Identification of salt treated proteins in sorghum using gene ontology linkage

Manoj Kumar Sekhwal · Ajit Kumar Swami ·
Renu Sarin · Vinay Sharma

Published online: 8 June 2012

© Prof. H.S. Srivastava Foundation for Science and Society 2012

Abstract *Sorghum bicolor* (L.) is an important crop of arid and semi arid zones with most of its varieties tolerant to drought, heat and salt stress. Functional identification of many salt tolerant proteins has been reported in *Arabidopsis*, rice and other plants, however only little functional information has been predicted in sorghum till date. A 2-D gel electrophoresis based proteomic approach with MALDI-TOF mass spectrometer was utilized to analyze the salt stress response of sorghum. Major changes in protein complement were observed at 200 mM NaCl in hydroponic culture after 96 h of salt-stress. Highly expressed five proteins were excised for functional identification. We developed shortest path (SP) analysis based method on Gene Ontology (GO) hierarchy using sum of GO-term's semantic similarities. In this study, we observed that majority of expressed proteins belonged to the functional category of energy production and conversion, signal transduction mechanisms and ribosome maturation. These identified functions suggest a distinct mechanism of salt-stress adaptation in sorghum plant. The proposed method in this paper potentially has great importance to further

understanding of newly identified proteins that can help in plant development.

Keywords Protein identification · 2-D Gel electrophoresis · Gene ontology · Salt-stress · *Sorghum bicolor*

Abbreviations

GO	Gene Ontology
PPI	Protein-protein interaction
SP	Shortest path
MICA	Most informative common ancestor
BP	Biological process
CC	Cellular component
MF	Molecular function
Pfam	Protein family

Introduction

The genome sequences of a large number of micro to higher organisms have been completed e.g. *Escherichia coli* (Blattner et al. 1997), *Arabidopsis thaliana* (The *Arabidopsis* Genome Initiative 2000), *Oryza sativa* (Goff et al. 2002), *Sorghum bicolor* (Paterson et al. 2009). This has provided a wealth of genomic information stored at online repositories such as Gramene (Jaiswal 2011), PlantGDB (Dong et al. 2004), TAIR (Huala et al. 2001) NCBI-Plant Genome (Wheeler et al. 2005) and Gene Ontology database (Harris et al. 2004). The genome sequencing of *S. bicolor*, the fifth most important cereal crop in the world for food, feed and biofuels, has been recently completed (Paterson et al. 2009). The plant is especially important in arid and semi arid zones as most of its varieties are drought, heat and salt stress tolerant (Mall et al. 2011). The environmental stresses such as drought, heat, salt etc. greatly affect the metabolism and growth of plants. For

Electronic supplementary material The online version of this article (doi:10.1007/s12298-012-0121-y) contains supplementary material, which is available to authorized users.

M. K. Sekhwal · V. Sharma (✉)
Department of Bioscience & Biotechnology,
Banasthali University,
P.O. Banasthali Vidyapith,
304022 Rajasthan, India
e-mail: vinaysharma30@yahoo.co.uk

A. K. Swami · R. Sarin
Department of Botany and Biotechnology,
University of Rajasthan,
JLN Marg,
Jaipur 302055 Rajasthan, India

example, the primary effect of salinity is to induce osmotic stress by its effect on the ionic homeostasis in the plant cell (Serrano and Rodriguez-Navarro 2001). These adverse conditions increase the formation of reactive oxygen species (ROS) by accumulating during an imbalance of cellular homeostasis. Excessive ROS production causes oxidative stress, which damages plant's photosynthetic pigments (Takahashi and Badger 2011), membrane lipids, proteins and nucleic acids (Pitzschke et al. 2006). To keep the levels of active oxygen species under control, plants have antioxidant defense systems (Gill and Tuteja 2010).

Functional identification of many salt tolerant proteins has been reported by molecular and genomic analyses in *Arabidopsis*, rice and other plants (Chao et al. 2005). The groups of genes related to salt stress response in other species have been identified in earlier studies (Ouyang et al. 2007). The stresses defence proteins network have been recently explained in *Arabidopsis* and rice using multi-parallel analysis techniques (Amme et al. 2006). The multi-parallel techniques for gene expression have been proposed in *Arabidopsis* under cold, drought, and salt-stress (Brotman et al. 2011). However, in contrast only little information has been predicted in sorghum till date (Gramene release 34b). This suggested that functional identification of proteins in sorghum is far from saturation. The combination of high-throughput technique with bioinformatics tools and databases provide an opportunity to proteins functional identification (Isokpehi et al. 2011). The genes expression in response to high salinity stress in rice was first analyzed using cDNA microarray technology (Shinozaki and Yamaguchi-Shinozaki 2007). Several attempts have been made earlier to annotate the function of proteins. The wet laboratory experiments identify function of proteins correctly (Buza et al. 2007). However, these are time-consuming and costly. Currently, several computational approaches for functional identification of proteins such as sequence similarity (Lee et al. 2009), phylogenetic profiles (Cokus et al. 2007), protein-protein interaction (PPI) (Turanalp and Can 2008) and gene expression (Ulitsky and Shamir 2009) are available. Early established sequence similarity based methods have limitations when there is lack of orthologous proteins or low sequence similarity occurs among known proteins. Hence, it is highly desirable to develop standardized and reliable methods for functional identification of proteins (Friedberg 2006). Recently, high-throughput technologies produced huge amount of genomic information (Mochida and Shinozaki 2010). Several methods have been proposed to use this genomic information in functional prediction of proteins (Huynen et al. 2000). Earlier, protein-protein interaction method based on the assumption that interacting proteins usually share same function was proposed (Karaoz et al. 2004; Letovsky and Kasif 2003). The function of unknown protein can be identified on the basis of their interacted known protein (Hu et al. 2010).

Here, we report shortest path (SP) analysis method on Gene Ontology (GO) hierarchy for functional identification of putative uncharacterized salt responsive proteins in sorghum. GO-term's semantic similarity provides functional relationship between biological processes, molecular function, and cellular component. The semantic similarity between two proteins is usually calculated based on the terms similarity (Jain and Bader 2010). For the semantic similarity between two GO terms, earlier several methods have been reported (Yu et al. 2010), viz. Resnik's (Resnik 1999), simGIC (Pesquita et al. 2008) and JiangConrath's (Jiang and Conrath 1997). Gene Ontology (GO) is a structured and controlled vocabulary, which identify the functional annotation of proteins using standardized terms. GO is composed of three independent ontologies: biological process (BP), molecular function (MF), and cellular component (CC). In directed acyclic graph (DAG) the GO terms are structured as 'is-a' and 'part-of' relationships in GO database (Harris et al. 2004). Previously, it has been reported that co-localized and co-expressed proteins have similar functional property (Gu et al. 2011) (e.g. *mitochondrion* is more similar to *ribosome* than *nucleolus*, because both are in *cytoplasm*).

In the present study, the 2-DE and mass spectrometer based proteomic techniques were used to separate and identify highly expressed five proteins in sorghum. The high salinity (200 mM NaCl) at 96 h was utilized to investigate cellular responses. To identify functional annotation of expressed five putative uncharacterized proteins, we developed shortest path analysis method on GO hierarchy.

Methodology

Plant growth in hydroponics

The seeds of *Sorghum bicolor* (L.) genotype csv-17 were grown hydroponically (plastic container with 2 l of distilled water). The seeds were surface sterilized and germinated in culture chamber at 28 °C temperature, approximately 70 % humidity and from 325 to 760 Wm⁻² light intensity. After 7 days of germination, the salinity (200 mM) stress was imposed on germinated seeds and their leaves were harvested after time intervals of 48, 96, and 120 h. Three hydroponic assemblies for control and stress (200 mM) NaCl were used for gel analysis; leaves from 10 to 15 seedlings from each assembly were pooled for the extraction of proteins.

Protein extraction

Proteins were extracted from plant sample as per the method reported by Damerval et al. (1986). Freshly cut leaves from control and stressed plants were ground to a fine powder in liquid nitrogen using 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) and 0.07 % 2-mercaptoethanol (w/v) in acetone. The suspension was incubated at -20°C for 60 min with intermittent mixing (every 10 min) using a cyclomixer (Bangalore Genie, India). Precipitated material was collected by centrifugation ($25,000\times g$, 4°C , 15 min). The pellet was washed twice with acetone containing 0.07 % (w/v) 2-mercaptoethanol and the precipitate was air dried for 20 min. The pellet was stored at -80°C until further use. Proteins were dissolved from the dried precipitate into lysis-buffer (8 M urea, 2 % CHAPS) by repeated pipetting. Insoluble material was removed by centrifugation ($20,000\times g$, 20°C , 20 min) and the supernatant was clarified by passing through a $0.22\text{ }\mu\text{m}$ syringe filter (Millipore, India). Protease activity was kept low by maintaining the cell material at 4°C during centrifugation and addition of protease inhibitors (protease inhibitor cocktail, Sigma–Aldrich, USA) to the lysis buffer. Total protein concentration was determined using Quick Start Bradford Protein Assay kit (Bio-Rad, USA) as per manufacturer's instructions. The protein concentration was calculated using bovine serum albumin (BSA) as standard.

2-DE gel electrophoresis

Proteins samples were purified using 2D-cleanup kit (Bio-Rad) and the protein pellet was finally resuspended in sample rehydration buffer (8 M urea, 2 % CHAPS, 15 mM DTT and 0.5 % IPG buffer pH 4–7). The isoelectric focusing was performed using immobilized pH gradient (IPG) strips (Bio-Rad, USA). 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 was used to rehydrate IPG strips 7 cm, respectively. The IPG strips were rehydrated overnight and then the proteins were focused for 10,000 Vhr at 20°C under mineral oil. After focusing, the strips were incubated for 10 min, in 1 ml (for 7 cm strip) or 4 ml (for 17 cm strip) of equilibration buffer I (6 M urea, 30 % glycerol, 2 % SDS and 1 % DTT in 50 mM Tris/HCl buffer, pH 8.8) followed by equilibration buffer II (6 M urea, 30 % glycerol, 2 % SDS and 4 % iodo-acetamide in 50 mM Tris/HCl buffer, pH 8.8). After the equilibration steps the strips were transferred to 12 % SDS-PAGE for the second dimension by the method of Blackshear (1984). Protein spots were visualized by staining with Coomassie Brilliant Blue G-250. Gel images were captured by GS800 densitometer (Bio-Rad, USA). Relative abundance of the spots and the differential protein expression were determined by PD Quest software (Bio-Rad, USA). Two independent experiments were carried out for the differential study and replicate gels ($n=4$) were generated from each independent experiment. Results were compared with control (no additional salt) and salt-stress treatment

(200 mM NaCl) groups. Leaves from 8 to 10 seedlings were pooled together for each sample preparation. All images were subjected to automatic spot detection according to the manufacturer's recommendations. The spots were checked manually to eliminate any possible artifacts and spots that were consistently reproducible in all gel images, including both the biological and technical replicates, were chosen for subsequent analysis. The quantity of each spot was normalized by total valid spot intensity. The paired *t*-test was used for statistical analysis of the data. Protein spots which were expressed differentially between two experimental groups ($|\text{ratio}|\geq 1.5$, $p\leq 0.05$) were marked and excised.

Protein spots identification by mass spectrometry

The protein spots were excised by thin-walled PCR tubes (200 μl) cut at the bottom with the help of fresh surgical scalpel blade. The gel spots were washed with proteomic grade de-ionized water and proteins identified by using MALDI-TOF mass spectrometer (Ultraflex III, Bruker Daltonics, Germany). The gel piece containing the protein was destained 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 α -cyano-4-hydroxycinnamic acid in 80 % (v/v) acetonitrile and 0.1 % (w/v) TFA) and 1 μl of this mixture was deposited onto the MALDI target. Spectrum was obtained in the mass range of 500–4000 Da and calibrated using a calibration mixture containing angiotensin I, substance P, ACTH (1–17), ACTH (18–39) and somatostatin (28). The expressed proteins were analyzed using mascot sequence matching software (<http://www.matrixscience.com>) using MSDB in the taxonomy group of green plants. While performing mascot the search parameters were: 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

All five expressed proteins were identified from closely related homologous proteins in *Oryza sativa*. Homology search was carried out using blastp at Uniprot database (<http://www.uniprot.org/blast/>) with the parameters as Database: UniprotKB; Threshold: 0.1; Matrix: blosum-62. The molecular weights of expressed proteins were determined using the Compute pI/Mw search tool at expasy server (http://web.expasy.org/compute_pi/). The COGnitor program was utilized to proteins functional and phylogenetic annotation that compares proteins sequence against the Clusters of Orthologous Groups of proteins (COGs) database (<http://www.ncbi.nlm.nih.gov/COG/>) (Tatusov et al. 2000). Protein families were obtained using InterProScan search tool

(<http://www.ebi.ac.uk/Tools/pfa/iprscan/>) (Zdobnov and Apweiler 2001), against hidden Markov model (HMM) Pfam database (<http://pfam.sanger.ac.uk/>) (Bateman et al. 2004). The GO-terms were obtained using Gene Ontology database (<http://www.geneontology.org/>) (Harris et al. 2004) that are cross linked at Uniprot database (<http://www.uniprot.org/>) (Bairoch et al. 2005). AmiGO (version 1.8) (amigo.geneontology.org) was employed to identify GO hierarchical tree of assigned terms of our expressed proteins in sorghum. Currently, a complete data on Gene Ontology database for sorghum is not yet established. Though, the availability GO hierarchies of *A. thaliana* were used to obtain GO directed acyclic graphs (DAG). The obtained GO hierarchical graph was illustrated using cytoscape 2.8.1 version (Shannon et al. 2003).

Shortest path algorithm

We developed a shortest path (SP) analysis based method by sum of term's semantic similarity from root to its associated terms in possible paths. The ProteInOn tool (<http://xldb.di.fc.ul.pt/tools/proteinon/>) was utilized to calculate the semantic similarities with JiangConrath's (Jiang and Conrath 1997) measurement without ignore inferred from electronic annotation (IEA) (Schug et al. 2002). In this study, the concept behind the developed algorithm was that, if GO term is near to the root term; the sum of term's semantic similarity in the paths was decreased. On other hand, if GO term far from the root term, the sum of term's similarity in path was increased. The shortest path was confirmed by analysis of distance values in their corresponding paths. The smallest distance value of analyzed path was referred as shortest path. The shortest path algorithm used to find out distance value is given as below:

$$d = \frac{\arctan = \text{pathA}[t1 + t2]}{\pi/2}$$

Where, *pathA* is the path of GO category that connected to the root term, and *t1* and *t2* are the terms located in path. The value of *arctan* was obtained by sum the term's semantic similarity in *pathA*. The distance (*d*) for shortest path was calculated by *arctan*/π/2.

Results

Salt-tolerant variety of *S. bicolor* (L.) was cultivated under saline (200 mM NaCl) and normal conditions without additional salt. Cultivation of *S. bicolor* leaves under salt-stress condition leads to reduced growth of leaves and chlorosis in comparison with the control. A 2-D gel electrophoresis based proteomic approach with MALDI-TOF mass spectrometer was utilized to separate and identify highly expressed five spots (Fig. 1). To detect highly expressed proteins under the salt-stress condition, 2-DE spots from control and the corresponding stress condition were compared. The gel electrophoresis image was analyzed using PD Quest software (Bio-Rad Laboratories Ltd, UK), which showed more than 1.5-fold protein expression under salt-stress over the control plants (ratio stress/normal ≥ |1.5|, *p* < 0.05). The proteomic data on sorghum was not represented when mascot search tool (<http://www.matrixscience.com/>) was used. Hence, the homologous proteins were obtained from closely related *Oryza sativa*. Recently, while the whole genome sequencing of *S. bicolor* has been completed (Paterson et al. 2009), annotation of various genes has not yet been done. Thus, in the absence of comparable information, the identified homologues in sorghum were described as hypothetical proteins (Table 1). In this study, several bioinformatics approaches were carried out to put functional context of high expressed salt responsive proteins (Table 1). The COGNITOR result revealed that differentially expressed proteins belonged to functional categories of energy production and conversion [Inorganic pyrophosphatase (spot S3)] and signal transduction mechanisms [Protein kinase (spot S1); universal stress protein (spot S5)]. Among

Fig 1 Portion of representative 2-DE gel showing spots differentially expressed (>1.5-fold difference) in leaves from *Sorghum bicolor* plants grown in the presence of 200 mM NaCl (a) as compared to the control (b). The identified spots are marked with arrows and are statistically significant to a level of 95 % per group (Student's *t*-test) using biological and analytical replicates (*n*=4)

Table 1 Functional annotation of five expressed proteins under high salinity stress (200 mM NaCl) in *S. bicolor*

Spot #	Protein name	Gene Ontology terms	InterProScan	MICA terms
S1	Putative uncharacterized protein [Sb04g020990]	ATP binding (GO:0005524) protein serine/threonine kinase activity (GO:0004674)	Protein kinase	(R1) “binding” (R2) “catalytic activity”
S2	Putative uncharacterized protein [Sb06g017270]	ribosomal small subunit biogenesis (GO:0042274) cytoplasm (GO:0005737)	Ribosome maturation factor, RimP	No MICA found
S3	Putative uncharacterized protein [Sb09g001530]	phosphate-containing compound metabolic process (GO:0006796) inorganic diphosphatase activity (GO:0004427) magnesium ion binding (GO:0000287) cytoplasm (GO:0005737)	Inorganic pyrophosphatase	(R1) “binding” (R2) “catalytic activity”
S4	Putative uncharacterized protein [Sb06g019320]	defense response (GO:0006952) response to biotic stimulus (GO:0009607)	Bet_v_1	(R3) “response to stimulus”
S5	Putative uncharacterized protein [Sb08g017660]	response to stress (GO:0006950)	Universal stress protein	(R2) “catalytic activity”

the expressed proteins, YhbC family (spot S2) showed unknown function on COGNITOR search. Notably, this YhbC family belongs to ribosome maturation factor, RimP identified by InterProScan search tool (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>). AmiGO (amigo.genontology.org) pathway analyses were employed to identify GO hierarchical tree as sorghum. GO hierarchy information is also not available yet at the Gene Ontology database. Thus, the model plant *A. thaliana* was used to find GO hierarchy for assigned terms of expressed proteins in sorghum. The GO hierarchy paths were connected to root term biological process (BP), molecular function (MF) and cellular component (CC).

In this investigation, we developed a very effective shortest path (SP) analysis method on hierarchical structured GO terms. The shortest path was identified on the basis of the analysis of distance values of their corresponding paths. In the results, the smallest distance value has been referred as shortest path. While finding shortest paths for our assigned GO terms, the *most informative common ancestor* (MICA) terms were noted. The identified MICA terms were facilitated to discover functional annotation of salt stress responsive proteins. Previously, it has been confirmed that co-localized proteins have similar functional property that provides strong evidence of the relationship between proteins (Gu et al. 2011). Thus, more semantic similar proteins have the probability to become interacted proteins (Pesquita et al. 2008). The interacted protein with their corresponding known protein can help to elucidate the function of proteins.

Shortest path analysis in molecular function

The GO hierarchy has two layers from their root term molecular function (GO:0003674). We identified shortest path and distance values using sum of term's semantic similarity. The root term was associated by terms (n1, n2, n3 and n4) (Fig. 2).

The distance scores were obtained to analyse shortest paths for these terms. The distance score 77.55 was identified for (n1) GO-term by the following calculation: $\arctan = 98.4 + 96.0 + 100 + 99.9 + 92.7$; distance = $\arctan/\pi/2$ (Supplementary file 1). The same calculation was performed to find shortest path for other associated terms. While finding the shortest paths, MICA [(R1) “binding” (GO: 0005488)] was noted for terms [(n1) “magnesium ion binding” (GO: 0000287) (spot S3)] and [(n2) “ATP binding” (GO: 0005524) (spot S1)] (Fig. 2). The second MICA [R2 “catalytic activity”] was noted for terms [n3 “inorganic diphosphatase activity” (GO: 0004427) (spot S3)] and n4 [“protein serine/threonine kinase activity (GO: 0004674) (spot S1)]. Thus it may be concluded that the spots (S1 and S3) are strongly associated proteins that are involved in binding and catalytic activity process.

Shortest path analysis in biological process

The GO hierarchy of root term biological process (GO:0008150) was associated with terms (n5–n9). To identify shortest path, the distances' scores were identified by sum of term's semantic similarity from root (Supplementary file 1) using method described in methodology. Here, we noted one MICA [(R3) “response to stimulus” (GO:0050896)] for associated terms [(n6) “response to biotic stimulus” (GO: 0009607)], [(n7) “defense response” (GO:0006952)] in (spot S4), and term [(n8) “response to stress” (GO: 0006950) (spot S5)] (Fig. 2). Therefore, spots (S4 and S5) confirmed to associate proteins that are involved in response to stimulus.

Shortest path analysis in cellular component

The “cellular component” (GO: 0005575) is the root term that descends directly from its other terms “cell” (GO:0005623), “cell part” (GO: 0044464), “intracellular” (GO: 0005622),

Fig 2 GO hierarchy shortest paths for GO-terms of five expressed protein under salt stress. *R1, R2, R3, *most informative common ancestor (MICA)*; *n1–n10, GO-terms of expressed proteins; *BP* biological process; *MF* molecular function; *CC* cellular component

“intracellular part” (GO: 0044424). To identify shortest path for term [(n10) “cytoplasm” (GO: 0005737) (spots S2, S3)] from its root, the sum of term’s semantic similarity was calculated. The distance score 47.26 was identified for term (n10) by the following calculation, $\arctan = 99.7 + 97.8 + 99.3$; distance = $\arctan/\pi/2$ (Supplementary file 1). As a result, the term cytoplasm (spots S2, S3) shared same path with its root (Fig. 2). Consequently, it has been confirmed that (spots S2, S3) are co-localized and formed the major role as “intracellular part”.

Discussion

The functional identification of proteins is one of the most important tasks in post genomic era (Saghatelian and Cravatt 2005). We used a 2-DE based proteomic approach with MALDI-TOF to separate and identify proteins that expressed on exposure to 200 mM NaCl in hydroponic cultured leaves of *S. bicolor* (L.). On this basis, we reported functional identification of five highly expressed proteins (Fig. 1). The pH range of 4–7 was selected for isoelectric focussing as described in methodology. The same range of pH (4–7) has been utilized in earlier studies on for 2DE based separation of proteins expressed on salt exposure (Swami et al. 2011). A large number of stress related proteins have been identified in cereal model plants (Todaka et al. 2012). However, the proteins in sorghum are described as hypothetical till date. Thus, elucidation of function of

proteins in sorghum was needed to be worked out. The Gene Ontology database provides important evidence about the function of gene products (Harris et al. 2004). The complete genome sequencing assignments provide huge amount of information. This genomic information utilizes to identify functional linkages and biological context of genes and their products (Huynen et al. 2000). In this investigation, we developed a shortest path (SP) analysis based method to identify function of proteins. The sum of term’s semantic similarity was calculated to analysis paths. The smallest distance value was referred as shortest path. Further we also noted *most informative common ancestor (MICA)* terms in different paths. The *MICA* terms suggested a common function for distinct proteins, and hence are very helpful in annotation of the function of proteins. We identified a total of three *MICA* terms for highly expressed proteins (Table 1). The *MICA* term [(R1) “binding”] was identified in shortest path of its associated terms [(n1) “magnesium ion binding” (spot S3)] and [(n2) “ATP binding” (spot S1)]. The second *MICA* term [(R2) “catalytic activity”] was identified in shortest path of their associated terms [(n3) “inorganic diphosphatase activity” (spot S3) and [(n4) “protein serine/threonine kinase activity” (spot S1)].

The *MICA* “binding” and their associated terms “magnesium ion binding” and “ATP binding” are involved in numerous functions in plants including in electron carriers, maintenance of charge balance, and enzyme activation. The nutrient ions play important role in minimizing the adverse effects of abiotic stresses (e.g. drought, salt) in crop

plant (Cakmak 2005). Here, we identified *MICA* “catalytic activity” and its associated term “protein serine/threonine kinase activity” and examined them for their role in salt tolerance. However, the overexpressed serine-threonine protein kinase gene *SAPK4* in rice accumulates less Na^+ and Cl^- . Rather, the regulated gene *SAPK4* that plays the important functional role in ion homeostasis and oxidative stress response was identified (Diedhiou et al. 2008). This suggested that identified *MICA* and its associated terms in spots (S1 and S3) were involved in energy production and conversion, and signal transduction mechanisms. Sorghum crop is highly tolerant to drought, salt, and heat stress, and their high expressed proteins were involved for adaptation to salt stress. We identified another *MICA* term [(R3) “response to stimulus”] in the shortest path of associated terms [(n6) “response to biotic stimulus”], [(n7) “defense response”] in (spot S4), and term [(n8) “response to stress”] (spot (S5)) (Fig. 2). On this basis, spots (S4 and S5) were confirmed to be involved in response to salt stimulus. Expression of several genes response to salt, osmotic, and cold stress stimulus has been reported (Kreps et al. 2002). Calcium dependent protein kinases (*CDPKs*) respond to environmental stresses stimuli. *CDPKs* are involved in signal transduction during stress conditions. Several functional studies have been revealed that signaling pathways leading to biotic and abiotic stress resistance are mediated by *CDPKs* (Ludwig et al. 2004). The *CIPK14*, protein kinase has been reported in responses to salt stress in *Arabidopsis* (Qin et al. 2008).

We also noted that (spot S2, S3) belong to same intracellular part, [(n10) “cytoplasm”] (Fig. 1). In salt and drought stress tolerant conditions, active transport of solutes is established by proton pumps that generate proton electrochemical gradients across cell membranes (Roberto et al. 2001). However, in cell “intracellular part” is a parent node of several child components such as “cytoplasm”, “proton-transporting ATP synthase F_1 ”. The mechanism of salt tolerance in plants is based on salt sequestration into the vacuole, and accumulation of potassium ions and organic solutes in the cytoplasm to maintain the osmotic and ionic balance. The cytoplasm plays an important role in accumulation of ions (e.g. Na^+ , Cl^-) in salt-tolerant plants (Munns and Tester 2008). In them, active sodium ions are transported in cells by Na^+/H^+ transporters, which are driven by an ATPase. The *At3g46430* protein which is induced by several salts, drought and cold stresses (Zhang et al. 2008) has been characterized from mitochondrial F_1F_0 -ATPase protein in *A. thaliana*.

Notably, [(n9) Ribosome maturation factor, RimP (spot S2)] observed in our results indicated an independent GO hierarchy (Fig. 1). Previously *RsgA* protein involved in maturation of the small ribosome small subunit in improved resistance to high salt stress has been reported (Hase et al. 2009).

Conclusions

In this study, we described the isolation and characterization of highly expressed five proteins in *Sorghum bicolor* using 2DE-MS proteomic approaches. We developed shortest path analysis method using sum of GO term’s semantic similarity for the functional identification of these proteins. The results revealed that in shortest paths, *MICA* terms were identified in GO hierarchy of expressed five proteins. The identified proteins belonged to functional categories of energy production and conversion, and signal transduction mechanisms. These categories of proteins suggest a distinct mechanism of salt-stress adaptation in sorghum plant. Here, the proposed method for functional analysis of proteins would provide further information of newly identified proteins that can help in plant development.

References

- Amme S, Matros A, Schlesier B, Mock HP (2006) Proteome analysis of cold stress response in *Arabidopsis thaliana* using DIGE-technology. J Exp Bot 57:1537–1546
- Bairoch A, Apweiler R, Wu CH, Winona C, Barker WC, Boeckmann B, Ferro S, Gasteiger E et al (2005) The universal protein resource (UniProt). Nucleic Acids Res 33:D154–D159
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Jones SG, Khanna A, Marshall M et al (2004) The Pfam protein families database. Nucleic Acids Res 32:D138–D141
- Blackshear PJ (1984) Systems for polyacrylamide gel electrophoresis. Methods Enzymol 104:237–255
- Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD et al (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453–62
- Brotman Y, Riewe D, Lisec J, Meyer RC, Willmitzer L, Altmann T (2011) Identification of enzymatic and regulatory genes of plant metabolism through QTL analysis in *Arabidopsis*. J Plant Physiol 168:1387–1394
- Buza TJ, McCarthy FM, Burgess SC (2007) Experimental confirmation and functional-annotation of predicted proteins in the chicken genome. BMC Genomics 8:425–434
- Cakmak I (2005) The role of potassium in alleviating detrimental effects of abiotic stresses in plants. J Plant Nutr Soil Sci 168:521–530
- Chao DY, Luo YH, Shi M, Luo D, Lin HX (2005) Salt-responsive genes in rice revealed by cDNA microarray analysis. Cell Res 15:796–810
- Cokus S, Mizutani S, Pellegrini M (2007) An improved method for identifying functionally linked proteins using phylogenetic profiles. BMC Bioinformatics 8:S7–S18
- Damerval C, Vienne DD, Zivy M, Thiellement H (1986) Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. Electrophoresis 7:52–54
- Diedhiou CJ, Popova OV, Dietz KJ, Goldack D (2008) The SNF1-type serine-threonine protein kinase *SAPK4* regulates stress-responsive gene expression in rice. BMC Plant Biol 8:49–61
- Dong Q, Schlueter SD, Brendel V (2004) PlantGDB, plant genome database and analysis tools. Nucleic Acids Res 32:D354–D359
- Friedberg I (2006) Automated protein function prediction-the genomic challenge. Brief Bioinform 7:225–242

- Gill SS, Tuteja N (2010) Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol Biochem* 48:909–930
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296:92–100
- Gramene release 34b (http://www.gramene.org/Sorghum_bicolor/Info/Index)
- Gu H, Zhu P, Jiao Y, Meng Y, Chen M (2011) PRIN: a predicted rice interactome network. *BMC Bioinform* 12:161–173
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S et al (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32:D258–261
- Hase Y, Yokoyama S, Muto A, Himeno H (2009) Removal of a ribosome small subunit-dependent GTPase confers salt resistance on *Escherichia coli* cells. *RNA* 15:1766–1774
- Hu P, Jiang H, Emili A (2010) Predicting protein functions by relaxation labelling protein interaction network. *BMC Bioinform* 11: S64–S71
- Huala E, Dickerman AW, Hernandez MG, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D et al (2001) The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res* 29:102–105
- Huynen M, Snel B, Lathe W 3rd, Bork P (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res* 10(8):1204–1210
- Isokpehi RD, Simmons SS, Cohly HHP, Ekunwe SIN, Begonia GB, Ayensu WK (2011) Identification of drought-responsive universal stress proteins in viridiplantae. *Bioinform Biol Insights* 5:41–58
- Jain S, Bader GD (2010) An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology. *BMC Bioinform* 11:562–575
- Jaiswal P (2011) Gramene database: a hub for comparative plant genomics. *Methods Mol Biol* 678:247–275
- Jiang JJ, Conrath DW (1997) Semantic similarity based on corpus statistics and lexical taxonomy. *Proc ROCLING X Taiwan* 19:33.
- Karaoz U, Murali TM, Letovsky S, Zheng Y, Ding C, Cantor CR, Kasif S (2004) Whole-genome annotation by using evidence integration in functional linkage networks. *Proc Natl Acad Sci USA* 101:2888–2893
- Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF (2002) Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiol* 130:2129–2141
- Lee BJ, Shin MS, Oh YJ, Oh HS, Ryu KH (2009) Identification of protein functions using a machine-learning approach based on sequence-derived properties. *Proteome Sci* 7:27–45
- Letovsky S, Kasif S (2003) Predicting protein function from protein/protein interaction data: a probabilistic approach. *Bioinformatics* 19:i197–i204
- Ludwig AA, Romeis T, Jones JDG (2004) CDPK-mediated signalling pathways: specificity and cross-talk. *J Exp Bot* 55:181–188
- Mall TK, Dweikat I, Sato SJ, Neresian N, Xu K, Ge Z, Wang D, Elthon T et al (2011) Expression of the rice CDPK-7 in sorghum: molecular and phenotypic analyses. *Plant Mol Biol* 75:467–479
- Mochida K, Shinozaki K (2010) Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol* 51:497–523
- Munns R, Tester M (2008) Mechanisms of salinity tolerance. *Annu Rev Plant Biol* 59:651–81
- Ouyang B, Yang T, Li H, Zhang L, Zhang Y, Zhang J, Fei Z, Ye Z (2007) Identification of early salt stress response genes in tomato root by suppression subtractive hybridization and microarray analysis. *J Exp Bot* 58:507–520
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U et al (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457:551–556
- Pesquita C, Faria D, Bastos H, Ferreira AE, Falcao AO, Couto FM (2008) Metrics for GO based protein semantic similarity: a systematic evaluation. *BMC Bioinform* 9:S5–S20
- Pitzschke A, Forzani C, Hirt H (2006) Reactive oxygen species signaling in plants. *Antioxid Redox Signal* 8:1757–64
- Qin Y, Li X, Guo M, Deng K, Lin J, Tang D, Guo X, Liu X (2008) Regulation of salt and ABA responses by CIPK14, a calcium sensor interacting protein kinase in *Arabidopsis*. *Sci China C Life Sci* 51:391–401
- Resnik P (1999) Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language. *J Artif Intell Res* 11:95–130
- Roberto A, Gaxiola RA, Li J, Undurraga S, Dang LM, Allen GJ, Alper SL, Fink GR (2001) Drought- and salt-tolerant plants result from overexpression of the AVP1 H⁺-pump. *Proc Natl Acad Sci USA* 98:11444–11449
- Saghatelian A, Cravatt BF (2005) Assignment of protein function in the postgenomic era. *Nature Chem Biol* 1:130–142
- Schug J, Diskin S, Mazzarelli J, Brunk BP, Stoeckert-Jr CJ (2002) Predicting gene ontology functions from ProDom and CDD protein domains. *Genome Res* 12:648–655
- Serrano R, Rodriguez-Navarro A (2001) Ion homeostasis during salt stress in plants. *Curr Opin Cell Biol* 13:399–404
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504
- Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. *J Exp Bot* 58:221–227
- Swami AK, Alam SI, Sengupta N, Sarin R (2011) Differential proteomic analysis of salt stress response in *Sorghum bicolor* leaves. *Environ Exp Bot* 71:321–328
- Takahashi S, Badger MR (2011) Photoprotection in plants: a new light on photosystem II damage. *Trends Plant Sci* 16:53–60
- Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucl Acids Res* 28:33–36
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Todaka D, Nakashima K, Shinozaki K, Shinozaki KY (2012) Toward understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. *Rice* 5:6–25
- Turanalp ME, Can T (2008) Discovering functional interaction patterns in protein-protein interaction networks. *BMC Bioinformatics* 9:276–293
- Ulitsky I, Shamir R (2009) Identifying functional modules using expression profiles and confidence-scored protein interactions. *Bioinformatics* 25:1158–1164
- Wheeler DL, Smith-White B, Chetvernin V, Resenchuk S, Dombrowski SM, Pechous SW, Tatusova T, Ostell J (2005) Plant genome resources at the national center for biotechnology information. *Plant Physiol* 138:1280–1288
- Yu G, Li F, Qin Y, Bo X, Yibo W, Wang S (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 26:976–978
- Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847–848
- Zhang X, Liu S, Takano T (2008) Overexpression of a mitochondrial ATP synthase small subunit gene (*AtMtATP6*) confers tolerance to several abiotic stresses in *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. *Biotechnol Lett* 30:1289–1294