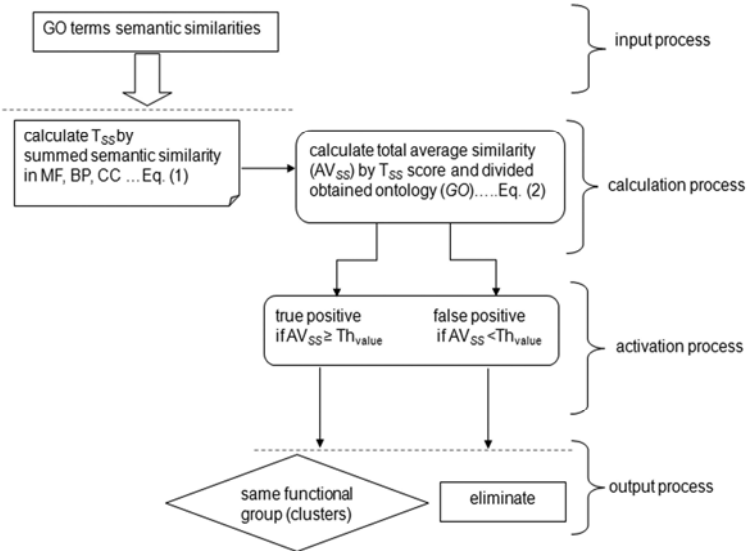

Fig 1. Schematic diagram of the proposed method. It is structured in four layers: input, calculation, activation and output.

The equation summed GO term semantic similarity in three ontologies: MF, BP and CC. The sum of the semantic similarity value was divided by total obtained for the GO terms. The algorithm is given as:

$$T_{SS} = \frac{\sum_{SS} MF}{T_T MF} + \frac{\sum_{SS} BP}{T_T BP} + \frac{\sum_{SS} CC}{T_T CC} \quad \text{Eq. (1)}$$

where $\sum_{SS}MF$, $\sum_{SS}BP$ and $\sum_{SS}CC$ are the total scores of semantic similarity divided by the total obtained GO terms (T_T) for MF, BP and CC. T_{SS} is the output of the summed semantic similarity.

Then the equation calculates the total average similarity (AV_{SS}) of the summed semantic similarity (T_{SS}) score that is divided by the obtained ontology (*GO*):

$$AV_{SS} = \frac{T_{SS}}{GO} \quad \text{Eq. (2)}$$

The calculated average similarity (AV_{SS}) value was used to compare with the threshold value. If the AV_{SS} is greater than or equal to the threshold value (90%), the total positive score (TP_{score}) will be 1. If the AV_{SS} is less than or equal to the threshold value (90%), the total positive score (TP_{score}) will be 0. An output value of 1 was considered as a true positive and 0 was a false positive.

Th_{value} is the threshold value. A threshold value of 90% was set as it has been validated earlier that ≥ 80 and ≥ 90% semantic similarities of GO terms have more probability to have same functional property [48]. For instance, SUC2-type transporters physically interact with other transporters, namely SUC3 and SUC4. Protein kinase SOS2 physically interacts with the calcium-binding protein SOS3, CBL10 and nucleoside diphosphate kinase in *Arabidopsis* [49]. These proteins have high (≥ 90%) semantic similarity with their corresponding proteins. The algorithm is given as:

$$TP_{score} = \begin{cases} 1 & if \quad AV_{SS} \geq Th_{value} \\ 0 & if \quad AV_{SS} < Th_{value} \end{cases} \quad \text{Eq. (3)}$$

RESULTS

Drought stress protein identification

Drought leads to reduced growth of leaves and increased chlorosis in comparison with the control (Fig. 2). Eight drought-inducible proteins were identified using MALDI-TOF/MS in sorghum seedlings (Fig. 3). The protein profiling of 2DE for the germinated seeds under drought conditions were analyzed and compared to those for the control *S. bicolor*. In response to drought stress, sorghum plants showed differences in traits such as reduced root length and leaf area (Fig. 2). Several differentially drought-induced protein spots were expressed. The identified eight proteins and their molecular weights are as follows (Table 1): 57.60 kDa, protein kinase (spot S1); 59.81 kDa, serine-threonine protein kinase (spot S2); 46.97 kDa, glycoside hydrolase (spot S3); 91.51 kDa, trehalose-phosphatase (spot S4); 85.98 kDa, raffinose synthase (spot S5); 24.24 kDa, inorganic pyrophosphatase (spot S6); 18.59 kDa, universal stress protein (spot S7); and 39.60 kDa, galactinol synthase (spot S8).

Transcription factor retrieval and analysis

In this study, all 176 transcription factors of different families in sorghum were identified as putative uncharacterized (Suppl. Table 1 in Supplementary material at http://dx.doi.org/10.2478/s11658-014-0223-3). These uncharacterized transcription factors from each family were assigned to find out their similar transcription factors from related and reference plant species: *S. bicolor* (64), *O. sativa indica* (154), *O. sativa japonica* (168) and *A. thaliana* (178) in the MYB family; *S. bicolor* (10), *O. sativa indica* (28), *O. sativa japonica* (34) and *A. thaliana* (23) in the AUX_ARF family; *S. bicolor* (17), *O. sativa indica* (64), *O. sativa japonica* (80) and *A. thaliana* (78) in the bZIP family; *S. bicolor* (51), *O. sativa indica* (67), *O. sativa japonica* (80) and *A. thaliana* (75) in the AP2 family; and (56), *S. bicolor* (34), *O. sativa indica* (66), *O. sativa japonica* (88) and *A. thaliana* (78) in the WRKY family. A total of 1,493 positive transcription factors were obtained at the ≥ 90% threshold value. Therefore, the threshold value (≥ 90%) in the proposed algorithms (methodology section) was applied to obtain highly semantically similar transcription factors.

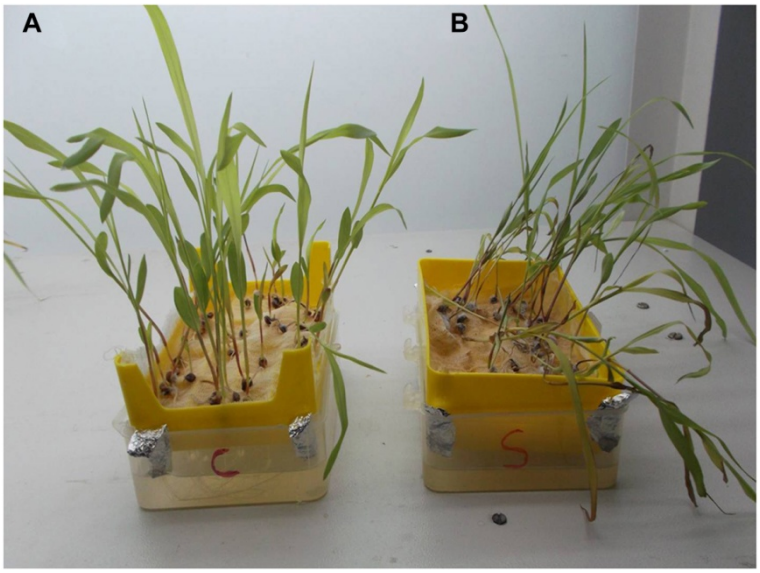

Fig. 2. Seven-day old *Sorghum bicolor* (L.) seedlings exposed to drought stress (no water supply) for 96 h (B) and the corresponding control (A).

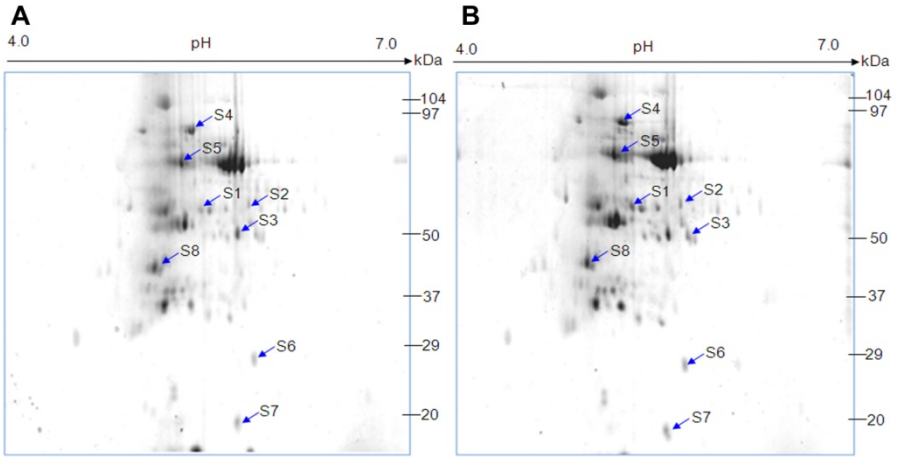

Fig. 3. 2DE gel of differentially expressed spots of *Sorghum bicolor* (L.) in control conditions (A) and under drought stress (no water supply; B). The identified spots are marked with arrows.

The PlantCare tool was used to identify *cis* elements and motifs of retrieved transcription factors using different reference plant species, such as *A. thaliana*, *Zea mays* and *Populus trichocarpa*. The majority of the transcription factors have same motifs that are involved in important pathways of plant development. For example, CAAT-box is a common *cis*-acting element in promoter and enhancer regions. It was found in all transcription factor families. The G-box,

LAMP-element, MNF1 and ACE motifs were involved in the *cis*-acting element in light responsiveness (Suppl. Table 2 in Supplementary material at http://dx.doi.org/10.2478/s11658-014-0223-3). Several other motifs with important functions were also identified in the majority of the clusters, e.g. the AuxRR-core involved in auxin responsiveness, C-repeat/DRE involved in cold and dehydration responsiveness, LTR involved in low temperature responsiveness, ABRE involved in abscisic acid responsiveness, RY-element involved in seed-specific regulation, MBS involved in drought inducibility, TC-rich repeats involved in defense and stress responsiveness, and HSE involved in heat stress responsiveness (Suppl. Table 2). The A-box motif, which was identified in all of the transcription factors, is a *cis*-acting regulatory element associated with P-box and L-box. It is involved in induced transcriptional activity. A-box in different families of transcription factors was found in the promoter regions.

Table 1. Eight protein families isolated from *Sorghum bicolor* leaves under drought stress.

Spot	Protein family	Transcription factor families
S1	Protein kinase	Myb, HSF, WRKY, AuxRE_ARF, bHLH
S2	Serine-threonine protein kinase	Myb, WRKY, AuxRE_ARF, bHLH
S3	Glycoside hydrolase	bZIP, Myb, HSF, AuxRE_ARF
S4	Trehalose-phosphatase	AuxRE_ARF, HSF, bHLH, bZIP
S5	Raffinose synthase	Myb, WRKY, AuxRE_ARF, DREB_AP2_EREBP, bZIP
S6	Inorganic pyrophosphatase	Myb, WRKY, DREB_AP2_EREBP, bZIP
S7	Universal stress protein	Myb, WRKY, AuxRE_ARF, DREB_AP2_EREBP, bZIP
S8	Galactinol synthase	Myb, WRKY, AuxRE_ARF, bHLH, bZIP

Transcription factor clusters

A total of 176 drought responsive putative uncharacterized transcription factors were identified from different families, such as MYB, AUX_ARF, bZIP, AP2 and WRKY (Suppl. Table 1). The identified transcription factors were classified in five different clusters based on their families, and each classified cluster enclosed highly (≥ 90%) semantically similar proteins. Clusters contained groups of highly semantically similar proteins that have same functional properties: family, GO terms, etc. The gene ontology of the newly identified 176 putative uncharacterized transcription factors in sorghum was mapped (Table 2). These transcription factors from each cluster were assigned to find their similar transcription factors from reference plant species, namely *A. thaliana*, *O. sativa indica* and *O. sativa japonica*. The obtained transcription factors were classified in five different clusters based on their family group. The developed algorithms (methodology section) suggested positive scores (≥ 90%) by the calculation of semantic similarity. Therefore, the transcription factors in the MYB and bZIP families were highly (100%) semantically similar in all three aspects of gene ontology: BP, MF and CC. The family members of ARF,

AUF/IAA and AP2 showed 90–100%, 80–100% and 100% semantic similarity in their molecular function, biological process and cellular component respectively with their reference plants.

Table 2. Functional identification of transcription factors classified in 5 distinct clusters and their semantic similarity.

Clusters	Transcription factor family	Gene Ontology	InterPro	(TP_{score}) %
Cluster 1	MYB	^{BP}[regulation of transcription, DNA-dependent] ^{CC}[nucleus] ^{MF}[DNA binding]	Myb, DNA-binding [pfam] SANT domain, DNA binding [smart] Homeodomain-related [superfam] Transcription regulator HTH, Myb-type, DNA-binding [prosite] Molecular chaperone, heat shock protein, Hsp40 [panther]	99
Cluster 2	AUX_ARF	^{BP}[response to hormone stimulus, regulation of transcription, DNA-dependent] ^{CC}[nucleus] ^{MF}[DNA binding, protein dimerization activity]	Auxin response factor [pfam DNA-binding pseudobarrel domain [superfam] Transcriptional factor B3 [pfam]	98
Cluster 3	bZIP	^{BP}[Involved in regulation of transcription, DNA-dependent] ^{CC}[nucleus] ^{MF}[sequence-specific DNA binding, sequence-specific DNA binding transcription factor activity]	DNA-binding WRKY [pfam,prosite,smart,superfam]	96
Cluster 4	AP2	^{BP}[regulation of transcription, DNA-dependent] ^{CC}[nucleus] ^{MF}[DNA binding, sequence-specific DNA binding transcription factor activity]	Pathogenesis-related transcriptional factor/ERF, DNA-binding [pfam,smart, superfam]	98
Cluster 5	WRKY	^{BP}[regulation of transcription, DNA-dependent] ^{CC}[nucleus] ^{MF}[sequence-specific DNA binding, sequence-specific DNA binding transcription factor activity]	DNA-binding WRKY [pfam,superfam, smart,]	99

BP, biological process; MF, molecular function; CC, cellular component

Transcription factor regulatory network and gene ontology

We found that the distribution of the semantic similarities between the identified transcription factors of sorghum in distinct clusters showed high (100%) functional or semantic similarity when we compared BP, MF and CC. Thus, these highly semantically similar transcription factors in distinct clusters may constitute highly interacted members. In this investigation, we noted that highly semantically similar proteins (in terms of GO terms) are involved in the same functions. However, several transcription factors have similar functional properties but do not interact. Thus, the proposed method considered the total average of the highly semantically similar true positive score (≥ 90%) and removed false positives (< 90%). The orthologous groups in *Arabidopsis*, rice and sorghum for distinct clusters showed the same functional properties that were identified by the InParanoid7 server. A total of 20 orthologs were retrieved from *Arabidopsis* and were analyzed for their gene ontology functional property. The genes of *Arabidopsis* were further utilized to find the regulatory network for distinct clusters of transcription factors in sorghum.

The regulatory network of the MYB family group of transcription factors were MYB, AGL1, LFY, GL2, HY5, SEPALLATA3, CAPRICE, WERWOLF, PIF4, bHLH15, AG, etc. The regulatory networks of the AUX_ARF family group of transcription factors were AP2, AGL15, bHLH15, LFY, PIF4, SEPALLATA3, ARF10, AGL15, etc. The regulatory network of the other family groups of transcription factors are shown in Table 3.

Table 3. Binding site and regulatory network of the functional group of transcription factors in distinct clusters**.**

Clusters	Binding site family	Binding site name	Binding site sequences	Family member	Regulatory network member
Cluster 1	MYB	MYB4 binding site motif	acctaac aacaaac	122	MYB, AGL15,LFY, GL2, HY5, SEPALLATA3, CAPRICE, WERWOLF, PIF4, bHLH15, AG
Cluster 2	AUX_ARF	ARF1 binding site motif	tgtctc	25	AP2, AGL15,bHLH15, LFY, PIF4, SEPALLATA3, ARF10, AGL15
Cluster 3	bZIP	ATB2/AtbZIP53/ AtbZIP44/GBF5 BS in ProDH	actcat	89	WRKY, AGL15, bHLH15, bZIP55, SEPALLATA3, AB15
Cluster 4	AP2	W-box promoter motif	ttgact	161	AGL15, CBF3, bHLH15, AP2, SEPALLATA3, WRKY, PIF4, TOE1
Cluster 5	WRKY	W-box promoter motif	ttgact	94	AGL15, PIF4, WRKY,AP2, SEPALLATA3, bHLH15, GL1

Family members: Gene model (http://plntfdb.bio.uni-potsdam.de/v3.0/index.php?sp_id=SBI)

Thus, it is clear that the distinct groups of transcription factors in different networks show the same gene ontology functional property as cellular component and molecular function. The transcription factors in clusters 1, 2, 3, 4 and 5 with their members in the regulatory network were located in the nucleus. Their molecular functions were DNA binding, chromatin binding, and sequence-specific DNA-binding transcription factor activity. In terms of biological processes, the regulatory network family members of MYB were AGL15 and LFY. They were involved in embryo development, seed dormancy, fruit abscission and dehiscence, gibberellin catabolic process, somatic embryogenesis, etc. GL2 was involved in epidermal cell fate specification, tissue development and regulation of transcription. HY5 was involved in calcium-mediated signaling, the gibberellic acid-mediated signaling pathway, mRNA export from the nucleus, positive regulation of cell proliferation, red or far-red light signaling pathway, regulation of flower development, regulation of photomorphogenesis, regulation of transcription, DNA-dependent, response to UV-B, response to abscisic acid stimulus, response to far red light, response to karrikin, response to red light, etc. SEPALLATA was mainly involved in carpel development, ovule development and specification of floral organ identity. WERWOLF was involved in cell fate specification, epidermal cell differentiation and root hair cell differentiation.

The regulatory network family members AB15 and AGL15 of bZIP were involved in the abscisic acid-mediated signaling pathway, embryo development ending in seed dormancy, the gibberellic acid-mediated signaling pathway, protein ubiquitination, regulation of flower development, regulation of transcription, response to abscisic acid stimulus, response to chitin, response to freezing, response to water deprivation, seed development, germination, the sugar-mediated signaling pathway, etc. AP2 and its regulatory network family members were involved in cell differentiation, flower development, ovule and tissue development, regulation of transcription, DNA-dependent seed development, sexual reproduction and specification of floral organ identity. The regulatory network family member GL of WRKY was involved in defense response to bacterium, negative regulation of transcription and DNA-dependent response to gibberellins stimulus. AUX_ARF and its regulatory network members were involved in response to hormone stimulus, regulation of transcription, DNA-dependent transcription, DNA-dependent response to auxin stimulus, response to cyclopentenone, response to ethylene stimulus, etc. We can therefore conclude that the proposed method confirmed that a group of highly semantically similar transcription factors in different families are regulated in the same pathway.

DISCUSSION

Drought stress causes several changes in plants, such as stomatal closure and decrease of turgor, which may act as signals to trigger the adaptation response [50, 51]. This stress includes several regulatory mechanisms that activate the expression of tolerance effector genes [50]. However, little is known about the mechanisms for sensing the changes of drought stress in sorghum. Here, we identified drought-responsive transcription factors of various families, such as MYB, AUX_ARF, bZIP, AP2/ethylene-responsive element binding proteins (EREBP) and WRKY, in sorghum.

It was previously reported that some transcription factors are expressed through abscisic acid pathways while others have independent expression to abscisic acid hormone [52]. Most ABA-inducible genes, such as rd22, contain a conserved *cis*-acting element like ABA-response elements (ABREs) in their promoter regions and are regulated by transcription factors such as bZIP, MYB and MYC [53]. In this investigation, ABREs (ACGTGGC) in AUX_ARF, AP2/EREBP and WRKY (clusters 2, 4 and 5) were noted in their promoter region. Dehydration responsive element-binding factors and DREB transcription factors are members of the AP2/ERF family, which consists of many important regulatory and stress-responding genes [54]. Transcription factors in the AP2/EREBP family (cluster 4) have been shown to regulate developmental processes and the response of plants to various types of biotic and environmental stress. Transcription factors in the AP2/EREBP family play important roles in the plant response and adaptation to abiotic stresses [55]. DREB proteins interact with DRE/CRT by their AP2 DNA-binding domain, thus mediating downstream gene expressed in the stress-responsive pathway [56].

In several studies, it has been reported that abscisic acid (ABA) is a major physiological signal that induces drought responses [57, 58]. ABA-dependent signalling systems activate bZIP, which mediates adaptation to drought, binding to ABREs and inducing their transcription [57]. Another ABA-dependent pathway requires protein biosynthesis of MYB transcription factors, which function to regulate the expression of target genes [59]. Among these genes, transcription factors play essential roles in stress responses by regulating their target genes through binding to the cognate *cis*-acting elements [60]. We identified several stress-responsive elements that were involved in drought inducibility: ABRE (ACGTGGC), anaerobic induction elements (AREs; TGGTTT) and the MYB-binding site (MBS; CAACTG); heat stress-responsive elements (HSEs; AAAAAATTTC); low temperature-responsive elements (LTRs; CCGAAA) and stress-responsive elements like TC-rich repeats (ATTTTCTTCA); etc. (Suppl. Table 2). A-box (CCGTCC) and CAAT-box, noted in all transcription factors families (clusters 1, 2, 3, 4 and 5) are common *cis*-acting elements in promoter and enhancer regions. CAAT-box has been reported to be responsible for the tissue-specific promoter activity of the pea

legumin gene (*LegA*) [61]. It has also been reported the CAAT and TATA boxes confer enhanced low temperature response in a barley shoot [62].

The G-box and I-box in the MYB, AUX_ARF, AP2/EREBP and WRKY families (clusters 1, 2, 4 and 5), and LAMP (CCTTATCCA) element in AP2/EREBP and WRKY (clusters 4 and 5) were involved in light-regulated expression [63]. The G-box element is essential for CMA5 responsiveness to light and chloroplast-derived signals [64]. Transcription factors in the MYB family are involved in plant development, secondary metabolism, hormone signal transduction, disease resistance and abiotic stress tolerance [65]. It has been reported that several R2R3-MYB genes are involved in regulating responses to environmental stresses such as drought, salt and cold [66]. In *Arabidopsis*, an R2R3-type MYB transcription factor (AtMYB96) regulates drought stress response by integrating ABA and auxin signals [67]. The present study isolated a novel R2R3-type MYB gene (OsMYB2) that greatly conferred tolerance of rice to salt, cold and dehydration stress [66]. *Arabidopsis thaliana* transgenic plants overexpressing OsMYB3R-2 showed increased tolerance to cold, drought and salt stress and the seed germination of transgenic plants was more tolerant to abscisic acid or NaCl than that of wild type [68].

GATA-motif (AAGGATAAGG) in the AUX_ARF, bZIP, AP2/EREBP and WRKY families (clusters 2, 3, 4 and 5) was noted for responsiveness to high-level, light-regulated and tissue-specific gene expression. GATA transcription factors are a group of DNA-binding proteins distinguished by a zinc finger motif, which have been implicated in light and nitrate-dependent transcription control [69]. GARE motif (AAACAGA) in AUX_ARF and WRKY families (clusters 2 and 5) *cis*-acting element and P-box (CCTTTTG) in basic leucine zipper (bZIP; cluster 3) is involved in gibberellin-responsiveness. The gibberellin (GA) pathway is regulated by endogenous signals and environmental cues such as light, temperature and salt stress [70]. Phytohormones abscisic acid (ABA) and GAs are well known to be involved in germination control [71].

AuxRR-core (GGTCCAT) in bZIP and WRKY family (clusters 3 and 6) promoter elements are bound and activated by plant-specific transcription factors, which are called auxin response factors (ARFs) [72]. At the molecular level, most plant growth and development processes, such as apical dominance, tropic responses, lateral root formation, vascular differentiation and embryo patterning, are controlled by the auxin response genes [73]. The transcription factors in Aux_ARF family are key regulators of auxin-modified gene expression. Auxin regulates diverse cellular and developmental responses in plants, including cell division, expansion, differentiation and patterning of embryo responses [74]. Auxins may regulate the gene expression of several families, including glycoside hydrolysis 3, SAUR and Aux/IAA [75]. The basic leucine zipper (bZIP) regulates diverse functions, such as plant development and stress response. In a previous study, a bZIP transcription factor gene (ZmbZIP72) expressed differentially in maize and was induced by abscisic acid,

high salinity and drought [76]. PtrABF is reported as a bZIP transcription factor that functions in positive modulation of drought stress tolerance [77].
The dehydration-responsive element C-repeat/DRE (TGGCCGAC) in AP2/EREBP and WRKY family (clusters 4 and 5) *cis*-acting element is involved in osmotic and cold stress inducible gene expression [52]. The noted TC-rich repeats (ATTTTCTTCA) in WRKY (cluster 5) and W1-box motif (TTGACC) in AUX_ARF, bZIP and WRKY families (clusters 2, 3 and 5) were tightly related to disease response motifs. Previously, analyzes have revealed that TC-rich repeats and W1-box motifs were related to disease response in *Populus* [78]. TC-rich repeats in the promoter of *AaERF1* were described in tobacco as *cis*-acting elements, which were involved in defense and stress responsiveness [79]. Several WRKY transcription factors have been shown to be involved in plant drought and salinity stress responses [80]. In *Arabidopsis*, the transcripts of two closely related WRKY transcription factors (AtWRKY25 and AtWRKY33) are increased by ABA, drought and salinity treatment [7]. The RY element (CATGCATG) in WRKY (cluster 5) plays a key role in seed-specific gene regulation in coordination with other *cis*-acting elements [81]. It was previously identified that three conserved motifs, two RY-like and one ACGT-like in *Brassicaceae* and *Fabaceae*, are involved in seed-specific *cis*-regulatory elements [82].

CONCLUSION

In this paper, we reported on a highly dedicated method to identify transcription factors of drought-induced proteins in sorghum. The method is based on the concept that highly GO semantically similar proteins may be involved in the same pathways. We used 2DE refined with MALDI-TOF to analyze the drought stress response proteins in sorghum. We identified 176 putative uncharacterized transcription factors for these proteins, which belong to the MYB, AUX_ARF, bZIP, AP2 and WRKY families. The members of these families regulate their target proteins based on endogenous signals and environmental cues, such as light, temperature and drought stresses. The regulatory network and *cis*-acting elements of identified transcription factors in distinct families are involved in auxin, abscisic acid, defense, stress and light responsiveness, and this may be highly important in the modulation of plant growth and development.

Acknowledgements. The authors gratefully acknowledge the use of the facilities of the DBT-supported Centre of Bioinformatics at Banasthali University, India.

REFERENCES

1. Tuteja, N. Abscisic acid and abiotic stress signaling. **Plant Signal Behav.** 2 (2007) 135–138.
2. Assefa, Y. and Staggenborg, S.A. Grain sorghum yield with hybrid advancement and change in agronomic practices from 1957 through 2008. **Agron. J.** 102 (2010) 703–706.
3. Nakashima, K., Ito, Y. and Yamaguchi-Shinozaki, K. Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. **Plant Physiol.** 49 (2009) 88–95.
4. Hu, H., You, J., Fang, Y., Zhu, X., Qi, Z. and Xiong, L. Characterization of transcription factor gene SNAC2 conferring cold and salt tolerance in rice. **Plant Mol. Biol.** 67 (2008) 169–181.
5. Ptashne, M. How eukaryotic transcriptional activators work. **Nature** 335 (1988) 683–689.
6. Zhang, J.Z. Over-expression analysis of plant transcription factors. **Curr. Opin. Plant Biol.** 6 (2003) 1–11.
7. Chen, L., Song, Y., Li, S., Zhang, L., Zou, C. and Yu, D. The role of WRKY transcription factors in plant abiotic stresses. **Biochim. Biophys. Acta** 1819 (2012) 120–128.
8. Lee, C. and Huang, C.H. LASAGNA: a novel algorithm for transcription factor binding site alignment. **BMC Bioinformatics** 14 (2013) 108–120.
9. Akhtar, M., Jaiswal, A., Taj, G., Jaiswal, J.P., Qureshi, M.I. and Singh, N.K. DREB1 /CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. **J. Genet.** 91 (2012) 385–395.
10. Umezawa, T., Fujita, M., Fujita, Y., Yamaguchi-Shinozaki, K. and Shinozaki, K. Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. **Curr. Opin. Biotechnol.** 17 (2006) 113–122.
11. He, K., Guo, A.Y., Gao, G., Zhu, Q.H., Liu, X.C., Zhang, H., Chen, X., Gu, X. and Luo, J. Computational identification of plant transcription factors and the construction of the PlantTFDB database. **Methods Mol. Biol.** 674 (2010) 351–368.
12. Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. Cluster analysis and display of genome-wide expression patterns. **P. Natl. Acad. Sci. USA** 96 (1999) 10943–10943.
13. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. and Church, G.M. Systematic determination of genetic network architecture. **Nat. Genet.** 22 (1999) 281–285.
14. Cramer, G.R., Urano, K., Delrot, S., Pezzotti, M. and Shinozaki, K. Effects of abiotic stress on plants: a systems biology perspective. **BMC Plant Biol.** 11 (2011) 163–177.
15. Todaka, D., Nakashima, K., Shinozaki, K. and Yamaguchi-Shinozaki, K. Toward understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. **Rice** 5 (2012) 6–15.

16. Dugas, D.V., Monaco, M.K., Olsen, A., Klein, R.R., Kumari, S., Ware, D. and Klein, P.E. Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. **BMC Genomics** 12 (2011) 514–534.
17. Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob-ur-Rahman, Ware, D., Westhoff, P., Mayer, K.F., Messing, J. and Rokhsar, D.S. The *Sorghum bicolor* genome and the diversification of grasses. **Nature** 457 (2009) 551–556.
18. Hirayama, T. and Shinozaki, K. Research on plant abiotic stress responses in the post genome era: past, present and future. **Plant J.** 61 (2010) 1041–1052.
19. Juncker, A.S., Jensen, L.J., Pierleoni, A., Bernsel, A., Tress, M.L., Bork, P., von Heijne, G., Valencia, A., Ouzounis, C.A., Casadio, R. and Brunak, S. Sequence-based feature prediction and annotation of proteins. **Genome Biol.** 10 (2009) 206–211.
20. Cokus, S., Mizutani, S. and Pellegrini, M. An improved method for identifying functionally linked proteins using phylogenetic profiles. **BMC Bioinformatics** 8 (2007) S7–S18.
21. Turanalp, M.E. and Can, T. Discovering functional interaction patterns in protein-protein interaction networks. **BMC Bioinformatics** 9 (2008) 276–293.
22. Ulitsky, I. and Shamir, R. Identifying functional modules using expression profiles and confidence scored protein interactions. **Bioinformatics** 25 (2009) 1158–1164.
23. Lee, D., Redfern, O. and Orengo, C. Predicting protein function from sequence and structure. **Nat. Rev. Mol. Cell Biol.** 8 (2007) 995–1005.
24. Gomez, A., Cedano, J., Amela, I., Planas, P., Pinol, J. and Querol, E. Gene ontology function prediction in mollicutes using protein–protein association networks. **BMC Syst. Biol.** 5 (2011) 49–59.
25. Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G.M., Blake, J.A., Bult, C., Dolan, M., Drabkin, H., Eppig, J.T., Hill, D.P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J.M., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S., Fisk, D.G., Hirschman, J.E., Hong, E.L., Nash, R.S., Sethuraman, A., Theesfeld, C.L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S.Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E.M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., White, R. and Gene Ontology Consortium. The gene

ontology (GO) database and informatics resource. **Nucleic Acids Res.** 32 (2004) D258–D261.
26. Resnik, P. similarity in a taxonomy: an information based measure and its application to problems of ambiguity in natural language. **J. Artif. Intell. Res.** 11 (1999) 95–130.
27. Wang, J.Z., Du, Z., Payattakool, R., Yu, P.S. and Chen, C.F. A new method to measure the semantic similarity of GO terms. **Bioinformatics** 23 (2007) 1274–1281.
28. Li, M., Wu, X., Pan, Y. and Wang, J. hF-measure: A new measurement for evaluating clusters in protein-protein interaction networks. **Proteomics** 13 (2013) 291–300.
29. Sharma, V., Sekhwal, M.K., Swami, A.K. and Sarin, R. Identification of drought responsive proteins using gene ontology hierarchy. **Bioinformation** 8 (2012) 595–599.
30. Sekhwal, M.K., Sharma, V. and Sarin, R. Identification of MFS proteins in sorghum using semantic similarity. **Theory Biosci.** 132 (2013), 105–113.
31. Sekhwal, M.K., Sharma, V. and Sarin, R. Annotation of glycoside hydrolases in *Sorghum bicolor* using proteins interaction approach. **J. Proteome Sci. Comput. Biol.** 2 (2013) 2.
32. Sekhwal, M.K., Swami, A.K., Sarin, R. and Sharma, V. Identification of salt treated proteins in sorghum using gene ontology linkage. **Physiol. Mol. Biol. Plants** 18 (2012) 209–16.
33. Damerval, C., Vienne, D.D., Zivy, M. and Thiellement, H. Technical improvements in two-imensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. **Electrophoresis** 7 (1986) 52–54.
34. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Anal. Biochem.** 72 (1976) 248–254.
35. Blackshear, P.J. Systems for polyacrylamide gel electrophoresis. **Methods Enzymol.** 104 (1984) 237–255.
36. Shameer, K., Ambika, S., Varghese, S.M., Karaba, N., Udayakumar, M. and Sowdhamini, R. STIFDB-*Arabidopsis* stress responsive transcription factor DataBase. **Int. J. Plant Genomics** 2009 (2009) 583429–583436.
37. Riano-Pachon, D.M., Ruzicic, S., Dreyer, I. and Mueller-Roeber, B. PlnTFDB: an integrative plant transcription factor database. **BMC Bioinformatics** 8 (2007) 42–51.
38. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res.** 25 (1997) 3389–3402.
39. UniProt Consortium. The Universal Protein Resource (UniProt). **Nucleic Acids Res.** 35 (2007) D193–D197.
40. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J.,

Yeats, C. and Eddy, S.R. The Pfam protein families database. **Nucleic Acids Res.** 32 (2004) D138–D141.
41. Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P. and Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. **Nucleic Acids Res.** 30 (2002) 325–327.
42. Ostlund, G., Schmitt, T., Forslund, K., Kostler, T., Messina, D.N., Roopra, S., Frings, O. and Sonnhammer, E.L. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. **Nucleic Acids Res.** 38 (2010) D196–203.
43. Yilmaz, A., Mejia-Guerra, M.K., Kurz, K., Liang, X., Welch, L. and Grotewold, E. AGRIS: the *Arabidopsis* Gene Regulatory Information Server, an update. **Nucleic Acids Res.** 39 (2011) D1118–1122.
44. Rhee, S.Y., Beavis, W., Berardini, T.Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M. and Huala, E. The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. **Nucleic Acids Res.** 31 (2003) 224–228.
45. Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R.D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A.F., Selengut, J.D., Sigrist, C.J., Thimma, M., Thomas, P.D., Valentin, F., Wilson, D., Wu, C.H. and Yeats, C. InterPro: the integrative protein signature database. **Nucleic Acids Res.** 37 (2009) D211–215.
46. Jiang, J.J. and Conrath, D.W. Semantic similarity based on corpus statistics and lexical taxonomy. **Proc. ROCLING X. Taiwan** (1997) 19–33.
47. Ding, X., Richter, T., Chen, M., Fujii, H., Seo, Y.S., Xie, M., Zheng, X., Kanrar, S., Stevenson, R.A., Dardick, C., Li, Y., Jiang, H., Zhang, Y., Yu, F., Bartley, L.E., Chern, M., Bart, R., Chen, X., Zhu, L., Farmerie, W.G., Gribskov, M., Zhu, J.K., Fromm, M.E., Ronald, P.C. and Song, W.Y. A rice kinase-protein interaction map. **Plant Physiol.** 149 (2009) 1478–1492.
48. Wippel, K. and Sauer, N. *Arabidopsis* SUC1 loads the phloem in suc2 mutants when expressed from the SUC2 promoter. **J. Exp. Bot.** 63 (2012) 669–679.
49. Quan, R., Lin, H., Mendoza, I., Zhang, Y., Cao, W., Yang, Y., Shang, M., Chen, S., Pardo, J.M. and Guo, Y. SCABP8/CBL10, a putative calcium sensor, interacts with the protein kinase SOS2 to protect Arabidopsis shoots from salt stress. **Plant Cell** 19 (2007) 1415–1431.
50. Anjum, S.A., Xie, X., Wang, L., Saleem, M. F., Man, C. and Lei W. Morphological, physiological and biochemical responses of plants to drought stress. **Afr. J. Agric. Res.** 6 (2011) 2026–2032.

51. Krasensky, J. and Jonak, C. Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. **J. Exp. Bot.** 63 (2012) 1593–608.
52. Nakashima K. and Yamaguchi-Shinozaki, K. Regulons involved in osmotic stress-responsive and cold stress-responsive gene expression in plants. **Physiol. Plant** 126 (2006) 62–71.
53. Sazegari, S. and Niazi, A. Isolation and molecular characterization of wheat (*Triticum aestivum*) dehydration responsive element binding factor (DREB) isoforms. **Aust. J. Crop Sci.** 6 (2012) 1037–1044.
54. Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K. and Yamaguchi-Shinozaki, K. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration and cold-inducible gene expression. **Biochem. Biophys. Res. Commun.** 290 (2002) 998–1009.
55. Sharoni, A.M., Nuruzzaman, M., Satoh, K., Moumeni, A., Attia, K., Venuprasad, R., Serraj, R., Kumar, A., Leung, H., Islam, A.K. and Kikuchi, S. Comparative transcriptome analysis of AP2/EREBP gene family under normal and hormone treatments, and under two drought stresses in NILs setup by Aday selection and IR64. **Mol. Genet. Genomics** 287 (2012) 1–19.
56. Zhao, L., Hu, Y., Chong, K. and Wang, T. ARAG1, an ABA-responsive DREB gene, plays a role in seed germination and drought tolerance of rice. **Ann. Bot.** 105 (2010) 401–409.
57. Mundy, J. and Chua, N.H. Abscisic acid and water-stress induce the expression of a novel rice gene. **EMBO J.** 7 (1988) 2279–2286.
58. Kizis, D., Lumbreras, V. and Pages, M. Role of AP2/EREBP transcription factors in gene regulation during abiotic stress. **FEBS Lett.** 498 (2001) 187–189.
59. Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T. and Shinozaki, K. Role of MYC and MYB homologs in drought and abscisic acid regulated gene expression**. Plant Cell** 9 (1997) 1859–1868.
60. Ma, H.S., Liang, D., Shuai, P., Xia, X.L. and Yin, W.L. The salt and drought inducible poplar GRAS protein SCL7 confers salt and drought tolerance in *Arabidopsis thaliana*. **J. Exp. Bot.** 61 (2010) 4011–4019.
61. Shirsat, A., Wilford, N., Croy, R. and Boulter, D. Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. **Mol. Gen. Genet.** 215 (1989) 326–331.
62. Dunn, M.A., White, A.J., Vural, S. and Hughes, M.A. Identification of promoter elements in a low-temperature-responsive gene (blt4.9) from barley (*Hordeum vulgare* L.). **Plant Mol. Biol.** 38 (1998) 551–564.
63. Blecken, J., Weisshaar, B. and Herzfeld, F. Two distinct *cis*-acting elements are involved in light-dependent activation of the pea elip promoter. **Mol. Gen. Genet.** 245 (1994) 371–379.
64. Lopez-Ochoa, L, Acevedo-Hernandez, G, Martinez-Hernandez, A., Arguello-Astorga, G. and Herrera-Estrella, L. Structural relationships

between diverse cis-acting elements are critical for the functional properties of a rbcS minimal light regulatory unit. **J. Exp. Bot.** 58 (2007) 4397–4406.
65. Katiyar, A., Smita, S., Lenka, S.K., Rajwanshi, R., Chinnusamy, V. and Bansal, K.C. Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. **BMC Genomics** 13 (2012) 544–562.
66. Yang, A., Dai, X. and Zhang, W.H. A R2R3-type MYB gene, OsMYB2, is involved in salt, cold, and dehydration tolerance in rice. **J. Exp. Bot.** 63 (2012) 2541–2556.
67. Seo, P.J., Xiang, F., Qiao, M., Park, J.Y., Lee, Y.N., Kim, S.G., Lee, Y.H., Park, W.J. and Park, C.M. The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in *Arabidopsis*. **Plant Physiol.** 151 (2009) 275–289.
68. Dai, X., Xu, Y., Ma, Q., Xu, W., Wang, T., Xue, Y. and Chong. K. Over-expression of an R1R2R3 MYB gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic *Arabidopsis*. **Plant Physiol.** 143 (2007) 1739–1751.
69. Reyes, J.C., Muro-Pastor, M.I. and Florencio, F.J. The GATA family of transcription factors in *Arabidopsis* and rice. **Plant Physiol.** 134 (2004) 1718–1732.
70. Sun, T.P. Gibberellin metabolism, perception and signaling pathways in *Arabidopsis*. **Arabidopsis Book** 6 (2008) e0103–0131.
71. Toh, S., Imamura, A., Watanabe, A., Nakabayashi, K., Okamoto, M., Jikumaru, Y., Hanada, A., Aso, Y., Ishiyama, K., Tamura, N., Iuchi, S., Kobayashi, M., Yamaguchi, S., Kamiya, Y., Nambara, E., and Kawakami N. High temperature-induced abscisic acid biosynthesis and its role in the inhibition of gibberellins action in *Arabidopsis* seeds. **Plant Physiol.** 146 (2008) 1368–1385.
72. Hagen, G. and Guilfoyle, T. Auxin-responsive gene expression: genes, promoters and regulatory factors. **Plant Mol. Biol.** 49 (2002) 373–385.
73. Xing, H., Pudake, R.N., Guo, G., Xing, G., Hu, Z., Zhang, Y., Sun, Q. and Ni, Z. Genome-wide identification and expression profiling of auxin response factor (ARF) gene family in maize. **BMC Genomics** 12 (2011) 178–190.
74. Reed, J.W. Roles and activities of Aux/IAA proteins in *Arabidopsis*. **Trends Plant Sci.** 6 (2001) 420–425.
75. Tian, Q., Uhlir, N.J. and Reed, J.W. *Arabidopsis* SHY2/IAA3 inhibits auxin-regulated gene expression. **Plant Cell** 14 (2002) 301–319.
76. Ying, S., Zhang, D.F., Fu, J., Shi, Y.S., Song, Y.C., Wang, T.Y. and Li, Y. Cloning and characterization of a maize bZIP transcription factor, ZmbZIP72, confers drought and salt tolerance in transgenic *Arabidopsis*. **Planta** 235 (2012) 253–266.
77. Huang, X.S., Liu, J.H. and Chen, X.J. Overexpression of PtrABF gene, a bZIP transcription factor isolated from *Poncirus trifoliata*, enhances

dehydration and drought tolerance in tobacco via scavenging ROS and modulating expression of stress-responsive genes. **BMC Plant Biol.** 10 (2010) 230–247.
78. Zhao, J.P., Jiang, X.L., Zhang, B.Y. and Su, X.H. Involvement of microRNA-mediated gene expression regulation in the pathological development of stem canker disease in *Populus trichocarpa*. **PLoS One** 7 (2012) e44968–44980.
79. Lu, X., Jiang, W., Zhang, L., Zhang, F., Zhang, F., Shen, Q., Wang, G. and Tang, K. AaERF1 positively regulates the resistance to *Botrytis cinerea* in *Artemisia annua*. **PLoS One** 8 (2013) e57657–57668.
80. Luo, X., Bai, X., Sun, X., Zhu, D., Liu, B., Ji, W., Cai, H., Cao, L., Wu, J., Hu, M., Liu, X., Tang, L. and Zhu Y. Expression of wild soybean WRKY20 in *Arabidopsis* enhances drought tolerance and regulates ABA signalling. **J. Exp. Bot.** 64 (2013) 2155–2169.
81. Fujiwara, T. and Beachy, R.N. Tissue-specific and temporal regulation of a beta-conglycinin gene: roles of the RY repeat and other cis-acting elements. **Plant Mol. Biol.** 24 (1994) 261–272.
82. Fauteux, F. and Stromvik, M.V. Seed storage protein gene promoters contain conserved DNA motifs in *Brassicaceae*, *Fabaceae* and *Poaceae*. **BMC Plant Biol.** 9 (2009) 126–136.