THE UNIVERSITY GRANTS COMMISSION (UGC)
NEW DELHI

Final Progress Report of

The Major Research Project
on

Isolation and Expression of Drought Tolerant Genes in Groundnut (*Arachis hypogaea* L.)

Submitted

By

Prof. D. Manohar Rao

DEPAPARTMENT OF GENETICS
OSMANIA UNIVERSITY
HYDERABAD-500 007
INDIA

Annexure – IX

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

PROFORMA FOR SUBMISSION OF INFORMATION AT THE TIME OF SENDING THE FINAL REPORT OF THE WORK DONE ON THE PROJECT

1.	TITLE OF THE PROJECT	**: Isolation and Expression of Drought Tolerant Genes in Groundnut (*Arachis hypogaea* L.)**
2.	NAME AND ADDRESS OF THE PRINCIPAL INVESTIGATOR	: **Prof. D. Manohar Rao (Retd.)** **Dept. of Genetics** **University College of Science** **Osmania University, Hyderabad** **Telangana State – 500 007**
3.	NAME AND ADDRESS OF THE INSTITUTION	: **Dept. of Genetics** University College of Science **Osmania University, Hyderabad** **Telangana State – 500 007**
4.	UGC APPROVAL LETTER NO. AND DATE	: **F. No. 41-828/2012 (SR) Dt.18-07-2012**
5.	DATE OF IMPLEMENTATION	: **01-09-2012**
6.	TENURE OF THE PROJECT	: **01-09-2012 to 31-12-2015**
7.	TOTAL GRANT ALLOCATED	: **Rs. 8,55,200/-**
8.	TOTAL GRANT RECEIVED	: **Rs. 7,70,000/-**
9.	FINAL EXPENDITURE	: **Rs.10,52,126/-**
10.	TITLE OF THE PROJECT	: **Isolation and Expression of Drought Tolerant Genes in Groundnut (*Arachis hypogaea* L.)**

11. OBJECTIVES OF THE PROJECT :

- Detection of differentially expressed protein spots in drought - stressed groundnut plants through 2D-PAGE and their sequencing by MALDI-TOF/TOF
- Identification of differentially expressed proteins under drought stress by 2D- DIGE
- Profiling of drought induced proteins in ICGV91114 groundnut by LC-MS/MS
- Validation of selected proteins for drought stress tolerance by qRT-PCR

12. WHETHER OBJECTIVES WERE ACHIEVED (GIVE DETAILS): Yes

Work done (Please give details)

- **Objective – 1:**

Detection of differentially expressed protein spots in drought - stressed groundnut plants through 2D-PAGE and their sequencing by MALDI-TOF/TOF:

A. Isolation of Drought Tolerant Proteins:

1. Plant material:

Four varieties of peanut (*Arachis hypogaea* L.) *viz.* ICGV 91114, ICGS 76, J 11 and JL 24 with varying degrees of drought tolerance were obtained from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, India. ICGV 91114 is a high yielding (2.5-3.0 ton/ha) bunch variety with 48% oil and 27% protein content, matures in 90-95 days and has an ability to withstand prolonged drought spells, tolerant to mid-season and end-of-season drought. ICGS 76 is another high yielding (1.3-1.8 ton/ha), Virginia bunch variety with an oil content of 43% and matures in 120 days. It is tolerant to bud necrosis and has good recovery from mid-season drought. J 11 is a Spanish bunch type, adaptable under a wide range of agro-climatic conditions and resistant to collar-rot and aflatoxin. JL 24 also a high yielding (1.5-2.0 ton/ha) Spanish bunch variety, matures in 90-95 days, and one of the most popular national varieties sown in the areas where end-season drought is common due to its drought susceptibility.

2. ***Imposition of drought stress***

Four seeds per pot, from each variety, were sown in pots filled with a mixture of soil and nutrients, at the plant Genetics experimental farm, Department of Genetics, Osmania University, Hyderabad. They were watered regularly thrice a week for 20 days until the water stress was imposed. Drought was imposed by withholding water to the 20-day-old seedlings for 10, 15 and 20 days to ICGV 91114 while 10 and 20 days to other three cultivars, maintaining their respective controls. Fully expanded fresh leaf samples were collected at random, from each of the stressed plants of all the four varieties along with their respective controls on 31st, 36th and 41st days respectively. Leaf samples were collected from all the 4 plants at random from each replicate, pooled variety-wise, quickly frozen in liquid nitrogen and stored at -80°C prior to protein extraction. The experiment was laid out in a completely randomized block design with three replications each and the seedlings / plants were maintained in the glass house during the experiment.

3. Relative water content (RWC)

The RWC was estimated in 10, 15 and 20 days drought stressed ICGV 9114 and 10 and 20 days water stressed plants in other 3 varieties along with their respective controls. After recording fresh weight of the collected leaves from 31st, 36th and 41st day-old seedlings, stored overnight in de-ionised water, in refrigerator, at 4^{o}C and the next day the turgid weight of the blotted leaves was recorded. Later, the dry weight of leaves, incubated in hot-air-oven at 60^{o}C for 24 h, was recorded. RWC was estimated as per the following formula.

$$\text{Relative Water Content (RWC)} = \frac{\text{Fresh weight} - \text{Dry weight}}{\text{Turgid weight} - \text{Dry weight}}$$

4. Isolation and purification of proteins:

Two grams of fresh leaf tissue (pooled from 3 replicates) of each variety was ground to fine powder in liquid nitrogen in a chilled mortar using pestle, homogenized twice with 20 ml of cold acetone containing 10% trichloro-acetic acid (TCA) and 0.07% β-mercapto-ethanol (β-ME) for 30 seconds and incubated overnight at -20^{o}C. The homogenate was centrifuged at 20,000 g for 30 minutes and the supernatant was discarded and the pellet was homogenized again with fresh acetone containing 10% TCA and 0.07% β-ME. This pre-protein extraction

procedure was repeated 3-4 times until all the pigments from the leaf tissue were removed. Later, this pellet was homogenized with cold acetone containing only 0.07% β-ME without TCA and centrifuged at 10,000 g for 20 minutes. This step was repeated twice and the pellet was air dried overnight.

The total leaf proteins were extracted by thoroughly vortexing and homogenising the pellet in 5 ml of protein extraction buffer (8.8 M urea, 2.0 M thiourea, 4% CHAPS (3-[3-cholamidopropyl dimethylammo-o]-1-propanesulfonate), 20 mM Di-thiotheritol (DTT), 10% protease inhibitors cocktail and 0.01% biolytes). The homogenate was incubated at room temperature for one hour, centrifuged at 20,000 g for 30 minutes and filtered through a 0.45 µM filter. This protein extraction was repeated thrice and all the supernatants were pooled and the protein was estimated using Bradford assay.

B. Resolution of total leaf proteins into different protein spots by 2D-PAGE (2-DE)

The above isolated and purified total leaf proteins were resolved by 2-DE in two steps; in the first dimension by iso-electric focusing (IEF) and in the second dimension by SDS-PAGE. Briefly, 250µg of leaf protein was loaded onto 7 cm IPG strips (Bio-Rad) with pH 4–7 and was kept overnight for rehydration. IEF of proteins was performed with 50 mA/strip with the following program: Step-I: 250 V for 20 min. at a linear slope, Step- II: 4,000 V for 2h, Step-III: 4,000 V, at a rapid slope for 10,000 Vh, and Step- IV: 4000 V at a linear slope for 30 min. Later, these strips were equilibrated with an equilibration buffer-I [6 M urea, 2% (w/v) SDS, 20% (v/v) glycerol, 0.375 M Tris-HCl (pH 8.8) and 2% (w/v) DTT] for 15 min and then for another 15 min with equilibration buffer-II [6 M urea, 2% (w/v) SDS, 20% (v/v) glycerol, 0.375 M Tris-HCl (pH 8.8) and 2.5% (w/v) iodoacetamide]. At the end of equilibration, the strips were loaded onto the SDS-PAGE for second dimension.

In the second dimension, SDS–PAGE of all the protein samples, along with a pre-stained broad range SDS-PAGE marker (Bio-Rad), was performed using 12 % (w/v) polyacrylamide gels with 5% (w/v) stacking gels. Electrophoresis was carried out (Bio-Rad PROTEAN unit) at a constant current of 20 mA/gel in 1-X Tris-Glycine electrophoretic running buffer. After electrophoresis, the gels were fixed in a mixture of 40% methanol and 10% glacial acetic acid, stained with 0.25% colloidal Coomassie Brilliant Blue R-250 (CBB) and de-stained in a mixture of 40% methanol and 10% glacial acetic acid solution.

C. Quantitative and qualitative analysis of protein spots using PD Quest Basic software:

The CBB-stained polyacrylamide gels containing protein spots were scanned as digitized images using GS-710 Calibrated Imaging Densitometer (Bio-Rad) with a scan density of 42.3 x 42.3 and saved as tagged image file formats. These protein spots were analyzed for quantification including spot detection, measurement, back ground subtraction and matching using PD Quest Basic software version 8.0.1 (Bio-Rad). The protein spots on the polyacrylamide gels obtained from control plants of all the 4 cultivars were differentiated into faint, small and large spots. These control gels were used as reference gels for comparison of spots on the polyacrylamide gels obtained from their respective stressed plants. The quantitative variations in the intensity of protein spots due to CBB staining of the polyacrylamide gels were corrected using normalization parameters and thus the Guassian images were created. The Guassian images exhibiting matching between the protein spots of controls and their respective stressed plants, of each of the 4 varieties, were retained and the rest unmatched spots were removed using spot editing tool. Protein spots across the gels of the control and stressed plants within a variety were subjected to auto- spot matching by the 'classic match tool' in all the 4 cultivars. The unmatched spots on the member gels of the stressed plants were added to the reference gel (control gel image). Qualitative analysis to know the presence of the spots and quantification of spots to know their intensity was performed by employing the 'analysis set manager' tool. For quantification of protein spots in a set of gels within a variety, outside limits were set as 0.5 and 1.5. The spots showing < 0.5 expression levels were considered as down-regulated, > 1.5 as more abundant and between these limits as unchanged or normal spots. On the basis of matching, differential spots were selected and analyzed as described below. Three replicate gels were analyzed for each sample to ensure reproducibility.

D. In-gel tryptic digestion of protein spots of interest for PMF analysis by MALDI - TOF:

1. ***In-gel tryptic digestion:***

The CBB stained 2 DE gels were washed with ultra-pure water, protein spots of interest were excised from gels, de-stained in 0.2 M NH_4HCO_3 in 40% acetonitrile (ACN) for 30 min. at room temp., kept for drying and subjected to in-gel tryptic (0.1 mg) digestion in an ice bath for 30 min. Later, 20 ml of 50 mM NH_4HCO_3 was added and digestion was allowed to continue at

37^{o}C overnight. After digestion, samples were extracted in 0.1 % trifluoroacetic acid (TFA) and 50 % ACN and stored at -20^{o}C for MALDI-TOF analysis.

2. *PMF analysis by MALDI-TOF:*

The tryptic digested peptides were loaded into a constricted GELoader tip (Eppendorf, Hamburg, Germany) packed with POROS R2 chromatographic resin (Perseptive Biosystems, Framingham, MA) with 5 mm porosity for desalting and concentration. The columns were equilibrated with 20 ml of 5% formic acid (FA) into which the digested samples were added. Later, the bound peptides were washed with 20 ml of 5% FA and eluted directly onto the MALDI target with 0.5 ml of CHCA solution (5 mg/ml in ACN, 0.1% TFA, 70:30 v/v). These samples were analyzed in Applied Biosystems 5800 MALDI-TOF Proteomics Analyzer. The instrument was equipped with a nitrogen laser and operated in a positive-ion delayed extraction reflector mode. External calibration was performed by using a standard peptide/protein mixture. Usually, 250 individual spectra of each spot were averaged to produce a mass spectrum. The atmospheric air was used as collision gas to fragment the peptides and obtain the spectra. Peptide fragmentation was performed using collision-induced dissociation (CID), and 50 laser shots from five sample positions were summed up for each parent ion.

E. Identification of peptides by comparing their mass with theoretical mass in the protein database (NCBI / SWISS-PROT)

Identification of proteins was performed by searching against National Centre for Biotechnology Information non-redundant (NCBInr) and SWISS-PROT databases using Mascot software (http://www.matrixscience.com), with *Viridiplantae* (green plants) as taxonomic category. The MASCOT search compares the experimental data with that of all the sequences in a database and returns list of hits with decreasing scores, a measure of reliability of identification. The following parameters were used for database search with MALDI-TOF PMF data: (NCBI nr 20070216, 4626804 sequences; 1596079197 residues; Taxonomy Viridiplantae 186963 sequences) , trypsin as digesting enzyme, 2 missed cleavages allowed, carbamido-methylation, methionine-oxidation and deamidation (NQ) (variable modifications), mono isotopic mass, peptide mass tolerance at ±100 ppm, unrestricted protein mass, 1+ peptide charge state. The search scores are represented as Probability based Mowse score -10*Log (P), where P is the probability that the observed match is a random event. Protein scores > 65 are considered significant ($p < 0.05$) in NCBI and scores >56 are considered significant in SWISS-

PROT database. For a positive identification in MALDI TOF- MS, the peptide score should exceed or equal to minimum significant score. For database search with MS/MS spectra, the following parameters were used: (CDS combined KBMS5.0.20050302, 1967674 sequences, 672312456 residues; Taxonomy Viridiplantae, 177633 sequences), trypsin enzyme, carbamido-methylation, methionine-oxidation, deamidation (NQ) (variable modification), monoisotopic mass value, unrestricted protein mass, peptide mass tolerance at ±1.0 Da, unrestricted protein mass, +1, +2, +3 peptide charge state with one missed cleavage allowed. The detected proteins were considered if they were identified by more than two peptides per spot.

F. ***Statistical analysis:***

The data on RWC was recorded and analyzed through two-way analysis of variance and the values were expressed as mean and standard error. The significance of the treatment effects was tested at 5% probability level (P= 0.05) using Tukeys test, which is one of the Post Hoc multiple comparisons of Two-way ANOVA of Windostat version 8.5.

G. Predicting sub-cellular localization and functional annotation through Bio - Informatics:

Three independent sets of gene ontology: a) Biological Process (BP) in which the gene product participates, b) Molecular Function (MF) describing the gene product's activities at the molecular level such as catalytic or binding and (c) Cellular Component (CC) indicating its sub-cellular localization, were defined for all the 30 proteins identified through PMF by searching the identified protein ID in TargetP program (www. cbs.dtu.dk / services / TargetP).

RESULTS

Objective – 1:

Detection of differentially expressed protein spots in drought - stressed groundnut plants through 2D-PAGE and their sequencing by MALDI-TOF/TOF:

A. Isolation of Drought Tolerant Proteins :

1. Isolation of total leaf proteins from drought stressed Groundnut plants:

The drought stress induced 20 day-old groundnut seedlings; ICGV 91114, ICGS 76, J 11 and JL 24, at vegetative phase under pot culture, the ICGV 91114, ICGS 76 and J 11 verities could withstand drought stress up to 15 days and later started showing the symptoms of wilting, while JL 24 wilted after 10 days of drought stress. Hence, the leaf proteome studies were conducted under mild (10 days) and severe (20 days) drought stress conditions in all the 4 varieties, while in ICGV 91114 cultivar studies were also made at moderate (15 days) drought stress condition, to investigate the effect of drought stress on differential expression of proteins in stressed plants.

2. Relative water content (RWC) of leaf:

In all the 4 varieties, there was a significant reduction in RWC with increase in the duration of stress from 10 to 20 days ($p<0.01$). In controls of all the 4 varieties, there was a steady decrease in RWC with increase in the age of the plant. The highest RWC was observed in ICGV 91114, followed by ICGS 76, J 11and JL 24 in 10, 15 and 20 day stressed plants (Fig. 1).

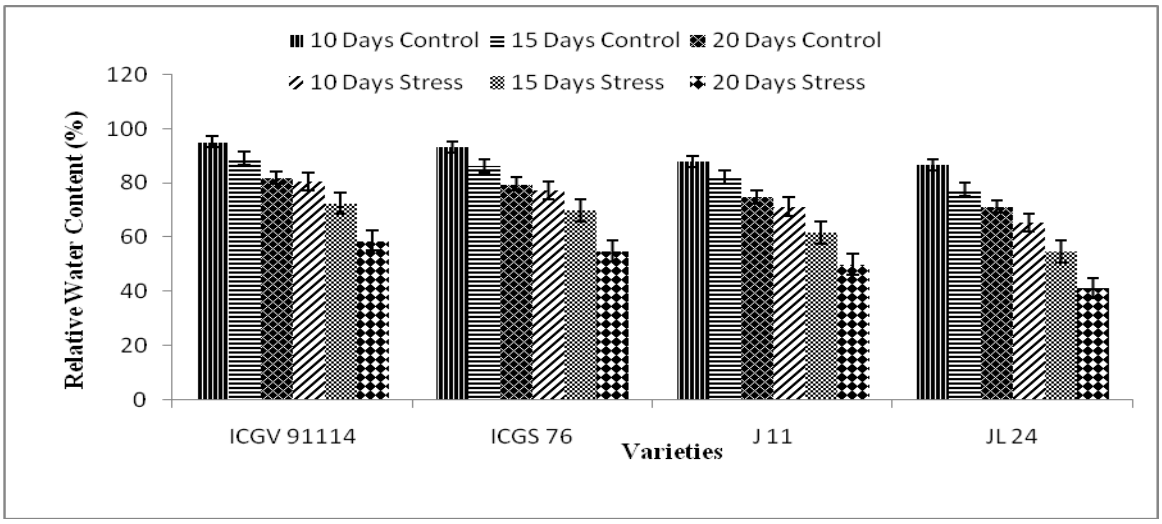


Fig. 1: RWC (%) in 10, 15 and 20 days drought stressed groundnut cultivars

3. Isolation and Purification of proteins:

In order to standardize the protocol for extraction of proteins from the leaves, 3 protocols; Tris – Hcl, Phenol, and TCA in acetone methods, were evaluated. The Tris – Hcl and Phenol extraction methods resulted in gels with very poor resolution and heavy streaking besides yielding only a few protein spots whereas the third, TCA in acetone method, resulted in gels with high resolution, little streaking and with more number of good quality protein spots. Therefore, in this present investigation, the TCA in acetone method was adapted for protein isolation for performing 2 D-PAGE analysis.

B. Quantification of differentially expressed protein spots:

Quantification of 280 differentially expressed protein spots observed on 2-DE gels in 4 groundnut cultivars was carried out using PD Quest Basic software. In ICGV 91114, of the 106 differential protein spots; 37 spots in 10 day-stressed (20 more abundant, 11 down-regulated and 6 new); 21 in 15 day-stressed (14 more abundant, 1 down-regulated and 6 new) and 25 in 20 day-stressed plants (13 more abundant, 9 down-regulated and 1 new) (Table-1 & Fig. 2) were found in addition to the 23 protein spots noticed in all the stressed plants, however, with variable levels of expression. In ICGS 76, of the 59 differential protein spots; 30 spots in 10 day-stressed (16 more abundant, 9 down-regulated and 5 new spots) and 8 in 20 day-stressed plants (4 more abundant, 1 down-regulated and 3 new proteins) in addition to the 21 spots observed in both 10 and 20 day-stressed plants (Table-2 & Fig. 3). In J 11 cultivar, out of 59 differential protein spots, 30 spots in 10 day-stressed (12 more abundant, 11 down regulated and 7 new spots) and 14 in 20 day-stressed plants (10 more abundant, 1 down-regulated and 3 new spots) were observed in addition to 15 protein spots found in both the 10 and 20 day-stressed plants (Table-3 & Fig. 4). In JL 24, out of 56 differential spots quantified, 28 were found in 10 day-stressed (12 more abundant, 14 down regulated and 2 new) while only 2 spots were found in 20 day-stressed plants (1 more abundant and 1 down regulated) and the remaining 26 spots were found both in 10 and 20 day-stressed plants (Table-4 & Fig. 5).

Table - 1. Quantification of differentially expressed protein spots in ICGV 91114

Sp. No.	pI	M.Wt.	Expression levels			Sp. No.	pI	M.Wt.	Expression levels			Sp. No	pI	M.Wt.	Expression levels		
			10	15	20				10	15	20				10	15	20
1	6.6	62.0	DR	DR	-	29	5.2	24.0	N	N	MA	57	5.3	19.0	-	MA	-
2	6.7	62.0	DR	DR	-	30	6.6	22.5	N	-	-	58	5.1	18.5	-	DR	-
3	6.1	48.0	DR	-	-	31	6.7	50.5	N	-	-	59	4.5	32.5	-	-	MA
4	6.2	47.0	DR	-	-	32	5.6	52.0	MA		-	60	4.7	32.5	-	-	MA
5	6.1	46.8	DR	-	-	33	5.8	28.5	MA	MA	-	61	4.6	33.0	-	-	MA
6	6.5	49.0	MA	-	-	34	5.5	47.0	MA	MA	-	62	5.0	45.0	-	-	DR
7	6.1	28.0	DR	-	-	35	5.5	13.5	MA	MA	-	63	5.1	44.0			DR
8	6.0	20.0	DR		-	36	4.3	23.5	MA	-	-	64	5.3	41.2			DR
9	5.9	21.5	DR	DR	-	37	4.6	17.5	MA	-	-	65	5.2	44.0			DR
10	4.5	37.0	DR	-	-	38	6.7	21.5	MA	MA	-	66	5.8	54.0			DR
11	4.6	41.0	MA	-	-	39	6.6	19.5	MA	MA	-	67	6.8	54.0			DR
12	4.7	37.5	MA	-	-	40	5.3	4.0	DR	DR	-	68	6.9	54.0			DR
13	4.8	35.0	MA	-	-	41	5.1	5.0	MA	MA	-	69	5.5	21.5	-	-	MA

14	4.9	17.5	MA	DR	-	42	6.6	17.5		N	N	70	5.0	18.6	-	-	DR
15	5.5	31.5	MA	MA	-	43	5.9	18.0	-	N	-	71	5.68	31.0	-	-	MA
16	5.36	24.0	MA	DR	-	44	5.5	22.0	-	N	-	72	6.4	26.5	-	-	DR
17	5.54	22.5	MA	MA	MA	45	5.1	30.0	-	N	-	73	4.9	17.0	-	-	MA
18	5.8	27.0	MA	-	-	46	6.4	43.0	-	MA	MA	74	4.6	21.1	-	-	MA
19	6.3	29.0	MA	-	-	47	6.0	39.0	-	MA	MA	75	4.8	32.0	-	-	MA
20	6.2	35.0	MA	MA	-	48	6.0	37.5	-	N	-	76	6.2	26.5	MA	-	MA
21	6.2	35.5	MA	MA		49	6.2	37.0	-	MA	-	77	6.1	29.5	MA	-	MA
22	6.1	30.5	MA	MA	MA	50	6.5	29.5	-	MA	-	78	5.4	32.5	-	-	MA
23	5.8	52.0	MA	-	-	51	6.1	35.5	MA	MA	-	79	5.4	34.8	-	-	MA
24	6.5	33.0	DR	-	-	52	5.0	27.0		MA	-	80	6.0	36.5	-	-	MA
25	6.5	35.0	N	-	-	53	6.0	23.0	N	MA	-	81	4.9	37.0	-	-	N
26	6.0	30.5	DR	-	-	54	5.7	26.0	N	MA	-	82	6.0	22.0	-	-	MA
27	6.4	25.0	N	-	-	55	6.6	20.5	N	MA	-	83	6.4	26.5	-	-	DR
28	4.6	26.0	N	-	-	56	4.8	32.5	-	MA	-						

(Sp.No.- Spot No., pI - isoelectric point, M.Wt.- Mol. Wt., MA-more abundant, DR- down-regulated, N- new)

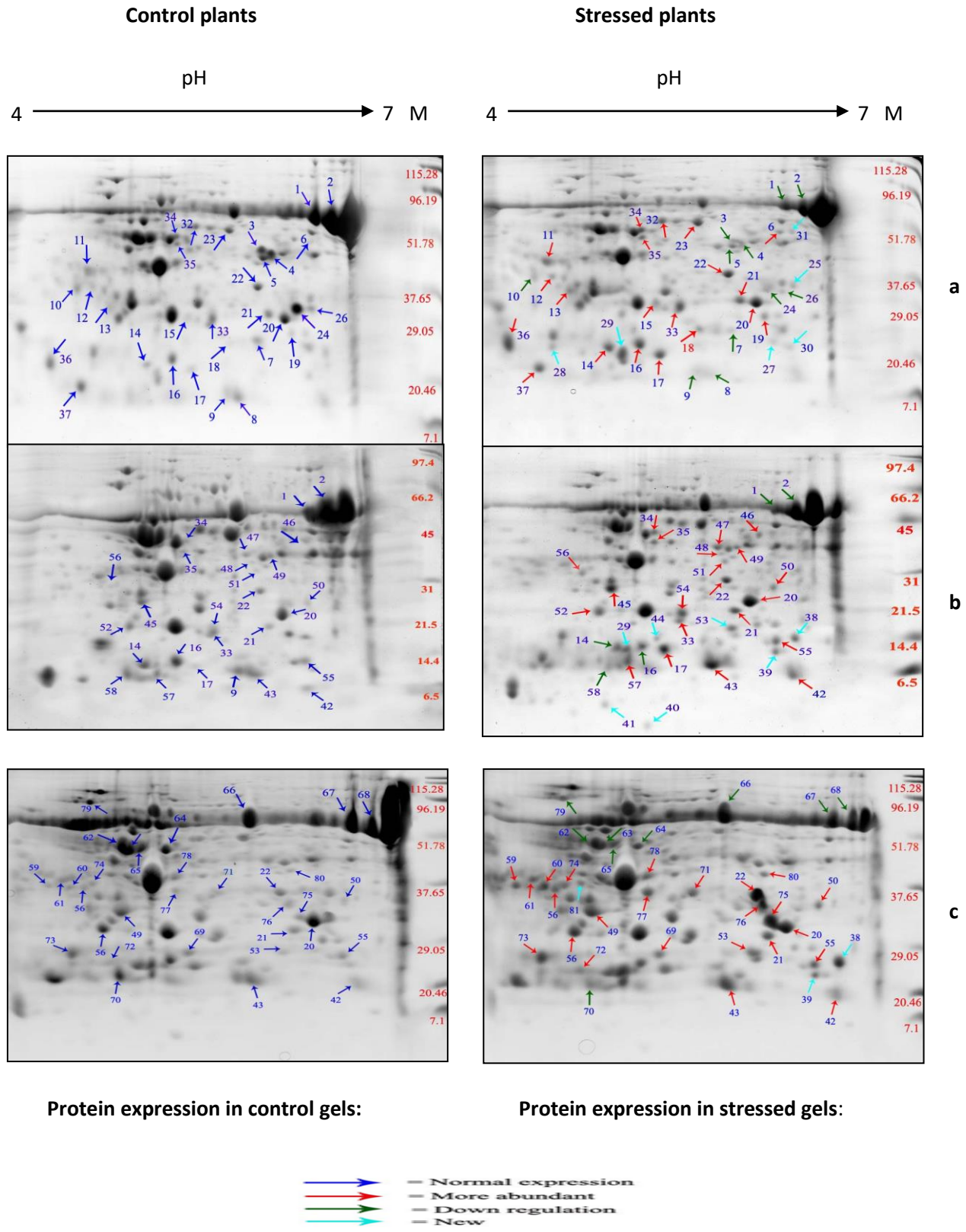


Fig. 2: 2 – DE gel images of 106 differentially expressed protein spots: a) 37 in 10, b) 21 in 15, and c) 25 in 20 day old stressed, and 23 common spots, seedlings of ICGV 91114

Table- 2. Quantification of differentially expressed protein spots in ICGS 76

Sp. No.	pI	M.Wt.	Expression levels		Sp. No.	pI	M.Wt.	Expression levels	
			10	20				10	20
1	6.6	54.0	DR	DR	20	6.0	18.0	N	-
2	6.7	54.0	DR	DR	21	6.7	17.5	N	-
3	6.8	54.0	DR	DR	22	6.5	31.0	N	-
4	6.1	31.0	MA	MA	23	4.9	24.5	N	-
5	6.14	27.0	MA	MA	24	6.5	24.0	MA	MA
6	6.3	27.5	MA	MA	25	5.7	31.0	MA	MA
7	5.4	20.5	MA	MA	26	5.6	35.0	MA	MA
8	5.5	20.5	MA	MA	27	5.1	31.0	DR	-
9	5.3	20.3	MA	MA	28	5.6	45.0	MA	MA
10	5.0	21.0	MA	MA	29	5.9	17.4	DR	DR
11	6.6	43.5	MA	MA	30	6.6	20.0	MA	MA
12	6.6	30.5	MA	-	31	6.2	27.5	-	N
13	5.9	45.0	MA	MA	32	6.5	33.0	-	N
14	5.9	21.0	DR	-	33	6.25	30.5	-	N
15	6.0	20.0	DR	-	34	6.18	35.0	-	MA
16	5.27	21.5	DR	MA	35	6.6	38.5	-	MA
17	4.7	23.5	DR	DR	36	5.7	38.5	-	MA
18	4.8	30.0	DR	-	37	5.1	35.0	-	MA
19	5.3	20.5	N	N	38	5.7	9.5	-	DR

(Sp. No.- Spot No., pI - isoelectric point, M.Wt.- Mol. Wt., MA-more abundant, DR- down-regulated, N- new)

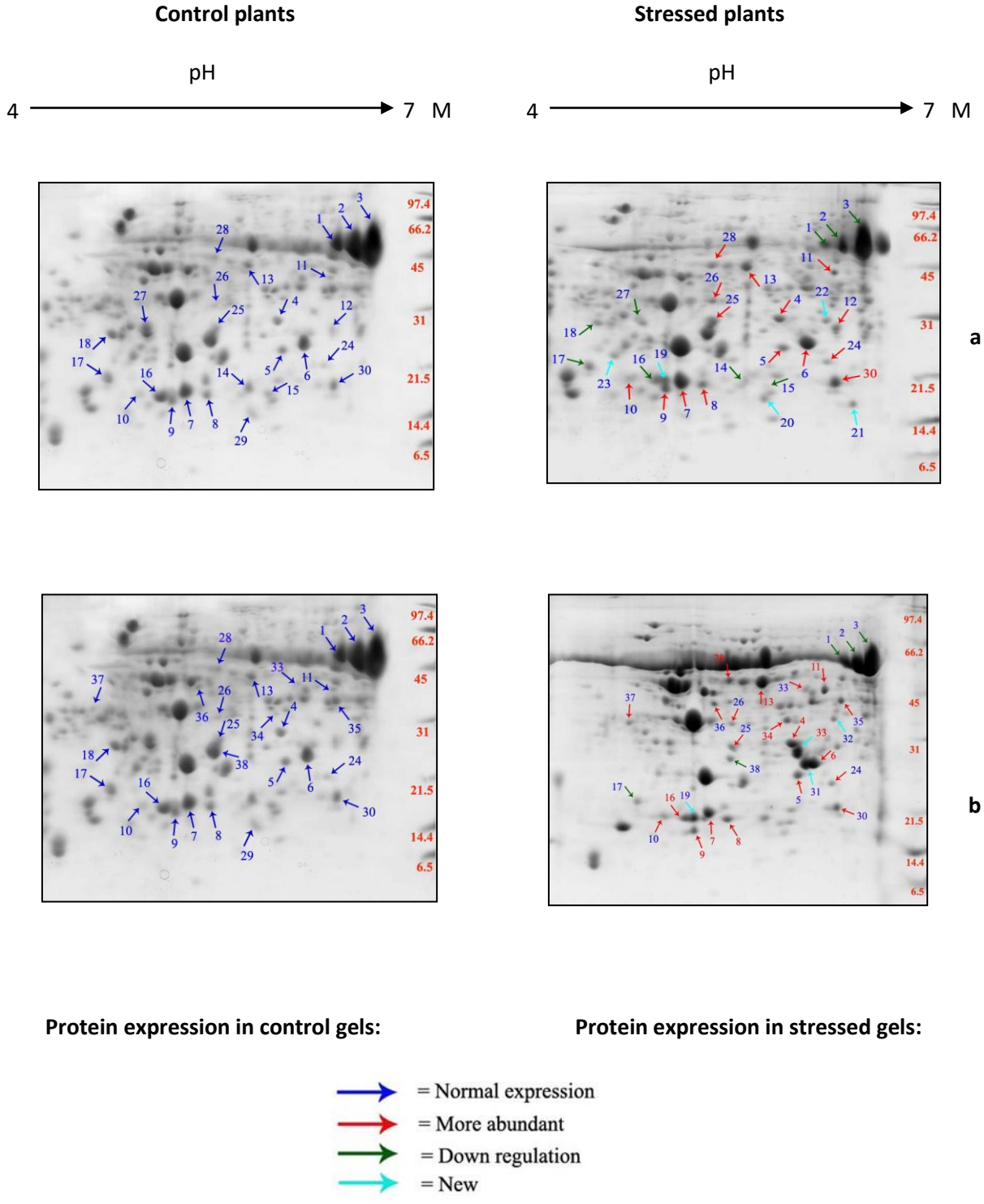


Fig. 3: 2 – DE gel images of 59 differentially expressed of protein spots: a) 30 in 10, and b) 8 in 20 day old stressed, and 21 common spots, seedlings of ICGS 76

Table -3. Quantification of differentially expressed protein spots in J 11

Sp. No.	pI	M.Wt.	Expression levels		Sp. No.	pI	M.Wt.	Expression levels	
			10	20				10	20
1	6.5	52.0	DR	DR	23	6.4	32.8	N	-
2	6.68	52.0	DR	DR	24	5.5	13.0	DR	MA
3	6.5	29.5	MA	MA	25	5.6	29.0	DR	DR
4	6.2	31.0	MA	DR	26	4.2	36.1	DR	DR
5	6.1	37.2	MA	MA	27	4.7	35.0	N	-
6	6.1	20.0	MA	-	28	5.4	37.6	N	-
7	5.8	21.0	MA	-	29	6.4	45.0	MA	MA
8	6.0	7.1	MA	MA	30	5.7	35.0	N	-
9	5.9	27.0	MA	MA	31	5.0	27.5	DR	DR
10	4.5	31.0	DR	DR	32	5.4	15.2	MA	MA
11	4.6	34.0	DR	DR	33	5.28	11.0	MA	MA
12	4.7	32.0	DR	DR	34	5.2	15.6	MA	MA
13	4.8	26.0	DR	DR	35	5.3	15.5	N	N
14	4.7	26.5	MA	DR	36	5.1	15.1	MA	MA
15	5.1	23.0	N	-	37	6.5	37.6	MA	MA
16	5.4	30.0	DR	-	38	5.5	39.5	MA	MA
17	5.5	33.0	N	-	39	6.0	29.0	MA	MA

18	6.8	52.0	DR	-	40	6.1	19.0	MA	MA
19	6.1	33.0	N	-	41	5.8	17.0	MA	MA
20	6.3	25.5	MA	MA	42	6.2	26.0	N	N
21	6.2	24.5	MA	MA	43	6.2	29.0	N	N
22	6.1	30.5	MA	MA	44	5.6	30.1	DR	MA

(Sp. No. = Spot No., pI = isoelectric point, M. Wt.= Mol. Wt., MA = more abundant, DR = down regulated, N = new)

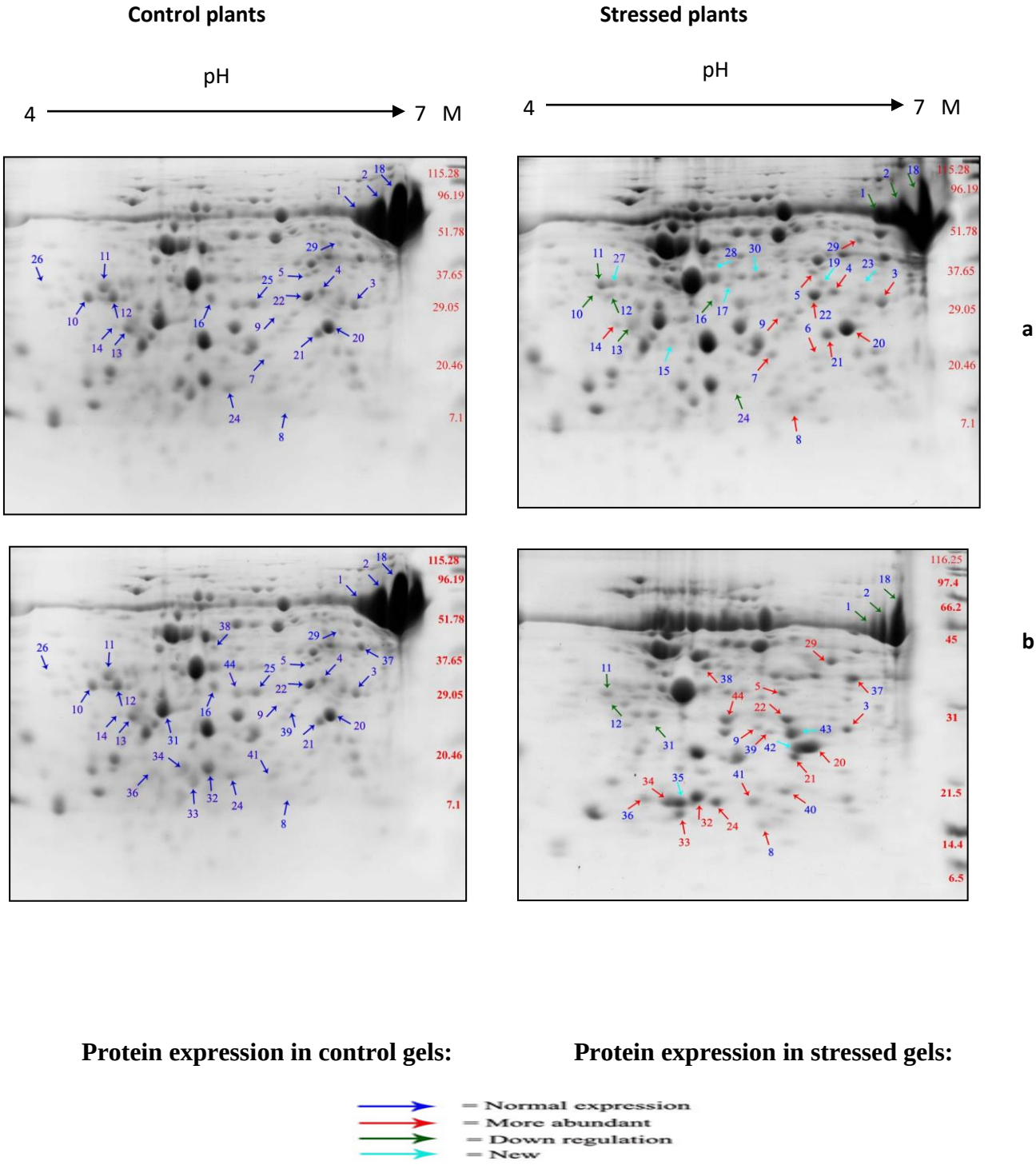


Fig. 4:– 2 – DE gel images of 59 differentially expressed protein spots: a) 30 in 10 and b) 14 in 20 day old stressed, and 15 common spots, seedlings of J 11

Table -4. Quantification of differentially expressed protein spots in JL 24

Sp. No.	pI	M.W.	Expression levels		Sp. No.	pI	M.W.	Expression levels	
			10	20				10	20
1	4.4	36.0	DR	DR	16	5.3	15.5	N	N
2	4.6	34.0	DR	DR	17	5.9	14.2	MA	DR
3	4.7	32.0	DR	DR	18	5.5	13.0	DR	DR
4	5.4	15.2	MA	MA	19	5.6	29.0	DR	DR
5	5.5	21.5	MA	DR	20	5.0	16.5	DR	DR
6	5.5	13.0	MA	DR	21	6.38	27.0	DR	-
7	4.8	26.0	MA	DR	22	5.9	38.0	DR	DR
8	4.7	26.5	MA	DR	23	6.68	52.0	DR	DR
9	5.2	5.0	MA	-	24	6.8	52.0	DR	DR
10	5.28	11.0	MA	-	25	6.5	29.5	MA	DR
11	5.7	23.0	DR	DR	26	6.5	19.0	DR	DR
12	6.3	25.5	MA	DR	27	4.9	23.0	MA	DR
13	6.1	30.5	MA	DR	28	6.4	45.0	N	N
14	6.2	24.5	DR	DR	29	5.3	23.0	DR	DR
15	5.9	16.3	DR	DR	30	5.4	30.0	DR	DR

(Sp. No. = spot number, pI = isoelectric point, M.Wt.= Mol. Wt., MA=more abundant, DR = down regulation, N= new)

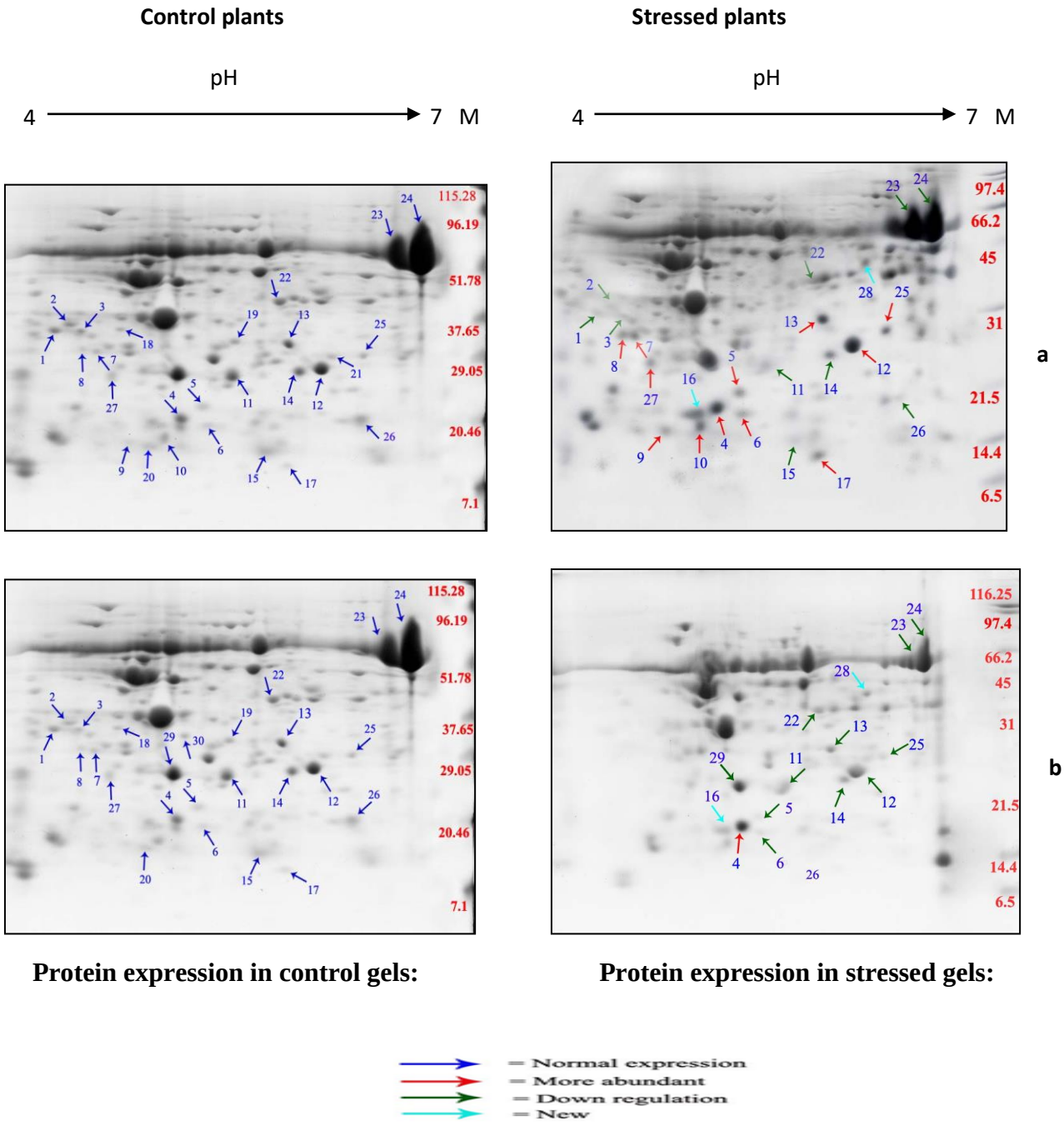


Fig. 4:– 2 – DE gel images of 56 differentially expressed of protein spots: a) 28 in 10 and b) 2 in 20 day stressed, and 26 common spots, seedlings of JL 24

C. Peptide Mass Fingerprinting (PMF) and sequencing of differentially expressed proteins:

Based on PD Quest quantification, 30 differentially expressed protein spots selected from 3 cultivars (ICGV 91114 = 17, ICGS 76 = 6 and J 11 = 7) were subjected to tryptic digestion for PMF analysis. These digested peptides were injected to MALDI-TOF analysis for obtaining the mass spectra (MS) of each protein. The MASCOT search in NCBI nr database for comparing mass of these peptides with that of the theoretical mass of proteins deposited in the taxonomy; *Viridiplantae* (green plants), resulted in the identification of 30 proteins (Tables - 5, 6 & 7). These 30 identified proteins exhibited homology with *Arabidopsis thaliana* (13), *Oryza sativa* (4), *Arachis hypogaea* (6), *Glycine max* (1), *Pisum sativum* (1), *Vigna unguiculata* (1), *Zea mays* (1), *Triticum aestivum* (1), *Cadellia pentastylis* (1) and *Daucus carota* (1) (Fig. 6). Of the 30 PMF identified differentially expressed proteins, 12 were sequenced by MALDI -TOF/TOF (Table-8).

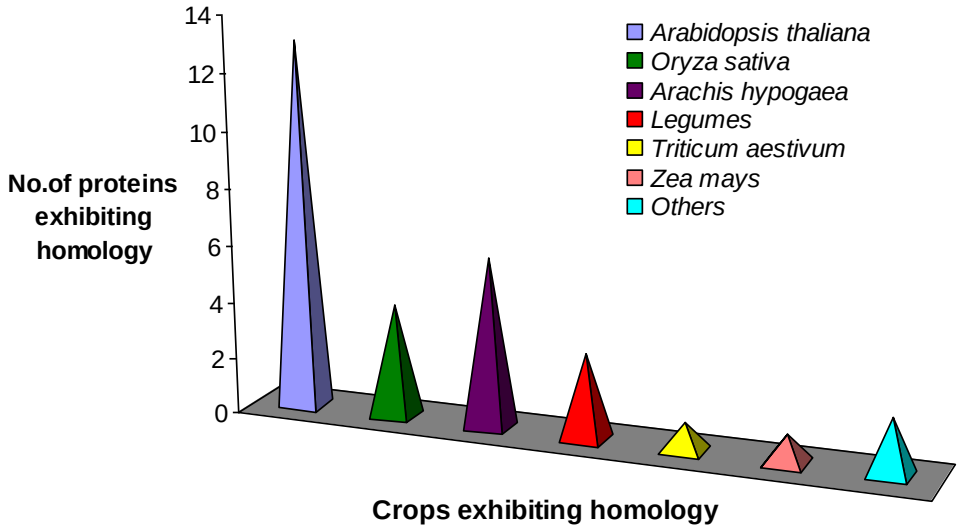


Fig. 6: PMF identified proteins in peanut exhibiting homology with crops listed in NCBI nr *Viridiplantae* database

Table – 5. Peptide Mass Fingerprinting (PMF) of differentially expressed protein spots in ICGV 91114

Sp. No.	Exptl. M.Wt.	Theo. M.Wt.	Homologous protein	Organism	Biological Process	Molecular function	Cellular component	Mascot Score	Accession No.
1	56.0	52.6	RUBP carboxylase large chain	*Cadellia pentastylis*	Photorespiration / Reductive pentose-phosphate cycle	Mg++ ion binding/ Mono-oxygenase activity / RUBP carboxylase activity	Chloroplast/ Plastid	519	gi|1351139
17	16.2	16.2	PR 10 protein	*Arachis hypogaea*	Defense response / Response to biotic stimulus	-	-	65	gi|52547774
22	31.5	100.5	Unnamed protein product	*Oryza sativa*	Carbohydrate metabolic process	Hydrolase activity / Hydrolyzing O-glycosyl compounds	-	74	gi|8099130
38	19.5	92.4	Sucrose synthase isoform 1	*Daucus carota*	Biosynthetic / Sucrose metabolism	Sucrose synthase / Glycosyl transferase	-	82	gi|1351139
38	19.5	140.8	Structural maintenance of chromosome 1	*Arabidopsis thaliana*	DNA recombination / DNA repair / Chromosome condensation	ATP binding	Nucleus	73	gi|45594277
39	26.5	49.2	Putative serine / Threonine protein kinase	*Oryza sativa*	Signalling	ATP binding/protein serine/threo-ne kinase activity	Plasma membrane	51	gi|30017556
40	14.4	14.1	Unknown protein	*Arabidopsis thaliana*	-	-	-	80	gi|30695529
40	14.4	14.2	Probable myosin heavy chain	*Arabidopsis thaliana*	-	-	-	79	gi|25408221
43	31.5	124.8	Gamma response I protein	*Arabidopsis thaliana*	Signalling	Sequence specific DNA binding transcription factor	-	70	gi|4678941
43	31.5	20.4	Ubiquitin conjugating enzyme	*Arabidopsis thaliana*	Post replication repair	Acid-aminoacid ligase activity/protein binding	UBC13-MMS2 complex cytosol	66	gi|18403085
47	41.5	66.2	Hypothetical protein	-	-	-	*Oryza sativa*	69	gi|51038186

48	19.5	22.3	OSJNBa0029C04.4	DNA integration / RNA dep. DNA replication	RNA binding / Ribonuclease H activity / RNA-directed DNA polymerase activity	-	*Oryza sativa*	92	gi|38346291
53	21.5	21.1	Calcium ion binding	Trichome branching	Calcium ion binding / Protein binding	-	*Arabidopsis thaliana*	75	gi|15233402
53	21.5	36.8	Late embryogenesis abundant protein	-	-	-	*Pisum sativum*	66	gi|56709428
69	21.5	6.5	Stress-induced protein KIN1	Stress response		-	*Arabidopsis thaliana*	50	KIN1_ARATH
74	33.0	47.5	26S protease regulatory subunit 7 (26S proteasome AAA-ATPase)	Protein catabolic process	ATP binding / Nucleoside-triphosphatase activity	Cytosol / Plasma membrane/proteosome	*Arabidopsis thaliana*	60	PRS7_PRUPE
75	26.5	21.3	Endoribonuclease Dicer homolog (EC 3.1.26.-)	Plant defense / RNA mediated gene silencing	ATP binding / Protein binding / Metal ion binding	Cytoplasm / Nucleus	*Arabidopsis thaliana*	73	DICER_ARATH
80	24.0	43.4	GSK-like kinase	-	ATP binding / Protein serine / threonine kinase activity	-	*Triticum aestivum*	67	gi|21745456
82	17.3	17.3	17.3 kDa class I heat shock protein	Stress response	-	Cytoplasm	*Glycine max*	70	gi|123534
83	26.5	84.9	Kinesin-3 (Kinesin-like protein C)	Celldivision / Mitosis / Microtubule based movement	ATP binding / ATPase activity / Microtubule binding / Microtubule motor activity	Cytoplasm / Microtubule	*Arabidopsis thaliana*	66	ATK3_ARATH

(Sp .No.= Spot number, Expl. M.Wt.= Experimental Mol. Wt., Theo. M.Wt. = Theoretical M.Wt.)

Table - 6. Peptide Mass Fingerprinting (PMF) of stress induced proteins in ICGS 76

Sp. No.	Exptl. M.Wt.	Theo. M.Wt	Homologous protein	Organism	Biological Process	Molecular function	Cellular component	Mascot score	Accession No.
4	30.5	27.2	Cytosolic ascorbate peroxidase	*Vigna unguiculata*	Hydrogen peroxide stress response	Oxidoreductase/ Peroxidase activity	Cytoplasm	113	gi|1420938
8	20.0	16.2	PR10 protein	*Arachis hypogaea*	Defense response / Response to biotic stimulus	-	-	76	gi|52547774
15	17.0	71.4	FACT complex subunit SSRP1	*Zea mays*	DNA repair / DNA replication / regulation of transcription- DNA dependent	DNA binding	Chromosome/ Nucleus	52	SSRP1_MAIZE
19	21.0	16.2	PR10 protein	*Arachis hypogaea*	Defense response / Response to biotic stimulus	-	-	52	gi|52547774
30	30.0	30.1	Lipoate protein ligase-like protein	*Arabidopsis thaliana*	Lipoate biosynthetic process / Protein modification process	Octanoyl transferase activity	Chloroplast	38	gi|7939551
31	27.0	22.3	Photosystem I reaction center subunit II, chloroplast precursor	*Arabidopsis thaliana*	Photosynthesis	Protein binding	Chloroplast Membrane Photosystem I Plastid Thylakoid	61	PSAD2_ARATH

(Sp. No. = Spot number, Expl. M.Wt.= Experimental Mol. Wt., Theo. M.Wt.= Theoretical M.Wt.)

Table – 7. Peptide Mass Fingerprinting (PMF) of stress induced proteins in J 11

Sp. No.	Exptl. M.Wt.	Theo. M.Wt.	Homologous protein	Organism	Biological process	Molecular function	Cellular component	Masco tScore	Accession No.
8	8.64	15.6	RUBP carboxylase small chain precursor	*Phaseolus vulgaris*	Photosynthesis	*Lyase / Monoxygenase*	Plastid	181	gi|123534
10	31.0	30.7	Putative lectin precursor	*Arachis hypogaea*	-	Kinase	Cell membrane	283	gi|15233402
11	34.0	29.2	14-3-3 protein	*Vigna angularis*	-	Protein domain specific binding	-	459	gi|45594277
12	32.0	28.3	Mannose/Glucose-binding lectin precursor	*Arachis hypogaea*	-	Sugar binding	-	233	gi|8099130
13	27.5	21.8	2 – cys peroxiredoxin – like protein	*Hyacinthus orientalis*	Stress response	Antioxidant activity / Peroxiredoxin activity	Chloroplast / Plastid	96	gi|47027073
24	21.0	16.2	PR10 protein	*Arachis hypogaea*	Defense response / Response to biotic stimulus	-	-	76	gi|52547774
32	19.5	16.2	PR10 protein	*Arachis hypogaea*	Defense response/ Response to biotic stimulus	-	-	52	gi|52547774

(Sp. No. = Spot number, Expl. M. Wt.= Experimental Mol. Wt., Theo. M. Wt. = Theoretical M. Wt.)

Table- 8. MS/MS sequences of MALDI-TOF identified drought stress tolerant proteins in groundnut

Sp. No.	Exptl. M.Wt.	Protein identified	Accession No.	Mascot Score	Sequence coverage	Matching peptide sequence in red
53	21.5/5.8	Late embryogenesis abundant protein	gi|56709428	66	50	MASRQDRREA RAEADARRAA EEIARAR**DER VMQAEVDAR**S AADEIARARA DR**GAATMGADTAHHAAGGGG ILESVQEGAK** SFVSAVGR**TF GGARDTAAEK TSQTADATRD K**LGEYKDYTA DK**ARETNDSV AR**KTNETADA SRDK**LGEYKD YTADKTR**ETK DAVAQK**ASDA SEATKNK**LGE YKDALARK**TR DAKDTTAQKA TEFKDGVK**AT AQETRDATAD TARKAK**DATK DTTQTAADK**A RETAATHDDA TDKGQGQGLL GALGNVTGAI KEK**LTVSPAA TQEHLGGGEE RAVK**ERAAEKAASVYFEEKD RLTRERAAER **VDKCVEKCVE GCPDATCAHR** HGKM
53	21.5/5.8	Calcium ion binding	gi|15233402	75	24	**MESNNNEKKK VARQSSSFRL RSPSLNALRL QR**IFDLFDKN GDGFITVEEL SQALTRLGLN ADLSDLKSTV ESYIQPGNTG LNFDDFSSLH KTLDDSFFGG ACGGGENEDD PSSAAENESD LAEAFKVFDE NGDGFISARE LQTVLK**KLGL PEGGEMERVE K**MIVSVDRNQ DGRVDFFEFK NMMRTVVIPS S
38	19.5/6.3	Sucrose synthase isoform 1	gi|1351139	82	33	**MGEPVLTRVH SLRERMDSTL ANHRNEILMF LSR**IESHGK**G ILKPHQLLAE YEAISKEDKL KLDDGHGAFA EVIK**STQEAI VSPPWVALAI RLRPGVWEYV RVNVHHLVVE ELSVPQYLQF KEELVIGSSD ANFVLELDFA PFTASFPRPTLTKSIGNGVE FLNRHLSAKM FHGK**DSMHPL LEFLRLHNYN GKTLMLNNRV QNVNGLQSML R**KAGDYLSTL PSDTPYSEFE HKFQEIGFER GWGDTAERVT EMFHMLLDLL EAPDASTLET FLGKIPMVFN VVILSPHGYF AQENVLGYPDTGGQVVYILD QVPALER**EMI KRIK**EQGLDI KPR**ILIVTRL LPDAVGTTCN QR**LEKVFGAE HAHILRVPFR **TEKGILRK**WI SRFEVWPYIE TFTEDVAKEI ALELQAKPDL IIGNYSEGNL VASLLAHKLG VTQCTIAHAL EKTKYPDSDIYWEKFDKKYH FSSQFTADLI AMNHTDFIIT STFQEIAGSK **DTVGQYESHT AFTMPGLYR**V VHGIDVFDPK F-VSPGADT SVYFSYKEKE KR**LTTLHPEI EELLYSSVEN EEHLCIIKDK NKPILFTMAR** LDNVKNLTGF

						VEWYAK**SPKLRELVNLVVVG GDR**RK**ESKDL EEQAQMKKMY ELIDTYKLNG QFRWISSQMN RVR**NGELYRY IADTKGAFVQ PAFYEAFGLT VVEAMTCGLP TFATLHGGPA EIIVHGKSGF HIDPYHGEQV AELLVNFFEK CKTDPSQWDA ISAGGLKRIQEKYTWQIYSE RLLTLAGVYG FWK**HVSKLDR LEIR**RYLEMF YALK**YRKLAE SVPLAKDE
38	19.5/6.3	Structural maintenance of chromosome 1	gi|45594277	73	31	M**PAIQSPSGK ILQLEMENFK SYK**GHQLVGP FKDFTAIIGP NGSGKSNLMD AISFVLGVRT GQLRGSQLKD LIYAFDDRDK EQRGRKAFVR **LVYQMDDGVE LRFTR**SITSA GGSEYRIDNR **VVNLDEYNGK LR**SLGILVKA R**NFLVFQGDVESIASKNPK**E LTGLLEEISG SEELKKEYEG LEEKKASAEE K**AALIYQKKK** TIGNEKKLKK AQKEEAEKHL RLQEELKALK RERFLWQLYN IENDIEK**ANE DVDSEKSNRK** DVMRELEKFE REAGKRKVEQ AK**YLKEIAQR EK**KIAEKSSKLGKIQPELLR FKEEIARIKA KIETNRKDVD KRKKEK**GKHS KEIEQMQKSI KELNK**KMELF NKK**RQDSSGK LPMLDSQLQD YFR**LKEEAGM KTIK**LRDEHE VLERQR**RTDL EALRNLEENY QQLINRKNDL DEQIKR**FKDR QGEIETSSSK** YKNETTSLKT ELRALQEKHV NAREASAKLK TRIAELEDQL SDLTAERYEN ERDSRLTQAV ESLK**RLFQGV HGRMTDLCRP NRKKYNLAVT VAMGR**FMDAV VVEDENTGKD CIKYLKEQRL PPMTFIPLQS VRVKQVFERLRNLGGTAKLVFDVIQFDPEL EKAVLYAVGN TLVCDELEEA K**VLSWSGER**F KVVTVDGILL TK**AGTMTGGT SGGMEAKSNK** WDDKKIEGLK KNKEDFEQQL ENIGSIREMQ MKESEISGKI SGLEK**KIQYA EIEKK**SIKDK **LPQLEQEER**N IIEEIDRIKP ELSK**ARTEVD KR**KTEMNKLE KR**MNEIVDRI YKDFSQSVGV PNIRVYEETQ LK**TAEKEAEE R**LELSNQPAK LKYQLEYEQN R**DVGSRIRKI ESSISSLETD LEGIQKTMSE RKETAVK**ITN EINNWKKEME ECK**QKSEEYE KEILDWKKQASQATTSITKL NRQIHSK**ETQ IEQLISQK**QE ITEKCELEHI TLPVLSDAME EDDSDGPQFD FSELGR**AYLQ ERRPSAREK**V EAEFRQKIES KTSEIER**TAP NLRALDQYEA IQEKEKQVSQ EFEAARKEEK** QVADAFNTVK QKRYELFMEAFNHIASNIDK IYKQLTKSNT HPLGGTAYLN LENEDDPFLH GIKYTTMPPT KRFRDMEQLS GGEKTVAALA LLFSIHSYRP SPFFILDEVD AALDNLNVAK VAKFIRSKSC

						QAARDNQDAE DGNGFQSIVI SLKDSFYDKA EALVGVYR**DTERSCSSTMSF DLR**NYQES
89	17.3/5.9	17.3 kDa class I heat shock protein	gi|123534	70	40	M**SLIPSFFGG R**RSSVFDPFS LDVWDPFKDF PFPSSLSAEN SAFVSTR**VDW KETPEAHVFK** ADIPGLKKEE VKLEIQDGRV LQISGER**NVE KEDKNDTWHR VER**SSGKLVR R**FRLPENAK**V DQVK**ASMENG VLTVTVPK**EE IKKPDVKAIDISG
4	30.5 / 6.1	Cytosolic ascorbate peroxidase	gi|1420938	113	20	MGKSYPTVSP DYQKAIEKAK RK**LRGFIAEK** KCAPLILRLA WHSAGTFDSK TKTGGPFGTI KHQAELAHGA NNGLDIAVRL LEPIKEQFPI VSYADFYQLA GVVAVEITGG PEVPFHPGRE DKPEPPPEGR **LPDATKGSDH LRDVFGKAMGLSDQDIVALS GGHTIGAAHK ER**SGFEGPWT SNPLIFDNSY FTELLTGEKD GLLQLPSDKA LLTDSVFRPL VEKYAADEDV FFADYAEAHL KLSELGFAEA
31	27/6.3	Photosystem I reaction center subunit II, chloroplast precursor	PSAD2_ARATH	61	9	MATQAAGIFN SAITTAATSG VKKLHFFSTT HRPKSLSFTK TAIRAEKTDS SAAAAAAPAT KEAPVGFTPP QLDPNTPSPI FAGSTGGLLR KAQVEEFYVI TWNSPKEQIF EMPTGGAAIM REGPNLLKLA RKEQCLALGT RLR**SKYKITY QFYR**VFPNGE VQYLHPKDGV YPEKANPGRE GVGLNMRSIG KNVSPIEVK**FTGKQSYDL**
10	31/4.5	Putative lectin precursor	gi|15233402	283	50	MAISKKILPL LSIATIFLLL LNK**AHSLGSL SFGYNNFEQG DERNLILQGD ATFSASK**GIQ LTKVDDNGTP AKSTVGR**VLH STQVRLWEKS TNR**LTNFQAQ FSFVINSPID NGADGIAFFI AAPDSEIPKN SAGGTLGLSD PSTAQNPSAN QVLAVEFDTF YAQDSNGWDP NYQHIGFDVD PIK**SAATTKW ERRNGQTLNV LVSYDANSKN LQVTASYPDG QSYQVSYNVD LRDYLPEWGR** VGFSAASGQQ YQSHGLQSWS FTSTLLYTSP HYLKLGRFMI
11	34/4.6	14-3-3 protein	gi|45594277	459	50	**MAAAPTPREE NVYMAKLAEQ AERYEEMVEF MEKVSAAADN EELNVEER**NL LSVAYKNVIG ARRASWR**IIS SIEQKEESR**G NEDHVTVIRD YRSKIESELS NICDGILKLL DSRLIPSASS GDSKVFYLK**M KGDYHR**YLAE FKTGAERKEA **AESTLAAYKS AQDIANAELP PTHPIRLGLA** LNFSVFYYEI **LNSPDRACNL AKQAFDEAIA ELDTLGEESY KDSTLIMQLL RDNLTLWTSD**

						MQDDGADEIK EAAPKQDDQ
12	32/4.7	Mannose/glucose-binding lectin precursor	gi|8099130	233	50	**LDSLSFSYNN FEQDDERNLI LQGDAK**FSAS KGIQLTK**VDD NGTPAKSTVG RVLHSTQVRL WEKSTNRLTN FQAQFSFVIK** SPIDNGADGI AFFIAAPDSE IPKNSAGGTL GLFDPQTAQN PSANQVLAVE FDTFYAQDSN GWDPNYQHIG IDVNSIK**SAA TTKWERRDGQ TLNVLVTYDA NSK**NLQVTAS YPDGQRYQLS YRVDLR**DYLP EWGR**VGFSAA SGQQYQSHEL QSWSFTSTLL YTSPHYLKLG RFMI
1	56/6.2	RUBP carboxylase large chain	gi|1351139	519	50	MSPQTETKAS VGFKAGVKDY K**LTYYTPEYE TKDTDILAAF RVTPQPGVPP EEAGAAVAAE SSTGTWTTVW TDGLTSLDR**Y KGRCYHIEPV AGEENQYIAY VAYPLDLFEE GSVTNMFTSI VGNVFGFKAL RALR**LEDLRI PTSYSKTFQGPPHGIQVER**D KLNKYGRPLL GCTIKPKLGL SAKNYGR**AVY ECLRGGLDFT KDDENVNSQP FMRWRDRFLF CAEALFK**AQA ETGEIK**GHYL NATEGTCEEM IKR**AVFAREL GAPIVMHDYL TGGFTANTSL AHYCR**DNGLL LHIHRAMHAVIDRQKNHGMH FR**VLAKGLR**L SGGDHIHAGT VVGK**LEGER**D ITLGFVDLLR DDFIEKDR**SR GIYFTQDWVS LPGVLPVASG GIHVWHMPAL TEIFGDDSVL QFGGGTLGHP WGNAPGAVAN R**VALEACVQA RNEGR**DLAR**E GNEIIR**EASK **WSPELAAACE VWKEIK**FEFE AMDTL
8	8.64	RUBP carboxylase small chain precursor	gi|123534	181	50	**TSVANNGGRV QCIQVWPTVG K**KKFETLSYL PPLTKQQLAK EVDYLLRKGW **VPCLEFELEH GFVYR**EHNK**S PGYYDGRYWT MWK**LPMFGCT DSSQVLKELY EAQTAHPDGF IR**IIGFDNVR** QVQCISFIAY KPPGY

(MS/MS sequenced peptides of *Arachis* shown in bold red are matching with the peptides of *Viridiplantae* taxonomic group)

D. Ontological classification of proteins:

The 30 identified proteins were grouped into different categories based on their sub-cellular localization, biological process and molecular function according to the annotation in the *Viridiplantae* taxonomic database. These proteins when queried using Targetp software revealed their localization in various cell compartments, mostly plastids, cell membrane, nucleus etc. Further, the data on biological process showed that over 20 % of proteins were defence related, 12 % stress responsive, 12 % signalling, 12 % DNA repair, 12 % carbohydrate metabolism, 12 % photosynthesis, 4 % in detoxification and other 4 % in lipid biosynthesis. The molecular function of these proteins indicated their involvement in protein binding, antioxidant activity, sugar binding, and kinase activity etc. (Fig. 7).

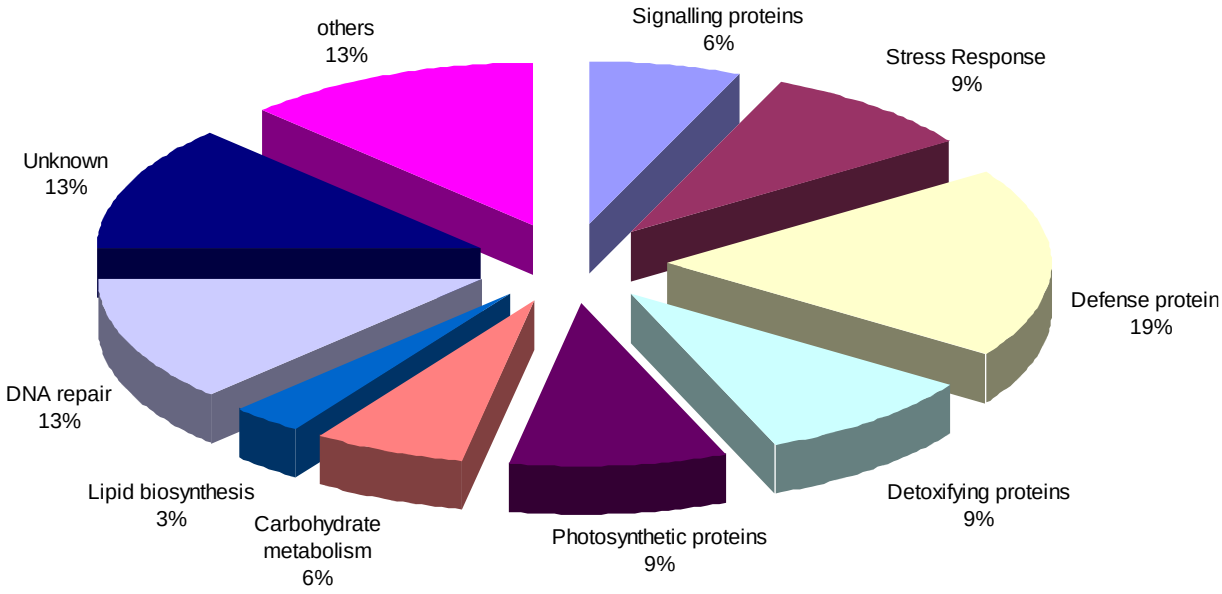
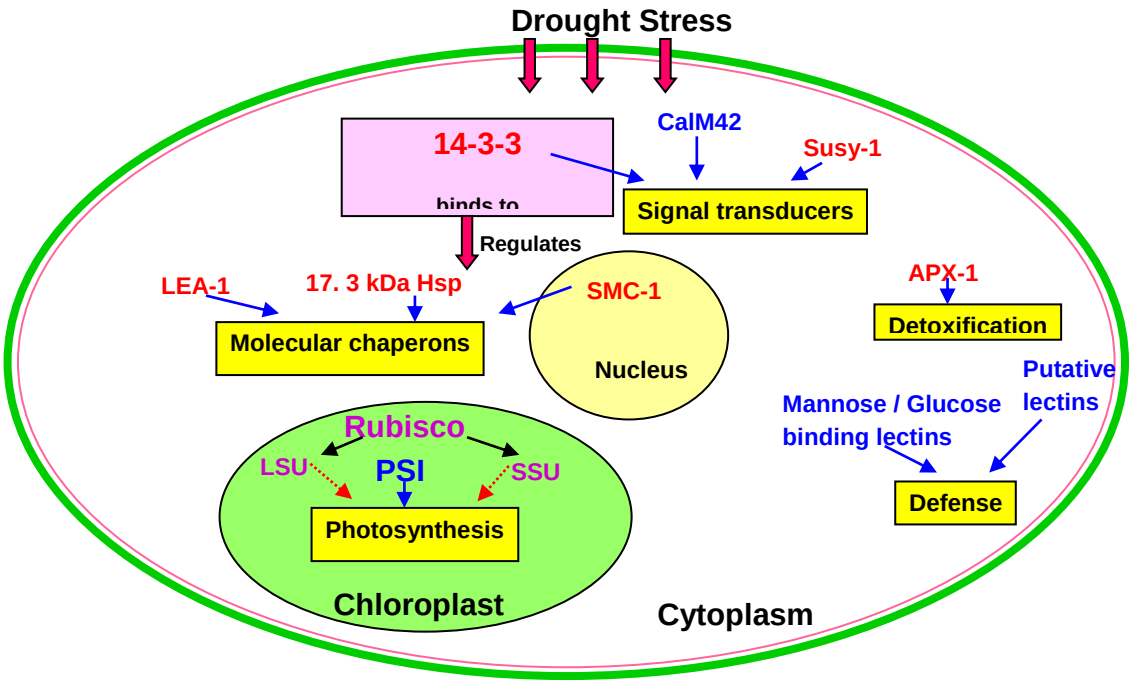


Fig. 7: Functional cataloguing of drought stress responsive protein

E. Proposed model for drought tolerance in groundnut:

In the present study, the following model is being proposed for drought tolerance in groundnut by interlinking the 12 identified and sequenced proteins under drought stress, as to how they might interact and coordinate with each other in conferring drought tolerance to groundnut. The LEA proteins, in dormant condition in the seeds expressed under drought stress, might act as a primary line of defence to tolerate stress by acting as molecular chaperones and also as space fillers preventing cellular collapse. The signalling proteins such as calcium ion binding protein perceives stress signals and transduce them to other signalling proteins. The regulatory protein; '14-3-3', interacts with H^+ ATPases, protein kinases, phosphatases *etc.* and regulate the expression of other drought stress responsive proteins like osmotins, water and ion channel transporters, antioxidants etc. Over-expression of '*Susy-1',* during stress, aids in the translocation of sugars under limited ATP supply for various metabolic pathways. When most of the proteins become dysfunctional due to stress, the '17.3 kDa Hsp' restores their function by preventing the aggregation and acting as molecular chaperons. Detoxifying proteins like 'APX-1', present in the cytoplasm, help in scavenging ROS produced due to oxidative burst in the plant cell and prevent lipid peroxidation and protein oxidation. 'Structural maintenance of chromosome protein-1' helps in preventing double strand breaks in DNA that might arise due to over production of ROS under severe stress conditions and further help in faithful chromosomal segregation. PS-I protein ensures the supply of reducing power, NADPH + H^+, to be fed into Calvin cycle for carbon fixation to carryout photosynthesis. All these factors may be attributed for making ICGV 91114 the most drought tolerant compared to other cultivars studied.



Red = New & More abundant, Blue = More abundant and
Violet = down-regulated, drought tolerant proteins

Fig.7. Proposed model for drought tolerance in peanut: Schematic diagram of a plant cell exhibiting localization of proteins and their co-ordinated function in conferring drought tolerance

Conclusions

In view of the importance of peanut crop, enormous shortage of its oil due to various constraints, especially drought stress, and surging population, an attempt has been made to identify drought tolerant proteins, probing the differentially expressed proteins under drought stress, through proteomic approach. In this investigation, the 30 drought tolerant proteins identified among the 280 differential protein spots could be used as a reference library for probing the drought tolerant proteins in other crops. Further, 6 of the 12 proteins identified and sequenced, for the first time, in peanut could be used as protein markers to evaluate the groundnut germplasm and other crop plants for drought tolerance besides using them as markers in breeding program for the development of new drought tolerant lines. Based on the peptide sequences of these proteins, primers could be designed to isolate their corresponding genes for their use in the development of transgenics in different crops. The results obtained bear significance at the present juncture where the countries all over the world are facing a food crisis, particularly, the oil front.

Objective – 2:

Identification of differentially expressed proteins under drought stress by 2D- DIGE:

Materials and Methods:

In addition, alternative proteomic approaches; 2D-DIGE and LC-MS, were adapted to achieve the above objectives by isolating the proteins from leaf and root tissues of the 40 day- old PEG- stressed ICGV 91114 groundnut plants at flowering stage after studying physiological parameters; the estimation of Relative Water Content (RWC) and the chlorophyll content.

A. Physiological Parameters

The RWC in leaves and roots, while chlorophyll content in leaves were estimated in 40 day- old PEG stressed seedlings with 100ml of 5, 10, 15 and 20% for 24 hours by maintaining the controls.

1. Estimation of RWC:

Fresh leaves, harvested separately from each treatment after 24 h PEG stressed plants, weighing 500 mg was recorded as Fresh Weight and stored in refrigerator for 24 hrs. Next day, the Turgid Weight was recorded and kept in hot air oven at 80º C for 24 hrs and the next day the Dry Weight was recorded. The Relative Water Content of these stressed samples and the controls were recorded as per the following formula

$$RWC = \frac{FW - DW}{TW - DW}$$

Where, FW= Fresh Weight, DW= Dry Weight, TW =Turgid Weight

2. Estimation of Chlorophyll content

The contents of ‘a’, ‘b’ and total Chlorophylls were estimated (Armon, 1949). Fresh leaf samples, weighing 500 mg were chopped into small pieces, suspended in 10ml of 80% Acetone, mixed well and kept for incubation at 4º C overnight in the dark. These samples were centrifuged at 10,000 g for 15 min at 4º C, supernatant was collected and the absorbance was recorded at 645 and 663 nm using spectrophotometer. Later, the contents of ‘a’, ‘b’ and total Chlorophyll were calculated using the following formulae:

Chlorophyll ‘a’ (mg/g) = [(12.7XA663-2.69XA645)V/W]

Chlorophyll ‘b’ (mg/g) = [(22.9XA645-4.68XA663)V/W]

Total Chlorophyll (mg/g) = [(20.2A645+8.02A663)V/W]

Where, A= Absorbance; V= final volume of chlorophyll extract in 80% Acetone, W=fresh weight; 12.7, 2.69, 22.9, 4.68, 20.2 and 8.02 constants

B. Isolation of total proteins from leaf and root tissues of PEG stressed plants:

Protein extraction from the leaf and root tissues was carried out (Saravanan and Rose, 2004). One g. of frozen leaf and root tissues were ground separately to fine powder in liquid nitrogen and suspended in 4 ml of the extraction buffer [0.5 M Tris- HCL (pH 7.5), 0.7 M sucrose, 0.1 KCl 50 mM EDTA, 2% β mercaptoethanol, 1 mM of PMSF] to which equal volume of phenol saturated with Tris – HCl (pH 7.5) was added, mixed for 30 min at 4 ^{o}C and centrifuged at 5000 *g* for 30 min at 4 ^{o}C. The upper phenolic phase was collected and an equal volume of extraction buffer was added to it. The above step was repeated and to the re-extract 4 volumes of 0.1M ammonium acetate in methanol was added and kept overnight at -20 ^{o}C for protein precipitation. The samples were centrifuged at 10,000 *g* for 30 min at 4 ^{o}C and the precipitate was washed thrice in ice cold methanol and twice in ice cold acetone and air dried for few minutes. The final pellet was solubilized in 0.2 ml solution of rehydration buffer [8M (W/V) urea, 2M (W/V) thio-urea 4(w/v) CHAPS, 30 mM DTT, 0.8% (V/V)], immobilized pH gradient (IPG) buffer (GE Healthcare, Uppsala and Sweden) at pH 4-7 and later the content of protein was estimated by Bradford reagent.

C. **Two dimensional-differential in gel electrophoresis (2D-DIGE):**

The analysis of leaf and root proteins was carried out by using Cy-Dye DIGE Fluor Minimal Dye Labelling Kit (GE Healthcare) following the manufacturer’s instructions. For convenience, leaf samples were labelled as; 1, 2, 3, 4, and 5, while root samples as A, B, C, D, and E. Following labelling, 50 µg of protein from leaf and root samples; control (sample-1 & A), 5% PEG treated (sample-2 & B), 10% (sample-3 & C), 15% (sample-4 & D), and 20 % (sample-5 & E). Gel 1 contains samples-1 and 2 labelled with Cy3 and Cy5 respectively. Gel-2 contains samples-3, 4 and 5 labelled with Cy3, Cy5 and Cy2 dyes. Gel 3 contains samples-A and B labelled with Cy3 and Cy5 respectively, while gel 4 consists of samples-C, D and E labelled with Cy3, Cy5 and Cy2 dyes respectively. Samples of 50

µg of internal standard (containing samples of 10 µg each of; 1 or A, 2 or B, 3 or C, 4 or D and 5 or E) were labelled with Cy2 dye. The 17cm IPG strips pH 4-7 (Bio-Rad) for leaf and pH 3-10 linear IPG Strips (GE-Healthcare) for root were rehydrated with leaf and root samples. In the first dimension, the iso-electric focusing (IEF) was used for protein separation and in the second dimension 10% SDS- PAGE gel was used for the separation of leaf and root proteins.

For the first dimension electrophoresis, samples of leaf and root proteins were adjusted to 300 µL with re- suspension solution [7M Urea, 2M thio-urea, 4% (w/v) CHAPS, 0.2 (v/v) IPG buffer pH 3-10, and 1.01% (w/v) bromophenol blue], and loaded onto IPG strips and rehydrated for 14 h at 50V and 20 ^{o}C. IEF was performed on the PROTEAN IEF CELL system (Bio-Rad, USA) by applying 200 V for 1 h, 500 V for 1 h, 1000 V for 1 h, and 10,000 V for longer than 5 h. A voltage of 10,000 V was maintained until a total of 80kV for 1 h was reached. After IEF, the strips were equilibrated for 1hr in 15 ml of reducing equilibration buffer [6M urea, 75 mM Tris-HCL pH 8.8, 29.3% (v/v) glycerol, 2% (w/v) SDS, and 0.002% (w/v) bromophenol blue] containing 1% DTT, followed by additional 1 h in SDS equilibration solution containing 2.5% (w/v) iodo-acetamide. SDS- PAGE (Ettan Dalt Six Gel Electrophoresis Unit) was performed at 14 ^{o}C and 5 mA/gel for 45 min and then in 20 mA/gel. Later, the gels were scanned using Typhoon 9410 scanner (GE healthcare) and the images were analyzed using De Cyder 2-D differential analysis software (GE healthcare). After analysis, the gels were stained overnight with silver stain (0.1% $AgNO_3$ in H_2O: 0.1 g $AgNO_3$ dissolved in 100 ml H_2O) (Morrissey, 1981), later developed by using 3% sodium carbonate (Na_2CO_3) in H_2O and added 50 µl formaldehyde/100 ml, and stirred for 2 min.

Gels were incubated in 50% Methanol for 15 min, later 5% Methanol was added and kept for incubation for 15 min at room temp. Later, 32 µM DTT (8 µl 1M DTT /250 ml water) solution was added and incubated for 15 min, the solution was discarded and washed the gels twice with double distilled water, followed by with a little quantity of silver solution, discarded the solution and added rest of the silver solution and incubated for 15 min. Later, the gels were washed twice for 5 to 10 sec with water and kept for incubation with developing solution until the bands appeared with the desired intensity. Citric acid was sprinkled into the solution containing the gel to stop staining, washed the gels thrice with water for de-staining the silver stained gels (Switzer *et al.*,1979).

Reagents:

Solution A:

To 0.633M sodium chloride (FW 58.44: 37 g/L), 0.231M cupric sulfate (FW 159.6: 37 g/L) concentrated aqueous ammonium was added until the precipitate was completely dissolved to give a deep blue solution and the volume was adjusted to 1 litre with water.

Solution B:

A quantity of 1.75 M sodium thio-sulfate pentahydrate (FW 248.18: 4.36 g/10 ml)

Stop solution:

Acetic acid (10%)

Procedure

The Gels were de-stained by freshly prepared de-staining solution (A and B solutions in 1:1), 300 µl of 10% acetic acid was added to stop the reaction, incubated for 5 min, the de-staining solution was discarded and the gels were washed twice with water. Three biological replicate samples of gels for each sample were maintained and were scanned by UMAX power look 2100 image scanner (UMAX Systems GmbH, Willich, Germany) at 300 dpi resolution. Image analysis was performed by using PD Quest 8.0.1 Software (Bio-Rad Laboratories, Hercules, CA). After automated detection, matching and normalization of gels further editing was performed manually to prevent the occurrence of discrepancy during spot selection. Comparative "analysis sets" based on statistical, quantitative and qualitative analysis was developed between the control and the stressed samples for three independent biological replicates. In the statistical sets, the Student's 't' test was performed at 95% level of significance. In the quantitative sets, the upper and lower limits were set to 1.5 and 0.66 respectively. Then, Boolean analysis sets were created between the statistical, quantitative and qualitative sets. Only the spots displaying reproducible change patterns were considered to be differentially expressed proteins for further mass identification.

D. In- gel trypsin digestion and matrix-assisted laser desorption/ionization- tandem time of flight (MALDI-TOF/TOF) analysis:

The manually excised differentially expressed protein spots from the gels, were subjected to in-gel trypsin digestion (Granvogl, B., 2007). The excised gel spots were de-stained twice with 100 mM NH_4HCO_3 in 30% aceto-nitrile (ACN) and were digested overnight

with 20µl of 50 mM NH_4HCO_3, containing 0.01 mg/ml sequencing – grade modified trypsin (Promega, Madison, WI, USA), at 37°C for 16 h. The peptides were extracted thrice with 0.1% trifluoroacetic acid (TFA) in 60% ACN, pooled together and lyophilized. Later, the peptides were dissolved in 5 mg/ml cyano-4-hydroxycinnamic acid consisting of 50% ACN and 0.1% TFA.

The analysis of tryptic digested peptides was performed on 480 plus MALDI TOF/ TOF Analyzer (Applied Biosystems, USA) and their mass maps were generated in positive ion reflector mode (2kV accelerating voltage) with 355 laser shots per spectrum. To define the mass peaks, the minimal criterion of signal to noise ratio of 50 and a scan area of 800 Da to 4000 Da of peptide mass fingerprint (PMF) were selected for MS/MS- positive ion mode operation, the 10 most intense ions with 2kV collision energy were selected. Using the individual PMF spectra, peptides exceeding a signal- to- noise of 20 and passed through a mass exclusion filter were subjected to fragmentation analysis. The parameters for peak matching were as follows: Min S/N: 20; Mass tolerance: 0.2Da; minimum peaks to match reference masses, 4; and maximum outlier error, 100 ppm. Total shots for each MS spectrum were 2000, whereas 3000 shots were used for MS/MS. The data were calibrated using the ABI-4700 calibration mixture (Applied Biosystems, CA). MS and the MS/MS spectra were searched against the NCBI database (*Viridiplantae*, 20140911) using the software GPS Explorer (Applied Biosystems) and MASCOT version 2.1 (Matrix Science, London, UK) with the following parameters: National Center for Biotechnology Information (NCBI), non-redundant protein database (released data Sep.11, 2014; including 45166402 proteins), and species restriction to *Viridiplantae* (green plants). The other parameters were as follows: trypsin cleavage; one missed cleavages allowed; peptide mass tolerance set to ± 0.4 Da. Credible results for the MALDI-TOF/TOF MS were the hits with high protein scores, similar molecular mass (*Mr*) and isoelectric point (pI) as experimental *Mr* and pI and protein score confidence interval (C.I.p%) of 95%. The criterion for identifying a protein as significant up- or down regulated was 1.5- fold threshold change in relative fluorescence signal intensity by pair wise comparison of the dye signal from the three co- electrophoresed samples. Protein MALDI-TOF/TOF MS results were a score above the identification threshold with significant value as specified in mascot page were considered significant for accuracy of identification.

Results

Objective – 2:

Identification of differentially expressed proteins under drought stress by 2D- DIGE:

In the present investigation, physiological, proteomic and genomic aspects of drought stress response were studied on a drought tolerant groundnut (*Arachis hypogaea* L.) accession, ICGV 91114, obtained from the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), at the Dept. of Genetics, Osmania University.

A. Physiological parameters:

Relative Water Content (RWC), total chlorophyll content, including chlorophyll '*a*' and chlorophyll '*b',* besides RNA content were estimated in the 40-day-old drought stressed seedlings with 100ml of different concentrations of 5, 10, 15 and 20% PEG 6000 while maintaining the controls.

In both the physiological parameters; Relative water content (RWC) and the chlorophyll content, besides the content of RNA in the drought stressed leaf and root tissues, was proportionately decreased with increase in the concentration of PEG 6000 compared to their controls.

A decrease in RWC was recorded in stressed leaf; 5% = 0.64, 10%= 0.52, 15%= 0.40 and 20%= 0.35 compared to control= 0.831, while in the root tissue the decrease in 5%= 0.59, 10%= 0.43, 15%=0.35 and 20%=0.093 compared to control= 0.88 (Table- 1 and Fig. 1).

A similar trend of decrease in the content of total chlorophyll with increase in the concentration of PEG was observed; 5%=1.81mg, 10%=1.67 mg, 15%= 1.39 mg and 20%= 1.04 mg compared to control=2.07 mg (Table- 1 and Fig. 2)

Similarly, results indicate a proportionate decrease in the content of RNA with the increase in the concentration of PEG treated leaf tissue was recorded; 5% = 2.67 µg/µl, 10 % = 2.05 µg/µl, 15%= 1.99 µg/µl and 20%= 1.56 µg/µl compared to the control= 2.87 µg/µl. (Table 3& Fig. 3). Similarly, a proportionate decrease in the content of RNA was observed with the increase in the concentration of PEG treated root tissue; 5% =2.08 µg/µl, 10 % =1.98 µg/µl, 15%=0.99 µg/µl and 20% = 0.58 µg/µl compared to the control= 2.56 µg/µl (Table- 1 and Fig.3).

Table- 1: Content of RWC, Chlorophyll and RNA in PEG stressed Leaf and Root

S. No.	Conc. of PEG (%)	RWC (%)		Chlorophyll content (mg/g)			RNA content (µg/µL)			
		Leaf	Root	'a'	'b'	Total	OD value		Leaf	Root
							Leaf	Root		
1.	Control	0.831	0. 879	1.60	0.47	2.07	1.85	1.88	2.87±0.005	2.56 ± 0.025
2.	5	0.640	0. 586	1.37	0.44	1.81	1.86	1.81	2.67±0.01	2.08 ±0.010
3.	10	0.523	0. 426	1.28	0.39	1.67	1.89	1.86	2.05±0.01	1.98 ±0.010
4.	15	0.400	0. 348	1.1	0.29	1.39	1.89	1.89	1.99±0.01	0.99 ±0.005
5.	20	0.353	0.093	0.83	0.21	1.04	1.87	1.83	1.56±0.01	0.58 ±0.005

(OD values recorded at A 260/A 280)

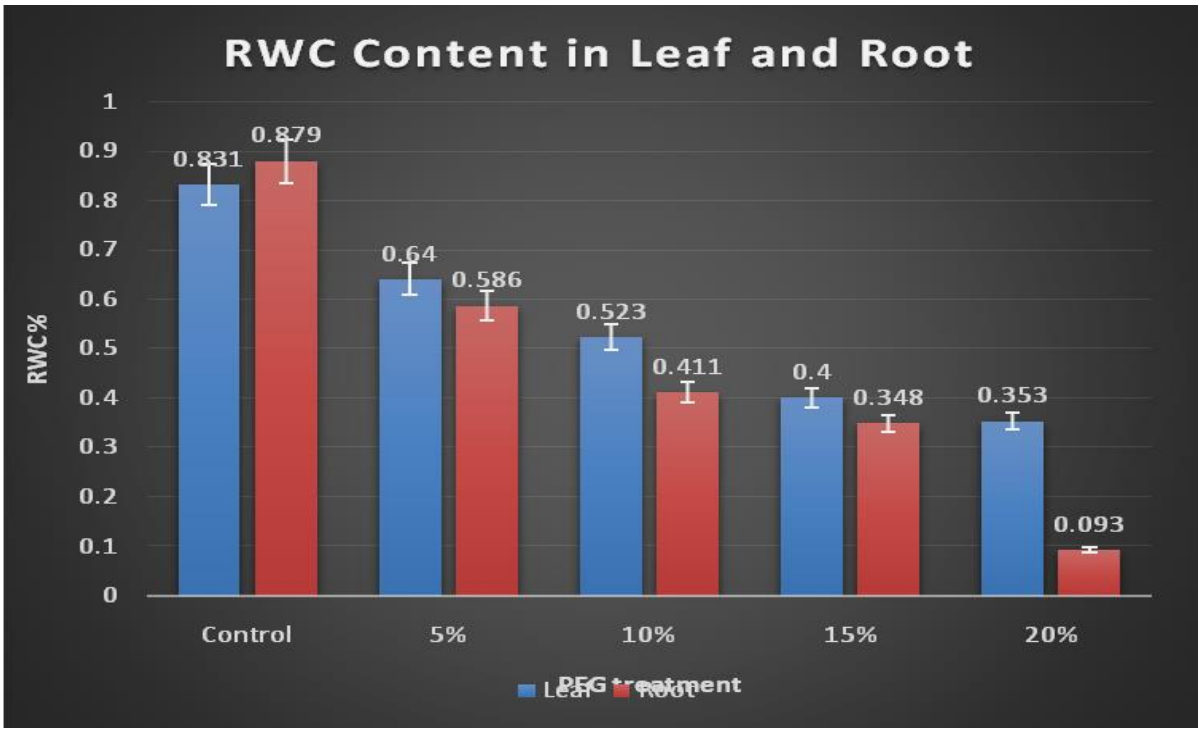


Fig.1: RWC in PEG stressed Leaf and Root tissues

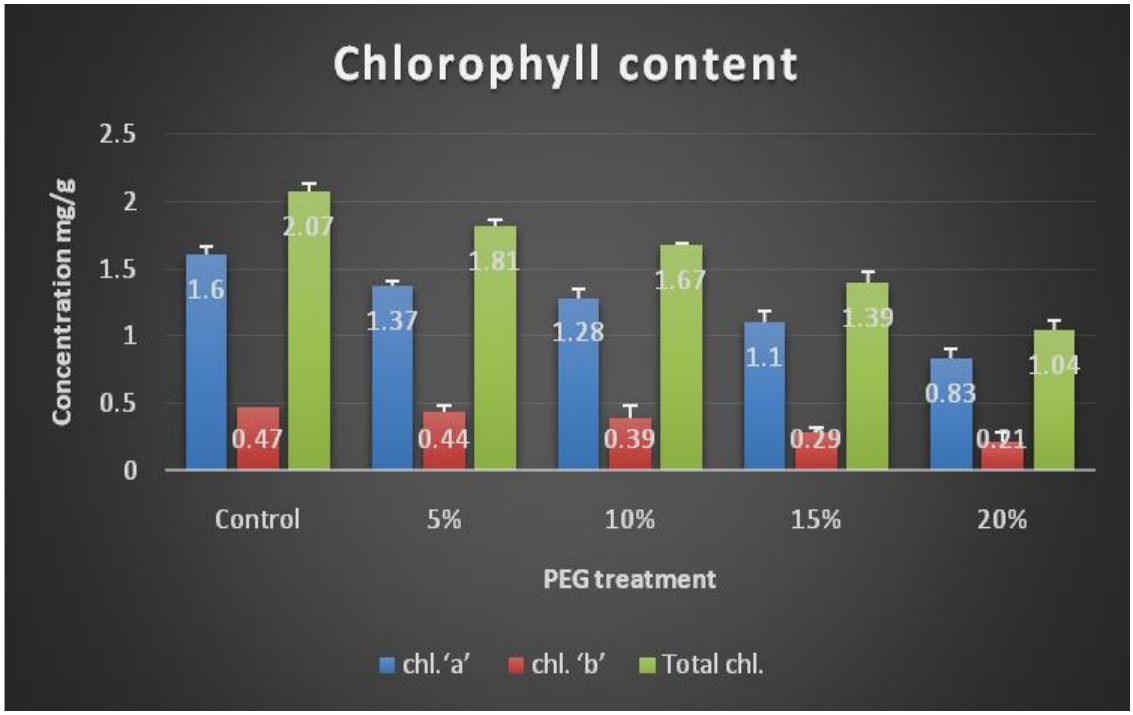


Fig.2: Content of 'a', 'b' and total Chlorophyll in PEG stressed Leaf

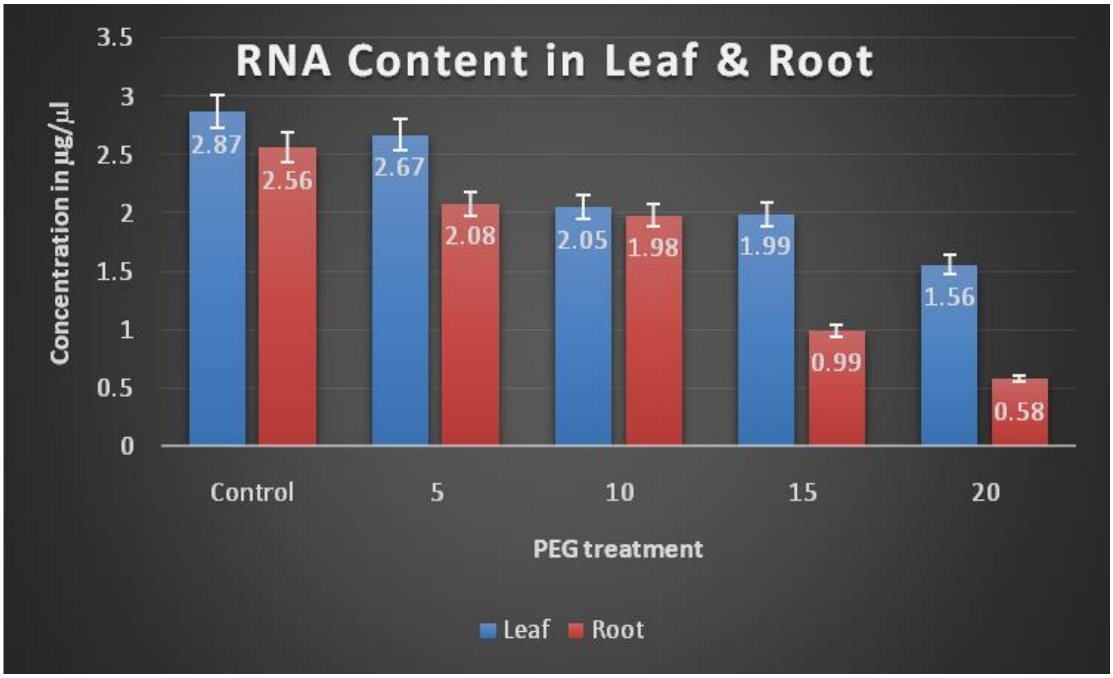


Fig.3: Content of RNA in PEG stressed Leaf and Root tissues

Proteomic analysis

The proteomic analysis for drought stress response was carried out by 2 methods: 1) 2-dimensional differential In- gel electrophoresis (2D-DIGE) method, and 2) Ultra-high performance liquid chromatography coupled to a Quadrupole QExactive™ Orbitrap Mass Spectrometer (MS) for identification of drought stress responsive proteins.

B. 2D-DIGE method:

After purification of isolated proteins from the leaf and root tissues were separately labelled with fluorescent dyes; cy2, cy3 and cy5 (Figs.4.1, 4.2, 4.3, 4.4, 5.1, 5.2, 5.3 and 5.4). After electrophoresis, the gels consisting of labelled proteins were subjected to PD Quest software to know the number of protein spots on the gel, their match Id, pI, MW, intensity, area, volume and percentage of the volume occupied by the protein spots. Out of 1022 protein spots identified from the leaf and 750 spots from the root tissues, 12 from leaf and 20 from root tissues were selected based on their intensity (Tables- 4 and 5). These selected protein spots were subjected for In-gel trypsin digestion to obtain peptides of different lengths and later subjected to MALDI TOF/TOF analysis to know their sequence, and molecular mass, and thereafter subjected to MASCOT search by blasting them against *viridiplantae* in a public domain, the NCBI.

In control leaf tissues, analyzed on the basis of their match ID, pI, MW, intensity, area, volume and volume%, 27 of the 169 protein spots were upregulated, while of a total of 1022 protein spots in the stressed samples; 24 of the 240 spots in 5%, 19 of the 192 spots in 10%, 22 of the 226 spots in 15%, and 23 of the 195 spots in 20%, exhibited upregulation.

A total of 750 root protein spots were analyzed based on their match ID, pI, MW, Intensity, Area, Volume, Volume% detected. In control 25 of the 146 spots showed upregulation, while 27 of the 176 spots in 5%, 27 of the 148 spots in 10%, 25 of the 152 spots in 15%, and 33 of the 128 in 20% stressed samples, exhibited upregulation.

Only 12 protein spots from the leaf tissue (Table-2) and 20 protein spots from the root tissues with 3- dimensional peaks (Table-3) were digested and analyzed.

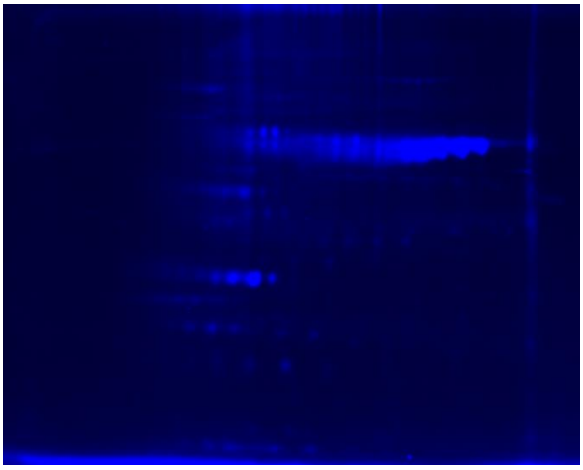


Fig. 4.1: cy2 labelled Leaf protein bands on 2D-DIGE gel (Control+Internal standard)

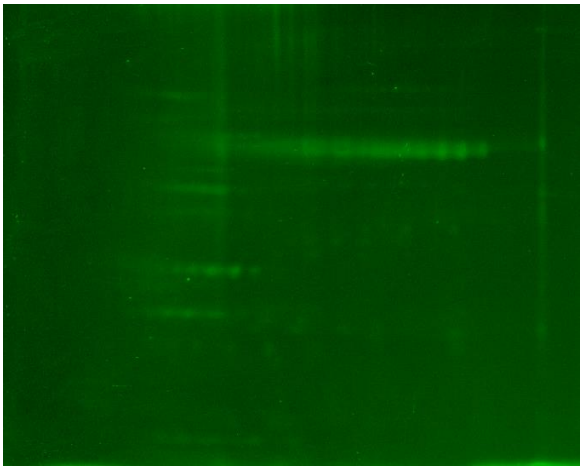


Fig. 4.2: cy3 labelled Leaf protein bands on 2D-DIGE gel (5% + 10% PEG)

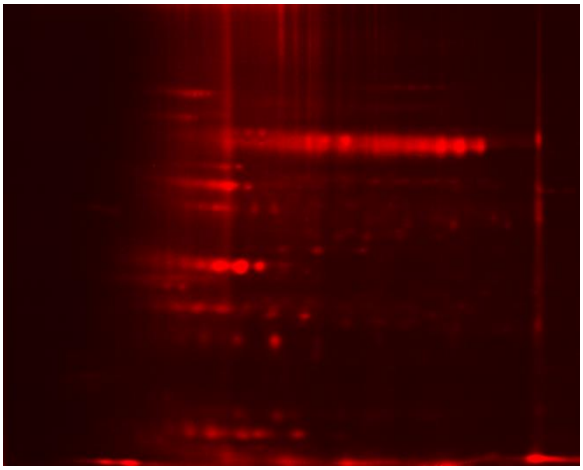


Fig. 4.3:cy5 labelled Leaf protein bands on 2D-DIGE gel(15% + 20% PEG)

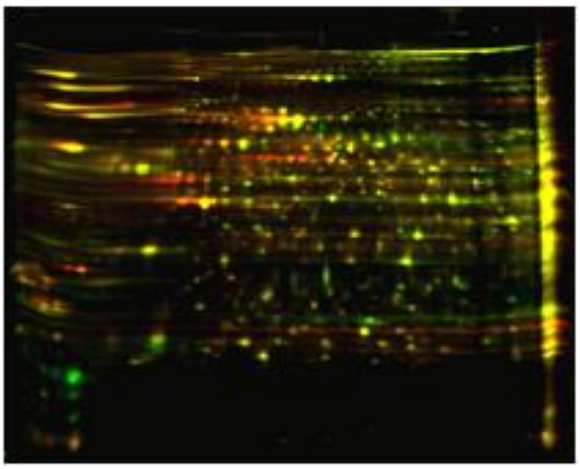


Fig.4.4: cy2, cy3 and cy5 labelled Leaf protein bands on 2D-DIGE overlaid gels (Internal standard + Control, 5+10 and15 + 20% PEG)

Fig. 4: Cy labelled Leaf proteins on 2D-DIGE gels

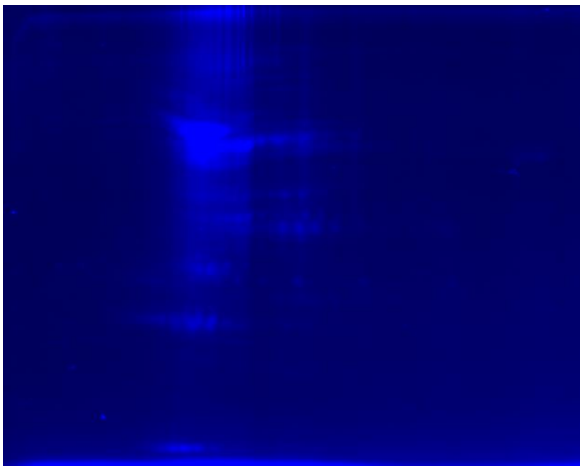


Fig. 5.1: cy2 labelled Root protein bands on 2D-DIGE gel (Control + Internal standard)

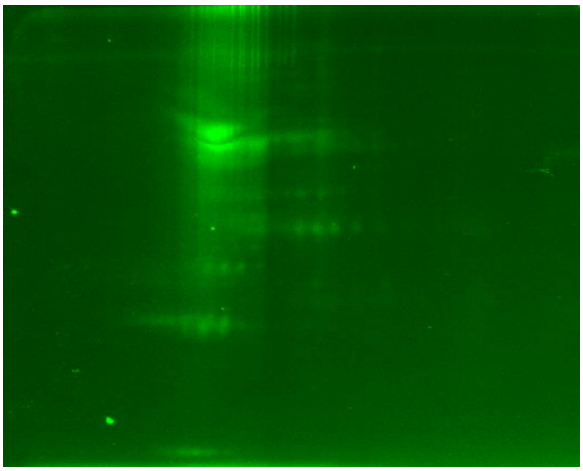


Fig. 5.2:cy3 labelled Root protein bands on 2D-DIGE gel (5% + 10% PEG)

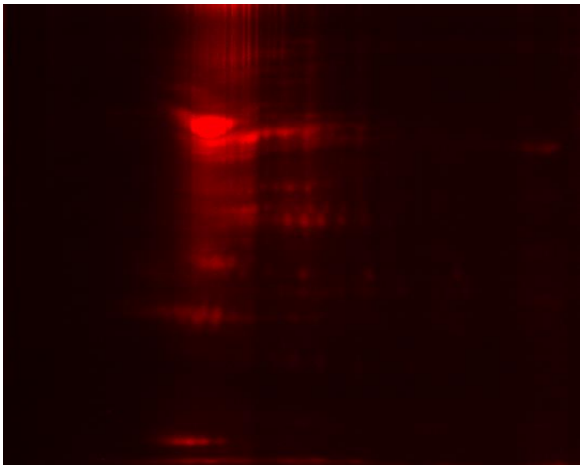


Fig. 5.3:cy5 labelled Root protein bands on 2D-DIGE gel (15% + 20% PEG)

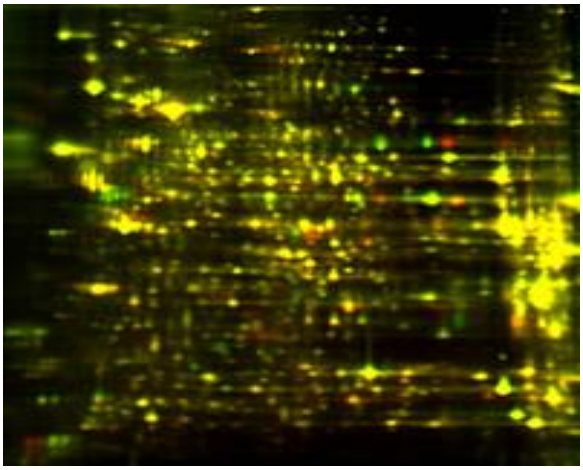


Fig.5.4: cy2, cy3 and cy5 labelled Root protein bands on 2D-DIGE overlaid gels (Internal standard + Control, 5+10 and 15 + 20% PEG)

Fig. 5: Cy labelled Root proteins on 2D-DIGE gels

1. Leaf proteomic analysis by 2D-DIGE:

The Proteins; glyceraldehyde-3 -phosphate dehydrogenase with match id 22 and the nucleoside diphosphate kinase with match id 2196 respectively exhibited an increase of 2.5 and 3.1 folds in their expression in 5% PEG stressed tissue. While the proteins, ferredoxin NADP reductase 9 with match id 2352, plastid-lipid-associated protein with match id 2382 and CARUB-v100157 with match id 2427 respectively exhibited a decrease of 0.507, 0.363, and 0.660 folds in their expression in 10% PEG stressed tissue. The proteins, oxygen-evolving enhancer with match id 21 and ceramidase with match id 2641 exhibited a decrease of 0.450 and 0.707 fold in their expression in 15% PEG stressed tissue. Proteins; Cell division Control 6 with match id 2057 and AtpA, with match id 2645 have exhibited a decrease of 0.944 and 0.318 fold decrease in 15% PEG treated sample. The proteins; catalase 2 with match id 59 and phosphatase PHPSI with match id 64 showed an increase of 2.23 and 3.10 fold in their expression in 15% PEG stressed tissues. While, the fructose 1, 6 bi-phosphate with match id 34 exhibited a decrease by 0.49 fold expression in 20% PEG stressed tissues (Table-2).

2. Root proteomic analysis by 2D-DIGE:

The proteins; methyl- transferase with match id 33 and SORBIDRAFT-01g020810 with match id 193 respectively, showed an increase of 2.02 and 1.1 fold, while a decrease of 0.73 fold was recorded in DNA-directed RNA polymerase with match id 188 in 5% PEG stressed root tissues. The proteins; catalase-2 and elongation factor- Tu with match ids of 48 and 145 respectively have exhibited 3.7 and 0.3 fold increase, whereas the proteins; phosphatase PHS1, Phospho-2-dehydro-3-deoxyheptonate aldolase-1 and protein JCGZ-15912 respectively with match ids 45, 136 and 73 have shown 0.470, 0.944 fold and 0.470 fold decrease respectively in 10% PEG stressed tissues. The 6 proteins; elongation factor- Tu, nucleoside di- phosphate kinase, DNA- directed RNA polymerase III, elongation factor- Tu, choline kinase and LOC104886667 with match ids 62, 36, 47, 180, 40 and 70 respectively have shown 2.31, 2.00, 4.64, 0.99, 2.23 and 2.30 folds increase in 15% PEG treated tissues. The proteins; BnaCnng08330D, 30S ribosomal protein S2, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and retrotransposon with match ids 95, 17, 32 and 50 have exhibited an increase of 1.16, 3.27, 2.30 and 1.30 folds, while the proteins M569-03092 and LOC101245158 with ids 0 and 168 showed a decrease of 0.44 and 0.71 folds respectively in 20% PEG treated tissues (Table-3).

Table- 2: Expression levels of 12 identified leaf protein spots by 2D-DIGE

PEG Conc	Match Id.	Expression level: Increase/ decrease	Protein Accession No.	No. of peptides matched	Peptide sequence matched with reference protein	Score	Coverage	MW Da.	Sequence homology to identified protein	Reference organism
5%	22	2.50 fold increase	gi|512374222 GenBank: BAN45710.1 LOCUSBAN45710	14	K.IGINGFGR.I.K.YDTVHGAWK.H K.DAPMFVVGVNEK.E + Oxidation (M)K.EYKPDIHILSNASCTTNCLAPL AKVINDR.FK.VINDRFGIVEGLMTT VHSITATQK.T.K.VINDRFGIVEGLM TTVHSITATQK.T + Oxidation (M) K.TVDGPSMK.D.K.TVDGPSMKDWR .G + Oxidation (M). R.AASFNIIPSSTGAAK.A. R.VPTVDVSVVDLTVR.L. K.AAIKEESEGK.L.K.LKGILGYTEDD VVSTDFIGDSR.S.K.GILGYTEDDVV STDFIGDSR.S.K.LVSWYDNEWGYS TR.V	100	48%	36462	Glycelaldehyde-3-phosphate dehydrogenase	*Pyrus pyrifolia var. culta*
	2196	vol % 3.1 fold increase	gi|71040669 GenBank: AAZ20283.1 LOCUS AAZ20283	7	R.GLVGEIISR.F.K.LITVDRPFAER.H. R.KLIGATNPLASEPGTIR.G. K.LIGATNPLASEPGTIR.G. K.LIGATNPLASEPGTIRGDFAIDIGR NR.GDFAIDIGR.N. R.GDFAIDIGRNVIHGSDSVESATK.E	104	40%	16429	Nucleoside diphosphate kinase I	*Arachis hypogaea*
10%	2382	vol% 0.507	gi|568826008 NCBI Reference Sequence: XP_006467367.1LOCUS XP_006467367	7	R.IEIGFCVFLVPLLVK.T. K.KALVDSFYGTDR.G. K.ALVDSFYGTDR.G R.FAGPLATTSISTNAK.F R.FAGPLATTSISTNAKFEVR.S R.VQIKFEEGIIGTPQVTDSLVIPENV EFLGQK.I.K.FEEGIIGTPQVTDSLVIP ENVEFLGQK.I	111	19%	43439	Plastid-lipid	5 10%

	2352	vol% 0.363	gi|323714331 PDB: 2XNC_A LOCUS 2XNC_A	13	R.AQVTTEAPAK.V.K.FKPKEPYVGR .C.K.ITGDDAPGETWHMVFSTEGEV PY.R.ER.LYSIASSAIGDFGDSKTVSL CVK.RK.RVPDGVCSNFLCDLKPGSE VK.IR.VPDGVCSNFLCDLKPGSEVKI TGPVGK.E.K.APENFRLDFAVSR.E.K .MYIQTR.M.R.MAQYAEELWELLK.K + Oxidation (M).K.KDNTFVYMCGLK.GK.DNTFV YMCGLKGMEK.G.K.DGIDWIEYKR. T.K.KAEQWNVEVYW.	79	52%	35783	Ferredoxin Nadp R9eductase (Fnr)	*From Pisum Sativum*
	2427	vol %0.66	gi|565483902 NCBI Reference Sequence: XP_006299591.1 LOCUS XP_006299591	6	R.LAHSQMSKDCK.L + Oxidation (M)K.LYAETALERLNLFVEVMMAR. D +Oxidation(M).R.GDFLSVRQDSKSVN R.S.R.QATLTVANSR.L.R.LEEFRFLP LLPWIILAVFMR.H.K.LVFPWENITH DLR.A.	78	11%	86980	CARUB_v100 15769mg	*Capsella rubella*
15%	59	2.23 flold increase	gi|332661063 GenBank: AEE86463.1 LOCUS AEE86463		K.TWPEDILPLQPVGR.M R.LGPNYLQLPVNAPK.C R.LGPNYLQLPVNAPK.C R.EGNFDLVGNNFPVFFIR.D	26	27%	54995	Catalase	**Arabidopsis thaliana**
	64	3.1 fold increase	gi|514772549 NCBI Reference Sequence: XP_004967523.1 LOCUS XP_004967523	9	R.LGYEFAR.L R.LLGVQTPQAR.V K.SSSPNVDR.LR.GSFRAALR.D R.NLRMTMK.L R.DFYKNPK.V R.LEHIIER.I R.AQPNDGFAK.A K.ALLALDKK.L	79	7%	103677	Phosphatase PHS1	*Setaria italica*
	2641	vol %0.707	gi|743929815 NCBIReferenceSequence: XP_011009147.1LOCUS XP_011009147	7	R.ARAFIVAEPQGSR.V. R.AFIVAEPQGSR.V. K.EAMEVAASFK.S+Oxidation (M). K.SSQGQPATR.Y.K.MVLTSGASK.E.	78	6%	86199	Ceramidase	*Populus euphratica*

					K.MVLTSGASK.E + Oxidation (M).R.IPQSAVSGVYR.I.					
	2645	vol %0.318	gi|408899391 GenBank: AFU94521.1 LOCUS AFU94521	16	R.EVKIVNTGTVLQVGDGIAR.I. K.IVNTGTVLQVGDGIAR.I.R.IAQIPV SEAYLGR.V .R.LIESPAPGIISR.R. R.LIESPAPGIISRR.S.R.SVYEPLQTGL IAIDSMIPIGR.G.R.SVYEPLQTGLIAI DSMIPIGR.G + Oxidation (M)K.ASSVAQVVTTLQER.GR.HTLII YDDLSKQAQAYR.Q.R.QMSLLLR.R. R.RPPGREAYPGDVFYLHSR.L R.EAYPGDVFYLHSR.L R.KFLVELR.TK.FLVELR.T K.TNKPQFQEIISSTK.I. K.IFTEEAEALLKESIQEQMER.F + Oxidation (M)	141	32%	55085	AtpA	*Erythroxylum areolatum*
	21	0.45 fold decrease	gi|702462073 NCBI Reference Sequence: XP_010028333.1 LOCUS XP_010028333	16	M.AASLQAAATLMQPTKVGR.I K.RLTYDEIQSK.T. K.GTGTANQCPTIDGGLDSFAFKPGK YNAK.KR.LTYTLDEIEGPFEVSPDGT VK.F K.FEEKDGIDYAAVTVQLPGGER.VK .DGIDYAAVTVQLPGGER.VK.DGID YAAVTVQLPGGERVPFLFTIK.QK.D GIDYAAVTVQLPGGERVPFLFTIK.Q R.VPFLFTIK.Q K.QLVASGKPESFSGEFLVPSYR.GR. GSSFLDPKGR.GR.GGSTGYDNAVAL PAGGRGDEEDLTK.EK.ENIKNASSS TGK.IK.NASSSTGKITLSVTK.T. K.ITLSVTKTKPETGEVIGVFESVQPS DTDLGAK.VK.DVKIQGVWYAQLES. -	112	66%	35204	Oxygen-evolving enhancer 1	*Eucalyptus grandis*

20%	34	0.49 fold decrease	gi|116786456 GenBank: ABK24111.1 LOCUS ABK24111	9	K.DAPGTHEFLLMDDGKWHHVK.EK .ETTEIGEGKLFSPGNLR.A K.LFSPGNLR.AK.LFSPGNLRATFDN PEYEK.L R.YTGGMVPDVNQIIVKEK.G + Oxidation (M)K.LRLLFEVAPLGMLVEK.A + Oxidation (M) R.TQVAYGSKNEIIR.F K.NEIIRFEETLYGSSR.L R.LNVPVGAGVKA.-	80	27%	45310	Fructose-1,6-bisphosphatase (FBPase) class 1	*Picea sitchensis*
	2057	vol % 0.944	gi|747053361 NCBI Reference Sequence: XP_011072840.1 LOCUS XP_011072840	13	R.SDNGLTESPMRK.S + Oxidation (M)R.SAQEQKLSENLLEKPIWDPR.D K.ILGQSHPQKKPDR.S. K.DRAVLHDLFMLTTMPFSK.C + Oxidation (M) K.LQSLNCKPMVITFRAYSK.D + Oxidation (M)R.AYSKDQIITILQER.L K.DQIITILQERLR.E. R.SAIEMLEAER.R + Oxidation (M) R.SAIEMLEAERR.D + Oxidation (M)R.IEDQQKPAACGTVKNQLNNM VR.I K.NQLNNMVRIDHVAAALSK.T + Oxidation (M) R.VLDDQGILKLGQSR.E R.VALKVDGADIAFALQGIR.F	82	33%	57772	Cell division control protein 6	*Sesamum indicum*

Table- 3: Expression levels of 20 identified protein spots in Roots by 2D-DIGE

PEG Conc	Match Id.	Expression Fold Increase / decrease	Accession No.	No. of peptides matched	Peptide sequence matched with reference protein	Score	Coverage	MW Da.	Sequence homology to identified protein	Reference organism
5%	33	2.02 fold Increase	gi|702367154 NCBI Reference Sequence: XP_010060917.1 LOCUS XP_010060917	10	MGGDGSEPRR.SR.ALARFSISPSAYYSR.F K.AHEALVGATELRSLMLK.MK.AHEALVGATELRSLMLKMK.G R.SLMLKMK.G + 2 Oxidation (M) K.MKGEVATAEDRK.L + Oxidation (M) R.KLAPPGVPPSFIELGR.L - K.VIPIFNNLVHNGTK.D. R.ESCFYMSDLQQIHR.M K.SRYPTLPNK.Y R.YPTLPNKYFLSLPVQQLAHR.E	79	37%	35805	Methyltransferase- 2 isoform X2	*Eucalyptus grandis*
	193	vol %1.190	gi|242039543 NCBI Reference Sequence: XP_002467166.1 LOCUS XP_002467166	9	R.AMRDLAQEHGPLMMLR.L R.AMRDLAQEHGPLMMLR.L + 3 Oxidation (M). R.DLAQEHGPLMMLR.L + 2 Oxidation (M) R.IREEEVAR.F. R.GEKTAHESLIGVLLR.L R.MHCPLPLLLPR.Q R.QCRETCQVMGYDIPK.G K.YWEDAEEFRPERFENTNLDYK.G K.KTGLILHPVTCIAPADA.-	104	20%	56815	SORBIDRAFT_01g020810	*Sorghum bicolor*

	188	vol %0.73	RPOC1_ANEMR UniProtKB/Swiss-Prot: B0YPL9.1 LOCUS RPOC1_ANEMR	7	R.YRMGYIELACPVTHVWYLK.R K.RLPSCIANLLDKPLK.E R.VIYRNNTLLDFLAR.S. R.SDSTPR.G. R.LVQEAVDALIDNGIRGQPMR.D+ Oxidation (M). K.ETYLK.NR.ID. EAIQGTCQASSRQTMLYIGNK.E	59	14%	81805	DNA-directed RNA polymerase subunit beta' OS= *GN=rpoC1 PE=3 SV=1*	*Aneura mirabilis*
10%	48	3.7fold increase	gi|17865693 UniProtKB/Swiss-Prot: P25819.3 LOCUS CATA2_ARATH		R.FSTVIHER.G. R.LGPNYLQLPVNAPK.C R.EGNFDLVGNNFPVFFIR.D R.LGPNYLQLPVNAPK.C R.EGNFDLVGNNFPVFFIR.D P.GVQTPVIVR.F R.GPILLEDYHLVEK.L	80	20%	56931	Catalase	**Gossypium raimondii**
	73	0.47fold decrease	gi|643734835 GenBank: KDP41505.1 LOCUS KDP41505	13	K.LSPGEWLAPPSEILPER.S K.LSPGEWLAPPSEILPERSK.LR.AC QTYFTEISTGK.E K.SEGELGLIGILVWLR.F R.IRDEILVIQR.N R.NNGCKTGMMEEWHQK.L K.TLNANGLMREK.L K.SLKAVHSGADFESAIETCMGGN LSLK.L + Oxidation R.IELHPALLTSNGR.A R.AGSAATLSMLINR.F + Oxidation (M) K.SMDLVISDVSGSNSLLNPSVSTTI PR.A + Oxidation(M) K.ANLNHDNIR.M K.FPIVTGYPSK.N R.LYGCSQNIEGVVK.D	78	15%	143516	JCGZ_15912	*Jatropha curcas*

	145 vol % 0.363	vol % 0.363	EFTU_PROWI UniProtKB/Swiss-Prot: Q9TJQ8.1 LOCUS EFTU_PROWI	4	K.KPHVNIGTIGHVDHGK.T K.NMITGAAQMDGAILVVSGADGP MPQTK.E + 2 Oxidation (M)K.IESFCTDAGEPIK.M R.EGGKTVGAGVVGK.I	59	16%	44793	Elongation factor- Tu, plastid OS= *GN=53ufa PE=3 SV=1*	*Prototheca wickerhamii*
	45	0.47 fold decrease	gi|514772549 NCBI Reference Sequence: XP_004967523.1 LOCUS XP_004967523	8	R.LGYEFAR.L R.LLGVQTPQAR.V K.SSSPNVDR.L R.GSFRAALR.D R.NLRMTMK.L R.NLRMTMK.L + Oxidation (M) R.DFYKNPK.V K.ALLALDKK.L	79	5%	103677	phosphatase PHS1	*Setaria italic*
	136	vol % 0.944	AROF_TOBAC UniProtKB/Swiss-Prot: P27608.1 LOCUS AROF_TOBAC	11	K.TEWTVESWK.S R.SLEERLGEAAMGR.A + Oxidation (M) R.AFLLQGGDCAESFKEFNANNIR.D R.ILLQMGAVLMFGGQMPVIKVGR.M + Oxidation (M) K.LPSYRGDNVNGDAFDAK.SR.AY CQSAATLNLLRAFATGGYAAMQR.I R.TRQLDGAHVEFLR.G R.GVANPLGIK.V K.VSDKMDPSALVK.L + Oxidation (M) R.MGAENMRVK.L + Oxidation (M) K.APCGLKTRPFDSIR.A	58	30%	60337	Phospho-2-dehydro-3-deoxyheptonate aldolase 1, OS= *GN=DHAPS-1 PE=1 SV=1*	*Nicotiana tabacum*
15%	36	2.02 Fold increase	gi|71040669 GenBank: AAZ20283.1 LOCUS AAZ20283	7	R.GLVGEIISR.F K.LITVDRPFAER.H R.KLIGATNPLASEPGTIR.G K.LIGATNPLASEPGTIR.G K.LIGATNPLASEPGTIRGDFAIDIG	104	40%	16429	Nucleoside diphosphate kinaseI	*Arachis hypogaea*

					R.N R.GDFAIDIGR.N R.GDFAIDIGRNVIHGSDSVESATK.E					
	47	4.64 fold increase	gi|565393217 NCBI Reference Sequence: XP_006362278.1 LOCUS XP_006362278	29	M.SMDCDSTMSSPSNNLNNGIDK.Q K.IVRANHEIRSSTR.D K.DVVIARMPIMLRSCLCVLYEK.D K.DEEQLAKLGECPLDPGGYFIVNG REK.V K.MEKEKVYLELNMFK.T + Oxidation (M) K.EKVYLELNMFKTK.V + Oxidation (M) R.RVMEAILNK.D R.LSYIGTLGYMTK.I + Oxidation (M) K.SPQRFANGMR.K + Oxidation (M) R.IREFVSIFVNDK.Q R.NTYQCAMGK.Q + Oxidation (M) R.DIYINKESPTVTITQVTSPMGLPD SAYK.S R.VALHSDENNNLSIKFMIRQTR.R R.VALHSDENNNLSIKFMIRQTR.R + Oxidation (M) R.QTRRPEIGDK.F K.FSSRHGQKGVCGIIVQQEDLPFS ER.G R.MTVGKMIELLGSK.A + Oxidation (M) R.MTVGKMIELLGSK.A + 2 Oxidation (M) K.LKHMVLDKMHAR.G + Oxidation (M) K.HMVLDKMHARGSGPR.V + Oxidation (M) K.MHARGSGPR.V	82	28%	131733	DNA-directed RNA polymerase III subunit RPC2	*Solanum tuberosum*

					K.MHARGSGPRVMMTR.Q + Oxidation (M) R.GSGPRVMMTRQPTEGR.S + 2 Oxidation (M) R.VMMTRQPTEGR.S + Oxidation (M) R.VMMTRQPTEGR.S + 2 Oxidation (M) R.SRNGGLRVGEMER.D + Oxidation (M) K.IGICSMCKNGENMSTLKLPYACK.L + 2 Oxidation(M) K.LLFQELQSMNIVPR.L K.LLFQELQSMNIVPRLKLTEA.-					
	180	vol % 0.944	gi|449440632 NCBI Reference Sequence: XP_004138088.1 LOCUS XP_004138088	16	M.AAISLASVSTSSK.L R.KKPHVNIGTIGHVDHGK.T K.TTLTAALTMALSK.A + Oxidation (M) K.KYDEIDAAPEER.A K.YDEIDAAPEER.A K.YDEIDAAPEERAR.G K.NMITGAAQMDGAILVVSGADGPMPQTK.E + 2 Oxidaion(M)K.EHILLAK.Q K.QVGVPNMVVFLNKK.D K.QVGVPNMVVFLNKK.D + Oxidation (M) K.IFELMDAVDSYIPIPER.Q + Oxidation (M) K.ILDEALAGDNVGLLLR.G K.ILDEALAGDNVGLLLRGVQK.A R.GMVLAKPGTITPHTKFSAVVYVLK.K + Oxidation (M) K.DEESKMVMPGDR.V + Oxidation (M) K.MVVELIMPVACEQGMR.F + Oxidation (M)	92	40%	51886	Elongation factor Tu	*Cucumis sativus*

	40	2.23 fold increase	gi|729434504 NCBI Reference Sequence: XP_010520093.1 LOCUS XP_010520093	13	K.ALQVIPLK.G K.GAMTNEVFQIKWPTR.A K.WPTRAGGPSR.K R.AGGPSRK.V R.IAAKMK.E + Oxidation (M) K.MKEFHNLEMSGEKK.A K.EFHNLEMSGEK.K K.KAWVWDRLR.N K.AWVWDR.L R.LASPVEAKSFGLDVIEEEIMLLQ K.Q K.SFGLDVIEEEIMLLQK.Q R.HFLQIYLSSSGNYFFPVVENFR.Q R.FQQYWLTKGR.L	81	30%	42965	Choline kinase 2 isoform X1	*Tarenaya hassleriana*
	70	vol % 2.63	gi|731317904 NCBI Reference Sequence: XP_010669435.1 LOCUS XP_010669435	8	.MWAASKINDASAIVSK.C + Oxidation (M) R.RLSVHHHTSR.T R.LSVHHHTSR.T R.NLQMLFVEGNTTLNIPIGALLK.L R.LYGNPANMLPK.K + Oxidation (M) R.LYGNPANMLPKK.Q + Oxidation (M) R.VLSSIYPDKDFESAVMK.G + Oxidation (M) K.SSTWSEDALNMLGKFPNLHILK.L	82	26%	43729	LOC104886667	*Beta vulgaris subsp. Vulgaris*
	62	2.31 fold increase	gi|659129067 NCBI Reference Sequence: XP_008464512.1 LOCUS XP_008464512	13	K.KPHVNIGTIGHVDHGK.T K.TTLTAALTMALSK.A K.TTLTAALTMALSK.A + Oxidation (M) K.KYDEIDAAPEER.A K.YDEIDAAPEER.A R.ARGITINTATVEYETESR.H K.NMITGAAQMDGAILVVSGADGP	86	34%	52094	Elongation factor Tu	*Cucumis melo*

					MPQTK.E + 2 Oxidation (M) K.EHILLAK.Q K.QVGVPNMVVFLNKK.D K.QVGVPNMVVFLNKK.D + Oxidation (M) K.IFELMDAVDSYIPIPER.Q + Oxidation (M) K.ILDEALAGDNVGLLLR.G R.GMVLAKPGTITPHTKFSAVVYVLK.K + Oxidation (M)					
20%	17	3.27fold increase	RR2_ACOAM UniProtKB/Swiss-Prot: A9LYH4.1 LOCUS RR2_ACOAM	4	R.MAPYISAKR.K + Oxidation (M)K.KAADSVASAAIR.A K.KWLGGMSTNWSTTETR.L K.WLGGMSTNWSTTETRLQNFR.D	58	17%	26799	30S ribosomal protein S2, OS=*GN=rps 2 PE=3 SV=1*	*Acorus americanus*
	32	2.3fold increase	gi|512374222 GenBank: BAN45710.1 LOCUS BAN45710	14	K.IGINGFGR.I K.YDTVHGAWK.H K.DAPMFVVGVNEK.E + Oxidation (M) K.EYKPDIHILSNASCTTNCLAPLAKVINDR.F K.VINDRFGIVEGLMTTVHSITATQK.T K.VINDRFGIVEGLMTTVHSITATQK.T + Oxidation (M) K.TVDGPSMK.D K.TVDGPSMKDWR.G + Oxidation (M) R.AASFNIIPSSTGAAK.A R.VPTVDVSVVDLTVR.L K.AAIKEESEGK.L K.LKGILGYTEDDVVSTDFIGDSR.S K.GILGYTEDDVVSTDFIGDSR.S K.LVSWYDNEWGYSTR.V	100	48%	36462	Glycelaldehyde-3-phosphate dehydrogenase	*Pyrus pyrifolia var. culta*

	50	vol % 1.379	gi|108706137 GenBank: ABF93932.1 LOCUS ABF93932	10	R.VLNLSSVEASLIPQGVLPLCCDPERAK.I K.ILTIMQVVGASR.E K.ILTIMQVVGASRER.A R.SEAEETATAEARR.R R.EAEAARAR.Q K.LDAAQGVLDAAAAR.E K.LDAAQGVLDAAAARER.R R.LASEVGPGMLR.D R.DLARGAVELVLASYQAR.D R.AQVRDAADHIVHSFEGTAPR.L	88	13%	103483	Retrotranspo son protein	*Oryza sativa Japonica Grou]*
	95	vol %1.164	gi|674895213 GenBank: CDY37618.1 LOCUS CDY37618	15	K.QDALQALHDLITSKR.Y K.IMFKYLDLCVDLK.R K.FLWETYRTVLEILR.N R.SVEDIYGLMCMVKK.T + Oxidation (M) K.LFSLQKNFNK.N K.DLQLIASSVVLAALSVPPFDR.SR. SALLSELVSR.G R.IESLSQLVPFFEFSVVEK.I R.AMLYPVPSKASK.L R.QQLEMEREEEQK.R + Oxidation (M) R.MLEFKETFQGEVISR.R K.ETFQGEVISRR.Q R.EEVLKGTDAPPARPAEPAVATA AGPPASGAGK.Y K.GTDAPPARPAEPAVATAAGPPA SGAGK.Y R.WVPGPRGSDRPK.R	84	20%	113162	BnaCnng083 30D	*Brassica napus*

	0	0.44fold decrease	gi|527205963 GenBank: EPS71667.1 LOCUS EPS71667	12	R.IVGGLIAGR.T K.VFVTSFMLSCPKLR.S K.VFVTSFMLSCPKLRSK.K + Oxidation (M) K.LPRGTVHGDFVVGLDMLIQK.Y + Oxidation (M) K.EEQVITIAEQMVVHGIR.M + Oxidation (M) K.IGYEYR.C R.CSKNPSTVVPPEQRIK.S K.SFQYGPQVVPISSAELDAMK.F K.GVKLLGFTDASNIMR.HR.AMKE MNKVAIVR.C R.FYATLDAK.S R.NPDAAIPPVEDTLRK.T	82	17%	97154	M569_03092	*Genlisea aurea*
	168	vol % 0.71	gi|460385033 NCBI Reference Sequence: XP_004238210.1 LOCUS XP_004238210	6	.MAINGR.L + Oxidation (M) K.QMGLGNNIGGQNQYYSLTNNYS LR.FK.KLYYFQTLPSLMGMK.S + 2 Oxidation (M) K.LYYFQTLPSLMGMK.S + Oxidation (M) K.LYYFQTLPSLMGMKSVMQAK.A + 3 Oxidation (M) K.SVMQAK.A + Oxidation (M)	78	19%	28250	LOC1012451 58 isoform X2	*Solanum lycopersicum*

Objective -3:

Profiling of drought induced proteins in ICGV91114 groundnut by LC-MS/MS (QExactive™ Orbitrap):

1. Idenification of drought induced proteins in ICGV91114 groundnut by LC-MS/MS:

Label free proteomic analysis was performed by Q Exactive^{TM} Orbitrap, Mass spectrometry for identification of leaf and root proteins associated with drought tolerance.

The tryptic peptide lysates were fractionated with a C18 reversed-phase column (C18 PepMap 100, 75 μm ID×15 cm, 3 μm particle size; Dionex/LC Packings, Vernon Hills, IL, USA) operating at a flow rate of 300 nL/min. Solvent A was 0.1% formic acid in water, and solvent B was 0.1% formic acid in acetonitrile. Peptides were eluted into the Q Exactive Orbitrap mass spectrometer (Thermo Scientific) directly using a reversed-phase gradient (0-35% solvent B over 40 min) through electrospray ionization. Full MS scans in the m/z range of 400-1800 at a nominal resolution of 70,000 were collected in the OrbiTrap, followed by data-dependent acquisition of MS/MS spectra for the 12 most abundant fragment ions. Employing a 15-sec dynamic exclusion time minimized repeated fragmentation of the same ion. The results of quantitative profiling were further improved by normalizing the calculated peak areas. (Chelius and Bondarenko, 2002 and Bondarenko *et al.,* 2002).

2. MS data analysis using Mascot and ProteoIQ, and pathway analysis:

Raw files from Orbitrap Q-Exactive and LTQ-Orbitrap Velos were searched by MS/MS spectra against the Uniprot viridiplantae protein database using the Mascot algorithm v.2.3.02 provided by Matrix Science, Inc. (Boston, MA, USA). The Uniprot viridiplantae database contained 5,51,385 protein entries (50% forward, 50% reversed). Mascot was performed with the following parameters: trypsin enzyme specificity, two possible missed cleavages, 10 ppm mass tolerance for precursor ions, and 20 mmu mass tolerance for fragment ions. Search parameters specified a differential modification of oxidation on methionine and a static modification of carbamido-methylation (+57.0215 Da) on cysteine. Protein quantification was performed using ProteoIQ software v. 2.3.05 (BioInquire, Bogart, GA, USA) with spectra count data. To provide high confidence on peptide sequence assignment and protein identification, data were filtered with

following stringent criteria: Mascot Ion score >20 for all charge states, 1% peptide false discovery rate (FDR), and 1% protein FDR. Bioinformatic and pathway analysis was performed with the aid of Ingenuity Pathway Analysis and Ingenuity iReport software (Ingenuity Sys, Inc., Redwood City, CA, USA) with a fold change cut off value of 1.

3. Data analysis

Finally the data was subjected to statistical analysis such as Student's 't'-test to determine the significant changes between multiple samples (Wiener *et al.,* 2004; Higgs *et al.,* 2005). Probability values (*P*-values) for the comparative proteomics data from *t*-test were considered significant if less than 0.05. All results are expressed as fold change ± standard deviation or standard error of the mean.

Results

Objective -3:

Profiling of drought induced proteins in ICGV91114 groundnut by LC-MS/MS (QExactive™ Orbitrap):

Identification of drought tolerant proteins by LC-MS/MS:

The LC-MS proteomic analysis resulted in the quantification and identification of 65 leaf and 67 root proteins (Tables- 1 and 2 and Figs. 1 and 2).

1. Leaf proteomic analysis by LC-MS/MS:

Results showed that of the 65 drought stressed proteins, 11 were found to be unnamed proteins, while 54 were the functional non- redundant proteins. Sequence alignments of these proteins showed > 97% nucleotide homologies with that of the corresponding proteins of *Arabidopsis thaliana* expressing under stress (Table- 6). In all the 4 concentrations under PEG stress, of the 54 functional proteins, 32 proteins; 1) At1g07930/T6D22_3, 2) ribulose-1 5-bisphosphate carboxylase/oxygenase activase, 3) glutamine synthetase, 4) 2-hydroxy-acid oxidase GLO2, 5) glyceraldehyde-3-phosphate dehydrogenase of plastid, 6) ATPase alpha subunit, 7) At2g39730, 8) photosystem II 47 kDa protein, 9) putative elongation factor beta-1, 10) Chain C Crystal Structure Analysis Of The Plant Protease Deg1, 11) Ferredoxin--NADP reductase, 12) 3-ketoacyl-CoA thiolase 1, 13) nucleoside diphosphate kinase 1, 14)protein disulfide isomerase-like1-1, 15)arginine-tRNA-protein transferase, 16)AT3G12780, 17)NSD 18) Phophoribulokinase, 19)catalase 2, 20)Ferredoxin-dependent glutamatesynthase1, 21)At1g11860/F12F1_30, 22)At1g42970/F13A11_3, 23)riblose-1,5-bisphosphatecarboxylase/oxygenase, 24)putative RNA helicase SDE3, 25)putative tumor suppressor, 26)ATP synthase CF1 beta subunit, 27) Cyclophilin 38 (atcyp38), 28)50S ribosomal protein L9, 29) ADP/ATP carrier 3 protein , 30)AT5g66570/K1F13_25, 31)Histone H3-like 5, and 32)60S ribosomal protein L17 exhibited their expression, while 10 proteins; 1)Ubiquitin, 2)Photosystem I iron-sulfur center; 3)Peroxisomal (S)-2-hydroxy-acid oxidase GLO5; 4)F1O19.10/F1O19.10, 5)putative carbonic anhydrase, 6)gb|U20808 auxin-induced protein of the zinc-binding dehydrogenase family, 7)photosystem I subunit D-1,

8)Ribosomal L10-2, 9)Histone H4 and 10)At2g19750/F6F22.22 expressed under 5, 15 and 20 % PEG stressed tissues and whereas, ribulose bisphosphate carboxylase/oxygenase activase and phosphoenolpyruvate carboxylase expressed under 5, 10 & 15% PEG stress, and however, the At2g44100/F6E13.23 expressed under 15 and 20% PEG stress (Table- 1 and Fig. 1).

The 4 proteins; 1) PSII 43 KDa protein, 2) PSII 32 KDa protein, 3) RecName: Full=Photosystem I P700 chlorophyll a apo-protein A1 and 4) At1g32990/F9L11_15, expressed under 5 and 15% PEG stress, whereas, 6 proteins; 1) photosystem II D2 protein, 2) photosystem II type I chlorophyll a, b binding protein, 3) the 40S ribosomal protein S5-1, 4) AT4g21280/F7J7_220, 5) AT2G21330 and 6) the 40S ribosomal protein S6-2, expressed under 5% PEG stress (Table- 1 and Fig. 1)

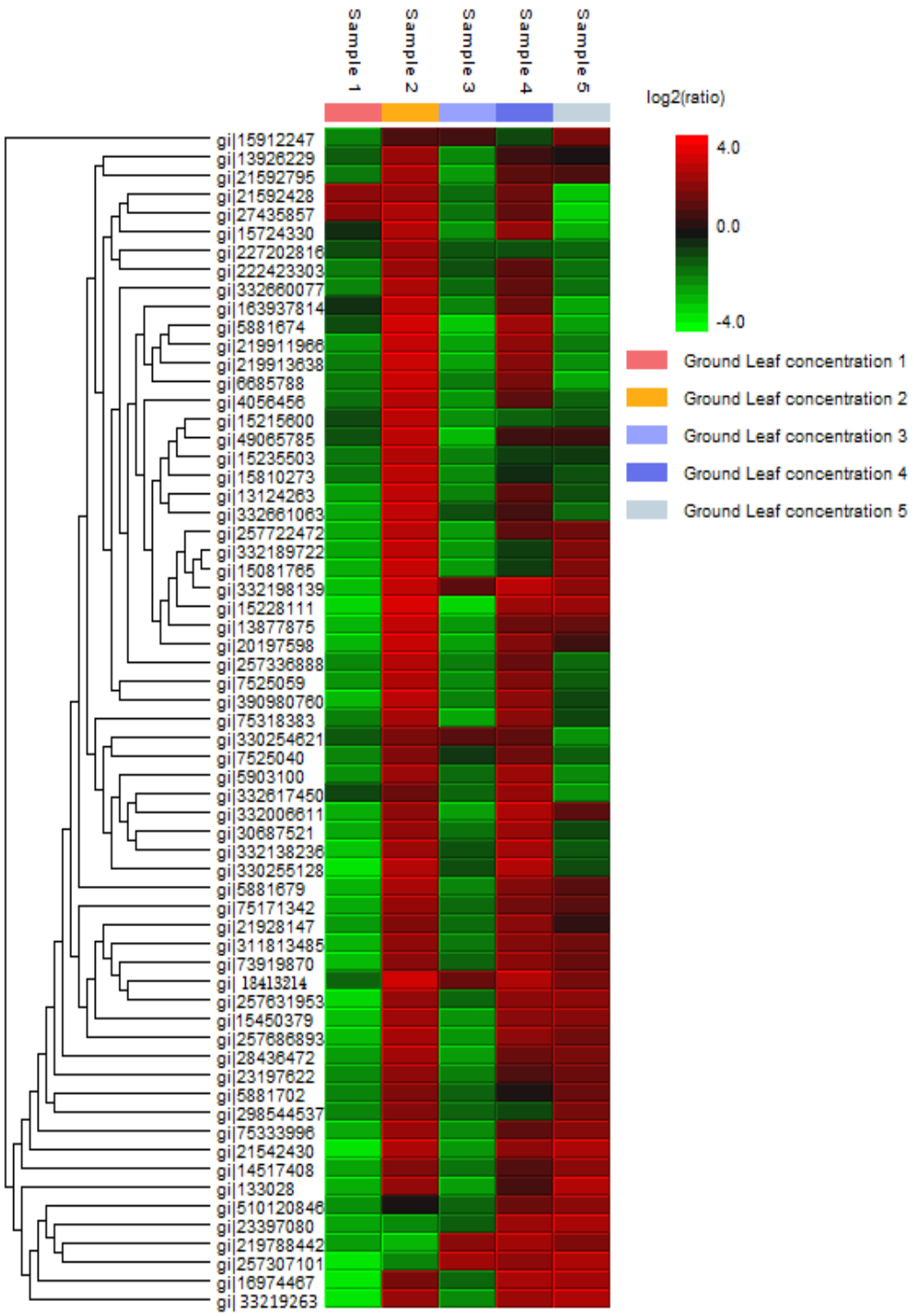


Fig. 1: LC-MS/MS analysis of differentially expressed proteins in PEG stressed

Table- 1: Expression levels of identified drought responsive proteins in PEG stressed leaf tissues by LC-MS/MS

S. No.	Protein ID	MW (Daltons)	Expression levels					Accession No.	Identified peptides	Homologous Gene in *Arabidopsis thaliana*
			control	5%	10%	15%	20%			
1	460	49451	1.00	16.20	1.37	3.06	5.96	gi|15081765	K.IGGIGTVPVGR.V. K.IGGIGTVPVGR.V R.LPLQDVYK.IR.STNLDWYK.G	At1g07930/T6D22_3
2	621	63342	1.00	11.42	1.70	8.40	6.02	gi|257686893	K.LADLVGVTLGPK.G K.VVAAGANPVLITR.G K.LADLVGVTLGPK.G K.VVAAGANPVLITR.G R.EVELEDPVENIGAK.L	unnamed protein product
3	4769	45916	1.00	11.20	1.20	3.93	4.53	gi|257722472	H.LPSYLLK.N	unnamed protein product
4	948	27880	1.00	5.29	1.45	5.38	1.05	gi|5903100	R.VPLIVTGNDFSTLYAPLIR.E R.VPLIVTGNDFSTLYAPLIR.E R.VPLIVTGNDFSTLYAPLIR.E	Highly similar to ribulose-1 5-bisphosphate carboxylase /oxygenase activase
5	966	47411	1.00	7.13	1.25	10.88	4.28	gi|332006611	K.AILNLSLR.H. K.AILNLSLR.H	glutamine synthetase
6	44	40307	1.00	10.88	1.35	3.26	2.11	gi|13124263	R.VPVFLDGGVR.R R.QLDYVPATISALEEVVK.A R.VPVFLDGGVR.R K.ALALGASGIFIGR.P K.ALALGASGIFIGR.P R.QLDYVPATISALEEVVK.A K.ALALGASGIFIGRPVVF.A	RecName: Full=Peroxisomal (S)-2-hydroxy-acid oxidase GLO2; AltName: Full=Glycolate oxidase 1; Short=AtGLO2; Short=GOX 1; AltName: Full=Short chain alpha-hydroxy acid oxidase GLO2
7	677	42442	1.00	7.09	1.13	2.82	1.37	gi|257336888	R.YTGGMVPDVNQIIVK.E R.YTGGMVPDVNQIIVK.E K.GFPGTHEFLLLDEGK.W	unnamed protein product
8	9390	44831	1.00	19.25	4.59	5.18	8.55	gi|332198139	K.LTGMAFR.V	glyceraldehyde-3-phosphate dehydrogenase of plastid
9	70	55328	1.00	10.92	1.93	6.61	4.57	gi|5881679	R.SVYEPLQTGLIAIDSMIPIGR.G R.LIESPAPGIISR.R	ATPase alpha subunit

									R.SVYEPLQTGLIAIDSMIPIGR.G K.IAQIPVSEAYLGR.V R.SVYEPLQTGLIAIDSMIPIGR.G R.LIESPAPGIISR.R K.IAQIPVSEAYLGR.V R.RSVYEPLQTGLIAIDSMIPIGR.G T.IVNTGTVLQVGDGIAR.I R.QMSLLLR.R. R.QMSLLLR.R R.EAYPGDVFYLHSR.L	
10	425	15733	1.00	7.81	2.21	6.80	5.55	gi|311813485	K.AMGIMNSFINDIFEK.L R.LVLPGELAK.H. R.LVLPGELAK.H K.IYIFK.V. K.IYIFK.V	unnamed protein product
11	20949	52039	1.00	11..06	1.89	8.48	8.29	gi|15450379	K.VPLILGIWGGK.G	At2g39730/T5I7.3
12	75	56037	1.00	7.46	1.12	4.02	1.72	gi|7525059	K.LAFYDYIGNNPAK.G K.LAFYDYIGNNPAK.G R.QGMFVIPFMTR.L R.YQWDQGYFQQEIYR.R R.AQLGEIFELDR.A M.GLPWYR.V. R.VHTVVLNDPGR.L	photosystem II 47 kDa protein
13	9170	25266	1.00	19.18	1.38	7.06	4.17	gi|20197598	K.LVPVGYGIK.K. K.LVPVGYGIK.K	putative elongation factor beta-1
14	9747	36744	1.00	11.74	3.87	13.20	3.71	gi|332138236	K.VFAIGNPFGLDHTL TTGVISGLR.R K.VFAIGNPFGLDH TLTTGVISGLR.R	Chain C Crystal Structure Analysis Of The Plant Protease Deg1
15	592	40327	1.00	7.42	2.20	4.95	4.03	gi|75171342	R.LYSIASSAIGDFGDSK.T R.MAEYAEELWELLKK.D	RecName: Full=Ferredoxin--NADP reductase leaf isozyme 1 chloroplastic; AltName: Full=Leaf FNR 1; Short=AtLFNR1; Short=FNR-1; Flags: Precursor

16	9624	46611	1.00	8.08	2.94	8.16	5.74	gi|73919870	K.GLPILGVFR.T	RecName: Full=3-ketoacyl-CoA thiolase 1 peroxisomal; AltName: Full=Acetyl-CoA acyltransferase 1; AltName: Full=Beta-ketothiolase 1; AltName: Full=Peroxisomal 3-oxoacyl-CoA thiolase 1; Flags: Precursor
17	205	18814	1.00	5.62	1.97	5.39	5.09	gi|18413214	R.GDFAIDIGR.N K.LITVDRPFAER.H R.KLIGATNPLASEPGTIR.G K.LIGATNPLASEPGTIR.G K.LIGATNPLASEPGTIRGDFAIDIGR.N R.GDFAIDIGR.N	nucleoside diphosphate kinase 1
18	10121	54159	1.00	6.60	1.90	7.44	2.65	gi|30687521	K.VVVVGIFPK.L	protein disulfide isomerase-like 1-
19	1369	68599	1.00	16.24	1.57	5.47	5.25	gi|13877875	K.LLSAMNKVGEFTGFSVK.V	putative arginine-tRNA-protein transferase
20	16	50112	1.00	4.23	1.45	2.23	1.11	gi|222423303	S.LVASLPEGGVLLLENVR.F K.FSLAPLVPR.L K.GVSLLLPTDVVVADK.F K.FLKPSVAGFLLQK.E K.IGVIESLLEK.C	AT3G12780 [
21	1790	9038	1.00	4.30	1.13	2.41	2.98	gi|23197622	K.FYGEVTQQMLK.H K.FYGEVTQQMLK.H	phosphoribulokinase precursor
22	407	54995	1.00	12.69	2.42	3.24	1.98	gi|332661063	K.TWPEDILPLQPVGR.M R.LGPNYLQLPVNAPK.C R.LGPNYLQLPVNAPK.C R.EGNFDLVGNNFPVFFIR.D	catalase 2
23	4183	176751	1.00	2.19	1.58	3.19	4.55	gi|510120846	R.WPLAQPMR.F R.WPLAQPMR.F R.WKPLTDVVDGYSPTLPHLK.G R.ETLSFWVK.A	RecName: Full=Ferredoxin-dependent glutamate synthase 1 chloroplastic/mitochondrial; AltName: Full=Fd-GOGAT 1; Flags: Precursor

24	608	44445	1.00	4.49	1.69	4.90	2.59	gi|21928147	K.MVPFAGWSMPIQYK.D	At1g11860/F12F1_30
25	165	47660	1.00	5.16	1.77	3.38	5.67	gi|14517408	R.AAALNIVPTSTGAAK.A K.YDSMLGTFK.A K.VVAWYDNEWGYSQR.V	At1g42970/F13A11_3
26	8	52955	1.00	3.29	1.35	1.87	2.68	gi|5881702	K.DTDILAAFR.V. K.DTDILAAFR.V K.TFQGPPHGIQVER.D. A.ALRLEDLR.I R.AMHAVIDR.Q. R.DNGLLLHIHR.A Y.VAYPLDLFEEGSVTNMFTSIVGNVF GFK.A. R.LSGGDHIHAGTVVGK.L	large subunit of riblose-1 5-bisphosphate carboxylase/oxygenase
27	36109	113363	1.00	12.44	1.26	2.70	5.20	gi|332189722	N.NPSWFNR.I	putative RNA helicase SDE3
28	1280	24917	1.00	1.39	2.18	7.20	8.75	gi|23397080	R.VAIGQVLLSVR.C	putative tumor suppressor
29	12	53934	1.00	3.43	1.79	2.82	1.45	gi|7525040	K.GIYPAVDPLDSTSTMLQPR.I R.FVQAGSEVSALLGR.M K.MPNIYNALVVK.G K.MPNIYNALVVK.G K.VVDLLAPYR.R. K.VVDLLAPYR.R K.LSIFETGIK.V. K.LSIFETGIK.V K.GIYPAVDPLDSTSTMLQPR.I K.IGLFGGAGVGK.T	ATP synthase CF1 beta subunit
30	22846	40378	1.00	15.16	2.11	7.85	3.40	gi|390980760	V.IPFNAFGTMAMAR.E	Chain A Crystal Structure Of Arabidopsis Thaliana Cyclophilin 38 (atcyp38)
31	15135	22134	1.00	7.42	1.25	3.73	11.30	gi|133028	K.LIFGSVTAQDLVDIIK.S	RecName: Full=50S ribosomal protein L9 chloroplastic; AltName: Full=CL9; Flags: Precursor
32	836	40718	1.00	5.95	1.32	2.52	1.21	gi|332660077	R.AVAGAGVLAGYDK.L	ADP/ATP carrier 3 protein
33	922	45162	1.00	15.51	4.87	14.30	13.95	gi|257631953	R.TLLYGGIYGYPR.D	unnamed protein product
34	825	35128	1.00	2.19	2.09	1.65	3.08	gi|15912247	R.LTYDEIQSK.T	AT5g66570/K1F13_25

35	4219	71102	0	1.00	5.14	4.06	6.02	gi|257307101	K.VQDLLLLDVTPLSLGLETAGGVMTV LIPR.N	unnamed protein product
36	9059	15591	1.00	7.65	1.54	4.07	6.04	gi|75333996	R.FRPGTVALR.D	RecName: Full=Histone H3-like 5
37	760	47719	1.00	3.58	1.38	1.67	3.11	gi|298544537	E.VQIVGDDLLVTNPK.R	unnamed protein product
38	2427	15027	1.00	11.06	1.89:	8.48	8.29	gi|15450856	K.VLPAVIVR.Q. K.VLPAVIVR.Q	Strong similarity to 60S ribosomal protein L17
39	4374	89393	1.00	0.68	5.30	7.23	4.68	gi|219788442	R.IVSQLLTLMDGLK.S R.IVSQLLTLMDGLK.S	unnamed protein product
40	14983	36466	1.00	7.08	0.99	3.68	4.23	gi|28436472	G.MQIFVK.T. K.ESTLHLVLR.L K.ESTLHLVLR.L. G.MQIFVK.T K.ESTLHLVLR.L. R.LIFAGK.Q	Ubiquitin
41	1720	9038	1.00	4.97	0.30	1.33	1.32	gi|49065785	R.VYLWHETTR.S R.VYLWHETTR.S	RecName: Full=Photosystem I iron-sulfur center; AltName: Full=9 kDa polypeptide; AltName: Full=PSI-C; AltName: Full=Photosystem I subunit VII; AltName: Full=PsaC
42	149	40482	1.00	5.19	0.62	3.62	1.59	gi|75318383	N.ISMPIMIAPTAMQK.M N.ISMPIMIAPTAMQK.M	RecName: Full=Peroxisomal (S)-2-hydroxy-acid oxidase GLO5; AltName: Full=Glycolate oxidase 3; Short=AtGLO5; Short=GOX 3; AltName: Full=Short chain alpha-hydroxy acid oxidase GLO5
43	4231	29824	1.00	10.42	0.79	4.71	1.27	gi|219911966	K.DSTLIMQLLR.D R.NLLSVAYK.N	unnamed protein product
44	385	14699	1.00	3.18	0.65	1.43	1.32	gi|13926229	R.YWTMWK.L. R.YWTMWK.L K.EVDYLIR.N. A.KEVDYLIR.N A.KEVDYLIR.N	F1O19.10/F1O19.10
45	733	36144	1.00	6.41	0.80	1.56	1.32	gi|15810273	T.NPALYGELAK.G	putative carbonic anhydrase chloroplast precursor

46	1201	32775	1.00	5.68	0.68	2.01	1.13	gi|4056456	K.ATDSPLPTVPGYDVAGVVVK.V	Strong similarity to gb|U20808 auxin-induced protein from Vigna radiata and a member of the zinc-binding dehydrogenase family PF|00107. ESTs gb|T43674 gb|H77006 and gb|AA395179 come from this gene
47	1040	22598	1.00	5.58	0.90	1.48	1.51	gi|15235503	K.EQIFEMPTGGAAIMR.E	photosystem I subunit D-1
48	256	11409	1.00	4.67	0.68	2.16	1.96	gi|21592795	K.IFLENVIR.D K.IFLENVIR.D K.TVTAMDVVYALKR.Q	histone H4-like protein
49	2019	24909	0	1.00	0.21	1.09	1.40	gi|332192636	R.VAIGQVLLSVR.C	60S ribosomal protein L10-2
50	30578	6887	0	1.00	0.42	1.76	1.60	gi|16974467	R.FVTAVVGFGK.K	At2g19750/F6F22.22
51	182	49100	1.00	2.16	1.60	1.68	0.56	gi|330254621	K.FYWAPTR.E. K.FYWAPTR.E K.LVVHITK.N	ribulose bisphosphate carboxylase/oxygenase activase
52	30698	51857	0	1.00	0.24	1.98	0.24	gi|330255128	R.IPAGFLEGVTNIVPALGGVNLK.Q	phosphoenolpyruvate carboxylase
53	42041	49819	0	1.00	0.21	1.09	1.40	gi|15215778	R.FQGGSPYIYPLYGLGELPQAFAR.L	At2g44100/F6E13.23
54	114	50303	1.00	4.17	0.55	1.49	0.36	gi|163937814	K.GIDRDFEPVLSMTPLN K.GIDRDFEPVLSMTPLN R.SPTGEVIFGGETMR.F R.FWDLR.A. R.FWDLR.A R.LGANVGSAQGPTGLGK.Y R.SAEYMTHAPLGSLNSVGGVATEINA VNYVSPR.S. R.LGANVGSAQGPTGLGK.Y	PSII 43 KDa protein
55	333	38937	1.00	7.16	0.23	3.05	0.43	gi|5881674	R.VINTWADIINR.A R.VINTWADIINR.A R.NAHNFPLDLA.A	PSII 32 KDa protein
56	486	83231	1.00	8.11	0.95	2.73	0.56	gi|6685788	R.DLLAQLYPSFAEGATPFFTLNWSK.Y R.DLLAQLYPSFAEGATPFFTLNWSK.Y R.DLLAQLYPSFAEGATPFFTLNWSK.Y K.ILVDRDPIK.TR.YNDLLDR.V	RecName: Full=Photosystem I P700 chlorophyll a apoprotein A1; AltName: Full=PSI-A; AltName: Full=PsaA

57	789	34976	1.00	1.65	0.80	2.68	0.52	gi|332617450	K.VAILGAAGGIGQSLSLLMK.M	unnamed protein product
58	9470	23148	1.00	3.48	0.47	2.29	0.31	gi|15724330	K.TPPASVLLLK.A	At1g32990/F9L11_15
59	322	35675	1.00	9.31	0.62	3.34	0.79	gi|219913638	K.MELVDAAFPLLK.G K.VLVVANPANTNALILK.E	unnamed protein product
60	122	35243	1.00	1.53	0.30	0.65	0.08	gi|27435857	R.AWMAAQDQPHENLIFPEEVLPR.G K.NILLNEGIR.A. K.NILLNEGIR.A	photosystem II D2 protein
61	313	28171	1.00	1.06	0.32	0.71	0.10	gi|21592428	K.YLGPFSGEPPSYLTGEFPGDYGWDT AGLSADPETFAR.N K.YLGPFSGEPPSYLTGEFPGDYGWDT AGLSADPETFAR.N. R.ELEVIHSR.W	putative photosystem II type I chlorophyll a b binding protein
62	9970	22991	0	1.00	0	0.35	0.34	gi|15228111	R.LTNSLMMHGR.N	40S ribosomal protein S5-1
63	2210	23795	1.00	4.65	0.54	0.85	0.96	gi|15215600	K.AWPYVQNDLR.S	AT4g21280/F7J7_220
64	710	40466	1.00	2.86	0.95	0.97	0.83	gi|227202816	R.LASIGLENTEANR.Q	AT2G21330
65	9607	28162	0	1.00	0.13	0.66	0.99	gi|21542430	R.LVTPLTLQR.K	RecName: Full=40S ribosomal protein S6-2; AltName: Full=Protein EMBRYO DEFECTIVE 3010

2. Root proteomic analysis by LC-MS/MS:

The results exhibited, that of the 67 drought stressed proteins, 60 were found to be the functional non- redundant proteins, while 7 were the unnamed proteins (Table-7). The sequence alignments of these proteins expressed under the PEG stress showed > 97% nucleotide homology with the corresponding proteins of *Arabidopsis thaliana*. Of the 60 functional proteins; 5 proteins; 1. NAD+ dependent isocitrate dehydrogenase subunit 1, 2. At2g27860/F15K20.4, 3. Catalase-2, 4. nucleoside diphosphate kinase 1, and 5. glyceraldehyde-3-phosphate dehydrogenase, expressed under all the 4 concentrations of 5, 10%, 15% and 20 % PEG stress (Table-2 and Fig. 2).

The 4 proteins; 1) AT5g41520/MBK23_4, 2) 2-phospho-D-glycerate hydro-lyase 1, 3) cytosolic malate dehydrogenase and 4) translation initiation factor 4A-3, expressed under 10, 15 and 20 % PEG stress, while only 1 protein, the At2g39730, expressed under 5, 10, and 20% of PEG stress, whereas, 1 protein, tubulin alpha-6 chain, expressed under 5 and 20 % PEG stressed samples. Six proteins; 1) pyruvate kinase, 2) 26S proteasome subunit 4, 3) clathrin heavy chain putative; 4) Sucrose synthase 1, 5) AT1G07930, and 6) RAN2 small Ras-like GTP-binding, expressed under 10 and 15 % PEG stressed tissues (Table-2 and Fig. 2).

The 4 proteins; 1) O-acetylserine (thiol) lyase (OAS-TL) isoform A2, 2) proteasome subunit alpha type-2-B, 3) Peroxiredoxin-2B and 4) fructokinase 1, expressed under 15 and 20 % PEG stressed samples, while 1 protein; phospho-enol pyruvate carboxylase, expressed under 5 and 10% PEG stressed tissues, whereas, 15 proteins; 1) Glycosyl hydrolase, 2) mitochondrial elongation factor Tu, 3) putative aldehyde dehydrogenase (NAD+), 4) 20S proteasome beta subunit G1, 5) F22C12.1, 6) RmlC-like cupins superfamily protein, 7) adenylate translocator, 8) proteasome subunit beta type-5-B, 9) Chaperonin CPN60, 10) T27G7.6, 11) putative sorbitol dehydrogenase, 12) asparagine synthetase [glutamine-hydrolyzing], 13) enolase (2-phospho-D-glycerate hydroylase), 14) putative disease resistance protein and 15) fructose-bisphosphate aldolase, expressed only in 15 % PEG stressed tissues (Table-2 and Fig. 2).

The following 23 proteins; 1) Putative 60S ribosomal protein L9, 2) Eukaryotic Initiation Factor 4A-2, 3) 40S ribosomal protein S15, 4) 60S ribosomal protein L13a-4, 5) AT5g35530/MOK9_14, 6) phosphoglycerate kinase putative, 7) ribosomal protein S18 putative, 8) 60S ribosomal protein L11-2, 9) sucrose synthase 4, 10) 40S ribosomal protein

S5-1, 11) alcohol dehydrogenase class-3, 12) S-adenosyl-L-homocysteine hydrolase 1 mutant, 13) Alcohol dehydrogenase class-P, 14) Serine hydroxy methyltransferase 4; 15) monodehydroascorbate reductase (NADH), 16) ATP synthase subunit beta-1, 17) putative fructokinase, 18) histone H3, 19) Phosphor gluco-mutase, 20) fatty acid beta-oxidation multifunctional protein MFP2, 21) AT5g35530/MOK9_14, 22) histone H4-like protein and 23) Heat shock 70 kDa protein 18, exhibited their expression only in the control samples but not in any of the PEG stressed root tissues (Table-2 and Fig. 2).

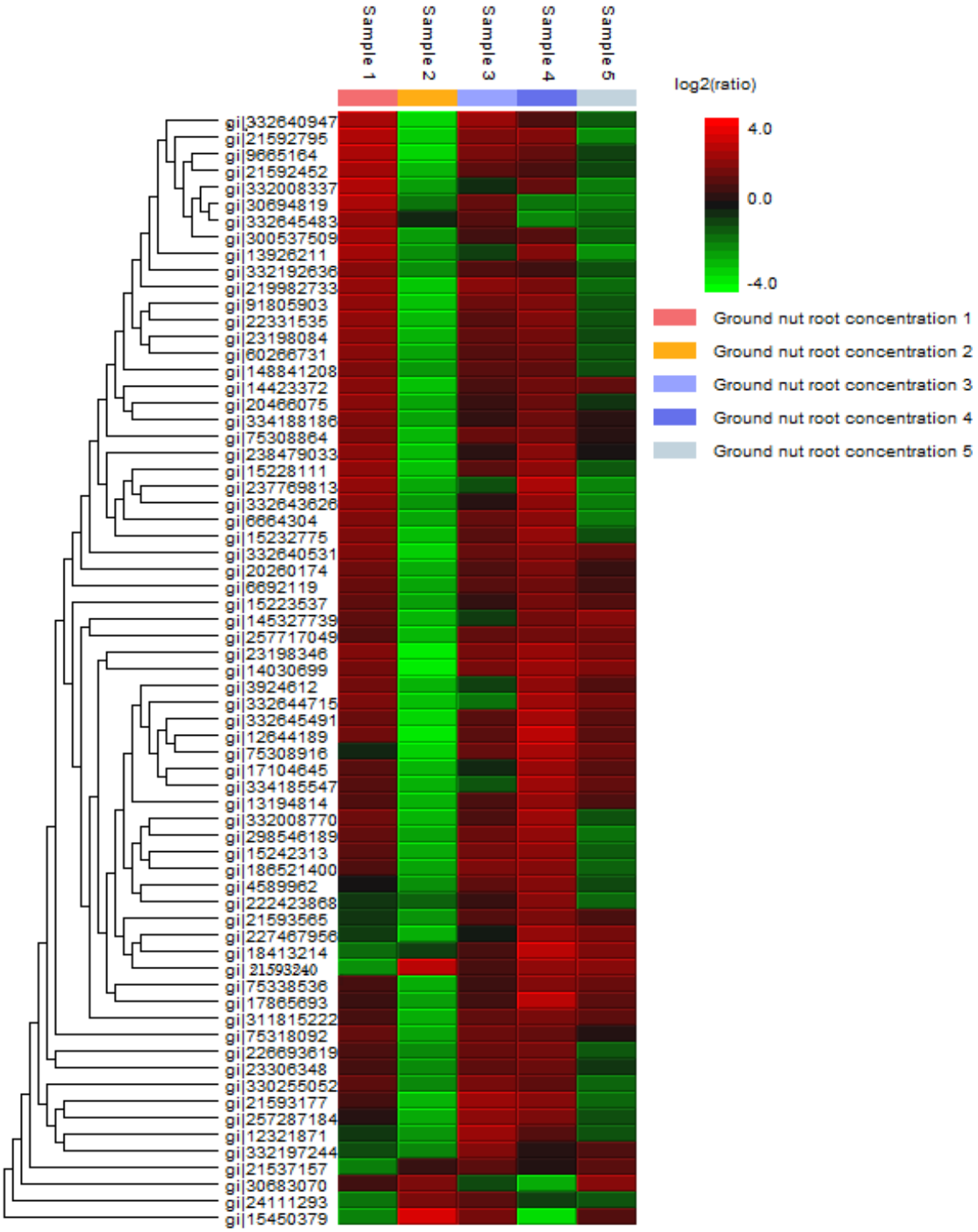


Fig.2: LC-MS/MS analysis of differentially expressed proteins in PEG stressed roots

Table- 2: Expression levels of identified drought responsive proteins in PEG stressed root tissues by LC-MS/MS

S. No.	Protein Id	MW (Daltons)	Expression levels					Accession No.	Identified peptides	Homologous Gene in *Arabidopsis thaliana*
			control	5%	10%	15%	20%			
1	4189	39657	1.00	1.87	2.26	1.78	2.24	gi|21537157	N.KANPVALLLSSAMMLR.H	NAD+ dependent isocitrate dehydrogenase subunit 1
2	3536	43638	1.00	2.71	2.03	1.45	1.29	gi|24111293	K.TSLWDLLESTLTYQHR.T	At2g27860/F15K20.4
3	249	56931	1.00	1.35	1.03	1.59	1.19	gi|17865693	R.FSTVIHER.G. R.LGPNYLQLPVNAPK.C R.EGNFDLVGNNFPVFFIR.D R.LGPNYLQLPVNAPK.C R.EGNFDLVGNNFPVFFIR.D	RecName: Full=Catalase-2
4	21145	18814	1.00	1.50	1.68	3.16	2.57	gi|18413214	R.GDFAIDIGR.N	nucleoside diphosphatekinase 1
5	4182	36989	1.00	5.62	1.97	5.39	5.09	gi|21593240	R.FGIVEGLMTTVHSITATQK.TK.TL LFGEKPVTVFGIR.N R.AASFNIIPSSTGAAK.A R.VPTVDVSVVDLTVR.L R.FGIVEGLMTTVHSITATQK.TK.TL LFGEKPVTVFGIR.N	glyceraldehyde-3-phosphate dehydrogenase of plastid
6	40228	19733	1.00	0	1.05	1.52	1.14	gi|14030699	R.TYLNLPSEIVPATLK.K R.TYLNLPSEIVPATLK.K	AT5g41520/MBK23_4
7	825	50975	1.00	0.32	1.08	2.45	1.74	gi|227467956	R.ISLAGLSLAK.C R.TEELQPYVLNVVK.K K.LNLGVGAYR.T. K.LNLGVGAYR.T. R.PMYSNPPVHGAR.I R.VATIQGLSGTGSLR.L K.AEMEMMAGR.I K.AEMEMMAGR.IK.WHVYMTK.D	unnamed protein product

8	2331	51475	1.00	0.17	1.41	2.97	1.49	gi|75308916	R.EGLVLLIDAIEK.A	RecName: Full=Enolase 1 chloroplastic; AltName: Full=2-phospho-D-glycerate hydro-lyase 1; AltName: Full=2-phosphoglycerate dehydratase 1; Flags: Precursor
9	647	35662	1.00	0.47	1.29	1.80	1.22	gi|21593565	K.MELVDAAFPLLK.G K.EFAPSIPEK.N. K.EFAPSIPEK.N R.VLVTGAAGQIGYALVPM.I	cytosolic malate dehydrogenase
10	1897	18464	1.00	0.21	1.10	1.27	1.22	gi|257717049	K.HVVFGQVVEGLDVVK.A K.HVVFGQVVEGLDVVK.A	unnamed protein product
11	309	46771	1.00	0.61	2.17	1.19	1.39	gi|332197244	R.GFKDQIYDIFQLLPSK.V R.DHTVSATHGDMDQNTR.D	translation initiation factor 4A-3
12	4180	20576	1.00	0.27	1.18	1.46	1.14	gi|311815222	K.MKDTDSEEELKEAFR.V K.MKDTDSEEELKEAFR.V	unnamed protein product
13	20949	52039	1.00	13.10	2.96	0	2.24	gi|15450379	K.VPLILGIWGGK.G	At2g39730/T5I7.3
14	2802	47235	1.00	1.56	0.78	0.28	1.77	gi|30683070	R.LVSQVISSLTASLR.F R.LVSQVISSLTASLR.F	tubulin alpha-6 chain
15	1049	54977	1.00	0.26	1.22	1.58	0.60	gi|15242313	K.ATDSEATEVIIEAALK.S	pyruvate kinase
16	2461	49328	1.00	0.24	2.17	1.68	0.62	gi|21593177	K.VLSVVGILQDEVDPMVSVMK.VK. VLSVVGILQDEVDPMVSVMK.V	26S proteasome subunit 4
17	851	193215	1.00 :	0.46	2.61	1.31	0.88	gi|12321871	R.LLTQFLEHLVSEGSQDVHVHNAL GK.I. R.LLTQFLEHLVSEGSQDVHVHNAL GK.IR.EATAFLLDVLKPNLPEHAFL QTK.V	clathrin heavy chain putative; 28833-19741
18	263	92998	1.00	0.43	1.20	1.32	0.67	gi|226693619	K.STQEAIVLPPWVALAVR.P K.STQEAIVLPPWVALAVRPR.PK.ST QEAIVLPPWVALAVRPR.PK.STQEA IVLPPWVALAVR.PR.VVHGIDVFDP	RecName: Full=Sucrose synthase 1; Short=AtSUS1; AltName: Full=Sucrose-UDP glucosyltransferase 1

									K.FR.PRPGVWEYLR.V R.PGVWEYLR.V K.STQEAIVLPPWVALAVRPR.PK.STQEAIVLPPWVALAVRPR.P	
19	930	55952	1.00	0.26	1.06	1.62	0.47	gi|298546189	R.AEATDVANAVLDGSDAILLGAETLR.GR.AEATDVANAVLDGSDAILLGAETLR.G R.VVDSMTDNLRPTR.A	unnamed protein product
20	368	49505	1.00	0.79	1.13	2.00	0.76	gi|222423868	K.IGGIGTVPVGR.V. R.LPLQDVYK.I K.IGGIGTVPVGR.V. R.LPLQDVYK.I R.STNLDWYK.G. R.QTVAVGVIK.S K.NGDAGMVMMTPTKPMVVETFSEYPPLGR.F. K.EVSSYLK.K K.EVSSYLK.K	AT1G07930
21	2358	25062	1.00	0.44	1.16	1.29	0.84	gi|23306348	K.LVIVGDGGTGK.T	RAN2 small Ras-like GTP-binding nuclear protein (Ran-2)
22	1276	44756	1.00	0.31	2.07	1.71	0.83	gi|257287184	R.ESIELPLMNPELFLR.V	unnamed protein product
23	4836	32339	1.00	0.23	0.65	2.06	1.10	gi|334185547	K.LYGVEPVESPILSGGKPGPHK.IK.LYGVEPVESPILSGGKPGPHK.I	O-acetylserine (thiol) lyase (OAS-TL) isoform A2
24	2256	25733	1.00	0.21	0.70	1.19	1.46	gi|145327739	K.LVQIEHALTAVGSGQTSLGIK.AK.LVQIEHALTAVGSGQTSLGIK.A	proteasome subunit alpha type-2-B
25	30377	17428	1.00	0.27	0.96	1.63	1.26	gi|75338536	R.FALLLDDLK.V	RecName: Full=Peroxiredoxin-2B; AltName: Full=Peroxiredoxin IIB; AltName: Full=Peroxiredoxin TPx1; AltName: Full=Thioredoxin reductase 2B; AltName: Full=Thioredoxin-dependent peroxidase 1
26	2888	37028	1.00	0.22	0.78	1.89	1.01	gi|17104645	R.LPLWPSEEAAR.K R.LPLWPSEEAAR.K	putative fructokinase 1

27	18	109753	1.00	3.12	1.98	0.78	0.48	gi|330255052	R.VPYNAPLIQFSSWMGGDR.DK.MASIDAQLR.L R.VPYNAPLIQFSSWMGGDRDGNPR.VR.VVPLFEK.L. R.FVEYFR.L R.VPYNAPLIQFSSWMGGDR.DR.VPYNAPLIQFSSWMGGDR.DR.LLLQVAGHK.D R.AIPWIFAWTQTR.F R.LLLQVAGHK.D K.DITPDDKQELDEALQR.E R.LLLQVAGHK.D R.LLLQVAGHK.D K.DITPDDKQELDEALQR.E K.QEVMIGYSDSGK.D	phosphoenolpyruvate carboxylase 2
28	5629	70583	1.00	0.13	0.37	1.49	0.91	gi|332644715	L.LTELLK.Q	Glycosyl hydrolase family protein
29	3155	49410	1.00	0.18	0.59	1.38	0.78	gi|3924612	K.VGEEVEILGLR.E K.EHILLAR.Q	elongation factor- Tu (mitochondrial)
30	2587	58589	1.00	0.24	0.97	1.77	0.99	gi|13194814	K.YGLAAGVFTK.N	putative aldehyde dehydrogenase (NAD+)
31	2757	27651	1.00	0.29	0.80	1.20	0.93	gi|15223537	K.HSLLGASGEISDFQEILR.Y	20S proteasome beta subunit G1
32	3331	377031	1.00	0.24:	0.89	1.06	0.78	gi|6692119	T.GSVLLLFDELRSK.F	F22C12.1
33	4237	24921	1.00	0.10	0.79	1.00	:0.77	gi|332640531	R.GVDLDALLDMSTEDLVK.H	RmlC-like cupins superfamily protein
34	634	41476	1.00	0.05	0.90	1.29	0.86	gi|23198346	K.LLIQNQDEMIK.A. R.MMMTSGEAVK.Y. R.MMMTSGEAVK.Y R.QFDGLVDVYR.K. R.QFDGLVDVYR.K. K.GAGANILR.A	adenylate translocator

35	4748	29485	1.00	0.22	0.53	1.07	0.29	gi|332643626	R.ASMGGYISSQSVK.K	proteasome subunit beta type-5-B
36	1312	61281	1.00	0	0.83	2.62	0.83	gi|12644189	K.IGVQIIQNALK.T. K.GEYVDMVK.A. K.GEYVDMVK.A.	RecName: Full=Chaperonin CPN60 mitochondrial; AltName: Full=HSP60; Flags: Precursor
37	30097	21729	1.00	0.19	0.78	1.12	0.33	gi|6664304	R.LLGPGLNK.A	T27G7.6
38	14786	39256	1.00	0.18	0.78	1.67	0.55	gi|332008770	K.IDVKPLITHR.F	putative sorbitol dehydrogenase
39	445	65621	1.00	0.14	0.72	1.29	0.48	gi|15232775	R.LAVIDPASGDQPLFNEDK.T	asparagine synthetase [glutamine-hydrolyzing]
40	326	47777	1.00	0.21	0.80	1.13	0.71	gi|20260174	E.VQIVGDDLLVTNPK.R K.EGLELLK.T. K.EGLELLK.T	enolase (2-phospho-D-glycerate hydroylase)
41	10850	155345	1.00	0.15	0.38	1.40	0.24	gi|237769813	E.IPGLSEASK.L	putative disease resistance protein
42	425	38540	1.00	0.11	:0.88 :	2.00	0.89	gi|332645491	K.GILAADESTGTIGK.R K.GILAADESTGTIGK.R K.EGGVLPGIK.V. K.EGGVLPGIK.V K.ANSEATLGTYK.G	fructose-bisphosphate aldolase
43	972	22018	1.00	0.05	0.53	0.44	0.29	gi|9665164	R.TALSHVDNLISGVTR.G R.TALSHVDNLISGVTR.G	Putative 60S ribosomal protein L9
44	320	46763	1.00	0.12	0.60:	0.78	0.70	gi|14423372	K.IQVGVFSATMPPEALEITR.KR.DH TVSATHGDMDQNTR.D	Eukaryotic Initiation Factor 4A-2
45	3063	8139	1.00	0.10	0.69	0.81	0.38	gi|91805903	R.GVDLDALLDMSTEDLVK.H R.GVDLDALLDMSTEDLVK.H R.GVDLDALLDMSTEDLVK.H	40S ribosomal protein S15
46	2128	23590	1.00	0.11	0.29:	:0.40	0.17	gi|332008337	R.MVIPDALK.V. R.MVIPDALK.V	60S ribosomal protein L13a-4
47	20840	27458	1.00	0.18	0.55	0.73	0.49	gi|20466075	K.LLGGLAVR.R. K.IMLDWDPK.G	AT5g35530/MOK9_14

48	287	49939	1.00	0.14	0.67	0.79	0.44	gi|23198084	T.LVASLPEGGVLLLENVR.F R.GVSLLLPTDVVIADK.F R.GVSLLLPTDVVIADK.F K.GVTTIIGGGDSVAAVEK.V K.IGVIESLLEK.C. K.IGVIESLLEK.C V.MSHISTGGGASLELLEGK.V	phosphoglycerate kinase putative
49	861	17545	1.00	0.10	0.47	0.42	0.32	gi|21592452	K.IPDWFLNR.Q K.YSQVVSNALDMK.L R.HYWGLR.V	ribosomal protein S18 putative
50	697	19775	1.00	0.20	0.45	0.20	0.19	gi|30694819	K.LVLNISVGESGDR.L K.VLEQLSGQTPVFSK.A K.VLEQLSGQTPVFSK.A	60S ribosomal protein L11-2
51	297	93003	1.00	0.12	0.58	0.82	0.38	gi|22331535	K.FQEIGLER.G. R.VVHGIDVFDPK.F	sucrose synthase 4
52	2928	22991	1.00	0.11	0.60	0.98	0.38	gi|15228111	R.QAVDISPLR.R	40S ribosomal protein S5-1
53	1858	17095	1.00	0.09	0.87	0.73	0.30	gi|219982733	K.AHGLAPEIPEDLYHLIK.K K.AHGLAPEIPEDLYHLIK.K	unnamed protein product
54	3117	42043	1.00	0.20	0.59	0.83	0.58	gi|334188186	K.FGVNEFVNPK.D K.KFGVNEFVNPK.D K.KFGVNEFVNPK.D	alcohol dehydrogenase class-3
55	195	53378	1.00	0.18	0.64	0.74	0.43	gi|60266731	R.LVGVSEETTTGVK.R R.LVGVSEETTTGVK.R R.HSLPDGLMR.A. R.TEFGPSQPFK.G. K.VYVLPK.H	S-adenosyl-L-homocysteine hydrolase 1 mutant
56	2280	16036	1.00	0.15	0.44	0.48	0.29	gi|300537509	K.IAGFSTHLMK.R	unnamed protein product
57	1141	41178	1.00	0.25	0.74:	0.77	0.50	gi|148841208	K.FITHTVPFSEINK.A K.FITHTVPFSEINK.A. K.AAVAWEAGKPLVIEEVEVAPPQ.K	RecName: Full=Alcohol dehydrogenase class-P

58	514	51718	1.00	0.26	1.05	0.99	0.73	gi|75318092	R.AVTLTLDIQK.T	RecName: Full=Serine hydroxymethyltransferase 4; Short=AtSHMT4; AltName: Full=Glycine hydroxymethyltransferase 4; AltName: Full=Serine methylase 4
59	1132	50160	1.00	0.46	0.57	0.24	0.34	gi|332645483	K.AVVVGGGYIGLELSAVLR.I	monodehydroascorbate reductase (NADH)
60	99	59631	1.00	0.30	1.46	1.56	0.62	gi|186521400	K.IGLFGGAGVGK.T K.KGSITSVQAIYVPADDLTDPAPAT TFAHLDATTVLSR.Q R.IPSAVGYQPTLASDLGALQER.I K.NLQDIIAILG MDELSEDDKLTVAR.A K.KGSITSVQAIYVPADDLTDPAPAT TFAHLDATTVLSR.Q K.KGSITSVQAIYVPADDLTDPAPAT TFAHLDATTVLSR.Q K.AHGGFSVFAGVGER.T	ATP synthase subunit beta-1
61	2509	35275	1.00	0.46	1.33	1.87	0.87	gi|4589962	R.NPSADMLLRPDELNLDLIR.S R.EFMFYR.N	putative fructokinase
62	1527	15268	1.00	0.17	0.33	0.68	0.17	gi|13926211	R.FRPGTVALR.E	histone H3
63	442	72558	1.00	0.13	0.52	0.98	0.51	gi|238479033	K.LVTVEDIVR.Q. K.LQSSLPEVNK.I	Phosphor gluco-mutase
64	3858	78840	1.00	0.05	0.81	0.40	0.26	gi|332640947	R.IVGAHFFSPAHIMPLLEIVR.T	fatty acid beta-oxidation multifunctional protein MFP2
65	20840	27458	1.00	0.18	0.55	0.73	0.49	gi|20466075	K.LLGGLAVR.R K.IMLDWDPK.G	AT5g35530/MOK9_14

66	1428	11409	1.00	0.06	0.51	0.54	0.15	gi|21592795	R.DNIQGITKPAIR.R K.IFLENVIR.D R.DNIQGITKPAIR.R R.DNIQGITKPAIR.R R.ISGLIYEETR.G R.ISGLIYEETR.G	histone H4-like protein
67	138	68357	1.00	0.16	0.84	0.92	0.60	gi|75308864	R.DNIQGITKPAIR.R K.IFLENVIR.D R.DNIQGITKPAIR.R R.DNIQGITKPAIR.R R.ISGLIYEETR.G R.ISGLIYEETR.G	RecName: Full=Heat shock 70 kDa protein 18; AltName: Full=Heat shock protein 70-18; Short=AtHsp70-18; AltName: Full=Heat-shock protein 70T-1

Objective -4:

Validation of selected proteins for drought stress tolerance by qRT-PCR:

1. Isolation of RNA from drought stressed leaves and roots:

The isolation of total RNA from the stressed leaves and roots of groundnut was carried out by using Trizol Reagent (in vitrogen) following the manufacturer's protocol. Leaf and root tissues, weighing 100 mg were homogenized separately in liquid nitrogen by adding 1ml of Trizol Reagent and kept for incubation at room temperature for 15 min. Later, 200 µl of chloroform was added, shook vigorously for 15 seconds, allowed to stand for 15 min at room temperature and centrifuged the mixture at 12,000 *g* for 15 min at 4 ºC. Later, the aqueous phase was transferred to fresh tubes, added 500 µl of isopropanol and kept for incubation for 15 min at room temperature. Thereafter, centrifuged at 12000 *g* for 10 min at 4 ºC, removed the supernatant and washed the RNA pellet by 70% ethanol, centrifuged again at 8000 rpm for 5 min at 4 ºC, removed the ethanol and the RNA pellet was dissolved in 30µl of DEPC water and later the concentration of RNA was checked by using Nano drop machine.

To remove DNA contamination, if any, samples were treated with DNAseI (BioLabs). RNA concentration and purity were determined before and after DNAseI treatment using a NanoDrop™ spectrophotometer - ND-1000 (Thermo Scientific) and RNA integrity was verified by electrophoresis using 1% agarose gel.

1. Formaldehyde gel electrophoresis:

Composition of gel:

a.1% agorose gel; b. 10X MOPS; c. 12.3 M formaldehyde

2. Electrophoresis running buffer:

a. 1X MOPS; b. 2.2 M formaldehyde

3. RNA denaturing buffer:

10 ml of 100% formaldehyde, and 1.5 ml of MOPS buffer

4. RNA loading buffer:

a.50% glycerol; b. 1mm EDTA; and c. 0.4% bromophenol blue

a. RNA sample preparation:

5 µl of RNA, 15 µl of RNA denaturation buffer and 1 µl of 10 mg/ml ethidium bromide were taken in 1.5 ml eppendorf tube, heated at 65 ºC for 10 min to denature the secondary structure of RNA, cooled on ice for 2 min and loaded on gel.

b. Preparation of gel:

A quantity of 0.5 g of 1% agarose powder was taken in a beaker, 37 ml of DEPC treated water was added and boiled in microwave and cooled it. Later, 10 ml of 10X MOPS and 8.9 ml formaldehyde were added, then poured it on the gel casting tray allowed it to solidify for about 10 min. After solidification, the tray containing the gel was transferred on to electrophoretic unit to carry out the electrophoresis.

c. Loading the RNA samples on gel:

The RNA samples from each of the leaf and root tissues were loaded separately in the wells and the electrophoresis was carried out. Later, the gel was examined under the gel doc for the presence and integrity of RNA. The total RNA isolated was diluted to 2.5 µg/µl concentration and aliquoted for use in synthesis of cDNA.

2. cDNA synthesis:

The previously isolated RNA samples from leaf and root tissues were converted to cDNA by reverse transcription through RT-PCR, by using Takara Prime script 1st strand cDNA synthesis kit. A 10 µl quantity of reaction mixture was prepared by taking template RNA from leaf/root, in a microtube in the following proportions.

Reagents	**Volume**
Oligo dT primer (50µm)	1µl
Random 6mer (50µm)	1µl
dNTP mixture (10mM)	1µl
Template RNA (leaf/root)	1.2 µl
RNAase free DH_2O	5.8 µl

The mixture of above reagents including the 1.2 µl of template RNA from leaf/ root were kept for incubation in water bath for 5 min at 65 ºC and cooled the mixture on ice. Later, 20 µl of the reaction mixture was prepared in the following proportions.

Template RNA primer mixture	10 µl
5X prime script buffer	4 µl
RNAase inhibitor (40 u/µl)	0.5 µl
Prime script RNAase (200 u/µl)	1.0 µl
RNAase free H_20	4.5 µl

The above reagents were mixed by gently taping and kept for incubation for 10 minutes at 30 ºC in PCR machine. Later, inactivated the enzyme by incubating at 95ºC for 5 min and cooled it on ice for 10 min.

The cDNA preparations were diluted 12- times with nuclease-free water (Qiagen, Valencia, CA, USA) to use as template in q-PCR analysis. To confirm the total absence of genomic DNA, cDNA was used as a template for PCR amplification using β-*actin, since* the β-*actin* gene was considered a valid option for real-time PCR analysis.

A quantity of 2 µg/µl synthesized cDNA concentration was diluted with milli Q water to obtain 1µg/µl for usage in qRT-PCR.

3. Selection of reference genes and primer design:

The reference candidate gene, β-actin was selected from the previous studies in groundnut (Murphy *et al.*, 2003). Since the genome sequence of groundnut is not known for designing of primers for selected 12 candidate genes, the alignments were made with relevant gene orthologs in *Arabidopsis* using BLASTN with optimization to 'similar sequences' to ensure the primer pairs span at least one intron. Primers were designed using primer analysis software PRIMER 3.0 (http://frodo.wi.mit.edu/primer3/) by considering; (a) product size ranging from 100-160 bp; (b) primer size from: 20-22 bp; (3) GC content 50 %. The accession number, primer sequence and amplicon lengths are listed in Table 1.

4. Quantitative Real-time Polymerase Chain Reaction (qRT-PCR):

All q-PCRs were carried out in Realplex (Eppendorf, Germany) Real-Time PCR system using SYBR Green in 96 well optical reaction plates (Axygen, USA) sealed with ultra-clear sealing film (Platemax). The PCR reaction was performed in a total volume of 10 µL containing 5 µl of 2X SensiFAST^{TM}SYBR No-ROX (Bioline, UK) master mix, 400 nM of each primer, 1.0 µg/µL diluted cDNA and nuclease-free water to make up the final volume. The reaction conditions for the RT-PCR performed were mentioned in Table- 2. No-template controls were included for each primer combinations. Pooled and diluted cDNA sample was used in q-PCR to check the specificity of all the primer pairs and verified by using 2% agarose gel electrophoresis with SYBR safe DNA gel stain (Invitrogen-life technologies, NY, USA). For expression profiling, all the cDNA samples were tested in qPCR with each primer pair and thus performed for each cDNA sample of each biological triplicate. The quantitative cycle (Cq) values were recorded using default settings of Real time PCR system where baseline was corrected automatically and threshold value was estimated by setting to noise band mode. Statistical analysis (mean and CV) of the Cq values was carried out using Microsoft Excel spreadsheet 2010. The PCR efficiency of each primer pair was evaluated by the dilution series method using a pooled cDNA sample of the groundnut variety ICGV 91114.

5. Data analysis:

Expression levels of three candidate genes in all the sample pools were determined by number of cycles needed for the amplification-related fluorescence to reach a specific threshold level of detection (quantification cycle Cq). The efficiency (E) of each primer pair was calculated based on slope of the line ($E = 10^{-1/slope}$) considering an ideal value range of 0.95 to 1.0. To carry out an in-depth data analysis, 13 diverse samples were categorized under experimental sets comprising condition-specific samplings with particular primers enlisted (Table 1) and RT-PCR Experimental conditions (Table 2). Experimental set included the leaf and root tissues of ICGV 91114 groundnut variety at flowering stage.

Table- 1: Gene name, accession number, primer sequence and amplicon length

Name of Gene	Accession No.	Primer sequence	Amplicon size
Nucleoside diphosphate kinase	gi|71040669	F: 5 - TTGTGAGTTTGTGACCGCAG - 3'	113
		R: 5'- TGGAACGAGTGTTTGAGGAA – 3'	
Catalase	gi|332661063	F: 5' -CATCACGCATGAAATCCGCA - 3'	100
		R: 5'-TGCTGAGACTTGAAGAACGAGA – 3'	
Glyceraldehyde-3-phosphatase dehydrogenase	gi|512374222	F: 5'-ATGGTACGACAACGAGTGGG - 3'	101
		R: 5'- CCTTGAATGCTCGAGGAGGG – 3'	
Phosphatase	gi|514772549	F: 5'–TCCAGGGCCTGATGAGTTTA– 3'	110
		R: 5'–CCCTTAGAGCTGCACGAAAT– 3'	
ATP synthase	gi|408899391	F: 5'– GTACTTCAAGTGGGTGACGGA– 3'	249
		R: 5'–AGCTGCAATTGGACCTCTCC– 3'	
Elongation factor -Tu	gi|659129067	F: 5'–TTGTCCTCTTGTTGGTTGGC– 3'	107
		R: 5'– AGCAGCATAAGAGAGGGGAA– 3'	
Phosphoenol pyruvate carboxylase	gi|330255052	F: 5'– CCGAAGGACACCTCCAACTC – 3'	113
		R: 5'– TTAATGCTGTGTCGATGCGG – 3'	
Ferredoxin-NADP reductase	gi|323714331	F: 5'– ATTGAAGAAGGCGGAGCAAT– 3'	120
		R: 5'–CTACAGAATGCAGGACGAGG– 3'	
Glutamine synthetase-2	gi|332006611	F: 5'–TCACCAATCTGGTCAAGAATCTG– 3'	100
		R: 5'–GGCACTCTCATCTGACATGTTG– 3'	
Carbonic anhydrase	gi|15810273	F: 5'– ACCCTGCTTTGTACGGTGAG– 3'	116
		R: 5'– GGCATCTCCTGGCTGAAAGT– 3'	
Ubiquitin	gi|28436472	F: 5'– TCCGGATCAGCAGAGGCTTA– 3'	114
		R: 5'– CCTCTGAGACGAAGCACCAA– 3'	
Methyltransferase	gi|702367154	F: 5'– TGCTTGCTAACTCAAGAGGAGT– 3'	120
		R: 5'– CAATGTGTGCTCGGCCAAC– 3'	
Retrotransposon	gi|108706137	F: 5'–GCTCTCCCCTCTGTCTTCCA– 3'	103
		R: 5'– CATCTGGGCACTTAACCCGT– 3'	

Table- 2: RT-PCR experimental conditions

Gene Name	RT-PCR Experimental conditions: Temp. & duration			
	Denaturation	Annealing	Extension	Final extension
Nucleoside diphosphate kinase	95°C, 30sec	57.4°C, 40 sec	72°C, 1 min 30 sec	72°C, 10 min
Catalase	95°C, 30sec	53°C, 40 sec	72°C, 1 min, 30sec	72°C, 10 min
Glyceraldehyde-3-phosphate dehydrogenase	95°C, 30sec	56°C, 40 sec	72°C, 1 min 30 sec	72°C, 10 min
Phosphatase	94°C, 30 sec	56.5°C, 30 sec	72°C 1 min	72°C, 10 min
ATP synthase	95 °C, 20 sec	60 °C, 20 sec	75 °C, 30 sec	72°C, 10 min
Elongation factor- Tu	95°C, 15 sec	55°C, 30sec	72°C, 30 s	72°C, 10 min
Phosphoenolpyru-vate carboxylase	95°C, 10 sec	55°C, 30 sec	72°C for 30 sec	72°C, 10 min
Ferredoxin-NADP reductase	95°C, 45 sec	60°C, 60 sec	72°C,1min,30 s	72°C, 10 min
Glutamine synthetase	95 °C, 15 sec	60 °C, 1 min	72°C, 30 sec	72°C, 10 min
Carbonic anhydrase	96°C, 30 sec	56°C, 60 sec	72°C, 60 sec	72°C, 10 min
Ubiquitin	95°C, 15 sec	60°C, 30 s	72°C, 30 sec	72°C, 10 min
Methyltransferase	95ºC , 20 sec	56ºC 15 sec	56ºC 20 sec	72ºC 10 min
Retrotransposon	97°C, 30 sec	55°C, 35 sec	72°C 30 sec	72°C 10 min

Results

Objective-4:

Validation of identified drought stress responsive proteins by qRT- PCR:

Among the 12 leaf and 20 root identified proteins by 2D-DIGE and 65 leaf and 67 root proteins identified by LC-MS methods under drought stress (Tables- 4, 5, 6 and 7), only a total of 13 proteins; 3 from 2D-DIGE (S. Nos. 1-3) and 4 from LC-MS (S. Nos. 4-7) expressed commonly both in leaf and root tissues, while 4 from leaf (S. Nos.8-11) and 2 from root (S. Nos. 12-13) (Table- 8) identified by LC-MS were selected for validation with β-actin as a reference gene. All the 13 identified genes exhibited high levels of expression compared to their respective controls indicating their stability under the tested experimental conditions (Table- 3 and Figs 1 and 2).

However, among the 13 identified genes validated, the Nucleoside diphosphate kinase in leaf, and the Phosphoenol pyruvate kinase and TU Elongation factor genes in root exhibited the lowest levels of expression by 1.07 folds, whereas, ATPsynthase exhibited highest levels of expression by 9.84 folds followed by the TU Elongation factor in leaf tissues and GAPDH in root exhibited expression levels by 8.00 folds compared to their respective controls (Table-3 and Figs 1 and 2).

Nucleoside diphosphate kinase in leaf showed 1.60 fold (20% PEG 6000) relative expression followed by 1.51-fold (10% PEG 6000), 1.30 (5%PEG) and 1.07 (15% PEG), whereas, in roots it showed 3.38 fold (15% PEG) relative expression followed by 2.60 (20% PEG), 2.49 (10%PEG) and 1.30 folds (5% PEG) expression (Table-3 and Figs 1 and 2).

In leaf tissues, Catalase showed a relative expression variation by 4.20 fold (15% PEG 60000) followed by 2.60 (20% PEG), 2.29 (10%PEG) and 2.14 (5% PEG), while, in roots it showed relative expression variation by 5.13 folds (20% PEG 6000) followed by 3.03 (15% PEG), 2.29 (10%PEG) and 1.51 folds (5% PEG) (Table-3 and Figs 1 and 2).

In leaf, GAPDH showed 2 fold relative expression variation (10% PEG 6000) followed by 1.6 (5% PEG 6000), 1.3 (20%PEG 6000) and 1.15 folds (15% PEG 6000), whereas, in roots it showed the expression by 8 folds (15% PEG 6000) followed by 3.03 (20% PEG 6000), 1.61 (5%PEG 6000) and 1.23 folds (10% PEG 6000) (Table- 3 and Figs 1 and 2).

Phosphatase in leaf showed relative expression variation 2.14- fold (5% PEG 6000) followed by 1.31-fold (15% PEG 6000), 1.14-fold (10%PEG 6000) and 1.14fold (20% PEG 6000), while in root it showed relative expression variation by 3.48 folds (15% PEG 6000) followed by 1.87 (20% PEG 6000), 1.51 (10%PEG 6000) and 1.41 (5% PEG 6000) (Table- 3 and Figs 1 and 2).

In leaf, Phosphoenol pyruvate kinase showed relative expression variation by 2.82 folds (20% PEG) followed by 2.14 (10% PEG 6000), 1.62 (15%PEG 6000) and 1.41 folds (5% PEG 6000), while in roots, it showed expression variation by 3.03 folds (10% PEG 6000) followed by 1.86 (20% PEG 6000), 1.51 (15%PEG 6000) and 1.07 (5% PEG 6000) (Table- 3 and Figs 1 and 2).

In leaf, TU Elongation factor showed relative expression variation by 8 folds (10% PEG 6000) followed by 5.27 (5% PEG 6000), 3.4 (15%PEG 6000) and 1.74 folds (20% PEG 6000), while in roots it showed expression variation by 6.06 folds (20% PEG 6000) followed by 2.46 (5% PEG 6000), 1.31 (10%PEG 6000) and 1.07 folds (15% PEG 6000) (Table- 3 and Figs 1 and 2).

In leaves, ATPsynthase showed relative expression variation by 9.84 folds (20% PEG 6000) followed by 5.89 (5% PEG 6000), 2.6 (10%PEG) and 1.74 folds (15% PEG), whereas, in roots it showed expression variation by 4.78 folds (15% PEG) followed by 3.73 (20% PEG), 1.86 (5%PEG) and 1.41 folds (10% PEG) (Table- 3 and Figs 1 and 2).

In leaf, Ferridoxin showed relative expression variation by 3.24 folds (5%) followed by 2.46 (15%), 2.29 (20%) and 1.62 folds (10%), while Glutamine synthase showed expression variation by 8.57 folds (20%) followed by 3.24 (10%), 1.74 (15%) and 1.62 folds (5%), whereas, Carbonic anhydrase showed expression variation by 6.49 folds (15%), followed by 3.73 (20%), 3.24 (10%) and 1.51 folds (5%) and Ubiquitin showed relative expression variation by 4.59 folds (10%) followed by 3.29 (15%), 3.24 (5%) and 1.74 folds (20%) (Table- 3 and Fig. 1).

In root, Methyltransferase showed relative expression variation by 3.63 folds (10%) followed by 2.63 (20%), 1.62 (15%) and 1.3 folds (5%), while Retrotransposon showed expression variation by 2.82 folds (15%) followed by 2.29 (10%), 1.74 (20%) and 1.31 (5%) (Table-3 and Fig. 2).

Table- 3: Expression levels of identified drought stress responsive genes in leaf and root tissues by qRT-PCR

S. No.	Name of gene	Level of expression									
		Leaf					Root				
		C	5	10	15	20	C	5	10	15	20
1.	Nucleoside diphosphate kinase	1	1.30	1.51	1.07*****	1.60	1	1.30	2.49	3.38	2.60
2.	Catalase	1	2.14	2.29	4.20	2.60	1	1.51	2.29	3.03	5.13
3.	Glyceraldehyde-3 - phosphate dehydrogenase	1	1.60	2.00	1.15	1.30	1	1.61	1.23	8.00#	3.03
4.	Phosphatase	1	2.14	1.14	1.31	1.14	1	1.41	1.51	3.48	1.87
5.	Phosphoenol pyruvate carboxylase	1	1.41	2.14	1.62	2.82	1	1.07*	3.03	1.51	1.86
6.	Elongation factor- TU	1	5.27	8.00#	3.40	1.74	1	2.46	1.31	1.07*	6.06
7.	ATPsyntase	1	5.89	2.60	1.74	9.84#	1	1.86	1.41	4.78	3.73
8.	Ferridoxin	1	3.24	1.62	2.46	2.29	NA	NA	NA	NA	NA
9.	Glutamine synthase	1	1.62	3.24	1.74	8.57	NA	NA	NA	NA	NA
10.	Carbonic anhydrase	1	1.51	3.24	6.49	3.73	NA	NA	NA	NA	NA
11.	Ubiquitin	1	3.24	4.59	3.29	1.74	NA	NA	NA	NA	NA
12.	Methyltransferase	NA	NA	NA	NA	NA	1	1.30	3.63	1.62	2.63
13	Retrotransposon	NA	NA	NA	NA	NA	1	1.31	2.29	2.82	1.74

N.A.: Expression levels have not been carried out. * lowest expression; # highest expression.

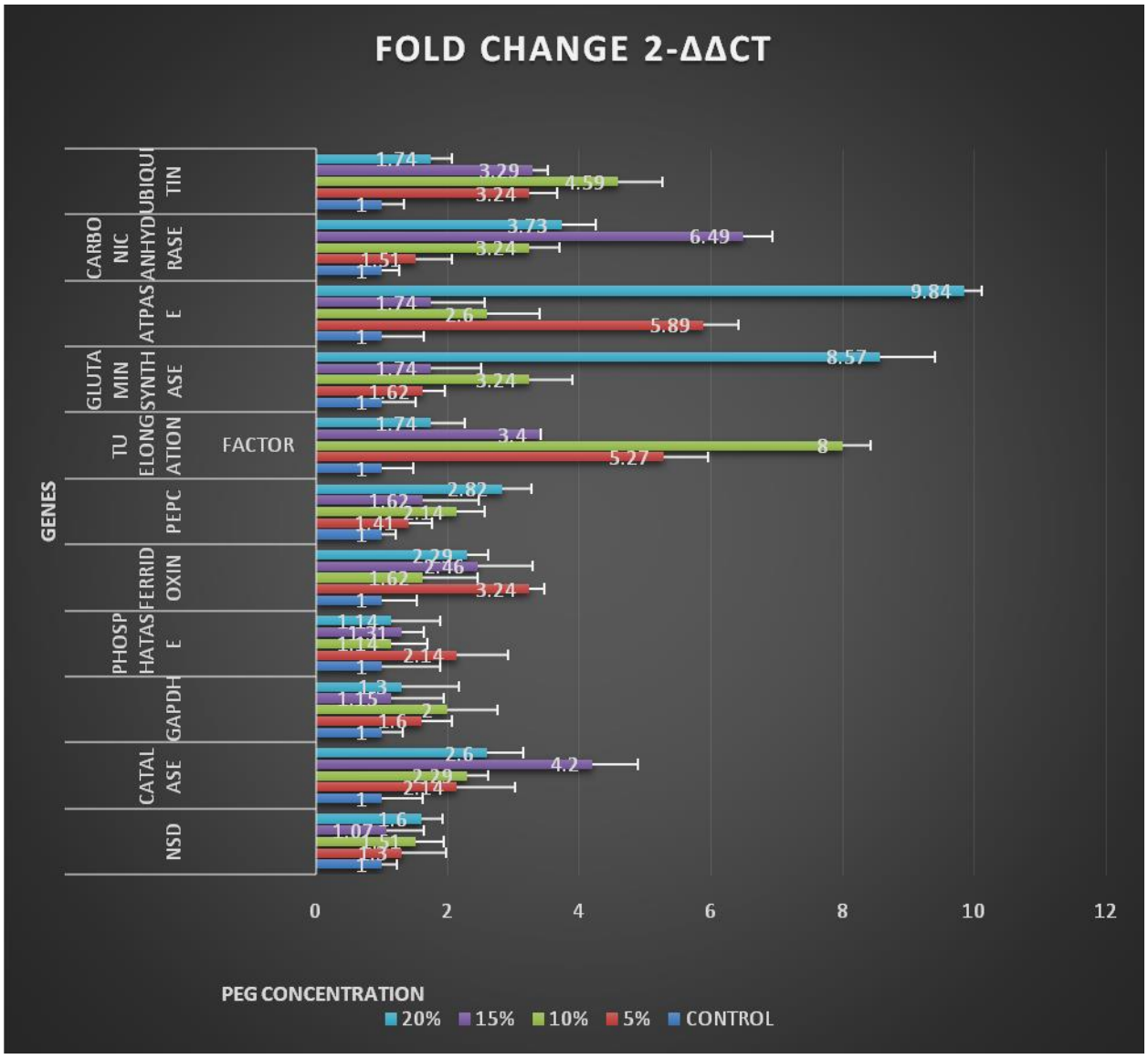


Fig. 1: Expression levels of identified genes in drought stressed leaf tissues by qRT-PCR

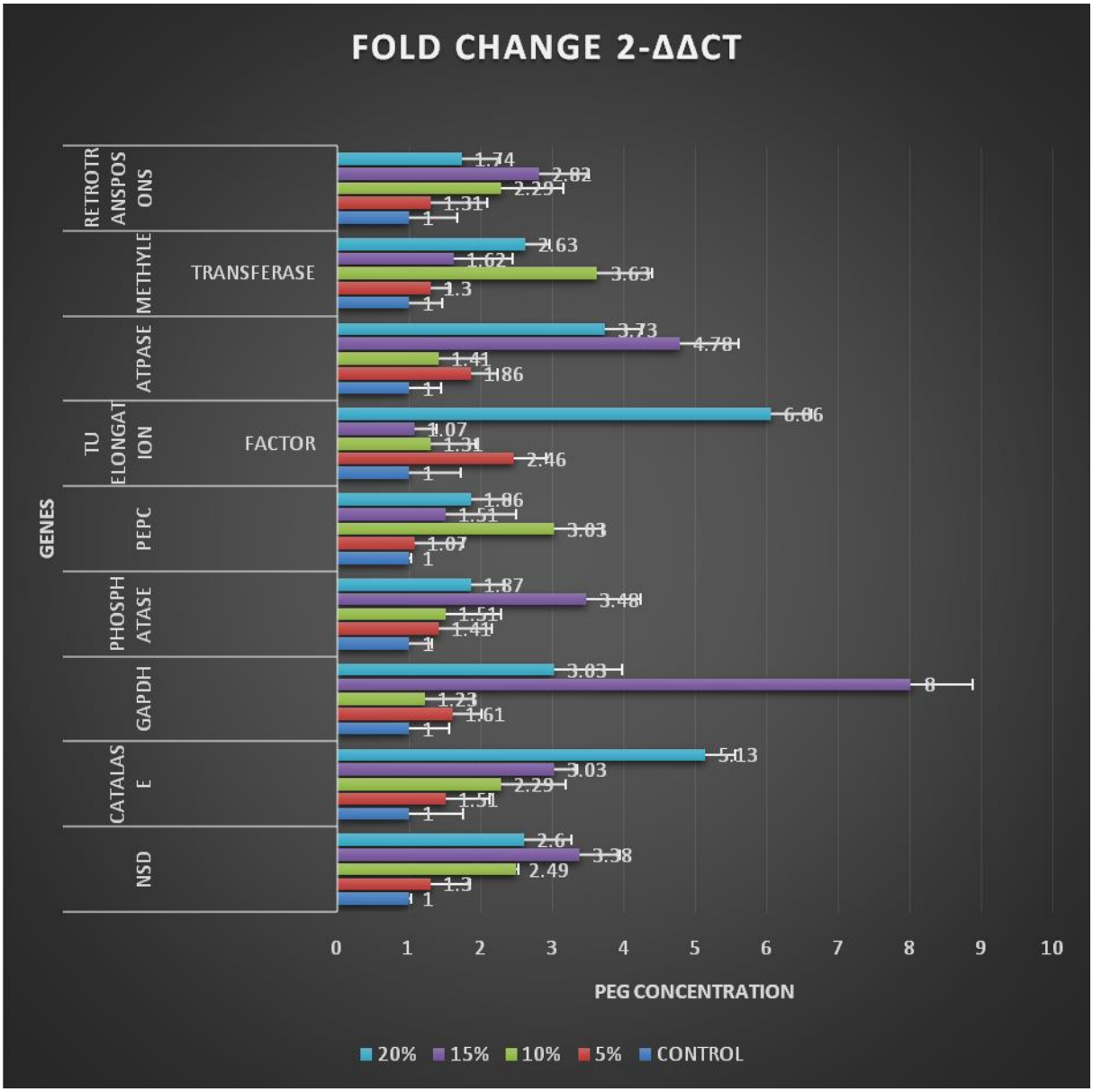


Fig. 2: Expression levels of identified genes in drought stressed root tissues by qRT-PCR

13. ACHIEVEMENTS FROM THE PROJECT:

1. Efficient and reproducible protocols for protein isolation, 2D- electrophoresis and 2D-DIGE were developed and standardized.
2. A total of 280 differentially expressed protein spots were identified from 4 peanut cultivars; ICGV91114, ICGS76, J11 and JL24 by 2DE.
3. Of a total of 30 protein spots selected for PMF analysis, only 12 were sequenced and functionally categorized into 5 groups; molecular chaperons, signal transducers, photosynthetic proteins, defence proteins and detoxification proteins.
4. Among the 12 sequenced, 6 proteins; LEA-1, CalM42, Susy-1, 17.3 kDa Hsp, PSI and SMC-1 protein, were identified for the first time in peanut under drought stress.
5. The results indicated that the PEG stressed seedlings exhibited a significant reduction in the relative water content (RWC), chlorophyll content and RNA content, in proportion to increase in the concentration of PEG, compared to the controls.
6. Of the 1022 differentially expressed proteins identified in leaves and 750 in roots, 12 and 20 proteins respectively were selected for PMF analysis by 2D-DIGE.
7. The molecular function of these 2D-DIGE analyzed proteins indicated their involvement in drought tolerance by way of protein binding, antioxidant activity, sugar binding and kinase activity.
8. A total of 65 proteins from leaf and 67 from root were quantified and characterized by LC-MS/MS.
9. Validation of 13 identified genes by 2D-DIGE and LC-MS/MS employing qRT-PCR (genomic approach) indicated their role in drought tolerance.
10. The results obtained from the proteomic and genomic analysis can be used for better understanding of the mechanisms related to drought response in different crop plants.

14. SUMMARY OF THE FINDINGS (IN 500 WORDS) :

Of a total of 280 protein spots obtained on 2D gel, 189 differentially expressed protein spots were identified by PD Quest Basic software in leaf proteome of all the 4 cultivars; 74 in ICGV 91114, 41 in ICGS 76, 44 in J 11 and 30 in JL 24. Of these, 30 spots were subjected to In-gel trypsin digestion followed by MALDI-TOF, which are functionally categorized into 5 groups: molecular chaperones, signal transducers, photosynthetic, defense and detoxifcation proteins. Of these, 12 proteins were sequenced; LEA protein, calcium ion binding protein, sucrose synthase isoform-1, 17.3 kDa heat shock protein and structural maintenance of chromosome proteins were

overexpressed only in 15 and 20 day- stressed plants of ICGV 91114 cultivar, while cytosolic ascorbate peroxidase was expressed with varying levels in the 10 and 20 day- stressed plants of all the 4 cultivars. Signalling protein like 14-3-3 and defence proteins like alpha-methyl-mannoside-specifc lectin and mannose/glucose-binding lectins were differentially expressed in the 4 cultivars. Photosynthetic protein like Rubisco was down-regulated in the stressed plants of all 4 cultivars while Photosystem-I reaction center subunit-II of chloroplast precursor protein was overexpressed in only 20 day- stressed plants of ICGV 91114, ICGS 76 and J11 cultivars. Further, the 30 identified drought-tolerant proteins including the 12 sequenced differentially expressed proteins play a significant role in conferring drought stress tolerance and could also be used as a reference library for probing the drought-tolerant proteins in other crops.

Further, identification of drought stress tolerant proteins by 2D-DIGE and LC-MS methods and their expression was validated by a genomic method, qRT-PCR. The 40 day- old seedlings of ICGV91114 were subjected to drought stress by 4 different concentrations; 5, 10, 15 and 20% PEG-6000 for 24 h maintaining controls. The results showed that the contents of RWC and RNA were significantly reduced in both leaves and roots with increased concentration of PEG, while a dose dependant decline in ‘a’, ‘b’ and total chlorophyll content was observed with increasing concentration. Reduction in chlorophyll ‘*a*’ was higher than the content of chlorophyll ‘*b*’.

Identification of drought responsive proteins in leaf and root tissues was carried out by gel based 2D-DIGE and non gel based LC-MS methods. By 2D-DIGE method of the 12 leaf proteins; 4 showed increased folds and 2 decreased fold levels, besides 6 unique spots with increase in percent volume, while of the 20 root proteins; 8 showed increased and 3 decreased folds, besides 9 unique spots showed increase in percent volume. A total of 65 leaf and 67 root proteins were identified by LC-MS.

Expression of 13 drought responsive genes; NSD, Catalase, GAPDH, Phosphatase, ATP synthase, efTu, Phosphoenolpyruvate carboxylase, Ferredoxin—NADP reductase, Glutamine synthetase 2, Carbonic anhydrase, Ubiquitin, Methyltransferase and Retrotransposon, identified by 2D-DIGE and LC-MS were validated through qRT-PCR with β-actin as reference gene, proved beyond doubt their positive role in the drought stress tolerance. The understanding of drought tolerant genes generated in the present investigation can be successfully used for breeding drought tolerant crops with improved yields.

15. CONTRIBUTION TO THE SOCIETY (GIVE DETAILS):

Though India is a self sufficient in most of the agricultural produce, still suffering from shortages, especially in oil and pulses. The groundnut is contributing more than 60% oil needs of the country besides providing the need of food, feed, fodder, medical and fertilizer industries. In view of the importance of groundnut crop, enormous shortage of its oil due to various constraints, especially drought stress, an attempt has been made to identify the drought tolerant proteins and thus to the identification of genes by both the proteomic and genomic approaches. Based on the functions of these genes, transgenics in different crops can be developed by enhancing abiotic stress tolerance and they can also be used as markers in molecular breeding. Hence, it is relevant to undertake this research project to meet the needs of the society and for sustainability of the country in the food sector.

16. WHETHER ANY Ph. D. ENROLLED / PRODUCED OUT OF THE PROJECT:

YES. Two Ph. Ds have been produced.

17. NO. OF PUBLICATIONS OUT OF THE PROJECT

Four Research Publications (Reprints are herewith enclosed at the end of the report)

(PRINCIPAL INVESTIGATOR)
(Seal)

(REGISTRAR / PRINCIPAL)
(Seal)

Name of the Institution: Osmania University

Statement showing the Utilization of the Grants received from University Grants Commission, New Delhi, towards Major Research Project on "Isolation.....Groundnut"

By Dr. D. Manohar Rao, Dept. of Genetics, Osmania University, Hyderabad – 500 007

Expenditure Incurred during the Period 01-09-2012 to 31-12-2015

UGC – MRP SANCTIONED LR. No. F. No. 41-828/2012 (SR) Dt.18-07-2012

S. No.	Particulars	Total Allocation Approved	Total Grant Released	Expenditure Incurred	Balance on hand, if any from the Grants Released
1	Honorarium @18,000/- p.m.	6,48,000/-	3,88,800/-	6,48,000/-	-2,59,200/-
2	Contingency	50,000/-	45,000/-	49,000/-	-4,000/-
3	Travel/fieldwork	50,000/-	45,000/-	33,329/-	11,671/-
4	Chemicals & Glassware	2,50,000/-	2,25,000/-	2,49,038/-	-24,038/-
5	Overhead	73,200/-	73,200/-	73,200/-	NIL
Total Rupees		10,71,200/-	7,77,000/-	10,52,567/-	-2,75,567/-

1. Certified that the expenditure is incurred in accordance with the terms and conditions attached to the Grant.
2. Certified that the Grant has been utilized for the purpose for which it was sanctioned.
3. If as a result of Audit/Check some irregularity is noticed at a later stage, action will be taken to adjust refund or regularize the objected amount.
4. Honorarium paid @ 18,000/- p.m. as per the XII Major Research Project Guidelines

Principal Investigator

Principal Investigator
UGC-MRP "Isolation....Groundnut"
Department of Genetics
Osmania University
Hyderabad - 500 007.

Finance Officer
O. U.

FINANCE OFFICER
Osmania University
HYDERABAD

Registrar
O. U.

Registrar
Osmania University,
Hyderabad-500 007.
INDIA

Statutory Auditor

DEPUTY DIRECTOR
STATE AUDIT
Govt. of Telangana,
Osmania University, Hyd.

Annexure - IV

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

Utilization certificate

Certified that an expenditure of Rs. **10,52,567**/- (Rupees **Ten Lakhs Fifty Two Thousand Five Hundred Sixty Seven** only) has been incurred against the Grant of Rs. **7,77,000/-** (Rupees **Seven Lakhs Seventy Seven Thousand** only) released by the University Grants Commission under the scheme of support for Major Research Project entitled "**Isolation and Expression of Drought Tolerant Genes in Groundnut (*Arachis hypogaea* L.)**" to **Dr. D. Manohar Rao, Dept. of Genetics, Osmania University, Hyderabad**, vide UGC letter No. **F. No. 41-828/2012 (SR)** dated **18-07-2012,** has been fully utilized for the purpose for which it was sanctioned and in accordance with the terms and conditions laid down by the University Grants Commission.

Further, Certified that excess expenditure of Rs. **2,75,567/-** (Rupees **Two Lakhs Seventy Five Thousand Five Hundred Sixty Seven** only) is provided by the University which is to be reimbursed by the University Grants Commission.

Principal Investigator	Finance Officer	Registrar	Statutory Auditor
	O.U.	O.U.	

Principal Investigator
UGC-MRP "Isolation....Groundnut"
Department of Genetics
Osmania University
Hyderabad - 500 007.

FINANCE OFFICER
Osmania University
HYDERABAD

Registrar
Osmania University,
Hyderabad-500 007.
INDIA

DEPUTY DIRECTOR
STATE AUDIT
Govt. of Telangana,
Osmania University, Hyd.

STATEMENT SHOWING THE HONORARIUM PAID TO PROF. D. MANOHAR RAO,
PRINCIPAL INVESTIGATOR
UGC-MRP "ISOLATION--------GROUNDNUT"
DEPARTMENT OF GENETICS, OSMANIA UNIVERSITY,
F.41-828/2012 (SR), DATED:18.07.2012.

Name of the Principal Investigator:- Prof. D. Manohar Rao, 01.09.2012 to 31.08.2015

S.No.	Month and year	Honorarium
1	Sep-12	18000
2	Oct-12	18000
3	Nov-12	18000
4	Dec-12	18000
5	Jan-13	18000
6	Feb-13	18000
7	Mar-13	18000
8	Apr-13	18000
9	May-13	18000
10	Jun-13	18000
11	Jul-13	18000
12	Aug-13	18000
13	Sep-13	18000
14	Oct-13	18000
15	Nov-13	18000
16	Dec-13	18000
17	Jan-14	18000
18	Feb-14	18000
19	Mar-14	18000
20	Apr-14	18000
21	May-14	18000

Contd.....2

22	Jun-14	18000
23	Jul-14	18000
24	Aug-14	18000
25	Sep-14	18000
26	Oct-14	18000
27	Nov-14	18000
28	Dec-14	18000
29	Jan-15	18000
30	Feb-15	18000
31	Mar-15	18000
32	Apr-15	18000
33	May-15	18000
34	Jun-15	18000
35	Jul-15	18000
36	Aug-15	18000
Grand Total		**648000**

PRINCIPAL INVESTIGATOR

Principal Investigator
UGC-MRP "Isolation....Groundnut"
Department of Genetics
Osmania University
Hyderabad - 500 007.

REGISTRAR

Registrar
Osmania University,
Hyderabad-500 007.
INDIA

Annexure - VI

UNIVERSITY GRANTS COMMISSION
BAHADUR SHAH ZAFAR MARG
NEW DELHI – 110 002

STATEMENT OF EXPENDITURE INCURRED ON FIELD WORK

Name of the Principal Investigator: Prof. D. Manohar Rao (Retd.)

Name of the Place visited	Duration of the Visit		Mode of Journey	Expenditure Incurred (Rs).
	From	To		
ICRISAT	O.U.	ICRISAT – One Day 03-07-2013	Taxi – Car	2,361/-
ICRISAT	O.U.	ICRISAT – One Day 18-07-2013	Taxi – Car	2,007/-
ICRISAT	O.U.	ICRISAT – One Day 10-01-2014	Taxi – Car	1,681/-
UGC, New Delhi	Hyderabad	New Delhi – Two Days 03-02-14 to 04-02-14	Flight – Air India	18,734/-
ICRISAT	O.U.	ICRISAT – One Day 12-03-2014	Taxi – Car	1,417/-
ICRISAT	O.U.	ICRISAT – One Day 21-05-2014	Taxi – Car	1,485/-
ICRISAT	O.U.	ICRISAT – One Day 02-09-2014	Taxi – Car	1,501/-
ICRISAT	O.U.	ICRISAT – One Day 18-09-2014	Taxi – Car	1,742/-
ICRISAT	O.U.	ICRISAT – One Day 05-06-2015	Taxi – Car	2,401/-
				33,329/-

Rupees Thirty Three Thousand Three Hundred Twenty Nine Only

Certified that the above expenditure is in accordance with the UGC norms for Major Research Projects.

D. Manohar Rao

Principal Investigator

Principal Investigator
UGC-MRP "Isolation....Groundnut"
Department of Genetics
Osmania University
Hyderabad - 500 007.

Registrar/Principal
O.U.

Registrar
Osmania University,
Hyderabad-500 007.
INDIA

3 Biotech (2018) 8:157
https://doi.org/10.1007/s13205-018-1180-8

ORIGINAL ARTICLE

Differential expression of leaf proteins in four cultivars of peanut (*Arachis hypogaea* L.) under water stress

Padmavathi A. V. Thangella[1,3] · Srinivas N. B. S. Pasumarti[2] · Raghu Pullakhandam[2] · Bhanuprakash Reddy Geereddy[2] · Manohar Rao Daggu[1]

Received: 18 June 2017 / Accepted: 20 February 2018

Abstract

Drought is a major constraint to the productivity of many crops affecting various physiological and biochemical processes. Seventy percent of the peanuts are grown in semiarid tropics that are frequently prone to drought stress. So, we analyzed its effect in 4 cultivars of peanut, with different degrees of drought tolerance, under 10 and 20 days of water stress using two-dimensional gel electrophoresis and mass spectrometry. A total of 189 differentially expressed protein spots were identified in the leaf proteome of all the 4 cultivars using PD Quest Basic software; 74 in ICGV 91114, 41 in ICGS 76, 44 in J 11 and 30 in JL 24. Of these, 30 protein spots were subjected to in-gel trypsin digestion followed by MALDI-TOF that are functionally categorized into 5 groups: molecular chaperones, signal transducers, photosynthetic proteins, defense proteins and detoxification proteins. Of these, 12 proteins were sequenced. Late embryogenesis abundant protein, calcium ion binding protein, sucrose synthase isoform-1, 17.3 kDa heat shock protein and structural maintenance of chromosome proteins were overexpressed only in the 15 and 20 days stressed plants of ICGV 91114 cultivar while cytosolic ascorbate peroxidase was expressed with varying levels in the 10 and 20 days stressed plants of all the 4 cultivars. Signaling protein like 14-3-3 and defense proteins like alpha-methyl-mannoside-specific lectin and mannose/glucose-binding lectins were differentially expressed in the 4 cultivars. Photosynthetic protein like Rubisco was down-regulated in the stressed plants of all 4 cultivars while Photosystem-I reaction center subunit-II of chloroplast precursor protein was overexpressed in only 20 days stressed plants of ICGV 91114, ICGS 76 and J11 cultivars. These differentially expressed proteins could potentially be used as protein markers for screening the peanut germplasm and further crop improvement.

Keywords Peanut · Water stress · 2-DE · PMF · MALDI-TOF · Differential expression

Abbreviations

LEA-1	Late embryogenesis abundant protein-1
CalM42	Calcium ion binding protein
Susy-1	Sucrose synthase isoform-1
SMC-1	Structural maintenance of chromosome-1
APX-1	Ascorbate peroxidase-1
PS I	Photosystem-I reaction center subunit-II of chloroplast precursor protein
10 DS	10 days water-stressed plants
20 DS	20 Days water-stressed plants
MALDI-TOF	Matrix-assisted laser desorption/ionization-time-of-flight

Introduction

Peanut is the most important crop for edible oil production and cultivated heavily in almost all the tropical and subtropical countries. The productivity levels of this crop are low due to a number of abiotic factors among which drought is the single largest factor affecting various physiological and biochemical processes (Jaleel et al. 2008; Farooq et al. 2009). Plants start responding and adapting to drought stress at various levels such as morphological, anatomical

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13205-018-1180-8) contains supplementary material, which is available to authorized users.

✉ Padmavathi A. V. Thangella
tpadma.nin@gmail.com

[1] Department of Genetics, Osmania University, Hyderabad, India

[2] National Institute of Nutrition, Hyderabad, India

[3] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA

Published online: 02 March 2018

and molecular level (Chaves et al. 2003; Timperio et al. 2008). Response to drought stress also depends on the type of species and genotypes. The major molecular response to drought stress is altered gene expression in relation to various pathways associated with stress perception, signal transduction and synthesis of stress related compounds (Ramanjulu and Bartels 2002). Plants also respond to stress by synthesizing novel proteins or up-regulating protein expression (Komatsu and Hossain 2013).

Proteomic analyses of various crops revealed numerous drought-stress responsive proteins involved in photosynthesis, oxidative stress response, redox regulation, signal transduction, protein folding and secondary metabolism in rice (Salekdeh et al. 2002; Ke et al. 2009), wheat (Faghani et al. 2015), sunflower (Castillejo et al. 2008), soybean (Kunert et al. 2016). Despite the agronomic and economic importance of peanut, very little is known about its molecular adaptive responses to drought (Luo et al. 2005; Kottapalli et al. 2009). Many biochemical processes in a cell are regulated by protein–protein interactions, post-translational modifications and enzymatic activities that cannot be identified by gene expression studies alone (Morris et al. 2007). Therefore, proteomics is a powerful tool to study the molecular response of plants to biotic and abiotic stresses (Kottapalli et al. 2009). Despite the development of high-density oligonucleotide microarray using 49,205 ESTs available in the public domain, a very little information is known regarding the molecular changes due to drought stress in peanut (Payton et al. 2009). Through genomic approach, Jain et al. (2001) identified 43 differentially expressed peanut transcripts responsive to drought (PTRD). Among them, 12 were completely suppressed under prolonged drought conditions, 2 were down-regulated and 2 up-regulated in response to drought stress. Guo et al. (2006) reported rapid induction of phospholipase D alpha (PLD) gene in the drought-sensitive lines under drought stress than in the drought-tolerant lines. Drame et al. (2007) investigated the involvement of phospholipids, proteases and LEA proteins in conferring drought tolerance in peanut plants and reported that accumulation of putative PLD and LEA proteins leads to increased drought tolerance. Devaiah et al. (2007), using differential display of mRNA transcripts, identified the increased expression of two genes encoding *Arachis hypogaea* serine-rich protein (*AhSrp*) and *Arachis hypogaea* leucine-rich protein (*AhLrp*) under drought stress. Using differential mRNA display, Govind et al. (2009) showed that nearly 700 genes were enriched in subtractive cDNA library in response to gradual water stress. These 700 candidates include genes encoding kinases, transcription factors and phosphatases, late embryogenesis abundant proteins, heat shock proteins, DnaJ like proteins, aldehyde reductase, proline rich protein and defensins, and phytohormones such as brassinosteroids, auxin and cytokinin responsive genes. Pruthvi et al. (2013) carried out expression analysis of few drought stress responsive ESTs from cultivated peanut and reported the association of genes like cyclin T, proline amino peptidase and choline kinase to drought tolerance.

In peanut, proteomic analysis identified differential expression of diverse seed storage proteins in drought-tolerant and susceptible genotypes besides characterization of arachin and methionine-rich proteins from cultivated peanut (Basha and Roberts 1981; Basha et al. 2007). Katam et al. (2007) evaluated genetic diversity in 200 peanut cultivars by studying their seed and leaf protein contents and their response to water stress. Kottapalli et al. (2009) studied the differential expression of leaf proteins during reproductive growth stage under drought stress, in U.S. peanut mini core collections, at maturity phase. They identified 49 non-redundant proteins and reported the overexpression of lipoxygenase and 1L-myo-inositol-1-phosphate synthase proteins in tolerant genotypes under water-deficit stress. Some of the important proteins playing a significant role in drought tolerance include acetyl –CoA carboxylase, several lectins and proteins involved in cellular detoxification, signal transduction and energy metabolism, etc. A reference leaf proteome map was developed in drought-tolerant peanut cultivar, Vemana by Katam et al. (2010). Recently, Katam et al. (2016) have done a comparative study on proteomic profiles of drought-tolerant and susceptible cultivars and also identified 42 unique protein interactions in tolerant cultivar and 20 interactions in susceptible cultivar. They reported that proteins like glutamine ammonia ligase, chitin class II, actin isoform B, and beta tubulin were unique to tolerant cultivar while serine/threonine protein phosphate PP1, choline monooxygenase, peroxidase 43, and SNF1-related protein kinase regulatory subunit beta-2 were not unique but induced upon drought stress in drought-tolerant cultivar. However, the data on protein expression in peanut under drought stress is still limited. Therefore, this study was undertaken with an aim of identifying and characterizing drought stress responsive proteins through 'proteomics approach' by subjecting 4 high yielding and locally adaptable peanut cultivars with varying degrees of drought tolerance, to different stress periods. The main aim of this study is to gain an understanding on the molecular basis of differential response of cultivars to water stress periods for the differential expression of proteins, thereby helping to improve productivity levels by developing high-yielding genotypes that can survive drought stress.

Materials and methods

Plant material

Four varieties of peanut (*Arachis hypogaea* L.), ICGV 91114, ICGS 76, J 11 and JL 24, with varying degrees of drought tolerance were obtained from the International

Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, Andhra Pradesh, India. ICGV 91114 is a high yielding (2.5–3.0 ton/ha), early maturing (90–95 days) bunch variety with an ability to withstand prolonged drought spells, tolerant to mid-season and end-of-season drought. ICGV 76 is another high yielding (1.3–1.8 ton/ha) virginia bunch variety that matures in 120 days and has a good recovery from mid-season drought. J 11 is a spanish bunch type adaptable under a wide range of agro-climatic conditions while JL 24 is also a high yielding (1.5-2.0 ton/ha) spanish bunch variety that matures in 90–95 days and is one of the most popular national varieties sown in areas where end-season drought is common.

Imposition of drought stress

The experiment was laid out in a completely randomized design with three replications for each variety and the seedlings/plants were maintained in the net house covered with water proof sheets during the experiment at the plant Genetics experimental farm, Department of Genetics, Osmania University, Hyderabad. Three independent pots were maintained for each control and stress treatment for each variety, considering each pot as a replication, with 4 seeds per pot. The seeds were sown at a depth of 4 cm in a pot measuring 225 mm height and 450 mm diameter. The pots were filled with a mixture of soil: sand at 2:1 ratio and nutrients like NPK fertilizers were added at a ratio of 2:3:2 for the growth of seedlings. They were watered regularly thrice a week for 20 days until the water stress was imposed. Drought was imposed by withholding water to the 20-day-old seedlings for 10, 15 and 20 days to ICGV 91114 while 10 and 20 days to other three cultivars, maintaining their respective controls. Fully expanded fresh leaf samples were collected at random from each replicate of the stressed plants of all the four varieties along with their respective controls on 31st and 41st days, respectively, quickly frozen in liquid nitrogen and stored at - 80 °C prior to protein extraction.

Relative water content (RWC)

The RWC was estimated in 10, 15 and 20 days water stressed plants along with their respective controls in all the 4 cultivars. After recording the fresh weight of the collected leaves from 31st, 36th and 41st day-old seedlings, they were stored overnight in deionised water, in a refrigerator, at 4 °C and the next day, the turgid weight of the blotted leaves was recorded. Later, the dry weight of leaves, incubated in hot air oven at 60oC for 24 h, was recorded. The leaf RWC was then measured and calculated as per the method proposed by Sharp et al. (1990) using the formula, RWC (%) = [(FW − DW)/(TW − DW)] × 100, where FW is the fresh weight, DW is the dry weight and TW is the turgid weight. The leaf RWC is expressed as a percentage of fully turgid water content.

Protein extraction

Two grams of fresh leaf tissue (pooled from 3 replicates) of each variety was ground to fine powder in chilled mortar and pestle with liquid nitrogen and homogenized twice with 20 ml of cold acetone containing 10% trichloroacetic acid (TCA) and 0.07% β-mercaptoethanol (β-ME) for 30 s. This homogenate mixture was incubated overnight at − 20 °C and then centrifuged at 20,000 g for 30 min. The supernatant was discarded and the pellet was homogenized again with fresh acetone containing 10% TCA and 0.07% β-ME. This pre-protein extraction procedure was repeated 3–4 times until all the pigments from the leaf tissue were removed. Later, this pellet was homogenized with cold acetone containing only 0.07% β-ME without TCA and centrifuged at 10,000 g for 20 min. This step was repeated twice and the pellet was air dried overnight.

Preparation of protein extract for 2-DE

The total leaf proteins were extracted by thorough vortexing and homogenization of the pellet in 5 ml of protein extraction buffer consisting of 8.8 M urea, 2.0 M thiourea, 4% CHAPS (3-[3-cholamidopropyl dimethylammo-o]-1-propanesulfonate), 20 mM Di-thiotheritol (DTT), 10% protease inhibitors cocktail and 0.01% biolytes. This homogenate was incubated at room temperature for one hour, centrifuged at 20,000 g for 30 min and filtered through a 0.45 QM filter. This protein extraction was repeated thrice and all the supernatants were pooled and the protein was estimated using Bradford assay (Bradford 1976).

Two-dimensional polyacrylamide gel electrophoresis (2-DE)

The above purified total leaf proteins were resolved by 2-DE in two steps: in the first dimension by isoelectric focusing (IEF) and in the second dimension by SDS-PAGE. Briefly, 250 g of leaf protein was loaded onto 7 cm IPG strips (Bio-Rad) with pH 4–7 and kept overnight for rehydration. IEF of proteins was performed with 50 mA/strip with the following program: Step-I: 250 V for 20 min. at a linear slope, Step-II: 4000 V for 2 h, Step-III: 4000 V, at a rapid slope for 10,000 Vh, and Step-IV: 4000 V at a linear slope for 30 min. Later, these strips were equilibrated with an equilibration buffer-I (containing 6 M urea, 2% (w/v) SDS, 20% (v/v) glycerol, 0.375 M Tris–HCl (pH 8.8) and 2% (w/v) DTT) for 15 min and then for another 15 min with equilibration buffer-II (containing 6 M urea, 2% (w/v) SDS, 20% (v/v) glycerol, 0.375 M

Tris–HCl (pH 8.8) and 2.5% (w/v) iodoacetamide). At the end of equilibration, the strips were loaded onto the SDS-PAGE for the second dimension. In the second dimension, SDS-PAGE of all the protein samples, along with a pre-stained broad range SDS-PAGE marker (Bio-Rad), was performed using 12% (w/v) polyacrylamide gels with 5% (w/v) stacking gels. Electrophoresis was carried out in a Bio-Rad PROTEAN unit at a constant current of 20 mA/gel in 1-X Tris–Glycine electrophoretic running buffer. After electrophoresis, the gels were fixed in a mixture of 40% methanol and 10% glacial acetic acid, stained with 0.25% colloidal Coomassie Brilliant Blue R-250 (CBB) and de-stained in a mixture of 40% methanol and 10% glacial acetic acid solution.

Image analysis

Protein profiles in the CBB-stained polyacrylamide gels were scanned as digitized images using GS-710 Calibrated Imaging Densitometer (Bio-Rad) with a scan density of 42.3 × 42.3 and saved as tagged image file formats. These protein spots were quantified using PD Quest Basic software version 8.0.1 (Bio-Rad) which includes spot detection, measurement, background subtraction and matching. The protein spots on the polyacrylamide gels obtained from control plants of all the 4 cultivars were differentiated into faint, small and large spots. These control gels were used as reference gels for comparison of spots on the polyacrylamide gels obtained from their respective stressed plants. The quantitative variations in the intensity of protein spots due to CBB staining of the polyacrylamide gels were corrected using normalization parameters and thus the gaussian images were created. These Gaussian images exhibiting the matching between the protein spots of controls and their respective stressed plants, of each of the 4 varieties, were retained and rest of the unmatched spots were removed using spot editing tool. Protein spots across the replicate gels of the control and stressed plants within a variety were subjected to auto spot matching by the 'classic match tool' in all the 4 cultivars. The unmatched spots on the member gels of the stressed plants were added to the reference gel (control gel image). Qualitative analysis, to know the presence of the spots, and quantification, of spots to know their intensity, was performed by employing the 'analysis set manager' tool. For quantification of protein spots in a set of gels within a variety, the outside limits were set as 0.5 and 1.5. The spots showing expression levels < 0.5 were considered as down-regulated > 1.5 as more abundant and between these limits as unchanged or normal spots. On the basis of matching, differential spots were selected and analyzed as described below. Three replicates gels were analyzed for each sample to ensure reproducibility.

In-gel trypsin digestion

The CBB stained 2 DE gels were washed several times with ultra-pure water and protein spots of interest were excised from gels. These spots were de-stained in a solution containing 200 mM NH4HCO3 in 40% acetonitrile (ACN) for 30 min. at 37 °C and dried using speedvac for 15 min. These dry spots were subjected to in-gel trypsin digestion by the addition of 0.5 µg trypsin and reaction buffer consisting of 40 mM NH4HCO3 and 9% ACN and digestion was allowed to continue at 37oC overnight. After digestion, the peptide samples were extracted in 0.1% trifluoroacetic acid (TFA) and 50% ACN and stored at -20oC for MALDI-TOF analysis.

Mass spectrometry analysis

The trypsin digested peptides were loaded into a constricted GELoader tip (Eppendorf, Hamburg, Germany) packed with POROS R2 chromatographic resin (Perseptive Biosystems, Framingham, MA) with 5 mm porosity for desalting and concentration. The columns were equilibrated with 20 ml of 5% formic acid (FA) into which the digested samples were added. Later, the bound peptides were washed with 20 ml of 5% FA and eluted directly onto the MALDI target with 0.5 ml of CHCA solution (5 mg/ml in ACN, 0.1% TFA, 70:30 v/v). These samples were analyzed in an Applied Biosystems 5800 MALDI-TOF Proteomics Analyzer. The instrument was equipped with a nitrogen laser and operated in a positive-ion delayed extraction reflector mode. External calibration was performed using standard peptide/protein mixture. Usually, 250 individual spectra of each spot were averaged to produce a mass spectrum. The atmospheric air was used as collision gas to fragment the peptides and obtain the spectra. Peptide fragmentation was performed using collision-induced dissociation (CID) and 50 laser shots from five sample positions were summed up for each parent ion.

Protein identification and database search

Identification of proteins was performed by searching against the National Center for Biotechnology Information non-redundant (NCBInr) and SWISS-PROT databases using Mascot software (http://www.matrixscience.com) (Perkins et al. 1999, Eriksson et al. 2000) with Viridiplantae (green plants) as the taxonomic category. The MASCOT search compares the experimental data to all the sequences in a database and returns the list of hits with decreasing scores, a measure of reliability of identification. The following parameters were used for database search with MALDI-TOF PMF data: (NCBI nr 20070216, 4626804 sequences, 1596079197 residues; Taxonomy Viridiplantae, 186963 sequences), trypsin as digesting enzyme, 2 missed cleavages allowed,

carbamido-methylation, methionine-oxidation and deamidation (NQ) (variable modifications), monoisotopic mass, peptide mass tolerance at ± 100 ppm, unrestricted protein mass, 1 + peptide charge state. The search scores were represented as probability-based Mowse score − 10*Log (*P*), where *P* is the probability that an observed match was a random event. Protein scores greater than 65 were considered significant ($p < 0.05$) in NCBI and scores > 56 were considered significant in SWISS-PROT database. For a positive identification in MALDI TOF-MS, the peptide score should exceed or equal to minimum significant score. For database search with MS/MS spectra, the following parameters were used: (CDS combined KBMS5.0.20050302, 1967674 sequences, 672312456 residues; Taxonomy Viridiplantae, 177633 sequences), trypsin enzyme, carbamido-methylation, methionine- oxidation, deamidation (NQ) (variable modification), monoisotopic mass value, unrestricted protein mass, peptide mass tolerance at ± 1.0 Da, unrestricted protein mass, + 1, + 2, + 3 peptide charge state with one missed cleavage allowed. Proteins were considered detected if they were identified by more than two peptides per spot.

Predicting sub-cellular localization and functional annotation

Gene ontology was predicted for all the 30 proteins identified through PMF, by searching the identified protein ID in TargetP program (www. cbs.dtu.dk/services/TargetP) (Emanuelsson et al. 2007).

Statistical analysis

The data for RWC were recorded and analyzed through two-way analysis of variance (ANOVA) and the values were expressed as mean and standard error. The significance of the treatment effects was tested at 5% probability level ($P = 0.05$) using Tukeys test, which is one of the Post Hoc multiple comparisons of Two-way ANOVA of Windostat version 8.5.

Results and discussion

After withholding water to 20-day-old seedlings at vegetative phase, ICGV 91114, ICGS 76 and J 11 cultivars started showing the symptoms of wilting after 15 days of water stress while JL 24 variety wilted after 10 days of water stress. Hence, the study was conducted under mild (10 days) and severe (20 days) water stress conditions while in ICGV 91114 cultivar, the leaf proteome was studied at moderate (15 days) stress also to investigate the effect of drought stress on differential expression of proteins in the 4 cultivars (Fig. 1 a, b, c, d).

Effect of drought stress on relative water content (RWC) of leaf

Decrease in RWC is one of the early symptoms manifested during drought stress and is considered to be the best integrated measure of plant water status that represents variations in water potential (WP), turgor potential (TP) and osmotic adjustment (OA) in plant tissues (Rampino et al. 2006; Sanchez-Rodriguez et al. 2010). In our experiment, the 4 varieties evaluated for their drought tolerance exhibited significant reduction in RWC with an increase in the duration of water stress from 10 to 20 days ($p < 0.01$). The highest RWC was observed in ICGV 91114, followed by ICGS 76, J 11 and JL 24 in the 10, 15 and 20-day stressed plants including the controls (Fig. 2). Groundnut is relatively a drought-tolerant crop with improved water use efficiency mechanisms that allow the plant to withstand water stress for certain period of time (Nautiyal et al. 2002). RWC was reported to be one of the best indicators of plant water status in maize (Ritchie et al. 1990), wheat (Valentovic et al. 2006) and tomato (Sanchez-Rodriguez et al. 2010) for separating tolerant and sensitive cultivars. The ability of ICGV 91114 to maintain high RWC might be a resistant mechanism which is a result of more osmotic regulation maintained by accumulation of osmolytes (Padmavathi and Rao 2013). On the contrary, JL 24 variety showed reduction in RWC to a greater extent, attributing to its inability to accumulate osmolytes for longer periods leading to decreased osmotic potential, thereby losing its turgor making it a susceptible variety.

Two-dimensional (2-DE) gel analysis of leaf proteins

On 2-DE gels, the total leaf proteins resolved in the range of 12–100 kDa between 4 and 7 pH with the majority of proteins separating between 20 and 40 kDa and 5.0–6.8 pH. Some proteins with molecular weights > 50 kDa did not show complete resolution resulting in clusters of multiple protein spots which is due to very similar molecular weights and a slight difference in pI values. Most of these clusters were concentrated between 4 and 7 pH with molecular weights in the range of 50 and 75 kDa. Quantification of protein spots on 2-DE gels, using PD Quest software, revealed the presence of 189 differentially expressed protein spots in the stressed plants of 4 cultivars against their controls. In ICGV 91114, a total of 74 differential protein spots were observed. Of them, 20 spots were seen in 10 DS (8 more abundant, 7 down-regulated and 5 new), 5 in 15 DS (3 more abundant and 2 new) and 14 in 20 DS plants (9 more abundant, 4 down-regulated and 1 new). A total of 14 protein spots were found in common among all the stressed plants, however, with variable levels of expression. 6 protein spots (spot No.23, 33, 34, 35, 51and 54) were common between

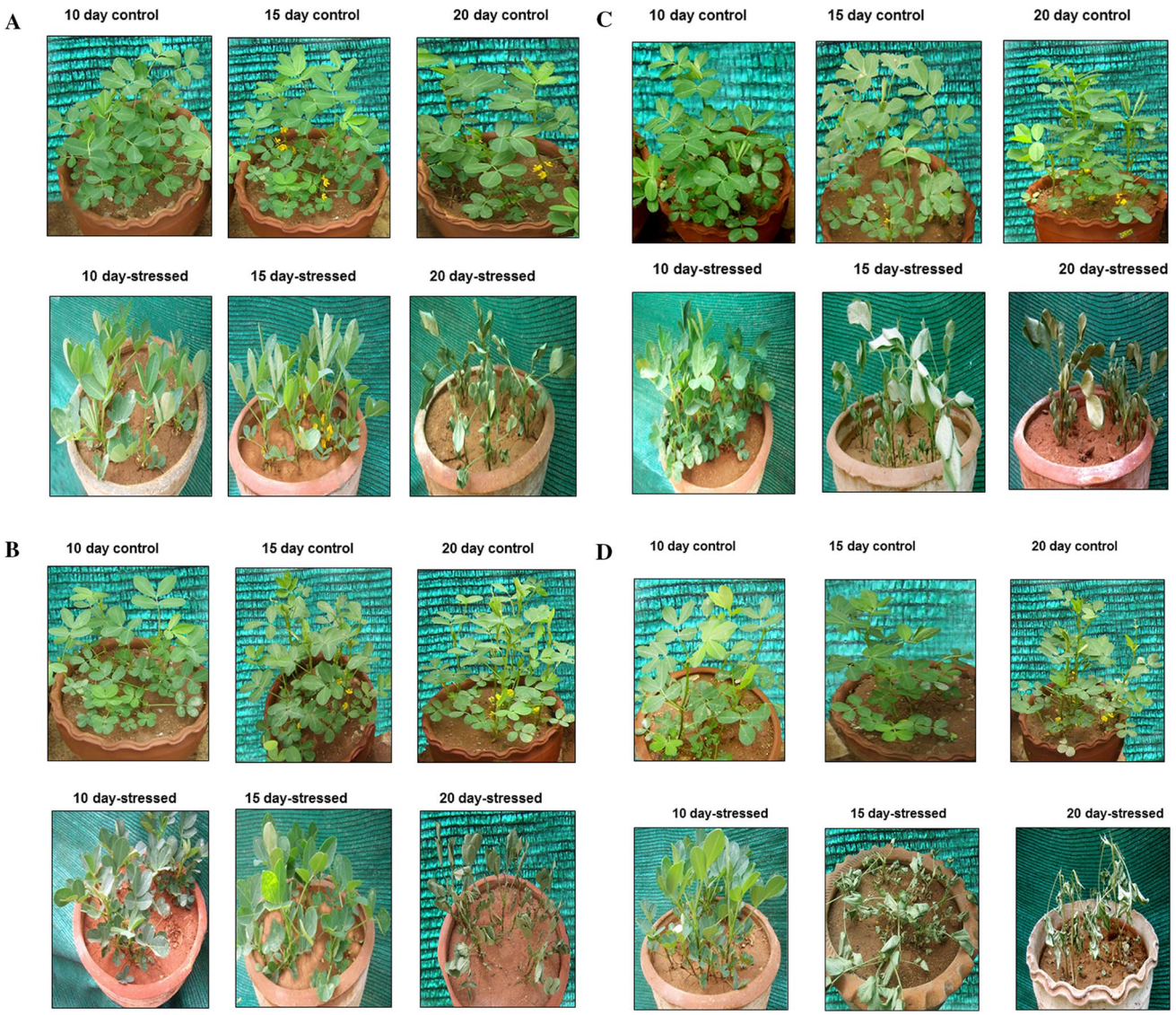


Fig. 1 **a** 10-, 15- and 20-day water stressed plants of ICGV 91114 along with their controls. Drought was induced to the 20-day-old peanut seedlings at vegetative phase under pot culture. The plants could withstand drought up to 15 days of water stress and then started showing the symptoms of wilting. **b** 10-, 15- and 20-day water-stressed plants of ICGS 76 along with their controls. Drought was induced to the 20-day-old peanut seedlings at vegetative phase under pot culture. The plants could withstand drought up to 15 days of water stress and then started showing the symptoms of wilting. **c** 10-, 15- and 20-day water-stressed plants of J11 along with their controls. Drought was induced to the 20-day-old peanut seedlings at vegetative phase under pot culture. The plants could withstand drought up to 10 days of water stress and then started showing the symptoms of wilting in 15-day stressed plants and completely wilted in 20-day stressed plants. **d** 10-, 15- and 20-day water-stressed plants of JL24 along with their controls. Drought was induced to the 20-day-old peanut seedlings at vegetative phase under pot culture. The plants could withstand drought up to 10 days of water stress and then started wilting completely in 15-day stressed plants and drying up in 20-day stressed plants

10 and 15 DS, 14 spots were common between 15 and 20 DS while only protein spot No. 10 is common between 10 and 20 DS (Fig. 3 & Supplementary Table 1). In ICGS 76, a total of 41 protein spots exhibited differential expression. In 10 DS, 17 spots were expressed more abundantly and 12 were down-regulated. Interestingly, 5 new spots; 19,20,21,22,23 appeared while spot No. 29 disappeared. Apart from the 26 common spots (spot No. 1–11, 13, 16, 17, 19, 20, 22, 24, 25, 26, 28, 30, 34, 38, 40 and 41 in 10 DS), in 20 DS, 9 protein spots were expressed additionally that includes 4 more abundant (spot No. 34,35,36 37), 2 down-regulated (spot No. 38, 41) and 3 new proteins (spot No. 31, 32, 33) (Fig. 4 and Supplementary Table 2). In J 11 cultivar, 44 protein spots showed differential expression, out of which 30 spots were expressed in 10 DS that include 9 spots that were down-regulated, 12 spots expressed more abundantly,

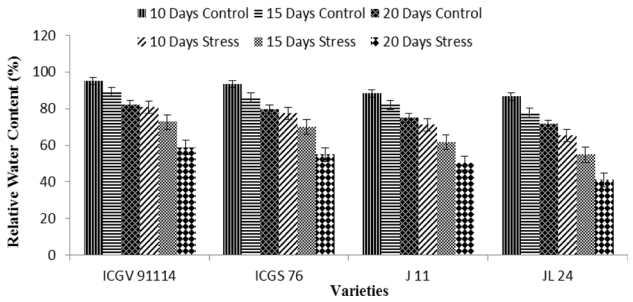


Fig. 2 Relative water content in 4 cultivars of peanut subjected to 10, 15 and 20 days of water stress. Four seeds from each variety were sown in pots filled with a mixture of soil and nutrients in a green house. Pots were watered regularly thrice a week for 20 days until the water stress was imposed. Drought was imposed by withholding water to the 20-day-old seedlings for 10, 15 and 20 days. At the end of experiments, the relative water content of leaves was estimated as described in methods. The bars represent mean ± SD and the bars that do not share common superscript differ significantly ($p < 0.05$)

7 new spots (15,17,19,23, 27, 28 and 30) and 2 spots (25 and 26) that disappeared. Both 10 and 20 DS showed 14 protein spots in common. There was no change observed in the expression of Spot No. 32, 33, 34, 38, 40, 41 and 44 in 10 DS in comparison with controls while in 20 DS, these spots were expressed more abundantly and also all the new spots observed in 10 DS did not show any expression in 20 DS (Fig. 5 and Supplementary Table 3). In JL 24, of the 30 protein spots, 28 spots were expressed in 10 DS (12 more abundant, 16 down-regulated and 2 new) and 2 spots in 20 DS. Among these 30 spots, 14 spots were common between 10 and 20 DS, however, with different levels of expression (spot No. 5,6, 12, 13 and 25 were more abundant in 10 DS while down-regulated in 20 DS. Similarly, spot No. 1,2,3,7,8,9,10, 15, 17, 18, 19, 20, 21, 26, 27 and 30 were expressed only in 10 DS and did not show any sort of expression 20 DS (Fig. 6 and Supplementary Table 4).

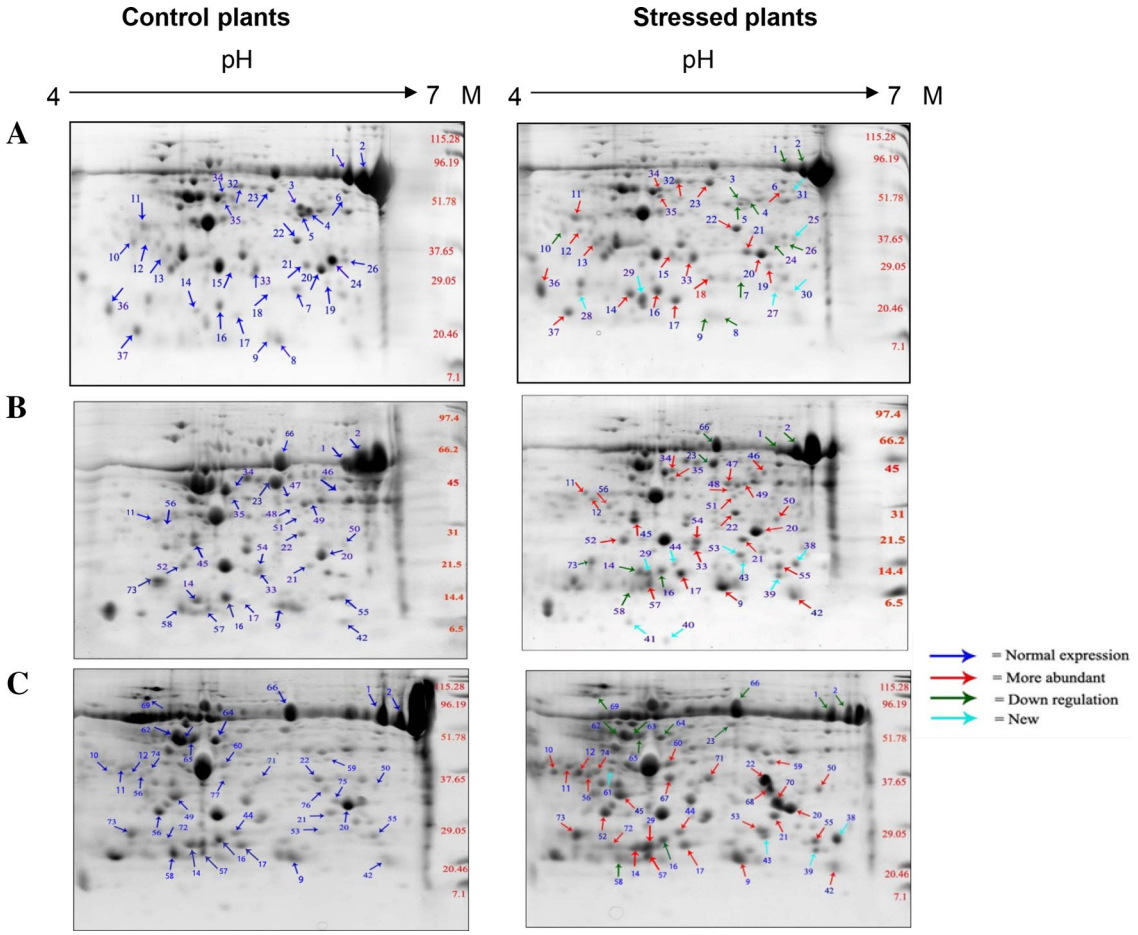


Fig. 3 The 2D gels showing the leaf proteome of peanut cultivar ICGV 91114. The leaf protein extracted from drought stressed and respective control plants from 10 (**a**), 15 (**b**) and 20 days (**c**) was subjected to 2D-gel electrophoresis as described in methods. The images were digitized and quantified densitometrically

Table 1 Peptide mass fingerprinting (PMF) of differentially expressed protein spots in ICGV 91114

Sp. No.	Exptl. M.Wt.	Theo. M.Wt.	Homologous protein	Organism	Biological Process	Molecular function	Cellular component	Mascot score	Accession no.
1	56.0	52.6	RUBP carboxylase large chain	*Cadellia pentastylis*	Photorespiration/ reductive pentose-phosphate cycle	Mono-oxygenase/ RUBP carboxylase activity	Chloroplast/plastid	519	gil1351139
17	16.2	16.2	PR 10 protein	*Arachis hypogaea*	Defense response/ response to biotic stimulus	–	–	65	gil52547774
50	30.0	100.5	Unnamed protein product	*Oryza sativa*	Carbohydrate metabolic process	Hydrolase activity	–	74	gil8099130
38	19.5	92.4	Sucrose synthase isoform 1	*Daucus carota*	Biosynthetic/sucrose metabolism	Sucrose synthase/Glycosyl transferase	–	82	gil1351139
38	19.5	140.8	Structural maintenance of chromosome 1	*Arabidopsis thaliana*	DNA repair/chromosome condensation	ATP binding	Nucleus	73	gil45594277
39	26.5	49.2	Putative serine/threonine protein kinase	*Oryza sativa*	Signaling	ATP binding/protein serine/threo-ne kinase activity	Plasma membrane	51	gil30017556
40	14.4	14.1	Unknown protein	*Arabidopsis thaliana*	–	–	–	80	gil30695529
40	14.4	14.2	Probable myosin heavy chain	*Arabidopsis thaliana*	–	–	–	79	gil25408221
9	31.5	124.8	Gamma response I protein	*Arabidopsis thaliana*	Signaling	Sequence specific DNA binding TF	–	70	gil4678941
9	31.5	20.4	Ubiquitin conjugating enzyme	*Arabidopsis thaliana*	Post replication repair	Acid-aminoacid ligase activity/protein binding	Cytosol	66	gil18403085
47	41.5	66.2	Hypothetical protein	–	–	–	*Oryza sativa*	69	gil51038186
48	19.5	22.3	OSJNBa0029C04.4	DNA integration/RNA dependent DNA replication	RNA binding/ribonuclease H activity	–	*Oryza sativa*	92	gil38346291
53	21.5	36.8	Late embryogenesis abundant protein	–	–	–	*Pisum sativum*	66	gil56709428
55	21.0	76.9	F3F9.11 protein	Detoxification/Regulation of growth	Enzyme binding/glutathione transferase activity	Cytoplasm/Nucleus	*Arabidopsis thaliana*	76	gil8052534
44	21.5	6.5	Stress-induced protein KIN1	Stress response		–	*Arabidopsis thaliana*	50	KIN1_ARATH
74	33.0	47.5	26S protease regulatory subunit 7	Protein catabolic process	ATP binding/nucleoside triphosphatase activity	Cytosol/Plasma membrane/proteosome	*Arabidopsis thaliana*	60	PRS7_PRUPE
70	26.5	21.3	Endoribonuclease Dicer homolog (EC 3.1.26.-)	Plant defense/RNA mediated gene silencing	ATP binding/protein binding/metal ion binding	Cytoplasm/Nucleus	*Arabidopsis thaliana*	73	DICER_ARATH

Table 1 (continued)

Sp. No.	Exptl. M.Wt.	Theo. M.Wt.	Homologous protein	Organism	Biological Process	Molecular function	Cellular component	Mascot score	Accession no.
59	24.0	43.4	GSK-like kinase	–	ATP binding/protein serine/threonine kinase activity	–	*Triticum aestivum*	67	gi|21745456
43	17.3	17.3	17.3 kDa class I heat shock protein	Stress response	–	Cytoplasm	*Glycine max*	70	gi|123534
24	26.5	84.9	Kinesin-3 (Kinesin-like protein C)	Celldivision/mitosis/microtubule based movement	ATP binding/ATPase activity/Microtubule binding/Microtubule motor activity	Cytoplasm/Microtubule	*Arabidopsis thaliana*	66	ATK3_ARATH

Sp.No., spot number, *Expl. M.Wt.*, experimental Mol. Wt., *Theo. M.Wt.*, theoritical M.Wt

Peptide mass fingerprinting (PMF) analysis

Based on the fold change and visual scoring, 30 differentially expressed protein spots; 17 from ICGV 91114, 6 from ICGS 76 and 7 from J 11 were selected and subjected to trypsin digestion to carryout PMF analysis. These digested peptides were injected to MALDI-TOF for obtaining the mass spectra (MS) of each protein. The MASCOT search in NCBInr database for comparing the mass of these peptides with that of the theoretical mass of proteins is deposited in the taxonomy; Viridiplantae (green plants) resulted in the identification of 30 proteins (Tables 1, 2 and 3). Of these, 28 protein spots were found to be with single protein identities while 2 spots, spot No. 38 & 53, were found to contain two proteins each on sequencing. These 30 identified proteins exhibited sequence homology with proteins from *Arabidopsis thaliana* (13), *Oryza sativa* (4), *Arachis hypogaea* (6), *Glycine max* (1), *Pisum sativum* (1) and *Vigna unguiculata* (1), while some proteins showed homology with cereal crops like *Zea mays* (1), *Triticum aestivum* (1) and with other crops: *Cadellia pentastylis* (1) and *Daucus carota* (1) (Fig. 7).

Ontological classification of proteins

The 30 identified proteins were grouped into different categories based on their sub-cellular localization, biological process and molecular function according to the annotation in the Viridiplantae taxonomic database. These proteins when queried using Targetp software revealed their localization in various cell compartments, mostly plastids, cell membrane and nucleus, etc. The data on biological process showed that over 20% of proteins were defense related, 12% stress responsive, 12% signaling, 12% DNA repair, 12% carbohydrate metabolism, 12% photosynthesis, 4% detoxification and other 4% lipid biosynthesis. The molecular function of these proteins indicated their involvement in protein binding, antioxidant activity, sugar binding, kinase activity, etc. (Fig. 8). Of these, 12 proteins were sequenced using MALDI TOF–TOF (Table 4). These 12 proteins were categorized into 5 different groups; (a) molecular chaperons (LEA-1, 17.3 kDa Hsp and structural maintenance of chromosome-1), (b) signal transduction proteins (Susy-1, Calmodulin-binding protein, 14-3-3), (c) photosynthetic proteins (PS I, small and large subunits of Rubisco), (d) defense-related proteins (putative lectins and M/G binding lectins) and (e) detoxification proteins (APX-1). Corresponding spot numbers of the proteins in each variety are provided in (Supplementary Table 5).

Molecular chaperons

Among the 3 molecular chaperons, LEA-1 proteins showed over expression in 15 and 20 DS plants of ICGV 91114.

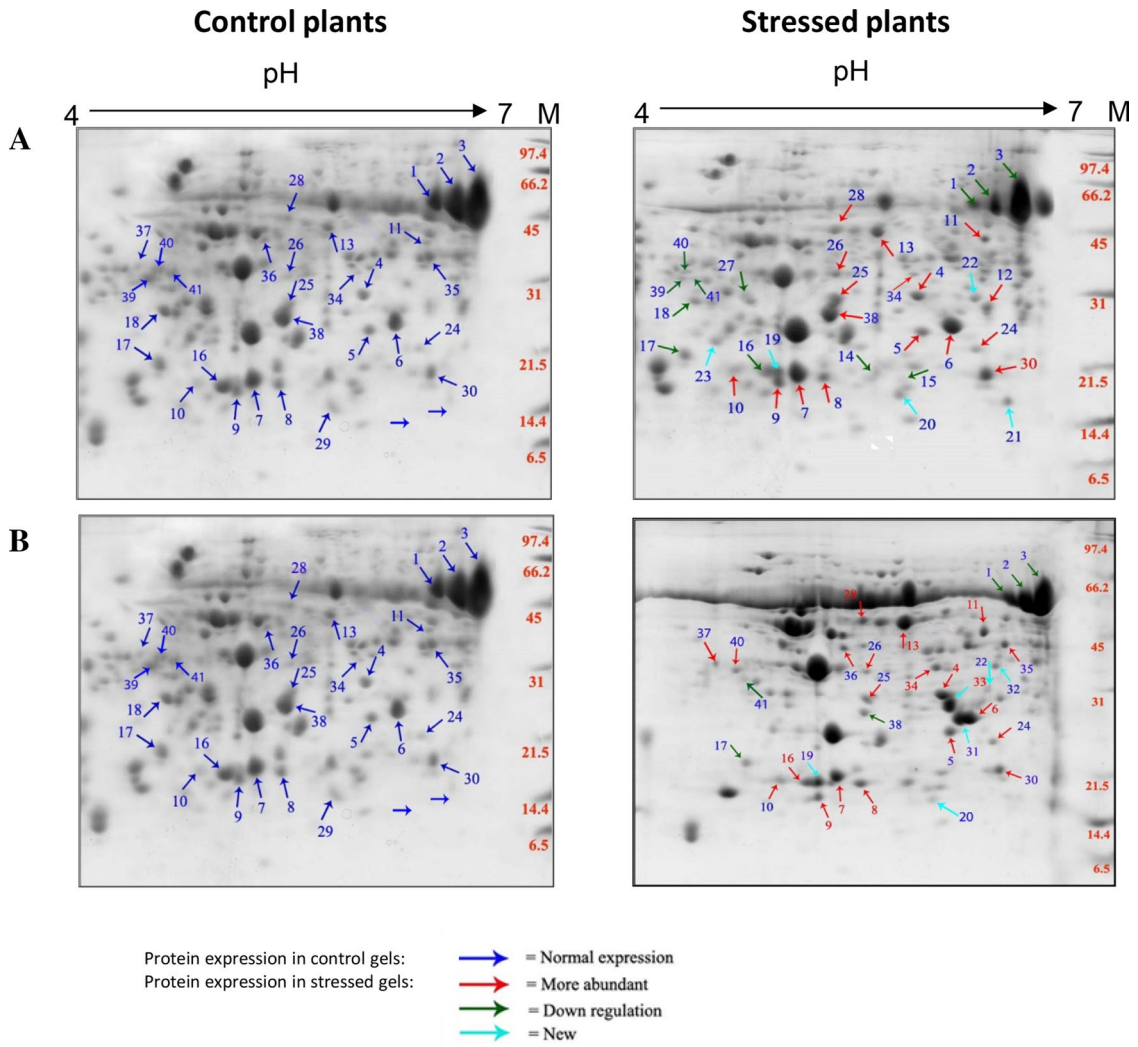


Fig. 4 The 2D gels showing the leaf proteome of peanut cultivar ICGS 76. The leaf protein extracted from drought stressed and respective control plants from 10 (**a**) and 20 days (**b**) was subjected to 2D-gel electrophoresis as described in methods. The images were digitized and quantified densitometrically

Plants adapt to water deficit by the induction of specific genes encoding LEA proteins (Granier 1988). An increased expression of LEA proteins in ICGV 91114 might be attributed to one of the primary defense mechanisms to prevent the loss of intercellular water during drought. During water stress, LEA proteins prevent cellular collapse by adjusting osmoticum along with other hydrophilic proteins and compatible solutes (Xiong and Zhu 2002; Tunnacliffe and Wise 2007) and also act as molecular chaperones (Wise and Tunnacliffe 2004). Another class of stress inducible proteins commonly expressed during environmental stresses are small heat shock proteins (sHSPs). In our study, sHSPs with 17.3 kDa M. Wt. were up-regulated only in the 15 and 20 DS of ICGV 91114 cultivar and the other cultivars did not exhibit any expression. Similarly, Katam et al. (2016) also reported abundant expression of heat shock proteins like disulfide isomerase-2 (PDI2) in drought-tolerant peanut cultivar, Vemana and reduced expression in susceptible cultivar. These proteins were known to assist in protein folding, assembly, translocation and degradation during plant growth and development (Park and Seo 2015). Drought stress causes unfolding and denaturation of proteins making them dysfunctional during which these sHSPs prevent the aggregation of denatured proteins and assists in refolding of non-native proteins by acting as molecular chaperones (Vierling 1991; Boston et al. 1996; Close 1996). Their function may be restored back by the interaction of 17.3 kDa HSPs with osmolytes, cell signaling molecules, cell cycling and cell death regulators (Wang et al. 2004). The other classes of proteins namely SMC-1 protein complexes have multiple functions like sister chromatid cohesion, condensation, repair of eukaryotic chromosomes and are essential

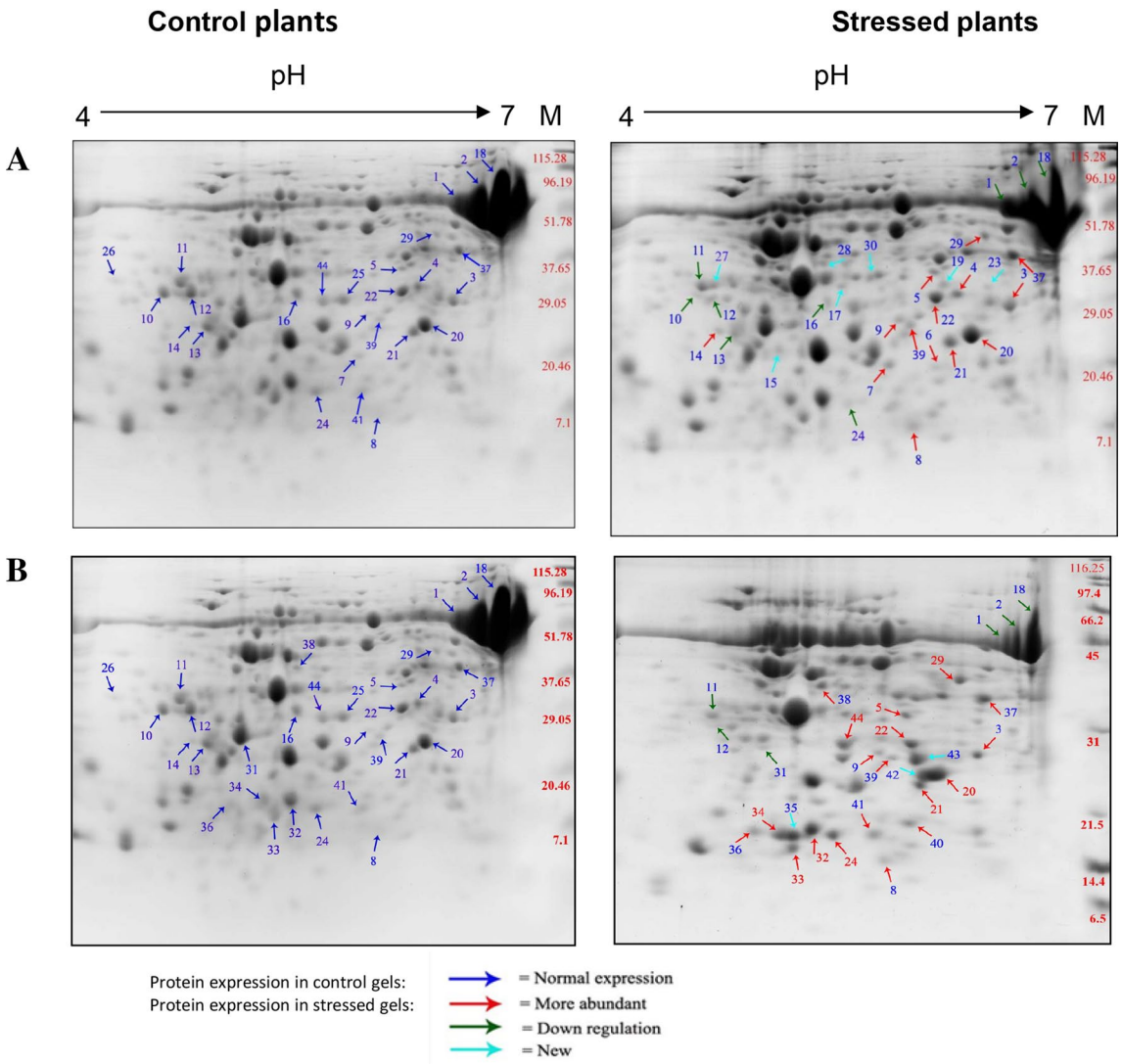


Fig. 5 The 2D gels showing the leaf proteome of peanut cultivar J 11. The leaf protein extracted from drought stressed and respective control plants from 10 (**a**) and 20 days (**b**) was subjected to 2D-gel electrophoresis as described in methods. The images were digitized and quantified densitometrically

for faithful chromosome segregation (Wang et al. 2004; Lehmann 2005). Overexpression of SMC-1 protein in 15 and 20 DS plants of ICGV 91114 implies better protection of its DNA from damages leading to double-strand breaks in DNA and mutations, reduced protein synthesis, cell membrane destruction and damage to photosynthetic proteins caused by reactive oxygen species (ROS) during drought stress (Britt 1999; Nasmyth and Haering 2005).

Signal transduction proteins

The 15 and 20 DS plants of ICGV 91114 exhibited over expression of Susy-1 while all other varieties did not exhibit its expression. It is an important enzyme belonging to the family glycosyltransferase-1 that participates in carbohydrate metabolism, catalyzing the reversible conversion of sucrose and UDP to UDP glucose and fructose in various metabolic pathways (Sebkova et al. 1995; Cooke et al. 2003). Similar results were reported in the phloem cells of leaves and roots of *Arabidopsis* (Strum et al. 1999). Phosphorylation of Susy-1, under severe drought stress, activates sucrose cleavage reaction enabling the plant to meet the increased demand for translocation of carbohydrates under limited ATP supply (Winter and Huber 2000). Calcium ion binding protein (CalM 42), also a signaling protein, was overexpressed only in the 15 and 20 DS of ICGV 91114 and did not show any expression in other 3 cultivars. In *Arabidopsis thaliana,* this protein is known to interact with calcium sensors and aid in trichome branching, the specialized epidermal structures that protect against drought stress (Dejardin et al.

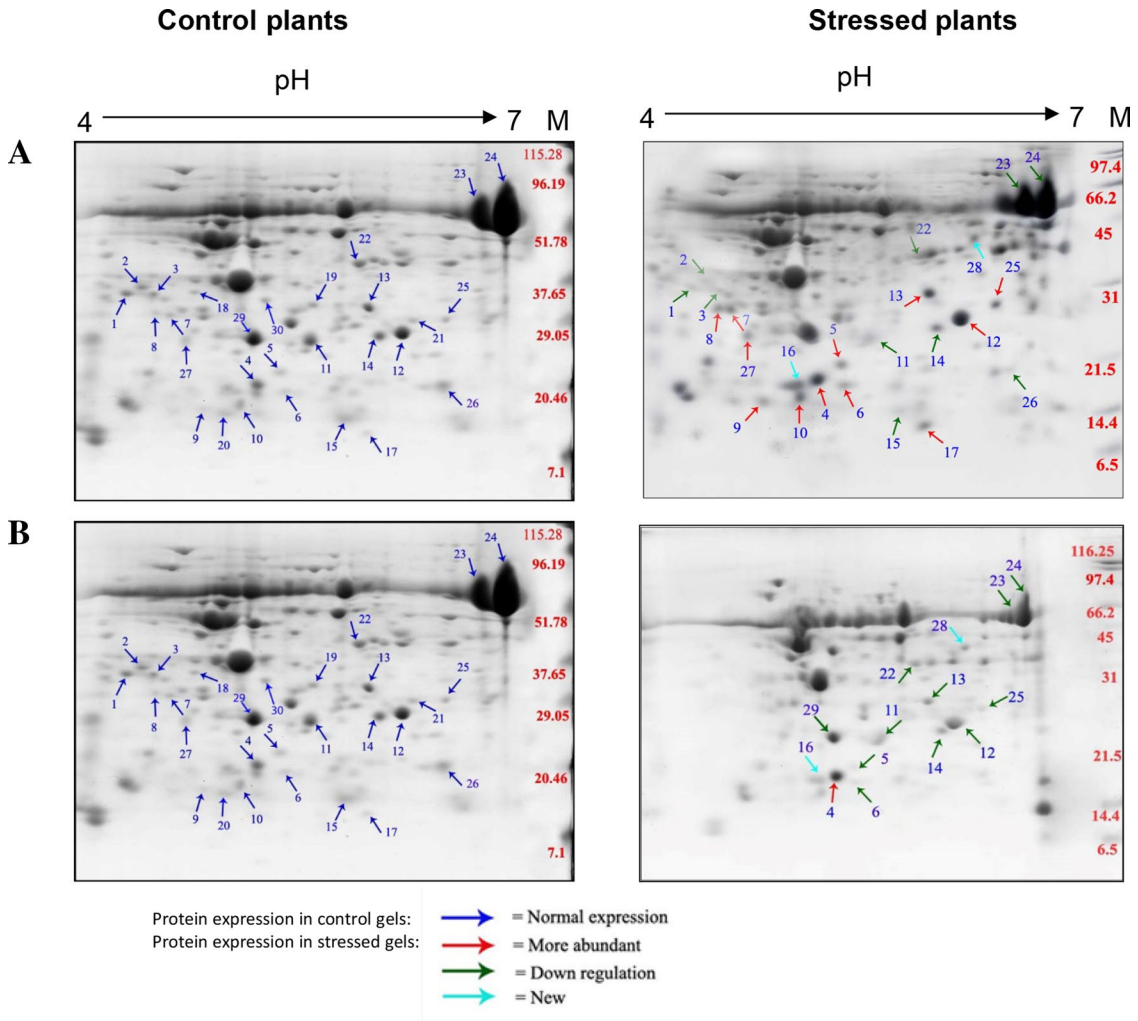


Fig. 6 The 2D gels showing the leaf proteome of peanut cultivar JL24. The leaf protein extracted from drought stressed and respective control plants from 10 (**a**) and 20 days (**b**) was subjected to 2D-gel electrophoresis as described in methods. The images were digitized and quantified densitometrically

1999; Schellmann and Hulskamp 2005; Ishida et al. 2008). The most significant class of signaling proteins, 14-3-3 proteins, were expressed more abundantly in 10, 15 and 20 DS of ICGV 91114. In ICGS 76, this protein was expressed in minute quantities compared to their controls in 10 DS and expressed more in 20 DS while it is down-regulated in 10 and 20 DS of J11. In the susceptible cultivar, JL24, the expression of this protein is almost negligible in 10 DS while 20 DS did not exhibit any expression. Katam et al. (2016) also reported the expression of these 14-3-3 proteins at higher levels in drought-tolerant peanut cultivar. These 14-3-3 proteins are a part of conserved regulatory protein family that binds to serine/threonine-phosphorylated residues expressed in all the eukaryotic cells with a striking ability to bind to signaling proteins like H + ATPases, protein kinases, phosphatases and transmembrane receptors that assist in regulation of stress responsive proteins (Ferl 1996). Binding of these 14-3-3 proteins to other phosphorylated drought stress responsive proteins might result in rapid adaptation of enzymatic activities and metabolic pathways under water stress (De Vetten and Ferl 1994).

Photosynthetic proteins

Water stress is one of the most important environmental factors inhibiting photosynthesis due to damage of chlorophyll pigments, thereby reducing their light harvesting capacity (Graan and Boyer 1990). During photosynthesis, chlorophyll pigments in photosystem-II are excited by sunlight releasing electrons and energy in the form of an ATP molecule required for the break down of water molecule (H_2O) into ½O_2 and $2H^+$ maintaining an electron gradient which is further excited in photosystem-I to produce $NADP^{++}$ H^+. This highly energetic NADPH molecule is

Table 2 Peptide mass fingerprinting (PMF) of stress-induced proteins in ICGS 76

Sp. No.	Exptl. M.Wt.	Theo. M.Wt	Homologous protein	Organism	Biological process	Molecular function	Cellular component	Mascot score	Accession no.
4	30.5	27.2	Cytosolic ascorbate peroxidase	*Vigna unguiculata*	Hydrogen peroxide stress response	Oxidoreductase/ peroxidase activity	Cytoplasm	113	gil1420938
8	20.0	16.2	PR10 protein	*Arachis hypogaea*	Defense response/ response to biotic stimulus	–	–	76	gil52547774
15	17.0	71.4	FACT complex subunit SSRP1	*Zea mays*	DNA repair/ DNA replication/regulation of transcription- DNA dependent	DNA binding	Chromosome/ nucleus	52	SSRP1_ MAIZE
19	21.0	16.2	PR10 protein	*Arachis hypogaea*	Defense response/ response to biotic stimulus	–	–	52	gil52547774
30	30.0	30.1	Lipoate protein ligase-like protein	*Arabidopsis thaliana*	Lipoate biosynthetic process/ Protein modification process	Octanoyl transferase activity	Chloroplast	38	gil7939551
31	27.0	22.3	Photosystem I reaction center subunit II, chloroplast precursor	*Arabidopsis thaliana*	Photosynthesis	Protein binding	Chloroplast membrane	61	PSAD2_ ARATH

Sp.No., spot number, *Expl. M.Wt.*, experimental Mol. Wt., *Theo. M.Wt.,* theoritical M.Wt.

then fed into the Calvin cycle to carry out carbon fixation. Hence, overexpression of Photosystem-I reaction center subunit-II of chloroplast precursor protein in the 20 DS plants of ICGV 91114, ICGS 76 and J 11 varieties might help in the uninterrupted supply of NADPH during photosynthesis to ensure the supply of substrates to carbon skeleton for various metabolic pathways conferring increased drought tolerance. However, the decline in the intracellular levels of CO_2 might create an oxidative stress leading to the destruction of photosynthetic apparatus, changes in the conformation of chloroplast proteins, imbalances in the ions and other macromolecules making JL 24 the most susceptible (Lauer and Boyer 1992; Bray et al. 2000). Ribulose-1,5-biphosphate carboxylase/oxygenase (RuBisCo) is the key photosynthetic enzyme that plays a significant role in the fixation of CO_2 in C3 plants which can be quickly remobilized under stress (Jensen and Bahr 1977). Many studies reported the reduced expression of RuBisCo under drought stress (Parry et al. 2002) while the studies in the drought-tolerant peanut cultivar, Vemana by Katam et al. (2016), reported high abundance of RuBisCo under stress conditions. Kottapalli et al. (2009) reported the reduction in photosynthesis-related proteins in tolerant genotypes during water stress which includes proteins like Rubisco LSU and SSU, oxygen evolving enhancer protein of PS II and chlorophyll a/b binding proteins. However, in our study, both larger and smaller subunits of Rubisco exhibited down-regulation in all the stressed plants of 4 varieties compared to their respective controls while PS I reaction centre subunit II protein showed overexpression in severely stressed plants of ICGV 91114, ICGS 76 and J11 cultivars. Overproduction of reactive oxygen species (ROS) during drought stress cleaves the large subunit of Rubisco directly or modify it to become more susceptible to proteolysis (Feller et al. 2008).

Table 3 Peptide mass fingerprinting (PMF) of water stress-induced proteins in J 11

Sp.No.	Exptl. M.Wt.	Theo. M.Wt.	Homologous protein	Organism	Biological process	Molecular function	Cellular component	Mascot score	Accession no.
8	8.64	15.6	RUBP carboxylase small chain precursor	*Phaseolus vulgaris*	Photosynthesis	*Lyase/Monoxygenase*	Plastid	181	gil123534
10	31.0	30.7	Alpha-methyl-mannoside-specific lectin	*Arachis hypogaea*	–	Cytokinin binding	–	283	gil15233402
11	34.0	29.2	14-3-3 protein	*Vigna angularis*	–	Protein domain specific binding	–	459	gil45594277
12	32.0	28.3	Mannose/Glucose-binding lectin precursor	*Arachis hypogaea*	–	Sugar binding	–	233	gil8099130
13	27.5	21.8	2-cys peroxiredoxinlike protein-	*Hyacinthus orientalis*	Stress response	Antioxidant activity/peroxiredoxin activity	Chloroplast/plastid	96	gil47027073
24	21.0	16.2	PR10 protein	*Arachis hypogaea*	Defense response/response to biotic stimulus	–	–	76	gil52547774
32	19.5	16.2	PR10 protein	*Arachis hypogaea*	Defense response/response to biotic stimulus	–	–	52	gil52547774

Sp.No., spot number, *Expl. M.Wt.*, experimental Mol. Wt., *Theo. M.Wt.*, theoritical M.Wt.

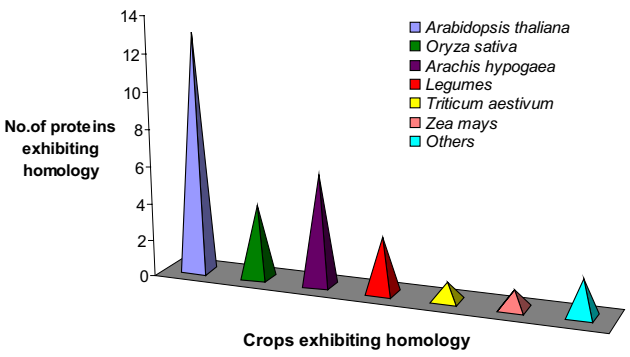


Fig. 7 The homology of differential peanut proteins with proteins from other crops listed in NCBInr *Viridiplantae* Database

Defense related proteins

Lectins are carbohydrate-binding proteins that specifically recognize diverse sugars and mediate a variety of biological processes (Lis and Sharon 1998). Several studies indicated the induction of different lectins with diverse functions from cell wall modification to regulation of gene expression in response to both abiotic and biotic stresses (Vijayan and Chandra 1999; Van Damme et al. 2004). In our study, alpha-methyl-mannoside-specific lectin was overexpressed only in the 20 DS plants of ICGV 91114 and either did not exhibit (or) showed almost negligible expression in other three cultivars. Similarly, the other classes of lectins like mannose/glucose-binding lectins also showed increased expression in all the stressed plants of ICGV 91114 cultivar only and showed reduced expression/negligible in other 3 cultivars. However, no expression is seen in 20 DS of JL24 cultivar. Similarly, Bhushan et al. (2007) reported expression

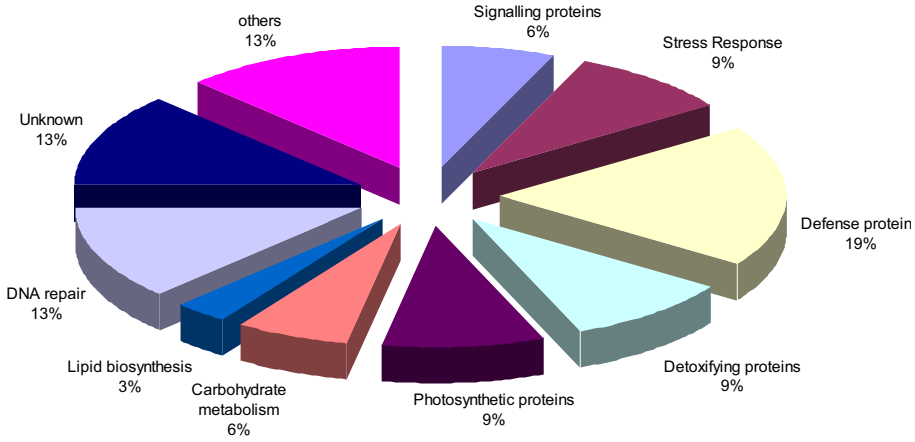


Fig. 8 Functional cataloguing of drought stress responsive proteins

of mannose lectin in stress-tolerant chickpea cultivar under water-deficit stress. In contrast, Kottapalli et al. (2009) reported overexpression of lectins like galactose-binding and mannose/glucose-binding isoforms only in the susceptible genotype and undetectable in tolerant genotypes. Lannoo and Van Damme 2014 in their extensive research on lectins reported that plants express minute amounts of specific lectins only upon environmental stresses, while most of the lectins involved in plant defense are constitutively expressed in high amounts in seeds and vegetative storage tissues.

Detoxification proteins

Cytosolic ascorbate peroxidase-1 (APX-1) enzyme that plays a key role in regulation and signaling of H_2O_2 in plant cells is yet another protein that accumulated in larger quantities in all the stressed plants of ICGV 91114, ICGS 76 and J 11 varieties under drought stress. During drought stress, ROS produced in the chloroplasts and peroxisomes traverse to the nuclei through cytosol (Davletova et al. 2005; Asada 2006; Van Breusegem and Dat 2006). APX-1 is required for the protection of chloroplasts against ROS and in its absence, the ROS scavenging machinery is significantly compromised (Mittler et al. 2004). Grimplet et al. (2009) reported that increased production of ascorbate by APX activity restored the oxidation levels during stress in the drought-tolerant cultivar. APX was significantly down-regulated in 20-day stressed plants of JL 24 leading to the accumulation of hydrogen peroxide in the cytosol triggering cell injury and death and thus making this variety the most drought susceptible. Similarly, Katam et al. (2016) reported negative interaction of APX in the drought susceptible peanut cultivar.

Comparative analysis of expression levels of identified proteins in 4 varieties

The following proteins: cytosolic ascorbate peroxidase, photosystem I reaction center subunit II of chloroplast precursor, alpha-methyl-mannoside-specific lectin, mannose/glucose binding lectin and 14-3-3 proteins were identified and analyzed for their differential levels of expression in 10 and 20 DS plants of all the 4 varieties. There was an increase in the expression levels of APX protein by 1.5 folds indicating its up-regulation in 10 DS plants compared to their respective controls in all the 4 varieties, while in 20 DS plants of ICGV 91114, ICGS 76 and J 11 there was an increase by twofold exhibiting over expression, whereas in JL 24 this protein exhibited a significant down-regulation. Photosystem I reaction center subunit II of chloroplast precursor protein exhibited an increased expression by > 2 folds in 20 DS plants of 3 varieties while this protein did not express in rest of the treatments and controls (Fig. 9a). The alpha-methyl-mannoside-specific lectin exhibited upregulation only in the 20 DS plants of ICGV 91114 and mannose/glucose binding lectin proteins exhibited upregulation in the 10, 15 and 20 DS of ICGV 91114 while 14-3-3 proteins exhibited upregulation in 10, 15 and 20 DS of ICGV 91114 and only 20 DS of ICGS 76 (Fig. 9b).

Conclusion

Considering the prominent role peanut crop plays in our economy as an oilseed crop, identification of genotypes tolerant to drought stress plays a significant role in improving productivity levels. The 30 drought-tolerant proteins identified in the present investigation could be used as a reference library for probing the drought-tolerant proteins in other crops. The twelve proteins that were sequenced namely LEA-1, calcium ion binding protein, Susy-1, 17.3 kDa heat shock protein, SMC-1, cytosolic ascorbate peroxidase, 14-3-3 proteins, alpha-methyl-mannoside-specific lectin, mannose/glucose-binding lectins, PS I and Rubisco were reported to confer a significant role in drought stress from previous studies and also the current study. A repository of these proteins provides a basis for exploring

Table 4 MS/MS sequences of MALDI-TOF identified drought-tolerant proteins in peanut

Sp. No.	Exptl. M.Wt.	Protein identified	Accession no.	Mascot score	Sequence coverage	Matching peptide sequence in red
53	21.5/5.8	Late embryogenesis abundant protein	gil56709428	66	50	MASRQDRREA RAEADARRAA EEIARAR**DER VMQAEVDAR**S AADEIARARA DR**GAATM-GAD TAHHAAGGGG ILESVQEGAK** SFVS-AVGR**TF GGARDTAAEK TSQTADATRD K**LGEYKDYTA DK**ARETNDSV AR**KTNETADA SRDK**LGEYKD YTADKTR**ETK DAVAQK**ASDA SEATKNK**LGE YKDALARK**TR DAKDTTAQKA TEFKDGVK**AT AQETRDATAD TARKAK**DATK DTTQTAADK**A RETAATHDDA TDKGQGQGLL GALGNVTGAI KEK**LTVSPAA TQEHLGGGEE RAVK**ERAAEK AASVYFEEKD RLTRERAAER **VDKCVEKCVE GCPDATCAHR** HGKM
53	21.5/5.8	Calcium ion binding	gil15233402	75	24	**MESNNNEKKK VARQSSSFRL RSPSLNALRL QR**IFDLFDKN GDGFITVEEL SQALTRLGLN ADLSDLKSTV ESYIQPGNTG LNFDDFSSLH KTLDDSFFGG ACGGGENEDD PSSAAENESD LAEAFKVFDE NGDGFISARE LQTVLK**KLGL PEGGEMERVE K**MIVSVDRNQ DGRVDFFEFK NMMRTVVIPS S
38	19.5/6.3	Sucrose synthase isoform 1	gil1351139	82	33	**MGEPVLTRVH SLRERMDSTL ANHRNEILMF LSR**IESHGK**G ILKPHQLLAE YEAISKEDKL KLDDGHGAFA EVIK**STQEAI VSPPWVALAI RLRPGVWEYV RVNVHHLVVE ELSVPQYLQF KEELVIGSSD ANFVLELDFA PFTASF-PRPT LTKSIGNGVE FLNRHLSAKM FHGK-**DSMHPL LEFLRLHNYN GKTLMLNNRV QNVNGLQSML R**KAGDYLSTL PSDTPYSEFE HKFQEIGFER GWGDTAERVT EMFHMLLDLL EAPDASTLET FLGKIPMVFN VVILSPHGYF AQENVLGYPD TGGQVVYILD QVPALER**EMI KRIK**EQGLDI KPR**ILIVTRL LPDAVGTTCN QR**LEKVFGAE HAHILRVPFR **TEKGILRK**WI SRFEVWPYIE TFTEDVAKEI ALELQAKPDL IIGNYSEGNL VASLLAHKLG VTQCTIAHAL EKTKYPDSDI YWEKFDKKYH FSSQFTADLI AMNHTDFIIT STFQEIAGSK **DTVGQYESHT AFTMPGLYR**V VHGIDVFDPK F-VSPGADT SVYFSYKEKE KR**LTTLHPEI EELLYSSVEN EEHLCIIKDK NKPILFTMAR** LDNVKNLTGF VEWYAK**SPKL RELVNLVVVG GDR**RK**ESKDL EEQAQMKKMY ELIDTYKLNG QFRWIS-SQMN RVR**NGELYRY IADTKGAFVQ PAF-YEAFGLT VVEAMTCGLP TFATLHGGPA EIIVHGKSGF HIDPYHGEQV AELLVNFFEK CKTDPSQWDA ISAGGLKRIQ EKYTWQIYSE RLLTLAGVYG FWK**HVSKLDR LEIR**R**YLEMF YALK**YRKLAE SVPLAKDE

Table 4 (continued)

Sp. No.	Exptl. M.Wt.	Protein identified	Accession no.	Mascot score	Sequence coverage	Matching peptide sequence in red
38	19.5/6.3	Structural maintenance of chromosome 1	gi|45594277	73	31	M**PAIQSPSGK ILQLEMENFK SYK**GHQLVGP FKDFTAIIGP NGSGKSNLMD AISFVLGVRT GQLRGSQLKD LIYAFDDRDK EQRGRKAFVR **LVYQMDDGVE LRFTR**SITSA GGSEYRIDNR **VVNLDEYNGK LR**SLGILVKA R**NFLVFQGDV ESIASKNPK**E LTGLLEEISG SEELKKEYEG LEEKKASAEE K**AALIYQKKK** TIGNEKKLKK AQKEEAEKHL RLQEELKALK RERFLWQLYN IENDIEK**ANE DVDSEKSNRK** DVMRELEKFE REAGKRKVEQ AK**YLKEIAQR EK**KIAEKSSK LGKIQPELLR FKEEIARIKA KIETNRKDVD KRKKEK**GKHS KEIEQMQKSI KELNK**KMELF NKK**RQDSSGK LPMLDSQLQD YFR**LKEEAGM KTIK**LRDEHE VLERQR**RTDL EALRNLEENY QQLINRKNDL DEQIKR**FKDR QGEIETSSSK** YKNETTSLKT ELRALQEKHV NAREASAKLK TRIAELEDQL SDLTAERYEN ERDSRLTQAV ESLK**RLFQGV HGRMTDLCRP NRKKYNLAVT VAMGR**FMDAV VVEDENTGKD CIKYLKEQRL PPMTFIPLQS VRVKQVFERL RNLGGTAKLV FDVIQFDPEL EKAVLYAVGN TLVCDELEEA K**VLSWSGER**F KVVTVDGILL TK**AGTMTGGT SGGMEAKSNK** WDDKKIEGLK KNKEDFEQQL ENIGSIREMQ MKESEISGKI SGLEK**KIQYA EIEKK**SIKDK **LPQLEQEER**N IIEEIDRIKP ELSK**ARTEVD KR**KTEMNKLE KR**MNEIVDRI YKDFSQSVGV PNIRVYEETQ LK**TAEKEAEE R**LELSNQPAK LKYQLEYEQN R**DVGSRIRKI ESSISSLETD LEGIQKTMSE RKETAVK**ITN EINNWKKEME ECK**QKSEEYE KEILDWKKQA SQATTSITKL NRQIHSK**ETQ IEQLISQK**QE ITEKCELEHI TLPVLSDAME EDDSDGPQFD FSELGR**AYLQ ERRPSAREK**V EAEFRQKIES KTSEIER**TAP NLRALDQYEA IQEKEKQVSQ EFEAARKEEK** QVADAFNTVK QKRYELFMEA FNHIASNIDK IYKQLTKSNT HPLGGTAYLN LENEDDPFLH GIKYTTMPPT KRFRDMEQLS GGEKTVAALA LLFSIHSYRP SPFFILDEVD AALDNLNVAK VAKFIRSKSC QAARDNQDAE DGNGFQSIVI SLKDSFYDKA EALVGVYR**DT ERSCSSTMSF DLR**NYQES
43	17.3/5.9	17.3 kDa class I heat shock protein	gi|123534	70	40	M**SLIPSFFGG R**RSSVFDPFS LDVWDPFKDF PFPSSLSAEN SAFVSTR**VDW KETPEAHVFK** ADIPGLKKEE VKLEIQDGRV LQISGER**NVE KEDKNDTWHR VER**SSGKLVR R**FRLPENAK**V DQVK**ASMENG VLTVTVPK**EE IKKPDVKAID ISG
4	30.5/6.1	Cytosolic ascorbate peroxidase	gi|1420938	113	20	MGKSYPTVSP DYQKAIEKAK RK**LRGFIAEK** KCAPLILRLA WHSAGTFDSK TKTGGPFGTI KHQAELAHGA NNGLDIAVRL LEPIKEQFPIV-SYADFYQLA GVVAVEITGG PEVPFHPGRE DKPEPPPEGR **LPDATKGSDH LRDVFGKAMG LSDQDIVALS GGHTIGAAHK ER**SGFEGPWT SNPLIFDNSY FTELLTGEKD GLLQLPSDKA LLTDSVFRPL VEKYAADEDV FFADYAEAHL KLSELGFAEA

Table 4 (continued)

Sp. No.	Exptl. M.Wt.	Protein identified	Accession no.	Mascot score	Sequence coverage	Matching peptide sequence in red
31	27/6.3	Photosystem I reaction center subunit II, chloroplast precursor	PSAD2_ ARATH	61	9	MATQAAGIFN SAITTAATSG VKKLHFFSTT HRPKSLSFTK TAIRAEKTDS SAAAAAAPAT KEAPVGFTPP QLDPNTPSPI FAGSTGGLLR KAQVEEFYVI TWNSPKEQIF EMPTGGAAIM REGPNLLKLA RKEQCLALGT RLR**SKYKITY QFYR**VFPNGE VQYLHPKDGV YPEKANPGRE GVGLNMRSIG KNVSPIEVK**FTGKQSYDL**
10	31/4.5	Alpha-methyl-mannoside-specific lectin	gil15233402	283	50	MAISKKILPL LSIATIFLLL LNK**AHSLGSL SFGYNNFEQG DERNLILQGD ATFSASK**GIQ LTKVDDNGTP AKSTVGR**VLH STQVRLWEKS TNR**LTNFQAQ FSFVINSPID NGADGIAFFI AAPDSEIPKN SAGGTLGLSD PSTAQNPSAN QVLAVEFDTF YAQDSNGWDP NYQHIGFDVD PIK**SAATTKW ERRNGQTLNV LVSYDANSKN LQVTASYPDG QSYQVSYNVD LRDYLPEWGR** VGFSAASGQQ YQSHGLQSWS FTSTLLYTSP HYLKLGRFMI
11	34/4.6	14-3-3 protein	gil45594277	459	50	**MAAAPTPREE NVYMAKLAEQ AERYEEMVEF MEKVSAAADN EELNVEER**NL LSVAYKNVIG ARRASWR**IIS SIEQKEESR**G NEDHVTVIRD YRSKIESELS NICDGILKLL DSRLIPSASS GDSKVFYLK**M KGDYHR**YLAE FKTGAERKEA **AESTLAAYKS AQDIANAELP PTHPIRLGLA** LNFSVFYYEI **LNSPDRACNL AKQAFDEAIA ELDTLGEESY KDSTLIMQLL RDNLTLWTSD MQDDGADEIK** EAAPKQDDQ
12	32/4.7	Mannose/glucose-binding lectin precursor	gil8099130	233	50	**LDSLSFSYNN FEQDDERNLI LQGDAK**FSAS KGIQLTK**VDD NGTPAKSTVG RVLHSTQVRL WEKSTNRLTN FQAQFSFVIK** SPIDNGADGI AFFIAAPDSE IPKNSAGGTL GLFDPQTAQN PSANQVLAVE FDTFYAQDSN GWDPNYQHIG IDVNSIK**SAA TTKWERRDGQ TLNVLVTYDA NSK**NLQVTAS YPDGQRYQLS YRVDLR**DYLP EWGR**VGFSAA SGQQYQSHEL QSWSFTSTLL YTSPHYLKLG RFMI
1	56/6.2	RUBP carboxylase large chain	gil1351139	519	50	MSPQTETKAS VGFKAGVKDY K**LTYYTPEYE TKDTDILAAF RVTPQPGVPP EEAGAAVAAE SSTGTWTTVW TDGLTSLDR**Y KGRCYHIEPV AGEENQYIAY VAYPLDLFEE GSVTNMFTSI VGNVFGFKAL RALR**LEDLRI PTSYSKTFQG PPHGIQVER**D KLNKYGRPLL GCTIKPKLGL SAKNYGR**AVY ECLRGGLDFT KDDENVNSQP FMRWRDRFLF CAEALFK**AQA ETGEIK**GHYL NATEGTCEEM IKR**AVFAREL GAPIVMHDYL TGGFTANTSL AHYCR**DNGLL LHIHRAMHAV IDRQKNHGMH FR**VLAKGLR**L SGGDHIHAGT VVGK**LEGER**D ITLGFVDLLR DDFIEKDR**SR GIYFTQDWVS LPGVLPVASG GIHVWHMPAL TEIFGDDSVL QFGGGTLGHP WGNAPGAVAN R**VALEACVQA RNEGR**DLAR**E GNEIIR**EASK **WSPELAAACE VWKEIK**FEFE AMDTL
8	8.64	RUBP carboxylase small chain precursor	gil123534	181	50	**TSVANNGGRV QCIQVWPTVG K**KKFETLSYL PPLTKQQLAK EVDYLLRKGW **VPCLEFELEH GFVYR**EHNK**S PGYYDGRYWT MWK**LPMF-GCT DSSQVLKELY EAQTAHPDGF IR**IIGFD-NVR** QVQCISFIAY KPPGY

Peptides of *Arachis* proteins sequenced through MS/MS are shown in bold which are matching with the peptides of Viridiplantae taxonomic group

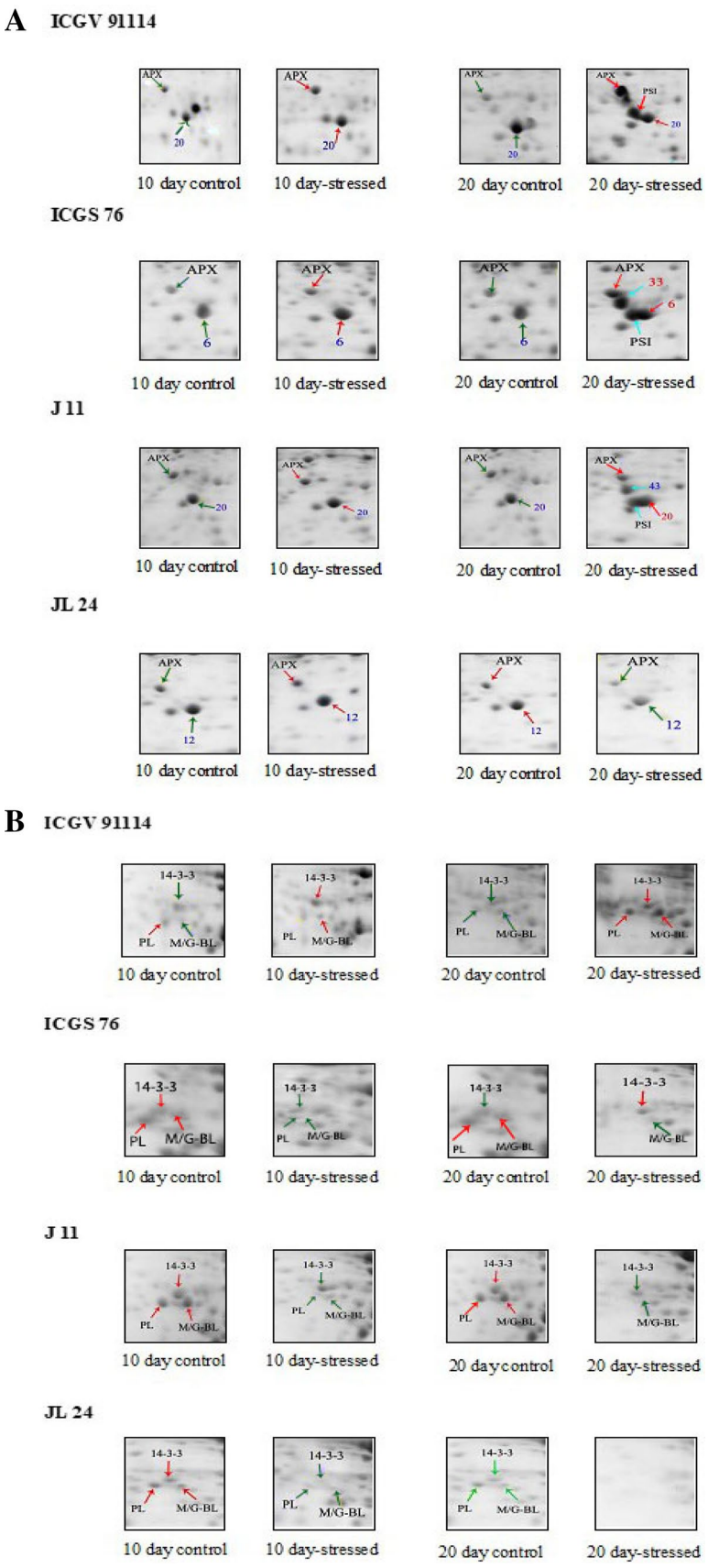


Fig. 9 **a** Comparison of 2DE gels (4–7 pH) exhibiting the differential expression of cytoplasmic APX and PSI proteins in 4 varieties of groundnut. **b** Comparison of 2DE gels (4–7 pH) exhibiting the differential expression of putative lectins in 4 varieties of groundnut

possible protein–protein interactions and for further functional genomic studies. These proteins could be exploited as protein markers to evaluate the peanut germplasm and other crop plants for drought tolerance.

Acknowledgements This work was supported by grants from Council of Scientific and Industrial Research (CSIR), New Delhi, India in the form of Fellowship and also University Grants Commission by providing financial assistance through the Major Research Project.

Author contribution PT, MRD and GBR conceived of this study and designed the experiments. PT isolated proteins from samples, done all proteomics experiments and analyzed the data and PNBS supported experiments. PT and MRD drafted the manuscript and GBR critically evaluated and proof-read the manuscript. RP supported manuscript corrections.

Complaince with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Asada K (2006) Production and Scavenging of reactive oxygen species in chloroplasts and their functions. Plant Physiol 141:391–396

Basha SM, Roberts RM (1981) The glycoproteins of plant seeds: analysis by two dimensional polyacrylamide gel electrophoresis and by their lectin-binding properties. Plant Physiol 67:936–939

Basha SM, Katam R, Naik KSS (2007) Differential response of peanut genotypes to water stress. Peanut Sci 34:96–104

Bhushan D, Pandey A, Choudhary MK, Datta A, Chakraborty S, Chakraborty N (2007) Comparative proteomics analysis of differentially expressed proteins in chickpea extracellular matrix during dehydration stress. Mol Cell Proteom 6:1868–1884

Boston RS, Viitanen PV, Vierling E (1996) Molecular chaperones and protein folding in plants. Plant Mol Biol 32:191–222

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Bray EA, Bailey-Serres J, Weretilnyk E (2000) Responses to abiotic stress. Biochemistry & molecular biology of plants. In: Gruissem W, Jones R (eds) American society of plant physiologists. Rockville, pp 1158–1203

Britt AB (1999) Molecular genetics of DNA repair in higher plants. Trends Plant Sci 4:20–25

Castillejo MA, Maldonado AM, Ogueta S, Jorrin JV (2008) Proteomic analysis of responses to drought stress in sunflower (*Helianthus annuus*) leaves by 2DE gel electrophoresis and mass spectrometry. Open Proteom J 1:59–71

Chaves MM, Maroco JP, Pereira JS (2003) Understanding plant responses to drought from genes to the whole plant. Funct Plant Biol 30:239–264

Close TJ (1996) Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol Plantarum 4:795–803

Cooke MS, Evans MD, Dizdaroglu M, Lunec J (2003) Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J 17:1195–1214

Davletova S, Rizhsky L, Liang H, Shenggiang Z, Oliver DJ, Coutu J, Shulaev V, Schlauch K, Mittler R (2005) Cytolosic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of Arabidopsis. Plant Cell 17:268–281

De Vetten NC, Ferl RJ (1994) Two genes encoding GF14 (14-3-3) proteins in Zea mays. Plant Physiol 106:1593–1604

Dejardin A, Sokolov LN, Kleczkowski LA (1999) Sugar/osmoticum levels modulate differential abscisic acid-independent expression of two stress-responsive sucrose synthase genes in Arabidopsis. Biochem J 344:503–509

Devaiah KM, Bali Geetha, Athmaram TN, Basha MS (2007) Identification of two new genes from drought tolerant peanut up-regulated in response to drought. Plant Growth Regul 52(3):249–258

Drame KN, Clavel D, Repellin A, Passaquet C, Zuily-Fodil Y (2007) Water deficit induces variation in expression of stress responsive genes in two peanut (*Arachis hypogaea* L) cultivars with different tolerance to drought. Plant Physiol Biochem 45:236–243

Emanuelsson O, Soren B, Gunnar VH, Henrik N (2007) Locating proteins in the cell using Target P, Signal P, and related tools. Nat Protoc 2:953–971

Eriksson J, Chait BT, Fenyo D (2000) A statistical basis for testing the significance of mass spectrometric protein identification results. Anal Chem 72:999–1005

Faghani E, Gharechahi J, Komatsu S, Mirzaei M, Ramzan et al (2015) Comparative physiology and proteomic analysis of two wheat genotypes contrasting in drought tolerance. J Proteom 114:1–15

Farooq M, Basra SMA, Wahid A, Cheema ZA, Cheema MA, Khaliq A (2009) Physiological role of exogenously applied glycinebetaine in improving drought tolerance of fine grain aromatic rice (Oryza sativa L). J Agron Crop Sci 194:325–333

Feller U, Anders I, Mae T (2008) Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. J Exp Bot 59:1615–1624

Ferl RJ (1996) 14-3-3 proteins and signal transduction. Annu Rev Plant Physiol Plant Mol Biol 47:49–73

Govind G, Harshavardhan VT, Patricia JK, Dhanalakshmi R, Senthil KM, Sreenivasulu N, Udayakumar M (2009) Identification and functional validation of a unique set of drought induced genes preferentially expressed in response to gradual water stress in peanut. Mol Genet Genom 281:591–605

Graan T, Boyer JS (1990) Very high CO_2 partially restores photosynthesis in sunflower at low water potentials. Planta 181:378–384

Granier F (1988) Extraction of plant proteins for two-dimensional electrophoresis. Electrophoresis 9:712–718

Grimplet J, Wheatley MD, Jouira HB, Deluc LG, Cramer GR, Cushman JC (2009) Proteomic and selected metabolite analysis of grape berry tissues under well watered and water-deficit stress conditions. Proteomics 9(9):2503–2528

Guo BZ, Xu G, Cao YG, Holbrook CC, Lynch RE (2006) Identification and characterization of phospholipase D and its association with drought susceptibilities in peanut (*Arachis hypogaea*). Planta 223:512–520

Ishida T, Kurata T, Okada K, Wada T (2008) A genetic regulatory network in the development of trichomes and root hairs. Annu Rev Plant Biol 59:365–386

Jain AK, Basha SM, Holbrook CC (2001) Identification of drought responsive transcripts in peanut (*Arachis hypogaea* L.). Electron J Biotechnol 4(2):59–67

Jaleel CA, Gopi R, Sankar B, Gomathinayagam M, Panneerselvam R (2008) Differential responses in water use efficiency in two varieties of Catharanthus roseus under drought stress. Comp Rend Biol 331:42–47

Jensen RG, Bahr JT (1977) Ribulose 1,5-bisphosphate carboxylase-oxygenase. Ann Rev Plant Physiol 28:379–400

Katam R, Basha SM, Hemanth KN, Vasanthaiah, Naik KSS (2007) Identification of drought tolerant groundnut genotypes employing proteomics approach. J SAT Agric Res 5(1):1–4

Katam R, Basha M, Suravajhala P, Pechan T (2010) Analysis of peanut leaf proteome. J Proteome Res 9:2236–2254

Katam R, Sakata K, Prashanth S, Pechan Tibor, Devaiah M et al (2016) Comparative leaf proteomics of drought-tolerant and -susceptible peanut in response to water stress. J Proteom 143:209–226

Ke Y, Han G, He H, Li J (2009) Differential regulation of proteins and phosphoproteins in rice under drought stress. Biochem Biophys Res Commun 379(1):133–138

Komatsu S, Hossain Z (2013) Organ-specific proteome analysis for identification of abiotic stress response mechanism in crop. Front Plant Sci 4:1–9

Kottapalli KR, Rakwal R, Shibato J, Burow G, Tissue D, Burke J, Puppala N, Burow M, Payton P (2009) Physiology and proteomics of the water-deficit stress response in three contrasting peanut genotypes. Plant Cell Environ 32:380–407

Kunert KJ, Vorster BJ, Fenta BA, Kibido T, Dionisio G, Foyer CH (2016) Drought stress responses in soybean roots and nodules. Front Plant Sci 7:1015. https://doi.org/10.3389/fpls.2016.01015

Lannoo N, Van Damme EJM (2014) Lectin domains at the frontiers of plant defense. Front Plant Sci 5:397

Lauer MJ, Boyer JS (1992) Internal CO_2 measures directly in leaves: abscisic acid and low leaf water potential cause opposing effects. Plant Physiol 98:1010–1016

Lehmann AR (2005) The role of SMC proteins in the responses to DNA damage. DNA Rep 4:309–314

Lis H, Sharon N (1998) Lectins: carbohydrate-specific proteins that mediate cellular recognition. Chem Rev 98:637–674

Luo M, Liang XQ, Dang P, Holbrook CC, Bausher MG, Lee RD, Guo BZ (2005) Microarray-based screening of differentially expressed genes in peanut in response to Aspergillus parasiticus infection and drought stress. Plant Sci 169:695–703

Mittler R, Vanderauwera S, Gollery M, Van Breusegem F (2004) Reactive oxygen gene network of plants. Trends Plant Sci 9:490–498

Morris JS, Clark BN, Gutstein HB (2007) Pinnacle: a fast, automatic and accurate method for detecting and quantifying protein spots in 2-dimensional gel electrophoresis data. Bioinformatics 24:529–536

Nasmyth K, Haering CH (2005) The structure and function of SMC and kleisin complexes. Annu Rev Biochem 74:595–648

Nautiyal PC, Nageswara Rao RC, Joshi YC (2002) Moisture-deficit-induced changes in leaf-water content, leaf carbon exchange rate and biomass production in groundnut cultivars differing in specific leaf area. Field Crops Res 74:67–79

Padmavathi TAV, Rao DM (2013) Differential accumulation of osmolytes in 4 cultivars of Peanut (*Arachis hypogaea* L.) under drought stress. J Crop Sci Biotechnol 16(2):151–159

Park CJ, Seo YS (2015) Heat shock proteins: a review of the molecular chaperones for plant immunity. Plant Pathol J 31(4):323–333. https://doi.org/10.5423/PPJ.RW.08.2015.0150

Parry MAJ, Andralojc PJ, Khan V, Lea PJ, Keys AJ (2002) RuBisCO activity: effects of drought stress. Ann Bot 89:833–839

Payton P, Kottapalli KR, Rowland D, Faircloth W, Guo B, Burrow M, Puppala N, Gallo M (2009) Gene expression profiling in peanut using high density oligonucleotide microarrays. BMC Genom 10:265

Perkins DN, Pappin DJC, Creasy DM (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3567

Pruthvi V, Rama N, Govind G, Nataraja KN (2013) Expression analysis of drought stress specific genes in Peanut (*Arachis hypogaea* L). Physiol Mol Biol Plants 19(2):277–281

Ramanjulu S, Bartels D (2002) Drought and desiccation-induced modulation of gene expression in plants. Plant Cell Environ 25:141–151

Rampino P, Pataleo S, Gerardi C, Mita G, Perrotta C (2006) Drought stress response in wheat: physiological and molecular analysis of resistant and sensitive genotypes. Plant Cell Environ 29:2143–2152

Ritchie SW, Nguyan HT, Holaday AS (1990) Leaf water content and gas exchange parameters of two wheat genotypes differing in drought resistance. Crop Sci 30:105–111

Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B, Bennett J (2002) Proteomic analysis of rice leaves during drought stress and recovery. Proteomics 2:1131–1145

Sanchez-Rodriguez E, Rubio-Wilhelmi M, Cervilla LM, Blasco B, Rios JJ, Rosales MA, Romer L, Ruiz JM (2010) Genotypic differences in some physiological parameters symptomatic for oxidative stress under moderate drought in tomato plants. Plant Sci 178:30–40

Schellmann S, Hulskamp M (2005) Epidermal differentiation: trichomes in *Arabidopsis* as a model system. Int J Dev Biol 49:579–584

Sebkova V, Unger C, Hardegger M, Sturm A (1995) Biochemical, physiological, and molecular characterization of sucrose synthase from *Daucus carota*. Plant Physiol 108:75–83

Sharp RE, Hsiao TC, Silk WK (1990) Growth of the maize primary root at low water potentials II role of growth and deposition of hexose and potassium in osmotic adjustment. Plant Physiol 93:1337–1346

Strum A, Lienhard S, Schatte S, Hardegger M (1999) Tissue-specific expression of two genes for sucrose synthase in carrot (*Daucus carota* L). Plant Mol Biol 39:349–360

Timperio AM, Eqidi MG, Zolla L (2008) Proteomics applied on plant abiotic stresses: role of heat shock proteins (HSP). J Proteom 71:391

Tunnacliffe A, Wise MJ (2007) The continuing conundrum of the LEA proteins. Naturwissenschaften 94:791–812

Valentovic P, Luxova M, Kolarovic L, Gasparikova O (2006) Effect of osmotic stress on compatible solutes content, membrane stability and water relations in two maize cultivars. Plant Soil Environ 52:186–191

Van Breusegem F, Dat J (2006) Reactive oxygen species in plant cell death. Plant Physiol 141:384–390

Van Damme EJM, Barre A, Rouge P, Peumans WJ (2004) Cytoplasmic/nuclear plant lectins: a new story. Trends Plant Sci 9:484–489

Vierling E (1991) The roles of heat shock proteins in plants. Annu Rev Plant Physiol Plant Mol Biol 42:579–620

Vijayan M, Chandra N (1999) Lectins. Curr Opin Struct Biol 9:707–714

Wang W, Vinocur B, Shoseyov O, Altman A (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci 9:244–252

Winter H, Huber SC (2000) Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. Crit Rev Biochem Mol Biol 35:253–289

Wise MJ, Tunnacliffe A (2004) POPP the question: what do LEA proteins do? Trends Plant Sci 9:13–17

Xiong L, Zhu JK (2002) Molecular and genetic aspects of plant responses to osmotic stress. Plant Cell Environ 25:131–140

Saudi Journal of Biological Sciences 25 (2018) 285–289

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

Effect of PEG-6000 imposed drought stress on RNA content, relative water content (RWC), and chlorophyll content in peanut leaves and roots

Meher [a], P. Shivakrishna [b], K. Ashok Reddy [b], D. Manohar Rao [a,*]

[a] Department of Genetics, Osmania University, Hyderabad, Telangana State, India
[b] Synteny Lifesciences Pvt. Ltd, Hyderabad, Telangana State, India

ARTICLE INFO

Article history:
Received 10 December 2016
Revised 18 April 2017
Accepted 23 April 2017
Available online 25 April 2017

Keywords:
Relative water content
RNA content chlorophyll content
PEG-6000
Peanut

ABSTRACT

Drought, one of the environmental stresses, plays crucial role in reduction in plant production on majority of agricultural fields of world, In order to evaluate drought stress on RNA content Relative water content (RWC), and chlorophyll content, Water deficit was induced by Polyethylene glycol (PEG) in peanut (*Arachis hypogaea*), accession number ICGV 91114. In this current study we evaluate RNA content and Relative water content (RWC) both in leaves and roots and chlorophyll content in leaf. The present study was undertaken with the aim to investigate the effect of water deficit imposed by PEG-6000, 40 old day seedlings were treated with varying concentrations of polyethylene glycol-6000 (PEG-6000; w/v-5%, 10%, 15% & 20%) for 24 h. The results showed that RNA content and Relative water content (RWC) content was significantly reduced in both leaves and roots with increased concentration of PEG, In leaves, a concentration dependent decline in chlorophyll content with increasing concentration of polyethylene glycol-6000 (PEG-6000). Reduction in chlorophyll '*a*' level was to a greater extent than the chlorophyll '*b*'. Thus, this attributes can be used as screening tool for drought tolerance in peanut.

1. Introduction

Due to unfavorable environmental conditions plants are subjected to various abiotic and biotic stresses affecting their growth, metabolism and yield (Kaur and Gupta, 2005). Drought is one of major abiotic stresses constraining crop productivity worldwide, it reduces plant productivity by inhibiting growth and (Singh et al., 2014) slows growth, induces stomatal closure, and therefore reduces photosynthesis (Németh et al., 2002). Extensive field studies have been conducted for understanding the plant tolerance and oxidative stress in response to water deficit. The stress caused due to water creates senescence and abscission in the plants (Karamanos, 1978). The effect of water stress in the leaf of the plant mainly reduces the bulk production of the biomass (Rawson and Turner, 1982; Saxena, 1993). The relative amount of Chlorophyll is directly connected to the photosynthetic capacity of the major plants (Fotovat et al., 2007). Besides chlorophyll content, drought stress play a major role in affecting the enzymes involved in the Calvincycle (Monakhova and Chernyadev, 2002). It is reported that the production of plants also affected by showing to reactive oxygen species (Horling et al., 2003). Polyethylene glycol(PEG-6000) generates osmotic stress which reduces photosynthetic rate, which later effects chlorophyll-*a* and chlorophyll-*b* contents, any stress to the plant effects mechanism of photosynthesis at cellular level which includes pigments, photosystems, the electron transport system and co2 reduction pathways and reduce photosynthesis.

PEG is mainly used for the determination of the drought stress related information's from the plants (Turkan et al., 2005; Landjeva et al., 2008). It is known that PEG does not enter the cell wall space (Rubinstein, 1982) and PEG molecules with a molecular weight greater than 3000 are apparently not absorbed (Tarkow et al., 1996). In the present study, PEG-6000 was used for drought. Simulation of drought stress by polyethylene glycol (PEG) induces drought stress on the plants (Jiang et al., 1995). It is reported that PEG induced significant water stress in plants and not having any toxic effects (Emmerich and Hardegree, 1990). The objective of this research was to determine, relative water content(RWC) and RNA content in leaves and root, chlorophyll content in leaves of peanut (Arachis hypogaea)ICGV 91114 under drought stress.

* Corresponding author.
E-mail address: meher06699@gmail.com (D. Manohar Rao).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.sjbs.2017.04.008
1319-562X/

2. Materials and methods

2.1. Sample collection

For carrying out the above three experiments, ICGV 91114 were obtained from the International Crops Research Institute for the Semi–Arid Tropics (ICRISAT), Patancheru, Hyderabad, Telangana, India. Four seeds per pot in the suitably sized pots filled with mixture of soil were sown to raise the seedlings plants, the seedlings/-plants were maintained in the net house covered with water proof sheets during the experiment. They were watered regularly thrice a week for 40 days until water stress was carried out at the flowering stage.

2.2. Drought treatment

40 old days seedlings were plucked from the pots and treated with different concentrations of PEG 6000 (Polyethylene Glycol) 100 ml of 5%, 10%, 15% and 20% along with controls (100 ml of water) for 24 h, later the leaf and root were harvested separately, Leaf Samples were named as 1(control) 2 (5% PEG treated) 3(10% PEG treated) 4 (15% PEG treated) & 5 (20% PEG treated), Root Samples were named as A (control) B (5% PEG treated) C (10% PEG treated) D (15% PEG treated) E (20% PEG treated). After naming, leaf and root were kept frozen under liquid N_2, and stored at −80 °C, until later experiment to measure chlorophyll, relative water content and RNA content.

2.3. RNA isolation and quantification

RNA was isolated from leaves and root by Trizol reagent. The Trizol reagent was developed by Chomczynski and Sacchi (1987) is a mixture of phenol and guanidine isothiocyanate. The freshly cultivated plants were used for the extraction of the total RNA. The standard extraction methodology were followed for the the extraction of the total RNA. Total RNA was quantified using micro-spectrophotometry (NanoDrop Technologies, Inc.). DNA was removed with Turbo DNA-free (Ambion, Inc.) using the rigorous protocol. RNA integrity was measured using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.).

2.4. Relative water content

To determine relative water content, 20 leaves from each group were weighed immediately (FW) after harvesting the plant. Leaves were then placed in distilled water for 4 h and then turgid weight (TW) was measured. Then the leaves were dried in oven at 80 °C for 24 h to obtain their dry weight (DW). The method proposed by Sharp et al. (1990) Relative water content was calculated by the following formula.

$$RWC = FW - DW/TW - DW \times 100$$

2.5. Chlorophyll estimation

Total Chlorophyll content, content of Chlorophyll '*a*' and Chlorophyll '*b*' were extracted and quantified the modified method of Arnon (1949). After the extraction and analysis, the relative amount of Chlorophyll '*a*', Chlorophyll '*b*' and the total content of Chlorophyll were calculated using the following formulae:

$$\text{Chlorophyll '}a\text{' (mg/g)} = [(12.7XA663 - 2.69XA645)V/W]$$

$$\text{Chlorophyll '}b\text{' (mg/g)} = [(22.9XA645 - 4.68XA663)V/W]$$

$$\text{Total Chlorophyll (mg/g)} = [(20.2A645 + 8.02A663)V/W]$$

where

A = Absorbance at specific wavelengths
V = final volume of chlorophyll extract in 80% Acetone
W = fresh weight of tissue extracted 12.7, 2.69, 22.9, 4.68, 20.2 & 8.02 are the constants

In the present experiment the volume (V) of 10 ml 80% Acetone and weight (W) of 500 mg fresh leaf tissue was used in all the 5 water stressed samples and also the control (Arnon, 1949).

3. Results and discussion

Drought tolerant with a high yield potential under drought stress; currently, drought is a major limiting factor in peanut cultivation, making irrigation necessary. However, peanut crop plants can adapt to water stress in various ways, Many agro-physiological parameters related to drought tolerance have been established, RNA content, Relative water content (RWC) and chlorophyll content with decrease in water supply (Deblonde et al., 1999). A fast screening tool would be helpful in selecting valuable genotypes with defined growth strategies that translate to drought tolerance and are suitable for experiments and/or breeding. In this present investigation the plants were short term drought-stressed by PEG 6000 (Polyethylene Glycol) for 24 h. In this experiment 40 day old seedlings were imposed water stress with different concentration; 100 ml of 5, 10, 15, 20%, of PEG 6000 (Polyethylene Glycol). The present data indicates the significant differences in RNA content, Relative water content (RWC) in leaf and root, chlorophyll contents in leaf when drought- stressed was induced by PEG 6000 (Polyethylene Glycol). A pronounced reduction in, RNA content (Tables 1 and 2; Figs. 1 and 2). Relative water content (RWC) (Tables 3 and 4; Figs. 3 and 4). And chlorophyll contents (Table 5; Fig. 5) with increasing water deficit was observed. There is reduced RNA synthesis with increased water stress, (He et al., 1999) chloroplast RNAase was upregulated which causes degradation of RNA during water stress. It was also been reported the decrease of ribosomes and polyribosomes during water stress ribosomes get cluttered on mRNAs to protect them from degradation, therefore ribosomes get disrupted hence another reason for mRNAs degradation (Mason et al., 1988) (Scott et al., 1979) This data indicates a pronounced reduction in RNA content in leaf and root with increasing water deficit (Tables 1 and 2 Figs. 1 and 2). To understand the dehydration tolerance which shows the metabolic activities in the tissues of plant, RWC is considered to measure water status in plant (Sinclair and Ludlow, 1986) it was also observed the decline of RWC with increased water stress was also observed in barley (Yuan et al., 2005) and tomato (Zgallaï et al., 2005) and pigeonpea plants (Kumar et al., 2011). It clearly evidence that severe stress clearly affect the relative water content as compare to the control of same age group plant the significant differences in RWC was observed as compare to control and stressed of leaf and root (40 days old). The sharp decrease in RWC with the increased PEG concentration was noted of same age group plants. As water stress effects photosynthesis, The highest content of chlorophyll '*a*' and '*b*' was observed in control leaves while both progressive stresses of PEG concentration. Chlorophyll content was also affected during the present investigation which shows that long progressive stress along with some other environmental factor may affect photosynthetic ability of the plant system. Water stress imposed by PEG-6000 effects Enzymes of chlorophyll metabolism and photosynthetic pigments, In our present report it was observed that Chla is more sensitive than Chlb to PEG induced water stress (Hsu and Kao, 2003). Also demonstrated that PEG induced water stress cause decrease in total chlorophyll content in rice leaves. Hassanzadeh et al. (2009)

Table 1
Effect of water deficit on RNA content in Leaf.

S. no	Samples	PEG concentration (%)	A260/A280	Concentration (μg/μl)
1	Leaf 1 (Control)	00	1.85	2.87 ± 0.005
2	Leaf 2	05	1.86	2.67 ± 0.01
3	Leaf 3	10	1.89	2.05 ± 0.01
4	Leaf 4	15	1.89	1.99 ± 0.01
5	Leaf 5	20	1.87	1.56 ± 0.01

Table 2
Effect of water deficit on RNA content in Root.

S. no	Samples	PEG concentration (%)	A260/A280	Concentration (μg/μl)
1	Root A (Control)	00	1.88	2.56 ± 0.025
2	Root B	05	1.81	2.08 ± 0.01
3	Root C	10	1.86	1.98 ± 0.01
4	Root D	15	1.89	0.99 ± 0.005
5	Root E	20	1.83	0.58 ± 0.005

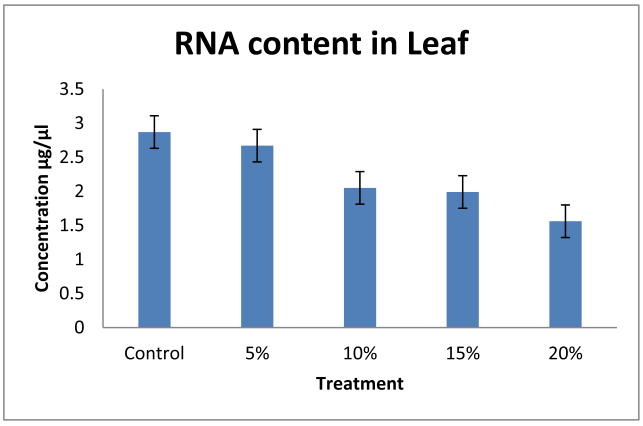


Fig. 1. Figure showing the RNA Concentration in Leaf.

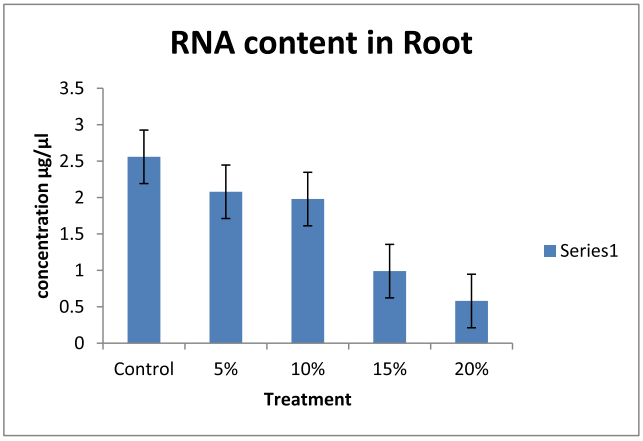


Fig. 2. Figure showing the RNA Concentration in Root.

Table 3
Effect of water deficit on relative water content (RWC) in Leaf.

PEG	RWC%
Control	0.831
5	0.64
10	0.523
15	0.4
20	0.35

Table 4
Effect of water deficit on relative water content (RWC) in Root.

PEG	RWC%
Control	0.879
5	0.586
10	0.426
15	0.348
20	0.093

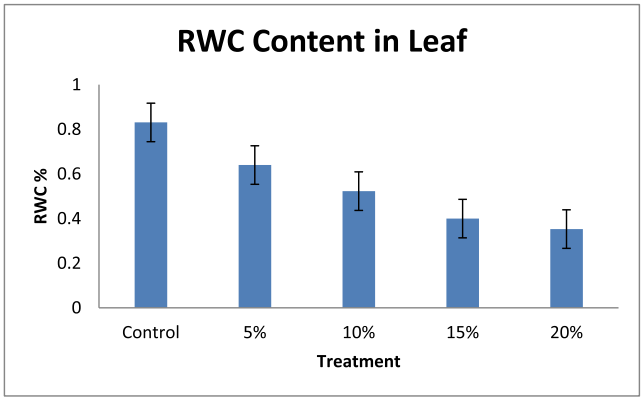


Fig. 3. Figure showing Relative Water Content (RWC) in Leaf.

revealed decrease in Chla but increase in Chlb content under drought stress in seasame. Decrease in the total chlorophyll content by PEG 6000 has also been noticed by Pratap and Sharma (2010) in black gram and Guo et al. (2013). A reason for decrease in chlorophyll content as affected by water deficit is that drought or heat stress by producing reactive oxygen species (ROS) such as O_2 and H_2O_2, can lead to lipid peroxidation and consequently, chlorophyll destruction also, with decreasing chlorophyll content due to the changing green color of the leaf into yellow, Several methods which range from withdrawal of water to plants to the use of chemicals such as polyethylene glycol, mannitol etc., have been employed to create water stress in plants. Plant exposes their root system to this solution and no other toxicities were observed at plant level following the addition of PEG-6000 (Scott et al., 1979). It is reported that PEG induced significant water stress in

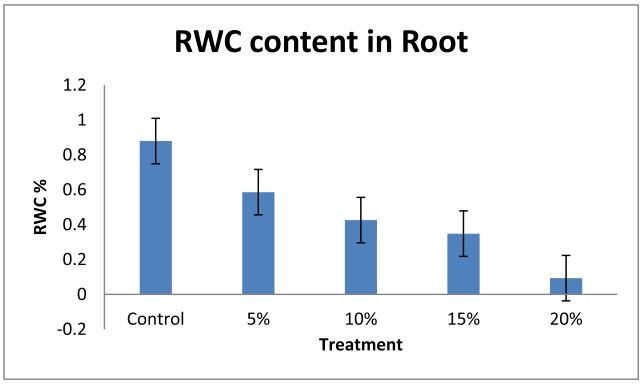


Fig. 4. Figure showing Relative Water Content (RWC) in Root.

Table 5
Effect of water deficit on Chlorophyll '*a*' '*b*' and total Chlorophyll.

PEG	Chlorophyll "*a*"	Chlorophyll "*b*"	Total
Control	1.606	0.474	2.08
5%	1.34	0.47	1.817
10%	1.1	0.56	1.673
15%	0.874	0.528	1.38
20%	0.66	0.386	1.048

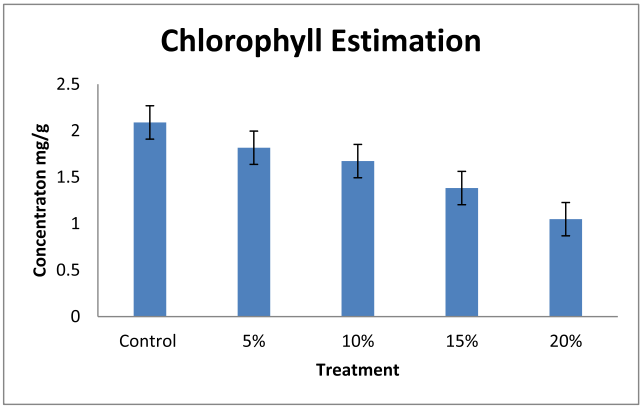


Fig. 5. Figure showing total chlorophyll content in Leaf.

plants and not having any toxic effects (Emmerich and Hardegree, 1990). RNA content, Relative water content (RWC) and chlorophyll content parameter can be used to select high yielding genotypes that maintain cell turgor under water stress environment to give relative high yield. A decrease in the RWC observed in both progressive mild and severe water stress. Relatively higher RWC was noted in progressive mild stress than severe stress indicating that plants have the ability to sustain their water content under mild stress, whereas this ability lost under severe stress treatment. Decrease in the RWC in PEG induced water stress was also reported in rice leaves (Hsu and Kao, 2003) and in Tomato (Zgallaï et al., 2005). According to results of (Bayoumi et al., 2008), RWC involved in absorbing more amount of water from the soil and/or the ability to control water loss through stomata and RWC parameter can be used to select high yielding genotypes that maintain cell turgor under water stress environment to give relative high yield.

4. Conclusions

Our present results indicate that a progressive water stress induced PEG-6000 cause significant physiological and biochemical changes in peanut (*Arachis hypogaea*) ICGV 91114 plant. RNA content, relative water content (RWC) and chlorophyll content, and parameter can be used to select high yielding genotypes that maintain cell turgor under water stress environment. This experiments can be used for other cultivars conditions to have more are more RNA content, relative water content (RWC) and chlorophyll content are more resistant to drought stress and their yield is stable. This attributes can be used as screening tool for drought tolerance in other cultivars. This study was following to find characters of resistant under drought stress and the results showed that RNA content, relative water content (RWC) and chlorophyll content made difference between control and stress of peanut (*Arachis hypogaea*) ICGV 91114. Thus, this attributes can be used as screening tool for drought tolerance.

Acknowledgements

The authors gratefully thanks the financial support provided by the University Grants Commission (UGC). Author also express deep gratitude to Scientist F of National Institute of Nutrition Dr Bhanu Prakash Reddy. Professor of CPMB (Center for Plant Molecular Biology) V.D Reddy and K.V.Rao.

References

Arnon, D.I., 1949. Copper enzymes in isolated chloroplasts. Polyphenol oxidase in Beta vulgaris. Plant Physiol. 24 (1), 1.
Bayoumi, T.Y., Eid, M.H., Metwali, E.M., 2008. Application of physiological and biochemical indices as a screening technique for drought tolerance in wheat genotypes. Afr. J. Biotech. 7 (14).
Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162 (1), 156–159.
Deblonde, P.M.K., Haverkort, A.J., Ledent, J.-F., 1999. Responses of early and late potato cultivars to moderate drought conditions: Agronomic parameters and carbon isotope discrimination. Eur. J. Agron. 11, 91–105.
Emmerich, W.E., Hardegree, S.P., 1990. Polyethylene glycol solution contact effects on seed germination. Agron. J. 82 (6), 1103–1107.
Fotovat, R., Valizadeh, M., Toorehi, M., 2007. Association between water-use-efficiency components and total chlorophyll content (SPAD)in wheat (*Triticum aestivum* L.) under well-watered and drought stress conditions. J. Food Agric. Environ. 5 (2007), 225–227.
Guo, R., Hao, W.P., Gong, D.Z., Zhong, X.L., Gu, F.X., 2013. Effects of water stress on germination and growth of wheat, photosynthetic efficiency and accumulation of metabolites. Soil Processes and Current Trends in Quality Assessment. InTech.
Hassanzadeh, M., Ebadi, A., Panahyan-e-Kivi, M., Eshghi, A.G., Jamaati-e-somarin, S. H., Saeidi, M., Zabihie-Mahmoodabad, R., 2009. Evaluation of drought stress on Relative water content and chlorophyll content of Sesame (*Sesamum indicum* L.) genotypes at early flowering stage. Res. J. Environ. Sci. 3 (3), 345–360.
He, T.X., An, L.Z., Lin, H.H., Liang, H.G., 1999. Evidence for transcriptional and post transcriptional control of protein synthesis in water-stressed wheat leaves: a quantitative analysis of messenger and ribosomal RNA. J. Plant Physiol. 155 (1), 63–69.
Horling, F., Lamkemeyer, P., Konnig, J., Finkemeir, I., Kandlbinder, A., Baier, M., Dietz, K., 2003. Divergent light, ascorbate and oxidative stressdependent regulation of expression of the peroxiredoxin gene family in Arabidopsis. Plant Physiol. 131, 317–325.
Hsu, S.Y., Kao, C.H., 2003. Differential effect of sorbitol and polyethylene glycol on antioxidant enzymes in rice leaves. Plant Growth Regul. 39 (1), 83–90.
Jiang, Y., Macdonald, S.E., Zwiazak, J.J., 1995. Effects of cold storage and water stress on water relations and gas ex-change of white spruce (Picea glauca) seedlings. Tree Physiol. 15 (4), 267–273.
Karamanos, A.J., 1978. Water stress and leaf growth of field beans (Vicia faba L.) in the field: leaf number and total leaf area. Ann. Bot. 42 (6), 1393–1402.
Kaur, N., Gupta, A.K., 2005. Signal transduction pathways under abiotic stresses in plants. Curr. Sci. 88 (11), 1771–1780.
Kumar, R.R., Karajol, K., Naik, G.R., 2011. Effect of polyethylene glycol induced water stress on physiological and biochemical responses in pigeonpea (Cajanus cajan L. Millsp.). Recent Res. Sci. Technol. 3 (1).
Landjeva, S., Neumann, K., Lohwasser, U., Borner, M., 2008. Molecular mapping of genomic regions associated with wheat seedling growth under osmotic stress. Biol. Plan. 52, 259–266.
Mason, H.S., Mullet, J.E., Boyer, J.S., 1988. Polysomes, messenger RNA, and growth in soybean stems during development and water deficit. Plant Physiol. 86 (3), 725–733.
Monakhova, O.F., Chernyad'ev, I.I., 2002. Protective role of kartolin-4 in wheat plants exposed to soil draught. Appl. Biochem. Microbiol. 38 (4), 373–380.

Németh, M., Janda, T., Horváth, E., Páldi, E., Szalai, G., 2002. Exogenous salicylic acid increases polyamine content but may decrease drought tolerance in maize. Plant Sci. 162 (4), 569–574.

Pratap, V., Sharma, Y.K., 2010. Impact of osmotic stress on seed germination and seedling growth in black gram (Phaseolus mungo).

Rawson, H.M., Turner, N.C., 1982. Recovery from water stress in five sunflower (Helianthus annuus L.) cultivars. II. The development of leaf area. Funct. Plant Biol. 9 (4), 449–460.

Rubinstein, B., Turner, N.C., 1982. Regulation of H+ excreation. Effects of osmotic shock. Plant Physiol. 99, 355–360.

Saxena, M.C. (Ed.), 1993. Breeding for Stress Tolerance in Cool-Season Food Legumes (p. 474). Wiley, Chichester, UK.

Scott, N.S., Munns, R., Barlow, E.W.R., 1979. Polyribosome content in young and aged wheat leaves subjected to drought. J. Exp. Bot. 30 (5), 905–911.

Sharp RE., Hsiao TC., Silk W.K. Growth of the maize primary root at low water potentials : II. role of growth and deposition of hexose and potassium in osmotic adjustment. Plant Physiol. 93(4), 1990, 1337–46.

Sinclair, T.R., Ludlow, M.M., 1986. Influence of soil water supply on the plant water balance of four tropical grain legumes. Funct. Plant Biol. 13 (3), 329–341.

Singh, N.P., Pal, P.K., Vaishali, S.K., 2014. Morpho-physiological characterization of Indian wheat genotypes and their evaluation under drought condition. Afr. J. Biotech. 13 (20).

Tarkow, H., Feist, W.C., Southerland, C.F., 1996. Interaction of wood and polymeric materials. Penetration versus molecular size. Forest Prod. J. 16, 61–65.

Turkan, I., Bor, M., Ozdemir, F., Koca, H., 2005. Differential responses of lipid peroxidation and antioxidants in the leaves of drought-tolerant *P. acutifolius* Gray and drought-sensitive *P. vulgaris* L. subjected to polyethylene glycol mediated water stress. Plant Sci. 168, 223–231.

Yuan, S., Liu, W.J., Zhang, N.H., Wang, M.B., Liang, H.G., Lin, H., 2005. Effects of water stress on major photosystem II gene expression and protein metabolism in barley leaves. Physiol. Plant. 125 (4), 464–473.

Zgallaï, H., Steppe, K., Lemeur, R., 2005. Photosynthetic, physiological and biochemical responses of tomato plants to polyethylene glycol-induced water deficit. J. Integr. Plant Biol. 47 (12), 1470–1478.

Plant Gene 13 (2018) 64–75

Contents lists available at ScienceDirect

Plant Gene

journal homepage: www.elsevier.com/locate/plantgene

Genome-wide *in silico* analysis of dehydrins in *Sorghum bicolor*, *Setaria italica* and *Zea mays* and quantitative analysis of dehydrin gene expressions under abiotic stresses in *Sorghum bicolor*

M. Nagaraju[a], Palakolanu Sudhakar Reddy[b], S. Anil Kumar[a], Anuj Kumar[c,d], Prashanth Suravajhala[c,e], Altaf Ali[a], Rakesh K. Srivastava[b], P.B. Kavi Kishor[a,*], D. Manohar Rao[a,*]

[a] *Department of Genetics, Osmania University, Hyderabad 500 007, India*
[b] *International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad 502 324, India*
[c] *Bioclues.org, Kukatpally, Hyderabad, Telangana 500 072, India*
[d] *Advance Centre for Computational & Applied Biotechnology, Uttarakhand Council for Biotechnology (UCB), Dehradun 248 007, India*
[e] *Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research (BISR), Jaipur, Rajasthan 302 001, India*

ARTICLE INFO

Keywords:
Abiotic stress
DHNs
Phosphorylation sites
qRT-PCR
Sorghum
YnS sub group

ABSTRACT

Dehydrins (*DHNs*) are highly hydrophilic, thermo stable, calcium dependent chaperons involved in plant developmental processes as well as in diverse abiotic stresses. A systematic survey resulted in the identification of 7 dehydrins (*DHNs*) in *Setaria italica* and *Zea mays*, but 6 in *Sorghum bicolor*. They are classified into 5 sub-groups, namely YnSKn, SKn, KnS, S, and YnS. *DHNs* of *Sorghum* exhibit 1 ortholog with *Oryza sativa* and *Z. mays* and 3 with *S. italica*. Unlike other *DHNs*, Sb*DHN*5 has been found as an ordered protein with many phosphorylation sites. Network analyses of novel YnS subgroup showed interaction with HSP70 and FKBP genes. *In silico* promoter analysis revealed the presence of abscisic acid (ABA), drought, salt, low temperature stress-responsive elements. The miRNA target analysis revealed *DHNs* are targeted by 51 miRNAs responsive to abiotic stresses. High transcript expressions of *DHNs* were observed in root, stem and leaf compared to inflorescence in *S. bicolor*. All *DHN* genes exhibited high levels of expression in stem under cold, heat, salt, and drought stresses. In contrast to other *DHNs*, the *SbDHN*2 of YnS subgroup, exhibited the highest expression, under multiple stresses in all the tissues indicating its involvement against a wide array of abiotic stresses.

1. Introduction

Dehydrins (*DHNs*) or group 2 LEA protein family members are expressed under cellular dehydration and play crucial roles in response to abiotic stresses. Due to their hydrophilicity and high glycine content, *DHNs* assist cells to withstand dehydration stress (Anchordoguy and Carpenter, 1996). *DHNs* are unstructured proteins and share many features with other types of intrinsically disordered/unstructured proteins. Due to their disordered state, *DHNs* escape from denaturing under abiotic stress conditions (Livernois et al., 2009; Hincha and Thalhammer, 2012). Under dehydration stress, tissue and developmental specific expressions of *DHNs* have been observed. Some *DHNs* are more responsive to the developmental stages of the plant than to abiotic stresses. They act as chaperons involved in developmental processes like late embryogenesis and stabilize macromolecules, denatured proteins, and membrane structures in stressed plants (Close, 1996; Hinniger et al., 2006). *DHNs* contain a consensus sequence of lysine rich residues (K-segment), representing a highly conserved 15 amino acid (EKKGIMDKIKELLPG) motif, with repeated glycine and polar amino acids forming amphipathic-helices. These helices interact with lipid components and hydrophobic sites of the partially denatured proteins of cell membranes and protect the proteins from denaturation (Koag et al., 2009). *DHNs* have a serine rich segment (S-segment), which can be modified by phosphorylation. The phosphorylated *DHNs* binding activity is generally conserved in the acidic subfamily of *DHNs* (Kovacs et al., 2008). *DHNs* consist of 1–3 tandem copies of the consensus Y-segment (V/T) DEYGNP, near the N terminus. They show similarity to the plant and bacterial chaperonin nucleotide binding site

Abbreviations: ABRE, abscisic acid-responsive elements; CK, casein kinase; *DHNs*, dehydrins; DRE, drought-responsive elements; HSE, heat shock-responsive elements; IDP, intrinsically disordered proteins; ILP, intron length polymorphism; LEA, late embryogenesis abundant proteins; LTR, low temperature-responsive elements; miRNA, micro RNA; MW, molecular weights; NJ, neighbour joining; pI, iso electric point; PKA, protein kinase A; PKC, protein kinase C; qRT-PCR, quantitative real-time PCR; SSR, simple sequence repeats; UTRs, untranslated sequence regions

* Corresponding authors.
E-mail addresses: pbkavi@osmania.ac.in (P.B. Kavi Kishor), dmanoharrao27@gmail.com (D. Manohar Rao).

https://doi.org/10.1016/j.plgene.2018.01.004
Received 4 December 2017; Received in revised form 23 January 2018; Accepted 23 January 2018
Available online 02 February 2018

motifs (Hanin et al., 2011). *DHNs*, based on the number and order of the Y-, S- and K-segments, are classified into several classes, such as KS, SK_3, YSK_2, Y_2SK_2, Kn, Y_2SK_3, and YSK_3 (Close, 1996). In barley, YSK_2-type *DHN* was up-regulated by drought, but not by cold stress. SK_3-, Kn-, and KS-type *DHNs* are induced by low temperature and drought (Tommasini et al., 2008). Overexpression of *Sorghum DHN1* (YSK_2) in tobacco displayed enhanced tolerance to high temperature and osmotic stress conditions (Halder et al., 2017). YnSKn-type *DHNs* are expressed during drought, salt, frost, ABA, gibberellic acid, methyl jasmonate, and salicylic acid (SA) treatments. KnS-type *DHNs* bind to metal and scavenge hydroxyl radicals and protect the membrane integrity (Hanin et al., 2011). KS-type *DHNs* on the other hand are small proteins expressed in the reproductive tissues like anthers during chilling stress (Wang et al., 2014). The SKn-type acidic *DHNs* consist of compositional and structural features, and membrane binding properties. They protect the membranes from freezing and desiccation by acting as molecular chaperones or ion sequestration agents to prevent the damage of membrane lipids (Alsheikh et al., 2003; Kovacs et al., 2008).

DHNs scavenge reactive oxygen species (ROS), causing enhancement in the antioxidative enzyme activity under dehydration stress (Kumar et al., 2014). *DHNs*/DHN-like proteins with ion (calcium in particular) binding activity might act either as calcium buffers or as calcium-dependent chaperones like calreticulin and calnexin (Alsheikh et al., 2003). Eight *DHNs* have been reported earlier in rice (Wang et al., 2007; Verma et al., 2017), 13 in barley (Tommasini et al., 2008), 10 in *Arabidopsis* (Hundertmark and Hincha, 2008), 11 in poplar (Liu et al., 2012), 9 in *Malus* (Liang et al., 2012), 4 in *Vitis* (Yang et al., 2012), and 23 in *Brassica napus* (Liang et al., 2016). However, the information regarding the number of diverse *DHN*-types in warm grasses like *Setaria italica*, *Sorghum bicolor*, and *Zea mays* is lacking. Hence, the present investigation was carried out with an objective to find out the number, type, distribution, characterization, motif and promoter analysis, phosphorylation sites and structure of *DHNs* in 3 economically important warm grasses; *S. italica*, *S. bicolor*, and *Z. mays* with special focus on *S. bicolor* and their evolutionary relationships with *Oryza* and *Arabidopsis*, besides tissue specific expression profiles.

2. Materials and methods

2.1. Plant material and stress conditions

To investigate the expression levels of *DHNs*, seeds of *Sorghum bicolor,* variety BTx623, were sown in pots containing 4.5 kg of black clay soil under glass house conditions at 28/20 °C day/night temperatures. After the emergence of inflorescence, the plants were subjected to drought stress by withholding water for 5-days, cold stress by keeping the plants at 4 °C for 4 h and heat stress by exposing the plants to 40 °C for 4 h in a growth chamber and salinity stress by treating the plants with 150 mM NaCl solution for 24 h. Respective controls were maintained under similar conditions. Roots, stems, leaves, and inflorescences were collected and snap frozen immediately in liquid nitrogen and stored at −80 °C until further use.

2.2. In silico identification of DHN genes

DHN gene sequences of *Arabidopsis*, *Oryza*, *Hordeum*, *Vitis*, *Lycopersicum*, and *Malus* were retrieved from NCBI database and searched against *Sorghum bicolor*, *Setaria italica,* and *Zea mays* genomes in Gramene database (http://www.gramene.org/) to find out their homologs. Edit plus (http://www.editplus.com/) and Genscan (http://genes.mit.edu/GENSCAN.html) programs were used to retrieve the *DHN* gene cds and protein sequences. Based on the homology, all the identified putative DHN protein sequences were subjected to SMART program (http://smart.embl-heidelberg.de/) to identify their conserved domains (Letunic et al., 2004).

2.3. Sequence analysis of DHNs

The identified *DHN* genes were mapped to their respective chromosomes based on the information provided in the Gramene Database. Gene Structure Display Server (http://gsds.cbi.pku.edu.cn) software was used for obtaining the *DHN* gene structures - exons, introns, and untranslated sequence regions (UTRs) based on the alignments of their coding sequences (Guo et al., 2007). Multiple sequence alignment was performed using ClustalX (Larkin et al., 2007) to explore conserved sequences and regulatory domains including their functional homology. Multiple Em for Motif Elicitation (MEME) software (http://meme-suite.org/) was employed with default parameters: number of motifs (1 − 10), motif width of (5–50) and the number of motif sites (5–10) to analyze sequence patterns and their significance (Bailey et al., 2006). Molecular weight (MW), isoelectric point (pI), and GRAVY (grand average of hydropathy) of *DHNs* were identified using ProtParam of Expasy tools (Gasteiger et al., 2005) (http://web.expasy.org/protparam), while phosphorylation sites were predicted by employing NetPhosK1 software of Expasy tools (Blom et al., 2004). Disorder tendencies of all the identified *DHNs* were analyzed using the IUPred (http://iupred.enzim.hu/) (Dosztanyi et al., 2005). The putative transmembrane helices within *DHNs* were identified using TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) (Moller et al., 2001). Subcellular localization of *DHNs* was identified using CELLO V2.5 (http://cello.life.nctu.edu.tw) (Yu et al., 2006) and WoLFPSORT programs (http://wolfpsort.org/) (Horton et al., 2007). Secondary structures of DHN proteins were predicted using PSIPRED v3.0 program (Jones, 1999). All the *DHNs* were queried against the Protein Data Bank (Berman et al., 2000) to identify the best template with similar amino acid sequences and known 3D structures for developing the homology models. Homology structures of *DHNs* were built by employing the Modeller 9.15 software (http://www.salilab.org/) (Webb and Sali, 2014) and validated by PROCHECK software to identify phi-psi angles of amino acids (Laskowski et al., 1993). Amino acids that were not found in the allowed regions were brought back into the allowed regions by loop building with the help of Swiss Protein Data Bank viewer programme. The protein-protein interaction of Sb YnS-subgroup was generated by employing STRING (http://string-db.org/) software.

2.4. In silico prediction of potential cis-regulatory elements

To predict the putative *cis*-regulatory elements of *DHN* promoter regions PLACE (Higo et al., 1999) and PLANTCARE (Lescot et al., 2002) software programs were used. Genomic sequence of length 2000 bp upstream to start codon was retrieved from *S. bicolor*, *S. italica*, and *Z. mays* and used for analysis.

2.5. Phylogenetic analysis of DHNs

The N-J phylogenetic tree was constructed with the DHN protein sequences of *S. bicolor*, *S. italica*, *Z. mays*, *O. sativa*, and *A. thaliana* using MEGA 6.2 software (Tamura et al., 2013) by employing the Poisson correction, pairwise deletion and bootstrap value (1000 replicates) parameters.

2.6. In silico prediction of gene specific molecular markers (SSRs and ILPs) and miRNAs targeting DHNs

Gene specific molecular markers including SSRs and ILPs were developed in genomic transcripts of identified *DHN* genes using BatchPrimer3v1.0 (http://probes.pw.usda.gov/batchprimer3/) server. Further, putative miRNAs in different plant species targeting the *DHN* genes were identified using psRNATarget server (Dai and Zhao, 2011) with default parameters.

2.7. Classification, signature amino acid analysis and evolutionary relationship of DHNs

DHN sequences belonging to different crops were retrieved from Gramene and Phytozome databases. They are further classified manually using MEME (Bailey et al., 2006) to identify the nature of motifs with default parameters; number of motifs (1–10), motif width (5–50), and the number of motif sites (5–10). The amino acid percentages were calculated by Protparam tool (Gasteiger et al., 2005), to identify the signature amino acids. To know the evolutionary relationship, phylogenetic tree for a set of 451 DHN sequences was constructed using MEGA 6.2 software (Tamura et al., 2013) by employing the Maximum Parsimony (MP) search method Tree-Bisection-Reconnection (TBR) that uses all sites and bootstrap value (1000 replicates) parameters.

2.8. RNA isolation and qRT-PCR analysis

Total RNA was extracted from different tissues of *S. bicolor* exposed to different abiotic stresses along with their respective controls using MACHEREY-NAGEL kit by following the manufacturer's instructions. A total of 2.5 µl RNA (2.5 µg concentration) was converted to cDNA using Superscript III first strand synthesis kit (Invitrogen) and used as template after diluting it with nuclease free water (1:12). The SYBR Green Master Mix (2×) was used according to the manufacturer's recommendations on the RealPlex (Eppendorf) to study the gene expression. Gene expression analysis was performed for 6 *SbDHNs* (*SbDHN1* to *SbDHN6*) with expected product sizes of 80–124 bp (Supplementary Table 1) in 96-well optical PCR plates. Three biological replicates were taken for qRT-PCR analysis with the following thermal cycles: 1 cycle at 95 °C for 10 min, followed by 40 cycles alternatively at 95 °C for 15 s and 62 °C for 1 min. Amplicon dissociation curves were recorded with fluorescence lamp after 40th cycle by heating from 58 to 95 °C within 20 min. Transcript levels of eukaryotic initiation factor4α (*SbEIF4α*) and protein phosphatase2A (*SbPP2A*) genes were used as internal controls (Reddy et al., 2016). Experiments were repeated thrice and average values are represented. Relative gene expressions were calculated by employing Rest software (Pfaffl et al., 2002).

3. Results

3.1. In silico identification of DHN genes

A total of 43 *DHN* nucleotide sequences; 10 from *Arabidopsis*, 4 from *Vitis*, 9 from *Malus*, 7 from *Oryza* and 13 from *Hordeum* were retrieved from NCBI database. Blast search of the DHN sequences against the genomes of *S. italica*, *S. bicolor*, and *Z. mays*, resulted in the identification of 17 putative genes in *S. italica*, 23 in *S. bicolor* and 19 in *Z. mays* (total 59). On testing these sequences by SMART software, a conserved domain search tool, only 20 of the 59 were confirmed to be *DHNs*; 7 each in *S. italica*, and *Z. mays* and 6 in *S. bicolor* (Table 1).

3.2. Chromosomal location and gene structure of DHNs

The 7 and 6 *DHNs* identified in *S. italica* and *S. bicolor* are localized on 3 different chromosomes each, while 7 *DHNs* identified in *Z. mays* are distributed on 6 different chromosomes. Among the 7 *DHNs* in *S. italica*, *SiDHN1* is mapped on chromosome 1, *SiDHN2*, *3*, and *4* on 5 and *SiDHN5*, *6*, and *7* on 8. Of the 6 *DHNs* in *S. bicolor*, *SbDHN1* and *2* are located on chromosome 3, *SbDHN3*, *4* and *5* on 9, and *SbDHN6* on chromosome 10. Out of the seven *DHNs* in *Z. mays*, *ZmDHN1* is tagged on to chromosome 1, *ZmDHN2* and *3* on 4, *ZmDHN4* and *5* on 5, *ZmDHN6* on 8 and *ZmDHN7* on chromosome 9 (Fig. 1 and Table 1). *DHN* gene structures revealed that only 2 of them contain one exon while all other *DHNs* contain 2 to 4 exons. In *S. italica*, 2 exons were identified in *SiDHN1*, *2*, *6* and *7* genes, 3 in *SiDHN4* and 4 in *SiDHN3* and *5*. In the case of *S. bicolor,* one exon was identified in *SbDHN3*; 2 in *SbDHN4* and *6*, 3 in *SbDHN1* and *2*, and 4 exons in *SbDHN5*. In *Z. mays*, only one exon was noticed in *ZmDHN4*; 2 in *ZmDHN1*, *3*, *5*, *6* and *7* genes and 3 exons in *ZmDHN2* (Fig. 2 and Table 1).

3.3. Conserved domains and motif analysis of DHNs

Multiple sequence alignment showed highly conserved domains of K- (lysine), S- (serine) and Y- (tyrosine) rich segments in all the *DHNs* of *S. italica*, *S. bicolor*, and *Z. mays* (Fig. 3A). The motif search by MEME software revealed that the K-rich domain is the most common among all the *DHNs*, while S- and Y-segments varied among these taxa. Motifs 1, 3, and 5 represented the K-segment, motif 2 the S-segment, and motifs 4 and 9 the Y-segment. In all the *DHNs*, the S- and Y-segments are present only once with an exception in *SiDHN4*, *SbDHN1*, and *SbDHN2*. In *Z. mays*, some *DHNs* have been found without Y-segments, while K-segment is repeatedly found 2–3 times similar to that of *Arabidopsis* (Fig. 3B and Supplementary Fig. 1). *DHNs* in these three crops are classified into YnSKn, SKn, KnS, S, and YnS types. The YnSKn group of *DHNs* is common among all the three crops, while in *S. italica* and *S. bicolor*, only the newly identified YnS group is present but absent in *Z. mays*. On the other hand, KnS and S groups appeared only in *Z. mays*. Based on the presence of 3 and 4 motifs, the *SbDHN*5 is grouped into SK-type (Table 1). The YnS sub group of *Sorghum* contained DnaJ domain, whereas it is absent in *Setaria*.

3.4. Analysis of DHN proteins in S. bicolor, S. italica, and Z. mays

*Zm*DHN3 is the smallest confirmed protein with 108 aa while the largest one (*Sb*DHN6) is 388 aa in length. MWs of *DHNs* in *S. italica* ranged between 14,126.34 and 33,741.27 Da and pI values from 4.79 to 10.11, while in *S. bicolor*, they ranged from 15,399.74 to 37,488.09 Da and pI from 5.79 to 9.25. In *Z. mays*, MWs ranged from 12,199.09 to 35,266.67 Da and pI from 5.51 to 9.92. Most of the identified *DHNs* are basic in nature. The *Zm*DHN6 of YnSKn type and all the SKn-type *DHNs* exhibited low isoelectric point, with an exception of SKn-type (*SiDHN*5 and *SbDHN*5) which have high pI compared to YnSkn *DHN*s. The GRAVY values of *S. italica* varied between −1.246 and −0.374, whereas in *S. bicolor DHNs*, they ranged from −1.282 to −0.330 and in *Z. mays DHNs* between −2.158 and −0.306 indicating their hydrophilicity. Both the WoLFPSORT and CELLO software predicted the subcellular localization of *DHNs* in nucleus, mitochondria, chloroplast, and extra cellular matrix, however, these software exhibited varied localization of *SiDHN*4, *SiDHN*5, *SbDHN*2, *ZmDHN*2, and *ZmDHN*6. The protein instability index of the *DHNs*, as explored by Protparam software, indicated that 5 of the 7 (71.4%) *SiDHNs* are stable, whereas 3 out of 6 (50%) *SbDHNs* and 3 out of 7 (42.8%) *ZmDHNs* are stable. The IUPred Server (http://iupred.enzim.hu/) predicted that all the *DHNs* are IDPs with the exception of *Sb*DHN5, which is an ordered or folded protein (Table 1). The NetPhos software predicted that all the *DHNs* contain higher number of PKC than CK1, CK2, and PKA types. The YnSKn-type *DHNs* contained more number of putative PKC sites than protein kinase CK2. In SKn *DHNs*, CK2 sites are more in number than PKC, with an exception of *SiDHN*5 and *SbDHN*5. Besides PKC and CK2 sites, SKn *DHN*s also contained PKA, DNAPK, RSK and CK1 sites, which are absent in YnSKn type. However, *ZmDHN*5 does not contain any PKC but contains more number of CK2 sites (Supplementary Table 2). The Psipred software analysis of secondary structures of all the DHN proteins of *Setaria*, *Sorghum*, and *Zea* exhibited highly disordered regions with less helix or strand motifs, except *SbDHN*5 which contained fewer disordered regions and consisted of high number of strand motifs. The helices are located within K-segments. Generally, YnSKn *DHN*s displayed less number of helices or strands, due to their disordered tendency, but *Si*DHN1 and *Si*DHN3 exhibited the highest amount (80%) of disordered tendency of motifs with high number of helices. Three SKn-type 3 (*SbDHN*3 and *ZmDHN*3, and *ZmDHN*5) proteins showed 90% of motifs with disordered tendency and the highest number of helices.

Table 1
List of identified dehydrins exhibiting chromosomal location, sub group, length, DNA binding domains (DBD), molecular weight (MW), iso-electric point (Pi), GRAVY, no. of exons, localization, instability index and disordered tendency.

ACC number	Common name	Sub-group	No. of amino acids	Chromosome number	DBD	pI/MW	GRAVY	No. of exons	Localization (CELLO)	Localization (PSORT)	Instability index	Disordered tendency
Si02g0669100	SiDHN1	SK3	290	1	156–285	5.76/31644.05	−1.246	2	N	N	57.93[a]	0.85
Si101756646	SiDHN2	YSK2	138	5	1–138	9.19/14126.34	−1.178	2	N	N	27.68	0.88
Si11g0454200	SiDHN3	YSK2	218	5	1–157	9.69/23222.86	−0.684	4	N	N	30.83	0.59
Si01g0702500	SiDHN4	Y2S	307	5	12–153	4.79/31173.12	−0.374	3	EC	C	44.42[a]	0.57
Si101755847	SiDHN5	SK2	150	8	1–150	10.11/15117.94	−0.829	4	NC	C	35.09	0.81
Si10g003700	SiDHN6	YSK3	347	8	13–287	8.99/33741.27	−0.804	2	N	N	11.92	0.76
Si11g0454000	SiDHN7	YSK2	169	8	14–169	8.81/16914.33	−1.050	2	NC	N	14.74	0.82
Sb01g20440	SbDHN1	Y2SK2	277	3	81–277	8.99/29881.14	−0.972	3	N	N	49.39[a]	0.72
Sb03g032255	SbDHN2	Y2S	188	3	12–178	8.37/19842.84	−0.535	3	N	C	52.57[a]	0.54
Sb03g037700	SbDHN3	SK3	283	3	141–279	5.79/31029.23	−1.282	1	N	N	59.66[a]	0.88
Sb09g018420	SbDHN4	YSK2	152	9	2–152	8.81/15399.74	−1.132	2	N	N	23.87	0.85
Sb05g50710	SbDHN5	SK	310	9	127–180	9.25/34491.4	−0.836	4	M	M	20.44	0.37[#]
Sb10g003700	SbDHN6	YSK3	388	10	293–388	8.50/37488.09	−0.330	2	N	N	35.69	0.77
Zm02g98750	ZmDHN1	K2S	326	1	9–264	7.37/31690.09	−0.798	2	N	N	24.17	0.62
Zm09g026210	ZmDHN2	S	325	3	9–266	8.56/31828.43	−0.745	3	N	EC	22.23	0.47
Zm04g032250	ZmDHN3	SK3	108	4	2–108	6.22/12199.09	−2.158	2	N	N	29.71	0.90
Zm01g013000	ZmDHN4	KS	290	5	2–164	6.05/31440.76	−1.250	1	N	N	54.20[a]	0.89
Zm03g037700	ZmDHN5	SK3	289	5	1–165	5.51/31466.47	−1.300	2	N	N	57.06[a]	0.89[b]
Zm01g20440	ZmDHN6	YSK3	143	8	39–123	6.06/15096.02	−0.306	2	EC	N	51.01[a]	0.72
Zm11g0840	ZmDHN7	YSK3	331	9	213–311	9.72/35266.67	−1.112	2	N	N	62.13[a]	0.76

Si:*Setaria italica*; Sb: Sorghum bicolour; Zm:*Zea mays*; N:Nuclear; EC: Extracellular; M: Mitochondrial; C: Chloroplast.

[a] Unstable.

[b] Ordered.

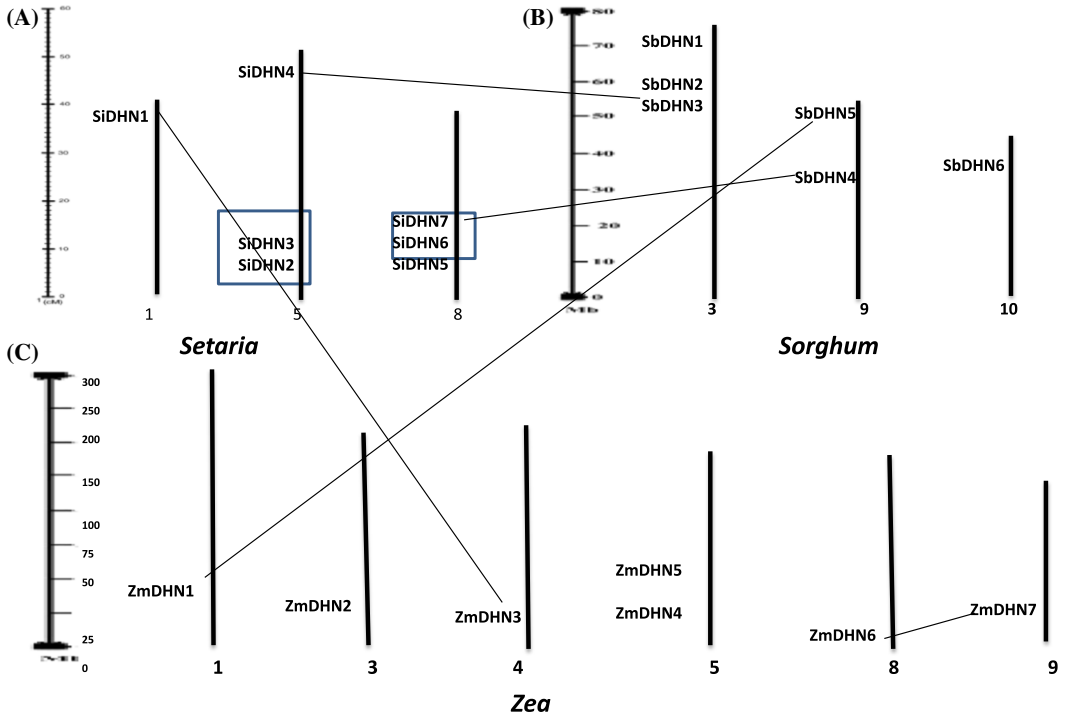


Fig. 1. Locations and duplications of *DHNs* in *Setaria* (A), *Sorghum* (B) and *Zea* (C); scale represents the mega bases. The chromosome numbers are indicated at the bottom of each bar.

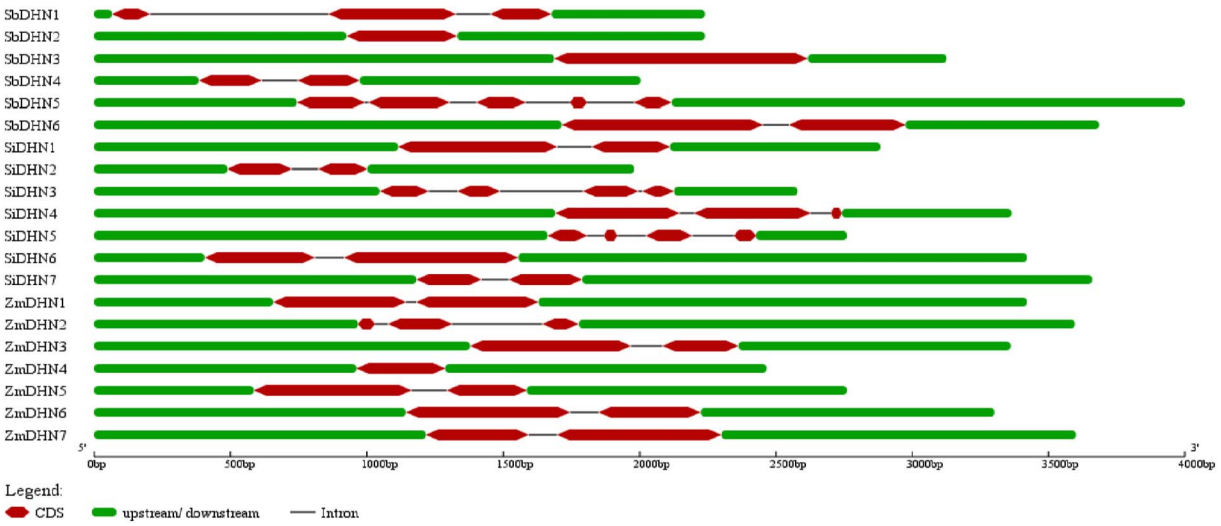


Fig. 2. Distribution of exons, introns, upstream and downstream regions in *DHNs*.

Further, it is observed that a KS-type *ZmDHN*4 and YSK-type *ZmDHN*7 lacked the strand motifs (Fig. 4A). Three-dimensional models of all 20 proteins generated at 80.4–95.2% confidence levels by similarity search software, the BLASTP are shown in Fig. 4B and Supplementary Table 3. Based on the highest homology, their structures have been visualized using Pymol tool (https://www.pymol.org/). The predicted 3-D structures of 20 *DHNs* revealed the presence of conserved DHN domain of nearly 150 amino acids. The β-sheets are absent in *SiDHN*1, 6 and *ZmDHN*1, while *ZmDHN*2, a KS-type *DHN*, lacked α-helices.

3.5. Identification of cis-regulatory elements of DHN promoters

Analysis of *cis*-acting elements revealed the presence of ABRE, DRE, DPBF, MYB and MYC, HSE, salt stress-responsive and LTR elements. *DHNs* are rich in Skn-1 type motifs, with endosperm specific expression elements that play an important role in seed development. They contain KST1 elements, involved in guard cell-specific gene expression, and pollen specific elements associated with pollen and anther development (Table 2). The motif analysis of *DHN* promoters revealed that 1, 3, 11, 14, 23, 25, 27 and 29 motifs have ABRE elements; 12, 13, 14, 21 and 23 have DRE; 2, 5, 6, 7, 9, 13 and 16 have HSE; 8, 25, 26 and 27 have LTRE; 3, 21, 23, 27 and 29 have CGCG (salt-responsive elements); 7, 14 and 24 have TAAG (endosperm-responsive elements); 6 has AGAAA (pollen and anther-responsive elements); 5 and 12 have GTCAT (guard cell-responsive elements); 9, 10, 11, 15, 17, 19, 27 and 30 have MYB; and 1, 6, 14, 21, 28 and 30 have MYC, the water stress-responsive elements (Supplementary Table 4; and Supplementary Figs. 2 and 3).

3.6. Phylogenetic and gene duplication analysis of DHNs

All *DHNs* of *Setaria, Sorghum*, and *Zea* were grouped into YnSKn, YnS, SKn, and KnS-types. While YnSKn has been found to be the largest

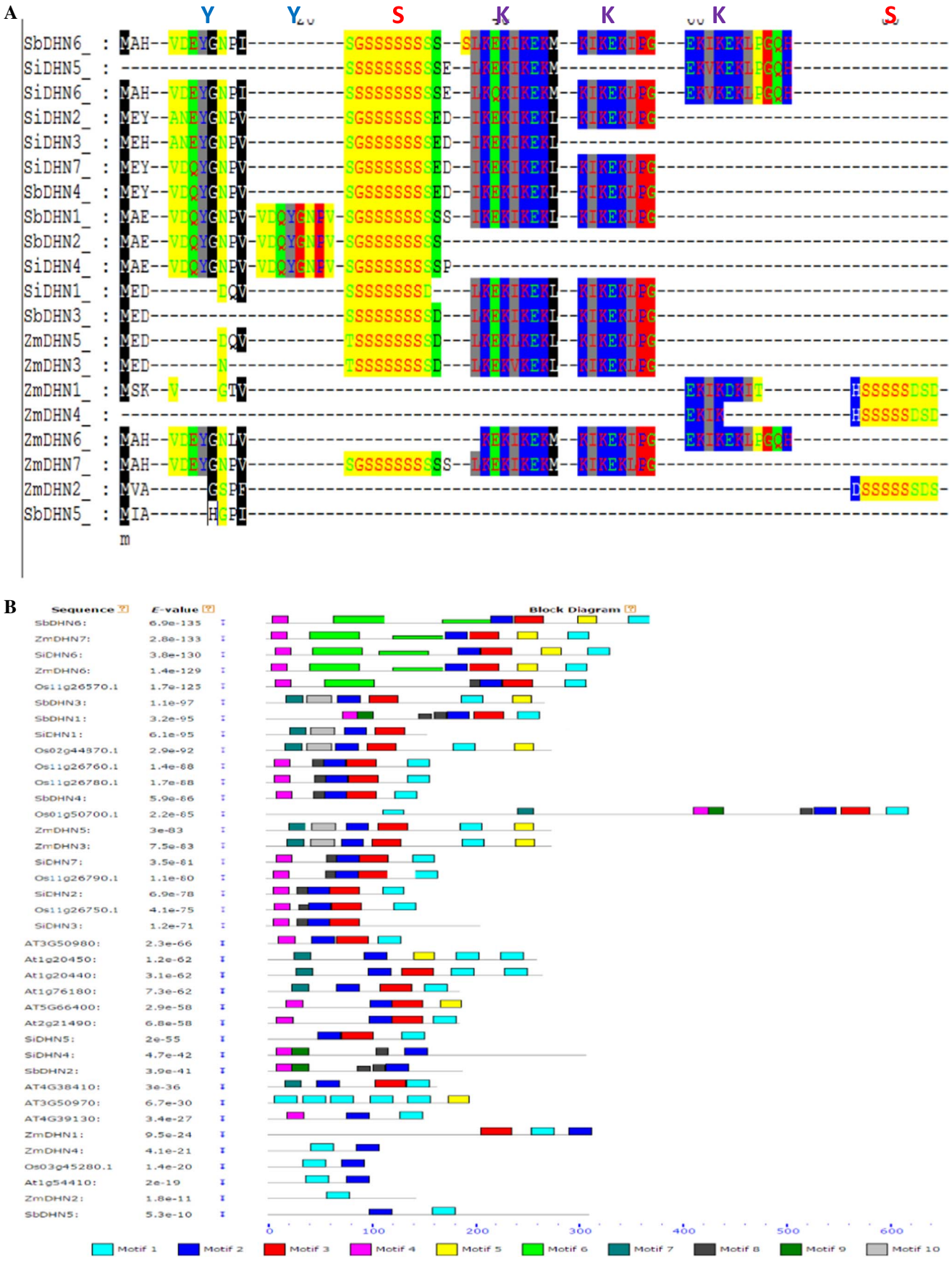


Fig. 3. A. *DHNs* exhibiting multiple sequence alignments and highly conserved Y, S, and K domains in *Setaria, Sorghum* and *Zea.* B. Distribution of 1–10 MEME identified *DHN* conserved motifs in *Setaria*, *Sorghum*, *Zea*, *Oryza*, and *Arabidopsis* are shown in colors. Gene clusters and p values are shown on the left side and motif sizes at the bottom of the figure.

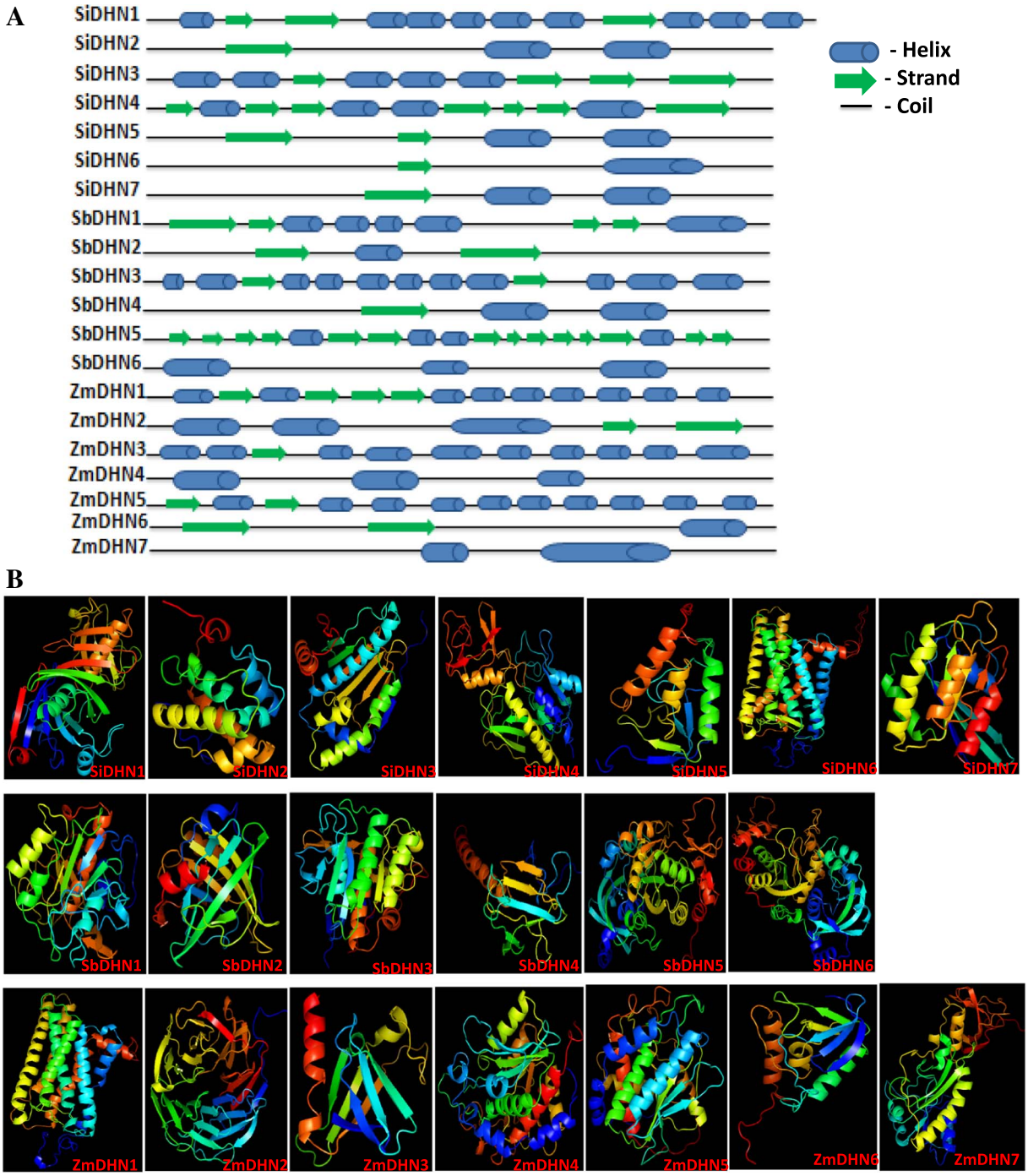


Fig. 4. A. *DHN* protein secondary structures of *Setaria*, *Sorghum* and *Zea*. B. Modelled 3D structures of *DHN* proteins; *Si* = *Setaria italica*, *Sb* = *Sorghum bicolor* and *Zm* = *Zea mays*.

subgroup (4/7 in *Setaria*, 3/6 in *Sorghum*, and 3/7 in *Zea*), followed by SKn subgroup (2/7 in *Setaria*, 2/6 in *Sorghum* and 2/7 in *Zea*), and the smallest has been KnS observed in *Zea* (2/7). But, only one KnS was found each in *Oryza* and *Arabidopsis*. A new subgroup, YnS, an intermediate type between YnSKn and SKn, and an ortholog clustered with YnSKn subgroup, has been noticed only in *Setaria* and *Sorghum* but absent in *Zea*. Two paralogs, the regional duplication events; *SiDHN*2 and *SiDHN3,* and *SiDHN5* and *SiDHN6*, were observed in *Setaria*, which might have resulted due to the gene duplication of their ancestral genes. One paralog *ZmDHN6* and *ZmDHN7* was reported as segmental duplication event in *Zea*, but no such paralogs were noticed in *Sorghum*. Out of the 4 common orthologs of *Setaria, SiDHN1* and *SbDHN3*, *SiDHN4* and *SbDHN2*, and *SiDHN7* and *SbDHN4*, are common to *Sorghum*, while only one *SiDHN1 and ZmDHN3* with *Zea*. Further, only one common ortholog, *SbDHN5* and *ZmDHN1* were found common among *Sorghum* and *Zea* (Figs. 1 and 5).

3.7. *In silico prediction of gene specific molecular markers and miRNAs targeting DHNs*

In the present study, a total number of 49 SSRs and 2 ILPs were discovered among genomic transcripts of identified *DHNs* (Supplementary Table 5). Tri-nucleotide SSR repeats (25/49) outnumbered the other repeats, while hexanucleotide SSR repeats (8/49) were found less than tri nucleotide SSRs. The dinucleotide and tetranucleotide SSR repeats were found more than pentanucleotide repeats,

Table 2
Conserved *cis*-acting elements in *DHN* promoters of *Setaria italica*, *Sorghum bicolor* and *Zea mays*.

DHNS	Cis elements										
	ABRE (CACGTG)	DRE (ACCGAC)	HSE (AGAAAATTCG)	LTR (CCGAAA)	*CGCGBOX* (*VCGCGB*)	*DPBF* (*ACACNNG*)	*GT1GMSCAM4* (*GAAAAA*)	*KST1* (*TAAAG*)	*MYB* (*WAACCA*/ *YAACKG*/ *CNGTTR*)	Myc (CANTTG)	SKN1(GTCAT)
SiDHN1	5	6	0	7	14	1	0	2	3	8	0
SiDHN2	7	1	2	1	2	2	1	4	24	8	2
SiDHN3	10	1	1	2	2	3	0	0	18	8	1
SiDHN4	14	4	2	3	4	5	6	0	20	6	3
SiDHN5	3	0	0	1	0	3	6	10	15	20	3
SiDHN6	13	5	5	7	10	8	2	12	44	24	2
SiDHN7	13	4	1	5	8	3	2	4	25	10	1
SbDHN1	10	8	0	6	12	4	0	2	3	10	1
SbDHN2	6	4	0	7	10	3	2	1	4	4	3
SbDHN3	17	1	0	5	34	1	2	3	14	21	0
SbDHN4	5	7	0	3	6	3	0	1	5	4	0
SbDHN5	9	5	0	8	32	4	3	1	12	15	0
SbDHN6	12	7	0	6	2	3	5	3	27	10	3
ZmDHN1	6	1	0	4	0	5	5	10	38	50	0
ZmDHN2	11	0	0	1	10	2	0	0	20	2	4
ZmDHN3	4	4	0	9	18	1	1	1	8	14	1
ZmDHN4	3	0	1	3	2	2	3	7	22	18	2
ZmDHN5	5	8	1	10	12	2	1	3	15	12	3
ZmDHN6	14	3	0	3	4	4	2	3	16	20	2
ZmDHN7	11	3	0	3	4	2	2	3	18	20	2

ABRECTAL: Response to ABA, CGCGBOX: Multiple signal transduction, DPBF: ABA, DRE: Dehydration responsive elements, GT1GMSAM4: Salt and pathogenesis related, LTRE: Low temperature and cold responsive, MYB: Response to drought and ABA, MYC: Response to drought, cold and ABA, POLLEN: Pollen and anther development, TKST1: Guard cell-specific gene expression.

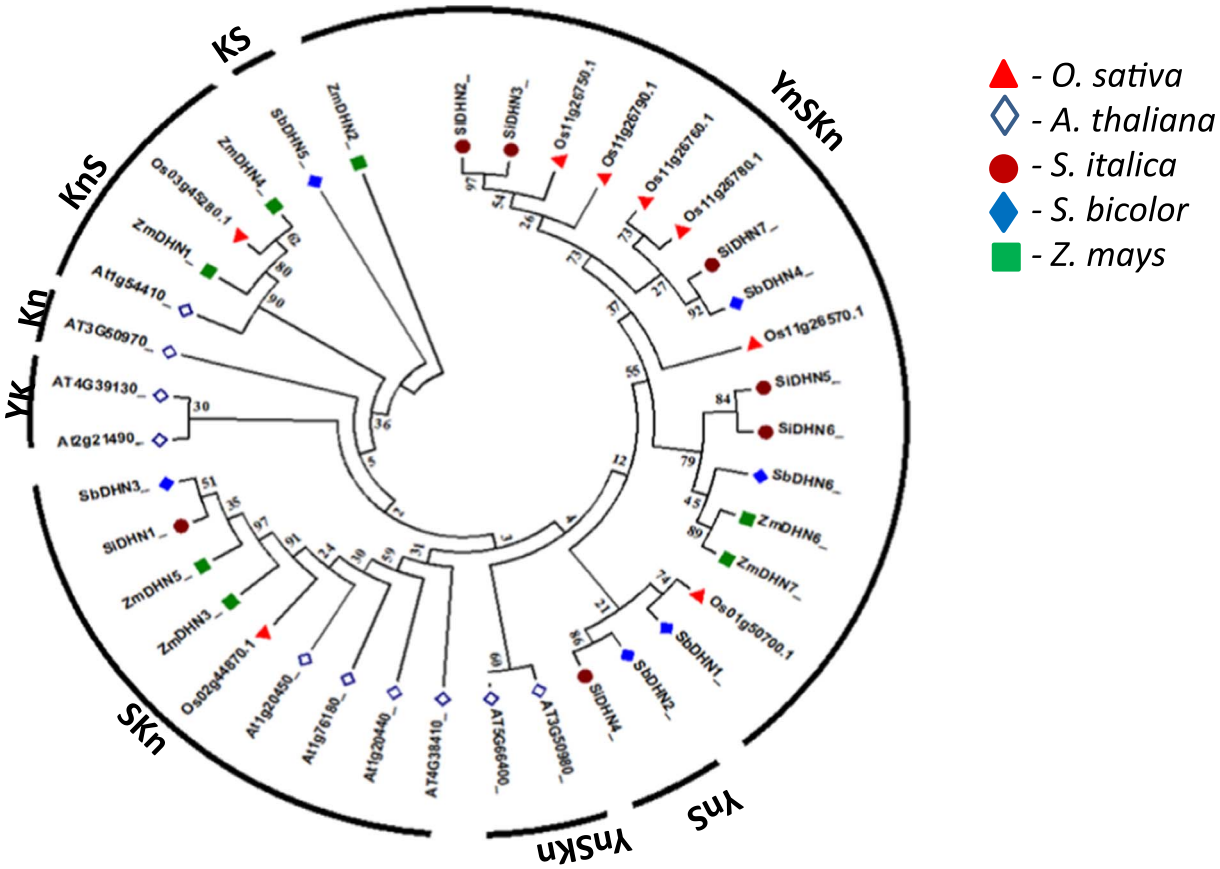


Fig. 5. Neighbor joining phylogenetic tree of YnSKn, YnS, SKn, KnS, YK and Kn *DHN* proteins of *O. sativa*, *A. thaliana*, *S. italica*, *S. bicolor* and *Z. mays*; *Os* = *Oryza sativa*, *At* = *Arabidopsis thaliana*, *Si* = *Setaria italica*, *Sb* = *Sorghum bicolor* and *Zm* = *Zea mays*.

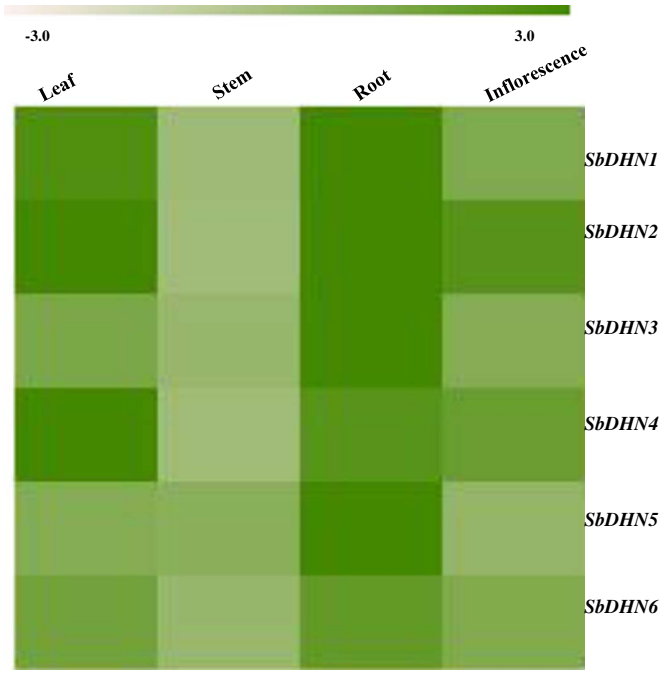


Fig. 6. Transcriptional profiling of *SbDHNs* in leaf, stem, root and inflorescence tissues of *Sorghum bicolor*.

but less than hexanucleotide repeats (Supplementary Fig. 4). Two ILP markers were also mined in *SbDHN2* gene. Besides, putative microRNAs (miRNAs) targeting the *DHNs* genes were identified using psRNATarget server. The analysis revealed 14 *DHN* (*ZmDHN1, ZmDHN3, ZmDHN4, ZmDHN6, SbDHN3, SbDHN5, SbDHN6, SbDHN1, SbDHN4, SiDHN2, SiDHN4, SiDHN6, SiDHN7,* and *SiDHN3*) genes are targeted by 51 miRNAs which belong to diverse classes of miRNA families responsive to various abiotic stresses (Supplementary Table 6).

3.8. Classification, signature amino acid analysis and evolutionary relationship of DHNs

A total of 451 *DHNs* belonging to 17 families and 53 crops were identified and classified into YnSKn, SKn, KnS, Kn, S, YnKn, and YnS based on their conserved characteristic domains. Out of them, 223 were divided into YnSKn, 123 SKn, 23 KnS, 47 Kn, 4 YnS, 23 YnKn, and 8 S sub-types. The YnSKn is the most common sub-group in all the families, while SKn members appeared less in number in monocots when compared to dicots. Both KnS and Kn appeared only in fewer species (Supplementary Fig. 5 and Supplementary Table 7), suggesting that they are evolved in particular genome or they might have lost during the course of evolution. Further evaluation of segments revealed several truncated segments especially with K segment. Out of 451 *DHNs*, it has been observed that they are absent in 24 K segments (Supplementary Table 8 and Supplementary Figs. 6 & 7). The amino acid composition analyses illustrated that *DHNs* are rich with glutamic acid, glycine, histidine, alanine, arginine, aspartic acid, leucine, proline, threonine, and valine along with lysine and serine. But, cysteine and tryptophan are completely absent. *DHNs* exhibited variations in glutamic acid and glycine percentages, and if glutamic acid residues are more, glycine residues are less and *vice versa*. Interestingly, all the *DHNs* exhibited the highest percentage of glycine, except SKn sub-type, which contained more amount of glutamic acid. The KnS and Kn sub-groups are rich with histidine. Further, proline levels also exhibited variations alongside glutamic acid (Supplementary Table 9).

3.9. Transcriptional profiling of SbDHNs

All the 6 identified and confirmed *DHN* genes in *S. bicolor* exhibited better expression in roots in comparison with leaves, inflorescences, and stems. Though *SbDHN2* and *SbDHN4* were constitutively expressed in all the four tissues, their expression levels were high in roots. The

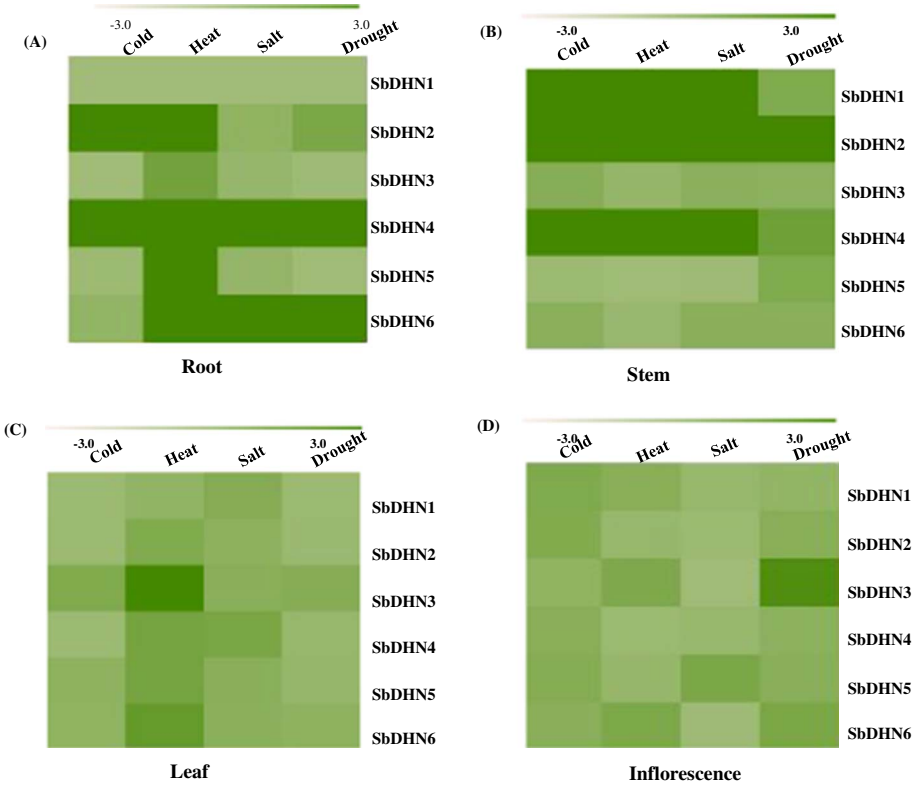


Fig. 7. Relative expression patterns of *SbDHNs* in different tissues under cold, heat, salt and drought stresses in *Sorghum bicolor*, (a) root, (b) stem, (c) leaf and (d) inflorescence.

other 4 *DHNs* exhibited upregulation in leaves and inflorescences but down-regulation in stems (Fig. 6 & Supplementary Table 10). Under drought, salt, heat, and cold stresses, *Sorghum DHNs* displayed differential expressions in roots, stems, leaves, and inflorescences. Among the 6 *DHN* genes, *SbDHN2* exhibited the highest expression levels under stress in roots, stems, leaves, and inflorescences followed by *SbDHN4* and *SbDHN6*. While *SbDHN1* and *3* did not exhibit upregulation in any one of the stress conditions, expression of *SbDHN2* in roots was 114.9-folds higher under high temperature compared to cold, salt, and drought. *SbDHN4* recorded upregulation (10.3-folds) in roots treated with high temperature stress followed by drought (7.6-folds). Fold-wise increase in the expression of *SbDHN5* was 3.5 under elevated temperature. In contrast, *SbDHN6* showed 18.8 and 5.5-folds higher expression levels in roots under salt and high temperature stresses respectively (Fig. 7A & Supplementary Table 10). In leaf, *SbDHN1*, *2, 4, 5*, and *6* did not exhibit any upregulation during stress. Expression level of *SbDHN3* was 3.2-folds better under high temperature stress in leaf (Fig. 7B & Supplementary Table 10). Contrarily, *SbDHN1* expression was 7.2, 3.8, and 3.8-folds higher under cold, high temperature, and drought stresses respectively in stem tissues. *SbDHN2* recorded 160.4-folds increase in its expression levels under cold followed by drought (136.4-folds), high temperature (123.6-folds), and salt stresses (50.8-folds). Thus, upregulation of *SbDHN2* appeared high in the stem compared to other *DHNs*. In contrast, *SbDHN3*, *5* and *6* did not exhibit higher levels of expression under stress conditions (Fig. 7C & Supplementary Table 10). In inflorescence, barring *SbDHN3*, other *DHNs* were not much upregulated. Only *SbDHN3* recorded 2.6-folds higher expression under salt stress (Fig. 7D & Supplementary Table 10).

3.10. Protein – protein interaction of SbYnS DHNs

To explore the functions of novel SbYnS sub-group *DHN*s, protein – protein interaction network map was constructed (Supplementary Fig. 8). The map showed interaction among the proteins Hsp 70, FKBP-type peptidyl-prolyl *cis*-trans isomerise, tankyrase 2, ser/thr protein phosphatase 6, ankyrin repeat, SOCS box protein 3 and ankyrin protein 3 that contained tetratricopeptide and ankyrin repeats with YnS type *DHN*s. Their functions are retrieved based on protein – protein interaction network, and it appeared that they participate in endocytosis, spliceosome and protein processing in the endoplasmic reticulum.

4. Discussion

4.1. Characterization of DHNs

Comparison between *Setaria, Sorghum*, and *Zea DHNs* revealed variation in the number and patterns of exons and introns. Presence of more than one intron in all the 3 cereals have been observed with an exception of *SbDHN3* and *ZmDHN4* which are devoid of any introns, similar to their common ancestor, the *Oryza* (Wang et al., 2007). The number of exons and introns in a gene revealed the divergent relationship between the gene families as pointed out by Cao (2012). More number of introns may be causing the delay of transcription by extending the length of nascent transcript and thus burdening the gene expression (Jeffares et al., 2008). All the *DHNs* exhibited distinct differences between pI and kinases. The positively charged YnSKn-type *DHNs*, with higher pI are bound to the cell membranes during stress, thereby protect the cells and thus confer stress tolerance (Yang et al., 2012). The present investigation revealed that YnSKn type *DHNs* are phosphorylated by PKCs, while SKn *DHNs* by CK2s, and both the types may be promoting the activity of *DHNs* for conferring tolerance against stress. *DHNs* are highly hydrophilic and unstructured, and due to this nature, they escape from stress and protect other proteins too (Hincha and Thalhammer, 2012). However, the *SbDHN5* has been found as an ordered protein with higher pI and phosphorylation sites.

The YnSKn *DHNs* are common in all the three crops (*Setaria* 4, *Sorghum* 3 and *Zea* 3), and are triggered in response to severe drought, salt, frost, ABA, methyl jasmonate, salicylic acid, and high temperature (Close, 1997; Rahman et al., 2010; Halder et al., 2017), compared to one SKn-type *DHN* in *Setaria*, 2 each in *Sorghum* and *Zea*. Two KnS-type chilling stress-responsive *DHNs* are present in *Zea*, but absent in *Setaria* and *Sorghum*. Expression of KnS-type *DHN* was reported in reproductive tissues in response to chilling stress (Wang et al., 2014). In the present investigation, a new group, called YnS-type *DHNs*, endowed with more number of phosphorylation sites and abiotic stress regulatory *cis*-acting elements, 1 each in *Setaria* and *Sorghum*, have been identified for the first time. Kn-type *DHNs* identified in *Arabidopsis* have not been observed in all the 3 crops (Close, 1996). It appears that LEA genes are highly conserved (Liang et al., 2016) among plants though gene losses or gains were noticed. Two regional duplication events were noticed in *Setaria*, while one segmental duplication event in *Zea*, mays. But, no duplication event was noticed in *Sorghum*, indicating less number of *DHNs* in *S. bicolor* compared to *Setaria* and *Zea*. Perhaps, it is lost in the evolution. The distributions of *DHNs* in monocots are crowded on only few specific chromosomes, with an exception of *Zea*, compared to dicots.

4.2. Phylogenetic analysis of DHNs

Analysis of phylogenetic tree revealed presence of 4 divergent subgroups of *DHNs* in *Setaria*, *Sorghum*, and *Zea* and on comparison exhibited similarity with *Oryza* (monocot), but wide variation was noticed with that of *Arabidopsis* (dicot), indicating that the *DHNs* are derived from their common ancestor *Oryza*. Expansion of *DHN* family generally occurs through tandem and genome duplication events. In *Arabidopsis*, 3 tandem duplications and 3 whole genome duplication events resulted into 6 *DHNs*, thus increasing the original 4 *DHNs* to a total of 10 *DHNs* (Hundertmark and Hincha, 2008). Similarly, 3 tandem duplication events in *Oryza* resulted into 3 *DHNs* and thus enhanced the original 5 *DHNs* to a total of 8 *DHNs* (Wang et al., 2007). It appeared that the whole genome duplication event must have occurred at least once in poplar, *Oryza*, and *Arabidopsis*, while such an event is unlikely in *Setaria*, *Sorghum*, and *Zea* (Jaillon et al., 2007), but resulted into a varied number of *DHNs* in these crops. The YnSKn *DHNs* are expressed during drought and salt stress, while SKn, KnS and Kn mostly during cold stress though some of them appeared to be associated with desiccation and salt stresses (Liang et al., 2012; Wang et al., 2014). This indicated that YnSKn *DHNs* are relatively associated with desiccation and salt stress, whereas SKn, KnS, and Kn seemed to be associated with plants like *Triticum* and members of Rosaceae family that grow in low temperatures.

4.3. Promoter analysis of DHNs

In the present study, endosperm specific *cis*-elements SKn-1 were noticed closer to translation start sites in majority of the *DHN* promoters and are upregulated during the late embryogenesis stage (Washida et al., 1999). Baker et al. (1994) demonstrated that *DHN* promoters rich in low temperature-responsive elements confer tolerance against cold, drought, and ABA-induced stresses. Heat shock elements (HSE) have been noticed in *Setaria* and *Zea DHNs*, but surprisingly not in *Sorghum*. The guard cell specific and stomatal conductivity regulating KST1 *cis*-elements noticed in *DHN* promoters may participate in K^+ influx and guard cell movement during stress (Plesch et al., 2001). This study also revealed the presence of AGAAA-rich POLLEN1LELAT52 *cis*-elements inferring the involvement of *DHNs* in anther and pollen development (Flicchkin et al., 2004).

4.4. In silico analysis of gene specific molecular markers and miRNAs

S. bicolor is sensitive to cold, drought and salt stresses. Several molecular markers (simple sequence repeats) have been identified in *S.*

bicolor for cold (Burow et al., 2011) and drought (Zhu et al., 2017). Therefore, the observed gene specific markers in *DHNs* might aid further in the development of drought, salinity, and cold stress tolerant sorghum cultivars using genotyping and marker-assisted selection approaches.

Most of the stress-responsive miRNAs target transcription factors. For example, miR164 targeted the NAC mRNAs in *Arabidopsis thaliana* and rice and altered the plant developmental and abiotic stress responses (Fang et al., 2014). In *Arabidopsis*, miR156-mediated down-regulation of SPL enhanced the abiotic stress tolerance, and heat stress memory (Stief et al., 2014). Since several miRNAs were shown to be upregulated under multiple abiotic stresses (Sunkar and Zhu, 2004; Zhou et al., 2010), Our *in silico* analysis points out that miRNAs target *DHNs* and therefore, they may be validated further for understanding the mechanism of DHN activities and for improving abiotic stress tolerance in *S. bicolor*.

4.5. Network analysis of SbYnS–type DHN

It has been observed that *SbYnS DHN* contained a Dnaj domain. DnaJs are co-chaperones which assist Hsp70 and bring about temperature stress tolerance. Work by Mulaudzi-Masuku et al. (2015) demonstrated that transfer of plant Hsp70 gene brought about thermal tolerance in *E. coli*. The Hsp 70 proteins along with DnaJ prevented aggregation of proteins, participated in protein translocation and mediated assembly or dis-assembly of multimeric proteins and targeted proteosomes for degradation (Hartl, 1996). Tetratrico peptide repeat (TPR) motifs are protein-protein interaction modules that are associated with the regulation of diverse cellular functions. Schapire et al. (2006) have identified TITAN LIKE protein (TTL1) containing TPR motifs. Such motifs have been found to be required for abscisic acid responses and osmotic stress tolerance. Therefore, proteins containing TPRs have emerged as essential determinants for signal transduction mediated by stress-related hormone. Association of *DHNs* with TPR protein indicates that this interacting partner is helping in signal transduction during stress. The SbYnS DHN along with other proteins in the network might maintain the membrane integrity, protect proteins from denaturation, and scavenge ROS under diverse abiotic stress conditions.

4.6. Transcript profiling of DHNs in different tissues under abiotic stress

High expression of *DHNs* was observed at the late embryogenesis stages compared to vegetative tissues and very limited expression at the seedling stages in *Arabidopsis* (Rorat et al., 2004). But upon exposure of plants to stress, higher amounts of *DHN* expressions were noticed in the vegetative tissues (Bray, 1994). Similar to *OsDHN3* in rice, *SbDHN*2 in *Sorghum* exhibited the highest expression levels under stress in roots, stems, leaves, and inflorescences (Verma et al., 2017). Higher expression levels of *SbDHN2*, *4*, and *6* observed in root, stem, leaf, and inflorescence indicated that they play an important role during vegetative as well as reproductive stages by their participation in plant development, pollen germination and seed filling; similar to that of higher expression levels recorded in *Arabidopsis* LEA gene (At5g27980) (Wang et al., 2008). *SbDHN2* and *4* were highly induced under all abiotic stresses in roots and stems in comparison with other *DHNs*. Massarelli et al. (2006) used a functional screening method based on random overexpression of a plant cDNA library in *E. coli* to identify plant genes related to salt tolerance. They found that DHN2 gene is induced by NaCl. This suggests that DHN2 protein is associated with salt stress and is conserved across prokaryotes as well as plants. Expression of three *DHN* genes was noticed in sugarcane under heat stress, but the expression was independent of changes in water relations in leaves (Wahid and Close, 2007). Similarly, grapevine *DHN2* was induced by both heat and cold stress with different expression profiles (Yang et al., 2012). Xu et al. (2008) found that expression of brassica *BjDHN2* and *BjDHN3* resulted in higher tolerance to Cd^{2+} and Zn^{2+} metals by attenuating lipid peroxidation and protecting cellular membranes. Thus, *DHN2* gene stands apart by associating with multiple stresses like salt, heat, cold and metal unlike that of other *DHN* genes. Significantly high expression of *SbDHN3* was observed in the inflorescence under salt and heat stresses, similar to the expression of grapevine *DHN1* during late embryogenesis under drought, cold and heat (Yang et al., 2012). High activity levels in different tissues under varied abiotic stress conditions inferred the involvement of *DHNs* during developmental processes also.

5. Conclusion

In the present study, a novel Y2S subgroup was identified in *Sorghum bicolor*. *SbDHN2* gene, belonging to Y2S subgroup, upregulated especially in stems under different abiotic stress conditions indicated its potential role in stress. *DHNs* expressed abundantly in roots, leaves, and stems, particularly *SbDHN2*, *4*, and *6* under cold, high temperature, salt, and drought stress conditions. The present investigation laid a foundation for further functional validation of *DHNs* and the development of cereals for abiotic stress tolerance.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plgene.2018.01.004.

Conflict of interest

The authors declare that no conflict of interest exists.

Author contribution

PBK and DMR conceived and designed the experiments. MN, PSR, SAK, AK, and AA performed the experiments. MN, PBK, SAK, PS, RKS and DMR prepared the manuscript. All authors have read and approved the manuscript.

Acknowledgments

MN is thankful to the UGC, New Delhi, for providing fellowship. PBK is thankful to the CSIR, New Delhi, for providing CSIR-Emeritus Scientist Fellowship. DMR is thankful to the UGC [F.41-828/2012 (SR)], New Delhi, for financial support through UGC-MRP.

References

Alsheikh, M.K., Heyen, B.J., Randall, S.K., 2003. Ion-binding properties of the dehydrin ERD14 are dependent upon phosphorylation. J. Biol. Chem. 278, 40882–40889.
Anchordoguy, T.J., Carpenter, J.F., 1996. Polymers protect lactate dehydrogenase during freeze-drying by inhibiting dissociation in the frozen state. Arch. Biochem. Biophys. 332, 231–238.
Bailey, T.L., Williams, N., Misleh, C., Li, W.W., 2006. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 34 (Suppl. 2), W369–W373.
Baker, S.S., Wilhelm, K.S., Thomashow, M.F., 1994. The 5′-region of *Arabidopsis thaliana* cor15a has *cis*-acting elements that confer cold, drought and ABA-regulated gene expression. Plant Mol. Biol. 24, 701–713.
Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., et al., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.
Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., Brunak, S., 2004. Prediction of post translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 4, 1633–1649.
Bray, E.A., 1994. Alterations in gene expression in response to water deficit. In: Basra, A.S. (Ed.), Stress Induced Gene Expression in Plants. Harwood Academic Publishers, Ludhiana, India, pp. 1–23.
Burow, G., Burke, J.J., Xin, Z., Franks, C.D., 2011. Genetic dissection of early-season cold tolerance in sorghum (*Sorghum bicolor* (L.) Moench). Mol. Breed. 28 (3), 391–402.
Cao, J., 2012. The pectin lyases in *Arabidopsis thaliana*: evolution, selection and expression profiles. PLoS One 7, e46944.
Close, T.J., 1996. Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol. Plant. 97, 795–803.
Close, T.J., 1997. Dehydrins: a commonalty in the response of plants to dehydration and low temperature. Physiol. Plant. 100, 291–296.
Dai, X., Zhao, P.X., 2011. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 39, 155–159.
Dosztanyi, Z., Csizmok, V., Tompa, P., Simon, I., 2005. IUPred: web server for the

prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21, 3433–3434.
Fang, Y., Xie, K., Xiong, L., 2014. Conserved miR164-targeted NAC genes negatively regulate drought resistance in rice. J. Exp. Bot. 65, 2119–2135.
Flicchkin, S.A., Leonard, J.M., Monteros, A., Liu, P.P., Nonogaki, H., 2004. A novel endo-beta mannanase gene in tomato LeMAN5 is associated with anther and pollen development. Plant Physiol. 134, 1080–1087.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., Walker, John M., 2005. Protein identification and analysis tools on the ExPASy server. In: The Proteomics protocols Handbook. Humana Press, pp. 571–607.
Guo, A.Y., Zhu, Q.H., Chen, X., Luo, J.C., Chuan, Y.I., 2007. GSDS: a gene structure display server. Yi Chuan 29, 1023–1026.
Halder, T., Upadhayaya, G., Ray, S., 2017. YSK2 type dehydrin (SbDhn1) from *Sorghum bicolor* showed improved protection under high temperature and osmotic stress condition. Front. Plant Sci. 8 (918). http://dx.doi.org/10.3389/fpls.2017.00918.
Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., Masmoudi, K., 2011. Plant dehydrins and stress tolerance. Plant Signal. Behav. 6 (10), 1503–1509.
Hartl, F.U., 1996. Molecular chaperones in cellular protein folding. Nature 381 (6583), 571–579.
Higo, K., Ugawa, Y., Iwamoto, M., Korenaga, T., 1999. Plant *cis*-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res. 27, 297–300.
Hincha, D.K., Thalhammer, A., 2012. LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. Biochem. Soc. Trans. 40, 1000–1003.
Hinniger, C., Caillet, V., Michoux, F., Ben Amor, M., Tanksley, S., Lin, C., McCarthy, J., 2006. Isolation and characterization of cDNA encoding three dehydrins expressed during *Coffea canephora* (Robusta) grain development. Ann. Bot. 97, 755–765.
Horton, P., Keun-Joon, P., Takeshi, O., Naoya, F., Hajime Harada, C.J., Adams, C., Kenta, N., 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35, W585–W587 Web Server issue.
Hundertmark, M., Hincha, D.K., 2008. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. BMC Genomics 9, 118.
Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., et al., 2007. French-Italian Public Consortium for grapevine genome characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463–467.
Jeffares, D.C., Penkett, C.J., Bahler, J., 2008. Rapidly regulated genes are intron poor. Trends Genet. 24, 375–378.
Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202.
Koag, M.C., Wilkens, S., Fenton, R.D., Resnik, J., Vo, E., Close, T.J., 2009. The K-segment of maize *DHN*1 mediates binding to anionic phospholipid vesicles and concomitant structural changes. Plant Physiol. 150, 1503–1514.
Kovacs, D., Kalmar, E., Torok, Z., Tompa, P., 2008. Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. Plant Physiol. 147, 381–390.
Kumar, M., Lee, S.C., Kim, J.K., Kim, S.J., Aye, S.S., Kim, S., 2014. Over-expression of dehydrin gene, *OsDHN1*, improves drought and salt stress tolerance through scavenging of reactive oxygen species in rice (*Oryza sativa* L.). J. Plant Biol. 57, 383–393.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK - a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.
Lescot, M.P., Dehais, G., Thijs, K., Marchal, Y., Moreau, Y., Van de Peer, P., Rombauts, R.S., 2002. 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, 325–327.
Letunic, I., Copley, R.R., Schmidt, S., Ciccarelli, F.D., Doerks, T., Schultz, J., Ponting, C.P., Bork, P., 2004. SMART 4.0: towards genomic data integration. Nucleic Acids Res. 32D, 142–144.
Liang, D., Xia, H., Wu, S., Ma, F., 2012. Genome-wide identification and expression profiling of dehydrin gene family in *Malus domestica*. Mol. Biol. Rep. 39, 10759–10768.
Liang, Y., Ziyi, X., Jianxiao, Z., Dongyang, X., Zeyang, Z., Jun, X., Jianping, G., Nadia, R., Yongtai, Y., Maoteng, L., 2016. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (*LEA*) gene family formation pattern in *Brassica napus*. Sci. Rep. 6, 24265. http://dx.doi.org/10.1038/srep24265.
Liu, C.C., Li, C.M., Liu, B.G., Ge, S.J., Dong, X.M., Li, W., Zhu, H.Y., Wang, B.C., Yang, C.P., 2012. Genome-wide identification and characterization of a dehydrin gene family in poplar (*Populus trichocarpa*). Plant Mol. Biol. Report. 30, 848. http://dx.doi.org/10.1007/s11105-11011-10395-11101.
Livernois, A.M., Hnatchuk, D.J., Findlater, E.E., Graether, S.P., 2009. Obtaining highly purified intrinsically disordered protein by boiling lysis and single step ion exchange. Anal. Biochem. 392, 70–76. http://dx.doi.org/10.1016/j.ab.2009.05.023.
Massarelli, I., Cioffi, R., Batelli, G., De Palma, M., Costa, A., Grillo, S., Leone, A., 2006. Functional screening of plant stress-related cDNAs by random over-expression in *Escherichia coli*. Plant Sci. 170, 880–888.
Moller, S., Croning, M.D.R., Apweller, R., 2001. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17, 646–653.
Mulaudzi-Masuku, T., Mutepe, R.D., Mukhoro, O.C., Faro, A., Ndimba, B., 2015. Identification and characterization of a heat-inducible Hsp70 gene from *Sorghum bicolor* which confers tolerance to thermal stress. Cell Stress Chaperones 20 (5), 793–804. http://dx.doi.org/10.1007/s12192-015-0591-2.
Pfaffl, W.M., Horganl, W.G., Leo, D., 2002. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 3, 900.
Plesch, G., Ehrhardt, T., Mueller-Roeber, B., 2001. Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression. Plant J. 28, 455–464.
Rahman, L.N., Chen, L., Nazim, S., Bamm, V.V., Yaish, M.W., Moffatt, B.A., et al., 2010. Interactions of intrinsically disordered *Thellungiella salsuginea* dehydrins *TsDHN*-1 and *TsDHN*-2 with membranes-synergistic effects of lipid composition and temperature on secondary structure. Biochem. Cell Biol. 88, 791–807.
Reddy, P.S., Reddy, D.S., Sivasakthi, K., Bhatnagar-Mathur, P., Vadez, V., Sharma, K.K., 2016. Evaluation of sorghum [*Sorghum bicolor* (L.)] reference genes in various tissues and under abiotic stress conditions for quantitative real-time PCR data normalization. Front. Plant Sci. 7, 529. http://dx.doi.org/10.3389/fpls.2016.00529.
Rorat, T., Grygorowicz, W.J., Irzykowski, W., Rey, P., 2004. Expression of KS-type dehydrins is primarily regulated by factors related to organ type and leaf developmental stage during vegetative growth. Planta 218, 878–885.
Schapire, A.L., Valpuesta, V., Botella, M.A., 2006. TPR proteins in plant hormone signaling. Plant Signal. Behav. 1, 229–230.
Stief, A., Altmann, S., Hoffmann, K., Pant, B.D., Scheible, W.R., Bäurle, I., 2014. *Arabidopsis* miR156 regulates tolerance to recurring environmental stress through SPL transcription factors. Plant Cell 26, 1792–1807.
Sunkar, R., Zhu, J.K., 2004. Novel and stress-regulated MicroRNAs and other small RNAs from *Arabidopsis*. Plant Cell 16, 2001–2019.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Tommasini, L., Svensson, J.T., Rodriguez, E.M., Wahid, A., Malatrasi, M., Kato, K., Wanamaker, S., Resnik, J., Close, T.J., 2008. Dehydrin gene expression provides an indicator of low temperature and drought stress: transcriptome-based analysis of barley (*Hordeum vulgare* L.). Funct. Integr. Genom. 8, 387–405.
Verma, G., Dhar, Y.V., Srivastava, D., Kidwai, M., Chauhan, P.S., Bag, S.K., et al., 2017. Genome-wide analysis of rice dehydrin gene family: its evolutionary conservedness and expression pattern in response to PEG induced dehydration stress. PLoS One 12 (5), e0176399. http://dx.doi.org/10.1371/journal.pone.0176399.
Wahid, A., Close, T.J., 2007. Expression of dehydrins under heat stress and their relationship with water relations of sugarcane leaves. Biol. Plant. 51 (1), 104–109.
Wang, X., Zhu, H., Jin, G., Liu, H., Wu, W., Zhu, J., 2007. Genome-scale identification and analysis of LEA genes in rice (*Oryza sativa* L.). Plant Sci. 172, 414–420.
Wang, Y., Zhang, W.Z., Song, L.F., Zou, J.J., Su, Z., Wu, W.H., 2008. Transcriptome analysis show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. Plant Physiol. 148, 1201–1211.
Wang, Y., Xu, H., Zhu, H., Tao, Y., Zhang, G., Zhang, L., et al., 2014. Classification and expression diversification of wheat dehydrin genes. Plant Sci. 214, 113–120.
Washida, H., Wu, C.Y., Suzuki, A., Yamanouchi, U., Akihama, T., Harada, K., Takaiwa, F., 1999. Identification of cis-regulatory elements required for endosperm expression of the rice storage protein glutelin gene GluB-1. Plant Mol. Biol. 40, 1–12.
Webb, W., Sali, A., 2014. Comparative protein structure modelling using modeller. In: Current Protocols in Bioinformatics. John Wiley & Sons, Inc. (5.6.1–5.6.32).
Xu, J., Zhang, Y.X., Wei, W., Han, L., Guan, Z.Q., Wang, Z., et al., 2008. *BjDHNs* confer heavy-metal tolerance in plants. Mol. Biotechnol. 38, 91–98. http://dx.doi.org/10.1007/s12033-007-9005-8.
Yang, Y., Mingyang, H., Ziguo, Z., Shuxiu, L., Yan, X., Chaohong, Z., Stacy, D.S., Yuejin, W., 2012. Identification of the dehydrin gene family from grapevine species and analysis of their responsiveness to various forms of abiotic and biotic stress. BMC Plant Biol. 12, 140.
Yu, C.S., Chen, Y.C., Lu, C.H., Hwang, J.K., 2006. Prediction of protein sub cellular localization. Proteins Struct. Funct. Bioinformat. 64, 643–651.
Zhou, L., Liu, Y., Liu, Z., Kong, D., Duan, M., Luo, L., 2010. Genome-wide identification and analysis of drought-responsive microRNAs in *Oryza sativa*. J. Exp. Bot. 61, 4157–4168.
Zhu, Y., Wang, X., Huang, L., Lin, C., Zhang, X., Xu, W., Peng, J., Li, Z., Yan, H., Luo, F., Wang, X., Yao, L., Peng, D., 2017. Transcriptomic identification of drought-related genes and SSR markers in sudan grass based on RNA-Seq. Front. Plant Sci. 8, 687. http://dx.doi.org/10.3389/fpls.2017.00687.

Send Orders for Reprints to reprints@benthamscience.ae

Current Genomics, 2015, *16*, 279-291

Genome-wide Scanning and Characterization of *Sorghum bicolor* L. Heat Shock Transcription Factors

M. Nagaraju[1,#], Palakolanu Sudhakar Reddy[2,#], S. Anil Kumar[1], Rakesh K. Srivastava[2], P. B. Kavi Kishor[1,*] and D. Manohar Rao[1]

[1]*Department of Genetics, Osmania University, Hyderabad 500 007, India;* [2]*International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru-502324, Hyderabad, India*

P. B. Kavi Kishor

Abstract: A genome-wide scanning of *Sorghum bicolor* resulted in the identification of 25 *SbHsf* genes. Phylogenetic analysis shows the ortholog genes that are clustered with only rice, representing a common ancestor. Promoter analysis revealed the identification of different *cis*-acting elements that are responsible for abiotic as well as biotic stresses. Hsf domains like DBD, NLS, NES, and AHA have been analyzed for their sequence similarity and functional characterization. Tissue specific expression patterns of *Hsfs* in different tissues like mature embryo, seedling, root, and panicle were studied using real-time PCR. While *Hsfs*4 and 22 are highly expressed in panicle, 4 and 9 are expressed in seedlings. *Sorghum* plants were exposed to different abiotic stress treatments but no expression of any *Hsf* was observed when seedlings were treated with ABA. High level expression of Hsf1 was noticed during high temperature as well as cold stresses, 4 and 6 during salt and 5, 6, 10, 13, 19, 23 and 25 during drought stress. This comprehensive analysis of *SbHsf* genes will provide an insight on how these genes are regulated in different tissues and also under different abiotic stresses and help to determine the functions of *Hsfs* during drought and temperature stress tolerance.

Keywords: Heat shock transcription factors, Phylogenetic analysis, *Cis*-acting elements, Paralogs, Molecular chaperones, Abiotic stress.

1. INTRODUCTION

High temperature and drought have adverse effects on water relations, photosynthesis and results in 50% crop reduction [1]. In response to heat stress, rapid accumulation of small heat shock proteins (Hsps) was observed in all eukaryotes and plants. Hsps act as molecular chaperones and prevent the aggregation and denaturation of proteins [2]. Heat shock transcription factors (*Hsfs*) transcriptionally regulate the *Hsp* genes. Plant *Hsfs* play a central role in the heat stress response. Tomato *HsfA1*, *A2*, and *A3* confer heat stress tolerance when overexpressed [3-5]. *LpHsfA1a* and *AtHsfA2* enhance thermotolerance upon overexpression but abolished when knocked-out or interfered [6, 7]. Transcription factor *A2* has been found as a key regulator in response to many environmental stresses [8]. In *Arabidopsis*, overexpression of HsfA4a leads to decreased production of cytosolic H_2O_2 scavenging ascorbate peroxidase (APX) and it was hypothesized that *Hsfs* may act as H_2O_2 sensors in the plants [9]. HSFA1D, HSFA2, and HSFA3 act as key factors in regulating *APX2* expression during diverse stress conditions [9]. Overexpression of *AtHsfA1b*-gusA in transgenic tomato plants led to the constitutive expression of Hsps, elevated levels of APX activity, with enhanced heat and chilling tolerance. *Hsfs* are also induced by other abiotic stresses like salinity, temperature, cold, and metal [10]. Overexpression of *OsHsfA2e* and *AtHsfA3* showed tolerance to salt stress [11, 12] but *HsfA3* conferred enhanced thermotolerance and salt hypersensitivity in germination in *Arabidopsis* [13]. While *HsfA1b* (*AtHsfA1b*) gene is involved in chilling tolerance in tomato [14, 15], *OsHsfA4a* is involved in cadmium tolerance in rice and wheat [16]. Besides imparting abiotic stress tolerance, several heat shock factors are also involved with disease resistance and developmental activities. *HsfB1* and *HsfB2b* are associated with pathogen resistance in *Arabidopsis* [17]. Further, HsfA9 was reported to be essential for embryogenesis and seed maturation in sunflower and *Arabidopsis* [18, 19]. *Hsfs* bind to the conserved *cis*-acting (5'-nGAAn-3') heat shock elements (HSE) of the promoters. At least 3 HSE are required for better interaction with Hsf. Based on homology and conservation of domains, plant *Hsfs* are classified into three classes. When compared with fungi and animals, plants have many *Hsf* genes [20, 21]. Genome-wide screening of many plants resulted in the identification of 16 to 35 *Hsfs* depending on the species [22-24].

The *Hsf* gene family has not been characterized in *Sorghum bicolor*. But, functional and evolutionary relationship between organisms can be studied only when multiple sequences of these families are available for alignment and phylogenetic analysis. Therefore, an attempt has been made in the present study to identify, classify and to characterize *Sorghum Hsf* genes and predict their evolutionary relationship with *Arabidopsis* and *Oryza*. Further, it is also not known where and when these *Hsf* genes are expressed in

*Address correspondence to this author at the Department of Genetics, Osmania University, Hyderabad 500 007, India; Tel: 91-40-2768-2335; Fax: 91-40-2709-5178; E-mail: pbkavi@yahoo.com

 # MN and PSR contributed equally.

1875-5488/15 $58.00+.00

Sorghum. Therefore, in the present investigation, tissue specific expression profiles of these *Hsfs* have been studied by carrying out quantitative real-time PCR under different abiotic stress treatments (by with-holding water for for 5-days for drought, by keeping at 4^0C for 4 h for cold, by exposing to 40^0C in a growth chamber for 4 h, by saturating potted plants with 150 mM NaCl and by collecting the tissue samples after 4 h treatment and by spraying 100 μM ABA and incubating the plants for 4 h for tissue collection). These results will be useful not only for studying the structure and function of *SbHsfs* but also for enhancing abiotic stress tolerance in this crop plant.

2. MATERIALS AND METHODS

2.1. Plant Materials and Stress Treatments

Sorghum bicolor variety cultivar Parbhani Moti, an improved desi variety was used for gene expression related experiments. *Sorghum* plants were grown in earthen pots containing 4.5 kg of black clay soil (Vertisol) under glass house conditions with $28/20^0$C day/night temperatures. Plants were maintained up to 28 days under well watered conditions and then used for different abiotic stress treatments. Drought stress was imposed by with holding the water supply for 5-days followed by leaf sample collection. For cold stress (low temperature) treatment, the plants were kept at 4^0C in a refrigerator for 4 h and was used for tissue collection. For heat stress (high temperature) treatment, plants were kept at 40^0C in a growth chamber and tissues were collected after 4 h of treatment. Salinity stress was induced by saturating the potted plants with 150 mM NaCl solution and leaf samples were collected after 24 h of treatment. For ABA stress, plant leaves were sprayed with 100 μM ABA solution and leaf sample was collected after 4 h. Different tissue samples like seedlings, leaf, flower, mature embryos, and roots were collected from different growth stages of Sorghum plants grown under normal growth conditions. For each sample, tissues were collected from three different plants grown under the same experimental condition ($28/20^0$C day/night temperature), to provide biological replicates. Tissues were snap frozen immediately in liquid nitrogen and stored at -80^0C until RNA extraction.

2.2. Identification and Localization of *Hsfs* in *Sorghum* Genome

Non-redundant nucleotide and amino acid sequences of *Arabidopsis* and rice *Hsfs* [25] were collected from TIGR and NCBI data bases. A total of 47 sequences were collected and each *Hsf* coding sequence (cds) was blasted against *Sorghum bicolor* genome in Gramene database by default settings. Gene sequences from the genome were retrieved using Edit plus (http://www.editplus.com/) and the sequences are subjected to Genscan (http://genes.mit.edu/GENSCAN.html) for coding sequences (cds) and amino acids. The redundant sequences which share the same chromosome location were eliminated and the remaining candidate genes were checked for *Hsf* DBD (DNA binding domain) in the Pfam database by employing SMART program [26], to identify coiled - coil structure and core of HR - A/B region. Sequences without the presence of DBD and coiled - coil regions have been eliminated.

2.3. Multiple Sequence Alignment

ClustalX2 [27] was used for multiple sequence alignment and domain prediction with default parameters. Bioedit (http://bioedit.software.informer.com/7.1/) and Genedoc (Free Software Foundation Inc.) were used for manually editing. For subcellular localization, WoLFPSORT [28], for finding out transmembrane helices TMHMM [29] and for gene characterization GSDS [30, Gene Structure Display Server http://gsds.cbi.pku.edu.cn] were used. NLS and NES were predicted with the help of NLStradamus [31], Nucleo [32], and Net NES [33]. Conserved motif analysis was carried out using MEME [34].

2.4. Promoter Analysis

In silico promoter analysis was carried out using 1 kb sequence upstream to all the *Sorghum Hsfs*. Promoter sequences were retrieved from the genome using Edit plus. PLACE [35] and Plant Care [36] softwares were used to identify the *cis*-acting elements in the promoter sequences. The distribution of *cis*-elements in promoter regions were further identified using MEME tool [34].

2.5. Phylogenetic Tree

Phylogenetic tree was constructed by MEGA 5.1 using the N-J method with 1000 boot strap replicates [37] on the basis of amino acid sequences of *Oryza sativa*, *Arabidopsis*, and *Sorghum*. Gene duplication events were also investigated using phylogenetic tree based on the 70% similarity and 80% coverage of aligned sequences [38, 39].

2.6. RNA Isolation and qRT-PCR

The list of primers used for the qRT-PCR analysis is shown in the supplementary (Table **1**). Total RNA was extracted from control and treated tissues using MACHEREY-NAGEL kit according to the manufacturer's instructions. A total of 2.5 μg RNA was reverse-transcribed into cDNA using SuperScript III First-Strand Synthesis Kit (Invitrogen) for qRT-PCR analysis. The cDNA was diluted into 1:12 with nuclease free water as template for qRT-PCR. The Bioline Master Mix (2X) was used to detect gene expression profile according to the manufacturer's recommendations on the RealPlex (Eppendrof). qRT-PCR was carried out in 96-well optical PCR plates, and the reaction was performed in a total volume of 10 μL containing 0.4 μM of each primer (1.5 μL), cDNA (1.0 μL) and Bioline Master Mix (2X) and nuclease free water was added upto 2.7 μL. qRT-PCR primers were designed using Primer3 software with GC content of 40-60%, Tm $>50^0$C, primer length 20-25 nucleotides, with expected product size of 90-180 bp (Table **1**). The thermal cycles performed were as follows: 95^0C for 10 min, followed by 40 cycles at 95^0C for 15 s and 62^0C for 1 min. Amplicon dissociation curves were recorded after cycle 40 by heating from 58 to 95^0C with fluorescence measured within 20 min. Three technical replicates were used for each gene. Expression levels of the *SbACP2*, *EIF4A*, and *S/T-PP* genes were used as internal controls. The experiments were independently repeated three times, and the data from these experiments were averaged. Relative gene expression calculations were carried using Rest software [40].

Table 1. List of *Sorghum* Hsf proteins. The identified Hsf proteins are listed according to their chromosome location. Hsf proteins are designated according to their locus id, protein sequence (AA) length, annotations, chromosomal locations their molecular weight (Mw), isoelectric point (pi), protein localization, and introns.

S. No.	Gene Name	Locus	AA Length	Annotation	Chromosome Location	Mw (Da)	pI	Localization	Intron
1	*SbHsf01*	Sb03g06630	467	RHsf 7	1	51515.94	5.48	Cytoplasm	2
2	*SbHsf02*	Sb03g12370	371	RHsf 8/Hsf 3	1	42186.32	4.96	Cytoplasm	2
3	*SbHsf03*	Sb03g53340	371	RHsf 4	1	40678.33	4.94	Nucleus	1
4	*SbHsf04*	Sb03g63750	477	-	1	52598.68	4.92	Nucleus	1
5	*SbHsf05*	Sb10g28340	328	RHsf 6	1	37535.13	5.00	Nucleus	1
6	*SbHsf06*	Sb3g02990	383	Putative Hsf sp 17	1	43217.74	5.58	Nucleus	2
7	*SbHsf07*	Sb3g63350	313	-	1	35158.51	7.23	Nucleus	1
8	*SbHsf08*	Sb01g042370	415	RHsf 8/ Hsf 3	1	46456.79	4.91	Nucleus	0
9	*SbHsf09*	Sb03g25120	302	RHsf 12 / Hsf 5	2	33727.63	6.78	Nucleus	1
10	*SbHsf10*	Sb08g36700	334	-	2	34544.70	9.71	Nucleus	0
11	*SbHsf11*	Sb09g28200	482	-	2	51240.71	10.10	Chloroplast	2
12	*SbHsf12*	Sb01g35790	561	-	2	59489.99	7.61	Chloroplast	1
13	*SbHsf13*	Sb02g004370	372	RHsf 5	2	41766.57	4.70	Nucleus	0
14	*SbHsf14*	Sb01g39020	456	Putative Hsf 8	3	49714.46	6.73	Chloroplast	4
15	*SbHsf15*	Sb01g53220	421	RHsf 11/Hsf 8	3	46415.83	9.60	Chloroplast	1
16	*SbHsf16*	Sb01g54550	434	RHsf 9	3	48351.37	5.13	Nucleus	1
17	*SbHsf17*	Sb03g028470	365	RHsf 13/Put. Hsf 1	3	39232.33	6.05	Lysosome	0
18	*SbHsf18*	Sb02g13800	347	-	4	37301.03	9.63	Chloroplast	2
19	*SbHsf19*	Sb02g29340	143	-	4	15257.53	8.07	Chloroplast	2
20	*SbHsf20*	Sb02g32590	176	-	4	19217.10	4.78	Chloroplast	1
21	*SbHsf21*	Sb4g13980	404	Putative Hsf sp 17	4	44957.02	5.34	Nucleus	1
22	*SbHsf22*	Sb04g48030	439	RHsf 1	6	46314.56	5.52	Chloroplast	1
23	*SbHsf23*	Sb09g026440	476	RHsf 10/ Hsf sp 17	9	52621.30	5.05	Nucleus	2
24	*SbHsf24*	Sb06g35960	279	-	10	29070.51	6.98	Cytoplasm	1
25	*SbHsf25*	Sb06g36930	439	-	10	47365.84	4.85	Cytoplasm	0

3. RESULTS

3.1. Identification and Localization of *Hsfs*

Screening of *Sorghum* genome resulted in the identification of 25 *SbHsfs* and are named according to their chromosomal locations (Table **1**). *Hsfs* are distributed on chromosomes 1, 2, 3, 4, 6, 9, and 10 and the number of *Hsfs* varied from chromosome to chromosome. Eight *Hsfs* were identified on chromosome 1, five on chromosome 2, four on 3 and 4, two on chromosome 10 and one on chromosomes 6 and 9 (Fig. **1**). WoLFPSORT was employed to identify subcellular localization of *Hsfs* and 12 of them are located in nucleus, 4 in cytoplasm, 8 in chloroplast and 1 in lysosome (Table **1**). Transmembrane helices were not observed in the *Hsfs* identified.

3.2. Sequence Analysis of *SbHsfs*

The length of the Hsf proteins varied from 143 to 561 amino acids, the molecular weights between 15.25 to 59.48 KDa and the pI from 4.7 to 10.10. Most of the *SbHsf* contain only 1 intron, 4 introns were noticed in *SbHsf14*, but no introns in *SbHsf8*, 10, 13, 17, and 25 (Table **1**). The multiple sequence alignment shows highly conserved DBD domains in *Sorghum bicolor Hsfs* (Fig. **2**). The N terminal DBD of *Hsfs* contains 3α and 4β folds, which is the specific location of HSE. The DBD is approximately 100 amino acids in length, but *SbHsf2*, 9 and 18 contain only 30 residues. HR-A/B domains in *Hsfs* are characterized by coiled - coil structures, which is the key feature containing Leu-Zipper protein interaction domains (Fig. **S1**). SMART program was used to

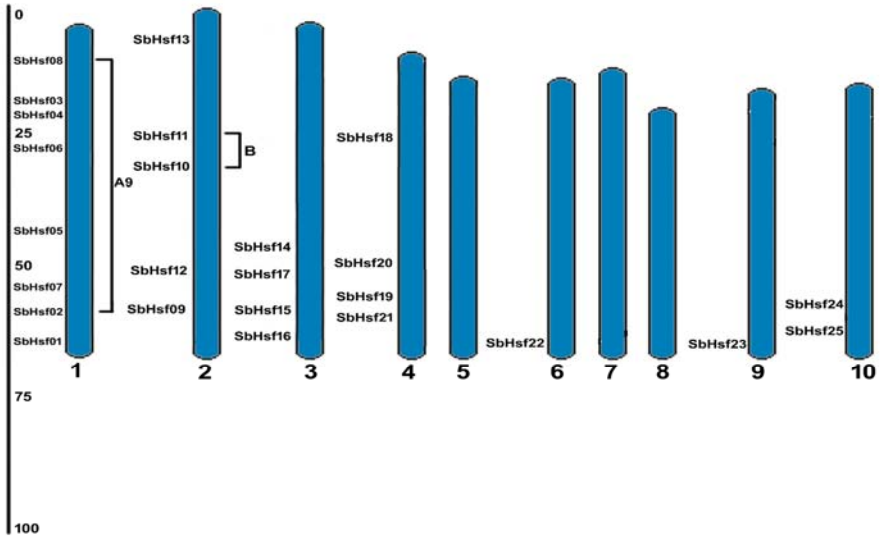


Fig. (1). Locations and duplications of *Sorghum* Hsf paralogs are shown on chromosomes 1-10. The scale represents megabases. The chromosome numbers are indicated at the top of each bar.

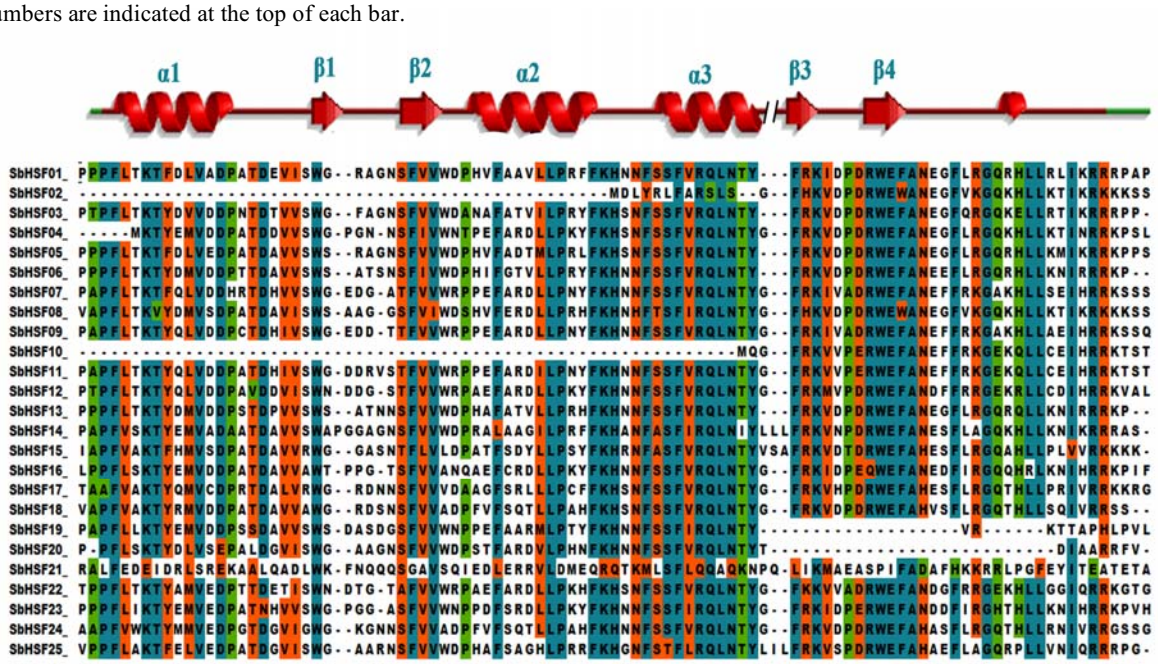


Fig. (2). Multiple sequence alignment of the DBD of the 25 members of *Sorghum Hsf* family is shown. The definition of *Hsf* number corresponds to order of alignment. The results clearly show the highly conserved DBD domains among all *Hsf* genes. The secondary structure elements of DBD (α1-β1-β2-α2-α3-β3-β4) are shown.

predict the DBD characteristic features of HR-A/B regions of *Hsfs* (Table **2**). NLS and NES are important for intracellular distribution of *Hsfs* between the nucleus and cytoplasm and was predicted by using cNLS, NLstradmus, and NET NES 1.1 tools. Most of the *Hsfs* contain two motifs of basic amino acids K/R. Previous comparisons from *Arabidopsis*, *Oryza*, and *Zea mays* show a wide range of NLS monopartite and bipartites found near C terminal of HR A/B regions of *Hsfs*. Only *SbHsfs* 2, 9, and 16 contain bipartite NLS (Table **2**). MEME tool was employed to explore motif distribution both in gene and promoter sequences. It supports the phylogenetic analysis and helps to determine conserved motifs which are species specific, class specific and group specific (Fig. **3**). The *Sorghum Hsfs* contain 30 highly conserved motifs with 5 to 43 residues in length (Fig. **S2**) and the number of motifs vary from 4 to 12. The SbHsf 18, 19, 20, and 24 contain 4 conserved motifs. Out of these, 2 and 3 are DBD, 15, 16, and 19 are coiled coil structure, 21, 24, and 25 are NLS, 12, 13, and 14 are AHA and 23 is NES motif. MEME finds the NLS motifs in SbHsf 11, 12, 17, 18, 22, 23, and 24 which could not be detected by NLS software.

3.3. Promoter Analysis

Table **3** shows the conserved *cis*-acting element motifs present in promoter regions. Motifs 12, 14, and 16 have ABA responsive elements; 16 and 24 have TATA box 2; 16, 26, and 29 have TATA box 3; 21 has LTRE which are low temperature and cold responsive elements, 26 have Myb and

Table 2. Functional domains and motifs of *Sorghum bicolor* Hsfs.

Gene	Group	DBD	NLS	NES
SbHsf01	A2	119-212	327 (ASRKRRRPIG)	384 (LENLALNI)
SbHsf02	A9	3-43	155 (DGNRKRRFQAL)	94 (LLMQQLLV)
SbHsf03	A2	55-148	143 (RTIKRRRPPS)	333 (VELLSLGL)
SbHsf04	A1	1-88	199 (ANKKRRLPKQ)	-
SbHsf05	A2	8-101	207 (ISKKRRRPID)	-
SbHsf06	A2	36-129	232 (ISKKRRRRIV)	165 (LLMTEVVKL)
SbHsf07	B4	44-137	280 (DGKKRRAQQV)	-
SbHsf08	A9	1-83	252 (DGNRKRRFQAL)	191 (LLMQQLVDL)
SbHsf09	B4	3-33	280 (GKKKKRAHQD)	-
SbHsf10	B4	87-181	-	313 (LALEGADLSLTV)
SbHsf11	B4	200-293	-	461 (LALEGADLSLTV)
SbHsf12	B2	107-232	-	90 (FFLVLLLLL)
SbHsf13	A2	155-288	246 (ISKKRRRRID)	-
SbHsf14	A10	10-103	227 (KNIKRRRASK)	-
SbHsf15	C	107-200	382 (PAPGKRRRIG)	366 (VVLRAML)
SbHsf16	A4	23-115	199 (HGKKRRLPIP)	166 (LEDKLIFL)
SbHsf17	C	63-135	-	11 (LHTELALGLL)
SbHsf18	C	2-36	-	-
SbHsf19	A4	140-233	-	113 (LVYDALLVL)
SbHsf20	A3	9-102	-	23 (MLLEPKLEDEDV)
SbHsf21	A5	88-203	137 (FHKKRRLPG)	97 (VSQIEDLERRV)
SbHsf22	B3	47-140	-	422 (LDVLTLSV)
SbHsf23	A4	30-123	-	279 (MELALVSL
SbHsf24	C	49-142	-	179 (MLAFLLKVV)
SbHsf25	A10	24-117	307 (AGRKRRLLD	336 (VLAFEELAL)

Number in brackets indicates the position of the putative localization signal (NLS), nuclear export signal (NES) and DNA Binding Domains (DBD).

28 have Myc waterstress responsive elements (Fig. **4** and Fig. **S3**). The promoter elements like ABRE, ANAERO, ARF, DPBF, DRE, LTRE, MYB, and MYC responsive to ABA, drought, low temperature, and cold are commonly present in all the 25 *Hsfs* along with high temperature responsive elements. The *Hsfs* also contained pathogenesis and salt stress responsive *cis*-elements GT1GMSCAM4 and WBOXNTERF3 for wound response and WBOXANTNPR1 for salicylic acid signal response. The CGCGBOX *cis*-elements present in *Hsfs* are involved in multiple signal transduction and KST1 is involved in guard cell-specific gene expression and pollen specific elements associated with pollen and anther development in different stress conditions. *SbHsfs* 9 and 13 contain a maximum of 15 ABRE *cis*-elements and *SbHsfs*2, 4, and 21 contain a minimum of one ABRE *cis*-elements (Table **4**).

3.4. Phylogenetic Analysis

Phylogenetic tree was constructed by using MEGA 5.1, and neighbour joining method was employed for multiple sequence alignment of 22 *Arabidopsis*, 25 rice, and 25 *Sorghum Hsfs*. Based on the bootstrap values and phylogenetic relationship, they were classified into 3 major *Hsf* classes A, B, and C. Phylogenetic analysis of rice, *Arabidopsis*, and *Sorghum* depicts a close relationship of rice and *Sorghum*, both being members of poaceae. While 10 subgroups are present in class A, 4 are seen in B and the least in C. The contrasting feature of the phylogenetic analysis is in the number of *Hsfs* that varied among the subclass A in rice, *Sorghum* and *Arabidopsis*. For example A2 (five) subgroup is present in the species rice and *Sorghum*, it is absent in *Arabidopsis*. While A6, A7, and A8 subgroups could not be found in monocot species like rice and *Sorghum*, 2, 2, 3

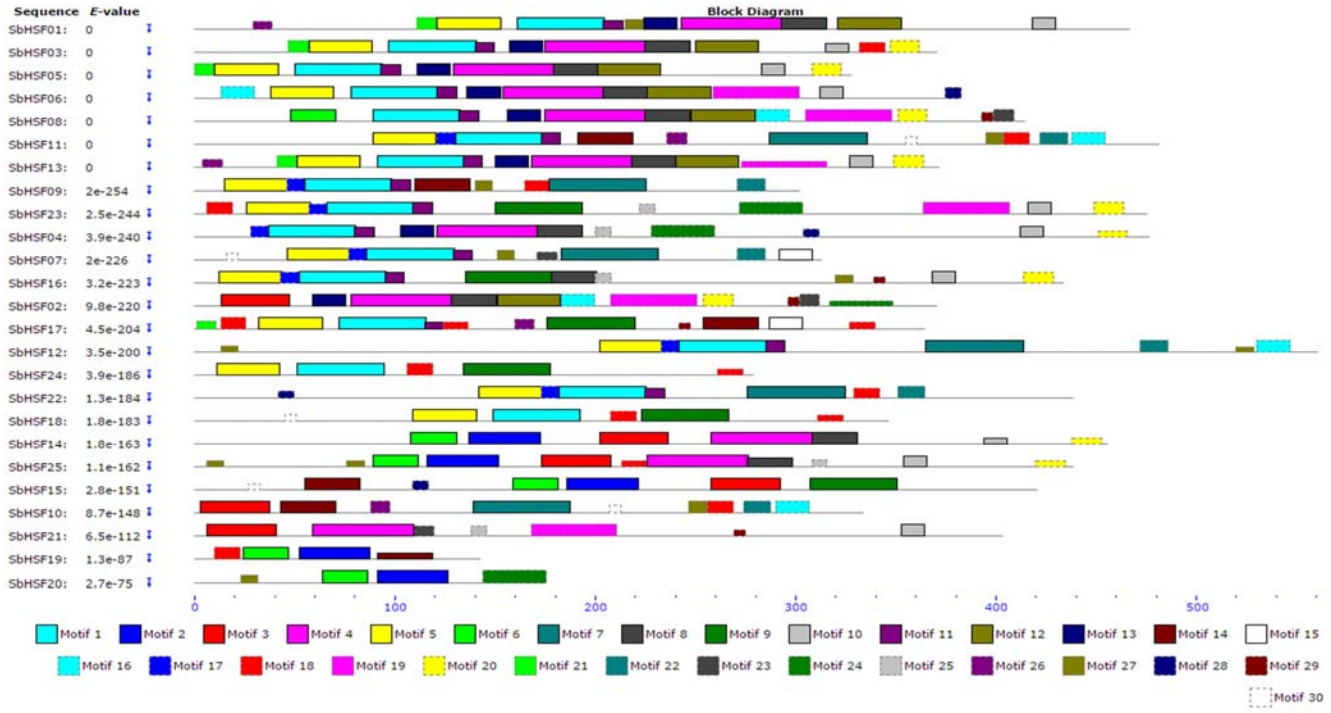


Fig. (3). Distribution of conserved motifs in the Hsf family members is shown. All motifs were identified by MEME using the 25 complete amino acid sequences of *Sorghum Hsf* genes. Names of all the members among the defined gene clusters and combined p values are shown on the left side of the figure, motif sizes are indicated at the bottom. Different motifs represented by different colours are numbered 1-30.

Table 3. Conserved *cis*-acting elements of *Sorghum bicolor* Hsfs. MEME motifs, *cis* elements, signal sequence and their functional roles.

S. No.	Motif	*Cis* Elements	Seq (signal)	Functions
1	12, 14, 16	ABA	ACGTG	Etiolation-induced expression (erd1)
2	17	Anaero 2	AGCAGC	Fermentative pathway
3	2,3,23, 24	ARR	NGATT	Response regulator
4	3, 25	CAAT	CAAT	Promoter of legumin gene
5	27,29,30	CACTT	CACT	Promote phosphoenolpyruvate carboxylase
6	6,17	CGC Box	VCGCGB	Ca++/calmodulin binding
7	1,3,21,24,25,26,27,28,29,30	DOF	AAAG	DNA binding proteins and carbon metabolism
8	2,12,28	DPBF	ACACNNG	ABA and embryo-specification
9	2,26,29	GATA	GATA	Chlorophyll a/b binding
10	6,13	GCC CORE	GCCGCC	G box high level constitution expression
11	9,25,27	GT1	GRWAAW	SA inducible
12	20,23,24,26	GTGA	GTGA	Late pollen gene g10, pectate lyase
13	5	HEXA	CCGTCG	Histone H4
14	3	I BOX CORE	GATAAG	Light regulated
15	21	LTRE	CCGACA	Low temperature and Cold
16	26	MYB	CNGTTR	Water stress
17	28	MYCONSES	CANNTG	erd1 (etiolation responsive to dehydration)
18	5,6	PAL BOX	CCGTCC	Phenylalanine ammonia-lyase
19	8,16,23	POLASI GI	AATAAA	Poly adenylation

(Table 3) contd….

S. No.	Motif	*Cis* Elements	Seq (signal)	Functions
20	4,20	POLASI G2	AATTAAA	Poly adenylation rice amylase
21	16,23,29	POLASI G3	AATAAT	Poly adenylation
22	25,27,28,29	POLLEN	AGAAA	Endo beta mannose, anther and pollen Development
23	21	PRE CONSES	SCGAYNRNN	Plastid responsive and light
24	3,16,29	ROOT MOTIF	ATATT	Promotes rol D
25	2	RYREPEATLE	CATGCAT	GLYCININE, ABA res., embryogenesis
26	11,13,14	SORLIP1	GCCAC	Phytochrome A, root development
27	5,12,24	SORLIP2AT	GGGCC	Light inducible
28	15	SORLREPSAT	TGTATATAT	Phytochrome A
29	2	SPH CORE	TCCATGCAT	Viviporous 1, seed specific development
30	1,12,14,20,23	SURE	GAGAC	Sulfur transporter
31	27	TAAAGSTKSTK1	TAAAG	Controlling guard cells and K+ influx
32	16,29	TATA2	TATAAAT	Accurate initiation for phaseolin
33	16,26,29	TATA 3	TATTAAT	Accurate initiation
34	8,16	TATA 4	TATATAA	Accurate inhibition G
35	8,16,23,26,29	TATA 5	TTATTT	lutamine synthase (non photo syn) ?
36	30	WBOXATNPR1	TTGAC	Response to SA signal
37	2, 11, 30	WBOXNTERF3	TGACY	Response to wound signal
38	2, 12, 30	WRKY	TGAC	Repressor for gibberellin signaling

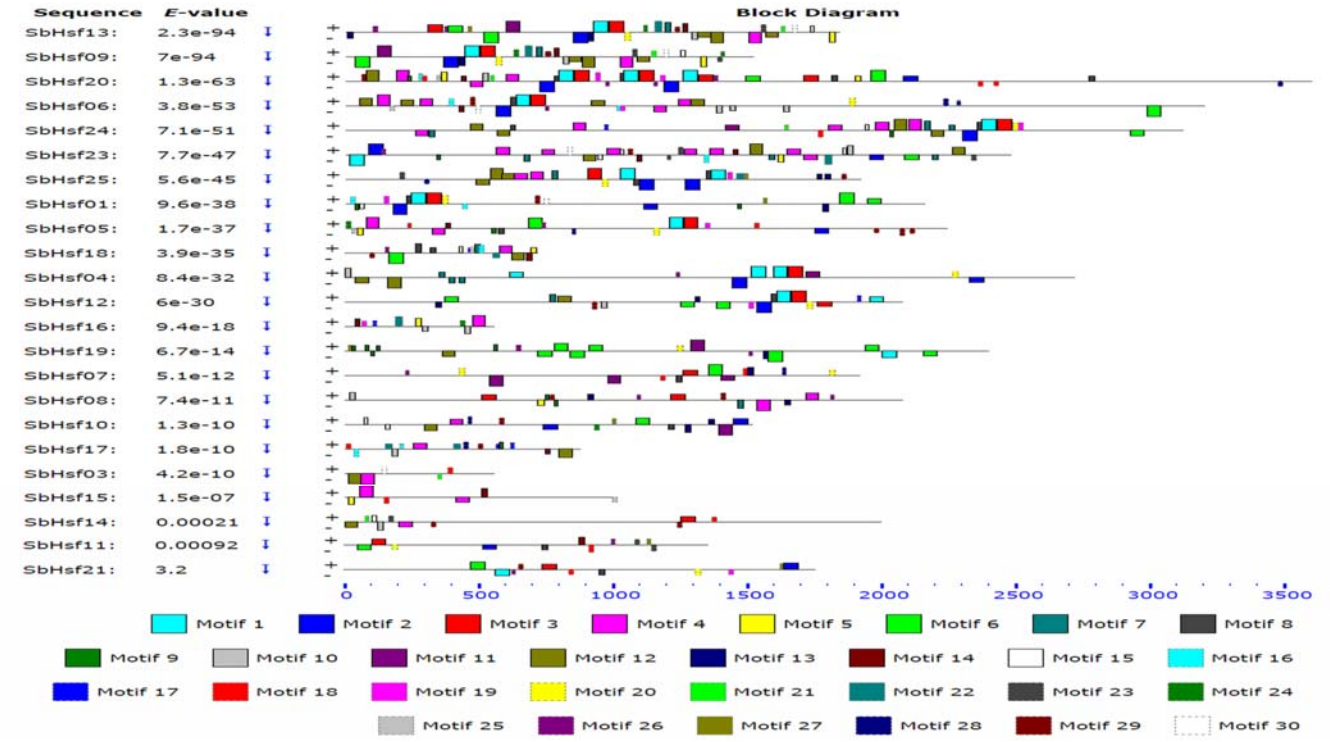


Fig. (4). Distribution of conserved motifs in promoter regions of *Hsf* family is shown. All motifs were identified by MEME using the promoter sequences of *Sorghum* Hsf family analyzed by PLACE and PlantCARE software. Different motifs are indicated by different colours and numbered 1-30, which represent the conserved *cis*-acting elements. For details of motifs refer to table **3**.

Table 4. Conserved *cis*-acting elements present in promoter of *Sorghum* Hsfs.

Gene	Cis Acting Elements														
	ABRECTAL (MACGYGB)	AN-AERO (AAACAAA)	ARF (TGTCTC)	CGCGBOXAT (VCGCGB)	CURE (GTAC)	DPBF (ACACNNG)	DRE (RCCGAC)	GT1GMSCAM4 (GAAAAA)	LTRE (CCGAAA)	MYB (WAACCA/YAACKG/CNGTTR)	MYC (CATGTG/CANNTG)	POL-LEN1LELAT52 (AGAAA)	TAAAGSTKST1 (TAAAG)	WBOXN-T ERF3 (TGACY)	WBOX-AT NPR1 (TTGAC)
SbHsf01	5	1	2	6	10	8	1	4	2	15	38	9	3	7	6
SbHsf02	1	2	1	0	8	0	0	0	2	16	8	4	3	5	3
SbHsf03	3	0	0	2	6	1	1	1	2	4	4	2	0	0	0
SbHsf04	1	2	1	10	20	6	4	3	4	23	24	8	7	8	3
SbHsf05	8	0	3	4	16	5	3	3	4	27	40	6	3	7	6
SbHsf06	9	9	0	26	10	4	2	4	7	30	38	9	10	5	4
SbHsf07	4	5	1	0	22	3	1	4	0	18	22	4	7	13	6
SbHsf08	4	7	0	6	6	3	2	3	5	16	24	4	5	3	6
SbHsf09	15	2	0	24	6	7	5	4	7	18	16	9	3	4	3
SbHsf10	3	3	2	0	12	4	0	1	0	7	14	7	6	3	0
SbHsf11	9	2	0	0	8	2	1	3	0	9	18	11	3	3	2
SbHsf12	7	4	2	4	14	1	0	5	0	23	10	5	1	3	2
SbHsf13	15	6	0	24	6	5	4	5	7	21	18	10	6	7	2
SbHsf14	3	3	2	4	12	3	0	0	1	29	22	11	4	7	1
SbHsf15	9	2	1	12	4	2	0	3	0	12	8	3	5	2	1
SbHsf16	2	1	0	10	0	0	0	0	0	10	0	2	0	2	2
SbHsf17	6	3	0	12	0	4	0	1	1	3	12	1	1	2	1
SbHsf18	12	3	0	14	4	1	3	1	3	10	4	0	0	3	1
SbHsf19	6	0	0	0	14	2	2	7	3	30	8	5	5	12	16
SbHsf20	2	6	1	10	8	4	4	10	6	25	36	16	6	9	7
SbHsf21	1	2	0	0	4	3	2	7	2	22	12	12	2	4	8
SbHsf22	4	0	0	4	8	3	0	5	4	14	10	12	2	6	2
SbHsf23	9	6	1	44	6	4	12	6	12	23	16	7	1	8	3
SbHsf24	8	4	3	42	12	4	3	5	7	35	30	13	3	9	6
SbHsf25	6	6	1	8	6	2	1	1	2	15	14	7	5	2	2

ABRECTAL: Response to ABA, ANAERO: Anaerobic conditions, ARF: ABA and auxin responsive, CGCGBOX: Multiple signal transduction, CURE: Cu and oxygen responsive, , DPBF: ABA, DRE: Dehydration responsive elements, GT1GMSAM4: Salt and pathogenesis related, LTRE: Low temperature and cold responsive, MYB: responsive to drought and ABA, MYC: Response to drought, cold and ABA, POLLEN: pollen and anther development, TKST1: Guard cell-specific gene expression, WBOXNTERF3: Wound signal and WBOXATNPR1: Salicylic acid responsive.

have been detected respectively in *Arabidopsis* (Figs. **5** and **6**). Among the four subclasses of B, B1 are absent in Sorghum, but one is detected in rice. Further, in class C, the genome of *Arabidosis* revealed only one *Hsf*, but four each could be identified in rice and *Sorghum*.

3.5. Gene Duplication Events

Two paralogs participated out of 25 *Sorghum Hsfs* in regional duplications within the chromosomes. These paralogs evolved from their common ancestral genes through gene duplication events. While no segmental duplication events were observed in *Sorghum* 8 and 7 were recorded in maize and rice respectively out of nine paralogs. Maize and rice Hsf family is expanding with large number of segmental duplications (Fig. **5**).

3.6. Transcript Profiling of *SbHsfs* in Different Tissues

SbHsf genes displayed differential expression in different tissues (Fig. **7a**). Out of four major tissues (mature embryo, seedling, root, and panicle), panicle showed higher levels of *Hsf* abundance than the mature embryos. No *Hsfs* were up- or down-regulated in the case of mature embryos (Fig. **7a**). While in seedling *Hsfs*4, 9 are highly expressed, 13 and 22 are moderately expressed. In the case of roots, only 4 and 13 are well expressed. Moderate expression levels were also recorded in *Hsfs*5, 6, 21, 23, and 25 in roots. On the other-hand, *Hsfs*4 and 22 are highly expressed, *Hsf1*, 3, 5, 9, 10, 13, 16, 19, 23, and 25 recorded moderate transcript levels in panicle tissues (Fig. **7a**).

3.7. Abiotic Stress Induced Expression of *Hsfs*

All *Hsfs* displayed a differential expression in response to various abiotic stresses (Fig. **7b**). Among the five treatments (ABA, cold, heat, salt, and drought), drought stress induced higher transcript abundance than the other treatments. ABA, did not enhance the levels of *Hsfs* except in Hsf23, where only minor increase was noticed. Expression was significantly upregulated in HSf1, 15, 19, and 25 under cold stress (Fig. **7b**). Moderate levels of expression was observed in *Hsfs* 2, 3, 4, 5, 6, 8, 10, 13, 16, 21, 23, and 24. During heat stress, Hsf1 was highly expressed, and moderate expression were displayed in *Hsfs* 6, 9, 13, 21 etc. During salt stress, *Hsfs*4, 6, 13, 16, 21, and 23 were up-regulated. In contrast, many *Hsfs* like 1, 5, 6, 10, 13, 18, 22, 23, and 25 were upregulated during drought stress (Fig. **7b**).

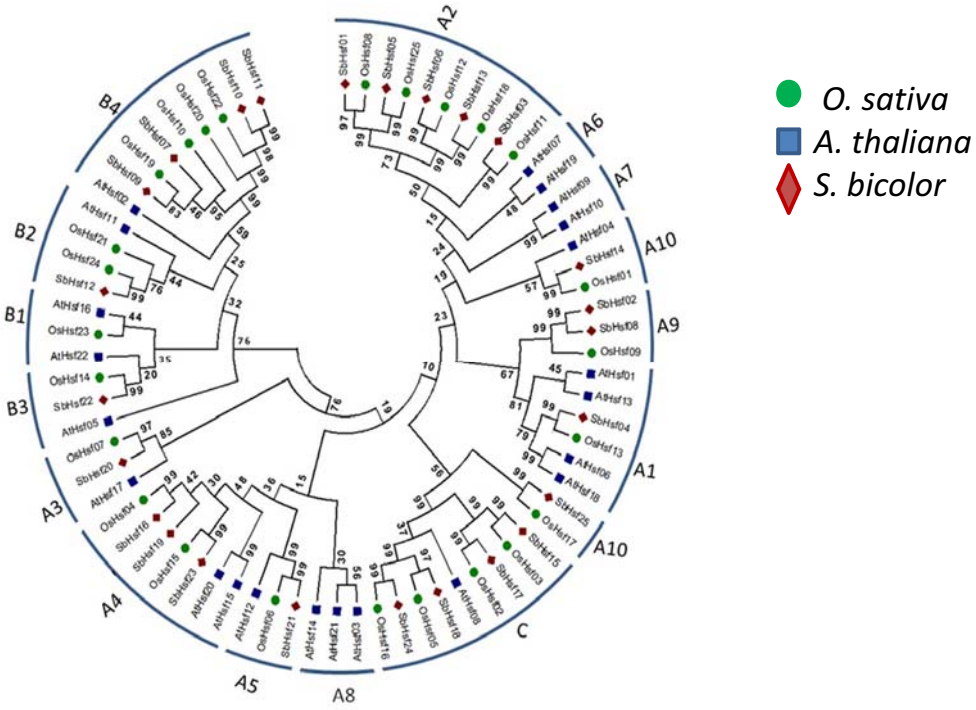


Fig. (5). The *Hsf* phylogenetic tree is constructed using neighbour joining method. The phylogenetic tree constructed with MEGA 5.1, has been generated on the basis of amino acid sequences of *Oryza sativa*, *Arabidopsis thaliana* and *Sorghum bicolor*. The Hsf proteins are classified into 3 major groups A, B and C, in which group A is subdivided into 10 groups, A1 to A10, and B is subdivided into 4 groups, B1 to B4. The abbreviations: Os = *Oryza sativa*, At = *Arabidopsis thaliana*, Sb = *Sorghum bicolor*.

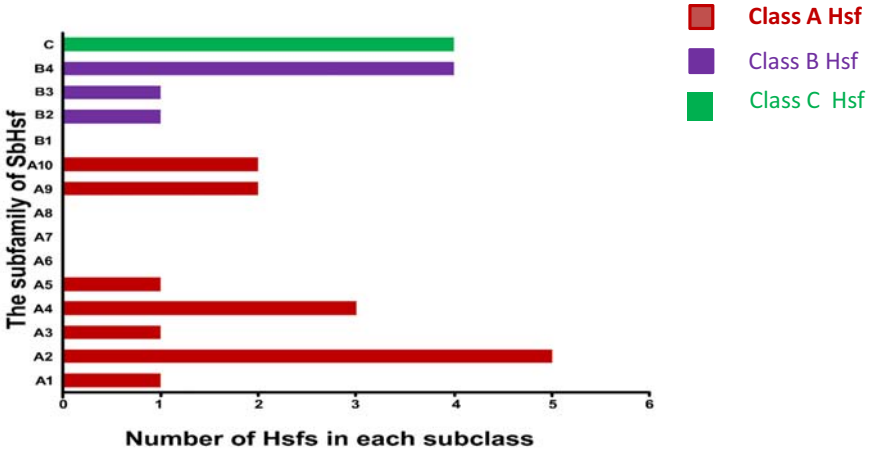


Fig. (6). The number of *SbHsf* subgroups in three classes are shown.

4. DISCUSSION

4.1. Sequence Analysis

Hsfs have been identified in several plants [21-24, 41-43] but not in *S. bicolor* which is often exposed to salt, drought, and temperature stresses. Genetic variability for drought tolerance exists in *Sorghum* [44] but the effects of high temperature and water stresses on reproductive biology and seed-set needs further investigations and identification of candidate genes for breeding programs aimed at crop improvement. While eight *Hsfs* are distributed on chromosome 1, no *Hsfs* could be detected on 5, 7 and 8. In the case of *Arabidopsis*, maize and rice, *Hsfs* are spread all over the chromosomes but chromosomes 11 and 12 lack them [47, 25]. Like rice and maize, *S. bicolor* has also the same number of *Hsfs*, which reflects that *Hsfs* are conserved during the process of evolution [47, 25]. The theoretical pI of *Hsfs* range between 4.7 to 10.10, which indicates that they contain both acidic and basic proteins. *Hsfs* 2, 9 and 18 contain 30 residue-length DBD, which may occur due to deletions in DBD regions of α and 4 β-helices and due to genetic diversity in *SbHsfs*. Class A requires AHA motifs for their functioning, but *SbHsf14* and 20 lack such motifs. *SbHsfs*18 and 24 belong to class C but do not contain AHA motifs. They

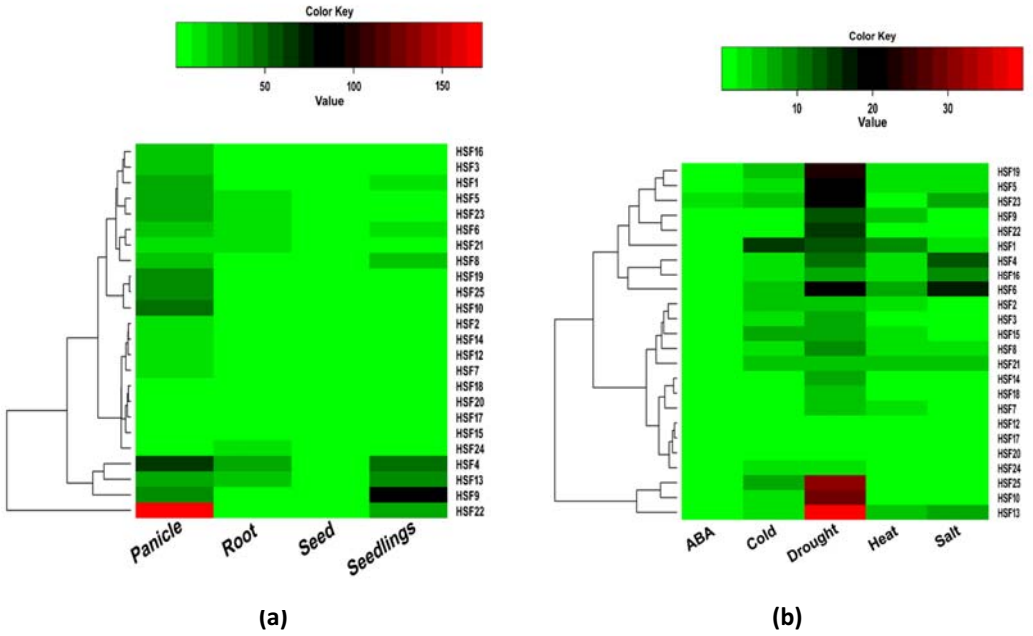


Fig. (7). (**a**). Relative expression of *SbHsfs* at the transcript level is shown in different tissues. Relative expression of *SbHsf* transcripts are shown during different abiotic stress conditions in comparison to its control as revealed by quantitative RT-PCR analysis. Values represent the expression values obtained after normalizing against control value. All samples were analyzed in triplicates, in three independent experiments. Names on the horizontal axis indicate the identified *SbHsfs*, and the vertical axis represents the various tissues, i.e., mature embryos, panicle, seedlings and root. Each color represents the relative expression levels. (**b**). Relative quantification of *SbHsfs* under diverse abiotic stress treatments is shown. Relative expression of *SbHsf* transcripts is shown during different abiotic stress conditions in comparison to its control as revealed by quantitative RT-PCR analysis. Values represent the expression values obtained after normalizing against control value. All samples were analyzed in triplicates, in three independent experiments. Names on the horizontal axis indicate the identified *SbHsfs*, and the vertical axis represent various treatments such as ABA, cold, heat, salt and drought. Each color represents the relative expression levels.

may bind to other classes of A and C *Hsf* types and form hetero oligomers and start their function [25]. *In silico* survey of the putative *cis*-elements of the *Sorghum Hsfs* showed the presence of HSE, ABA responsive elements, ARR, Anaero, CACTT, low temperature responsive elements (LTRE), pollen specific *cis*-regulatory (AGAAA) and desiccation responsive elements. This indicates that *Hsfs* are not only expressed during high temperature but also during other environmental stresses. The presence of HSE *cis*-elements in the promoter regions is correlated with the expression of *Hsf* genes under high temperature stress in *Arabidopsis*, rice, maize, and wheat [47-50]. Bate and Twell [51] observed that transcriptional activation of late pollen gene (*lat52*) is controlled by a pollen-specific *cis*-regulatory elements AGAAA and TCCACCATA to attain high gene expression levels associated with pollen maturation. Promoter analysis of the endo-β-mannanase gene demonstrated pollen-specific *cis*-acting elements POLLEN1LELAT52 (AGAAA) which are associated with anther and pollen development [52]. In the present study also, such AGAAA elements were detected in the promoter regions of *Hsfs* indicating that these *Hsfs* may be involved in anther and pollen development in *Sorghum*. Promoter analysis of the *KST1* gene, (an inward rectifying potassium channel) revealed a sequence motif TAAAG and the involvement of these elements suggests a role for Dof transcription factors in guard cell-specific gene expression and stomatal conductivity [53]. Such TAAAG elements have been observed in our promoter analysis, raising scope for speculation of Hsf promoters in K^+ influx and guard cell movement. *Hsfs* are not only expressed during abiotic stress, but also biotic stress since their promoter regions contain potential *cis*-elements such as WBOXNTERF3 and WBOXATNPR1 which are responsive to biotic stresses like wound, pathogen, and salicylic acid [54, 55]. While *ERF3* gene is activated by wounding in tobacco [55], the disease resistance regulatory protein NPR1 has been found to be required to activate AtWRKY18 [56]. Detecting ABA and salicylic acid response elements in the promoter regions of *Hsfs* provide valuable clues on the underlying regulatory mechanisms of *Hsfs* that may further lead to development of plants with biotic and abiotic resistance.

4.2. Phylogenetic Analysis

The phylogenetic tree revealed that proportion of the three *Hsf* classes differed considerably among the three species. While class A contained the large number of *Hsfs*, class B accounted for small number, and class C the minimum. *Hsfs* with three distinct classes A, B, and C appeared to be more in number in majority of angiosperms except in *Medicago truncatula* (class C absent), when compared to lower plants that contain classes like A and B as in the case of *Picea abies*, *Selaginella moellendorffii*, *Physcomitrella patens*, *Chlorella* sp. NC64 etc. [21]. Differences in different subgroups of A4, A9, B1 and B2 were observed between rice and a relatively temperature and drought tolerant *S. bicolor*, which is a C_4 plant. Subgroup B1 is absent in *Sorghum* while it is present in rice. Perhaps these differences in different

subclasses of A and B play critical roles during various types of abiotic stresses and developmental activities in these two contrasting plants. However, such an assumption needs to be validated experimentally. In plants, gene duplication events play an important role in evolution [57]. In polyploidization, gene duplicates accumulate [58] and these processes involve several transcription factors [59]. Recently, Song *et al.* [24] observed duplication events in the expansion of *Hsf* genes in Chinese cabbage. These observations clearly indicate that Hsf transcription factor family contributed to polyploidy [24, 59]. In the present study, segmental gene duplication events could not be observed in *Sorghum* unlike that of maize and rice [47, 25].

4.3. Transcript Analysis in Different Tissues and During Different Abiotic Stress Conditions

The expression patterns of different *Hsf* genes may differ depending on the plant species [21]. Yamaguchi-Shinozaki and Shinozaki [60] have shown that transcription factors interact with each other. It appears that each of the *Hsf* genes respond differentially to different abiotic stresses and developmental stages. Several transcriptome studies show that *Hsf* transcription plays significant roles in response to abiotic stress [23, 24, 48]. This type of unique expression patterns of *Hsf* transcripts were observed in response to both abiotic stresses and developmental stages also [9, 21, 43]. The varied patterns of *Hsf* expressions in different tissues may relate to the differences in *cis*-acting elements present in different promoter sequences. In the present study, *Hsfs* that are expressed during one type of abiotic stress, did not up-regulate when exposed to the other type of stresses, the exception being *Hsf1* for cold and drought, and *Hsf6* for salt and drought. Cross-talk exists between abiotic stress signal and plant growth and the expression of different transcription factor gene families [24, 41, 45, 60] indicating that these *Hsfs* play critical roles in maintaining drought and temperature stress tolerance and also play a vital role during development [21, 41, 46].

Six out of 21 *Hsfs* in Arabidopsis and 8 and 9 out of 25 in *Oryza* and *Sorghum* were induced by heat stress respectively [61, 62]. In many plants, intron-mediated enhancement (IME) of gene expression was noticed as in the case of *Alcohol dehydrogenase 1*, and *Bronze 1* as reported by Callis *et al.* [63], *Shrunken 1* in maize [64] and *Phosphoribosylanthranilate transferase 1* in *Arabidopsis* [65]. Introns increased the transcription initiation and mRNA levels in these cases [66]. While in rice, intron mediated enhanced gene expression was observed, in *Sorghum*, exceptions were noticed in *SbHsf08*, 10, 13 and 25. These *Hsfs* in *S. bicolor* showed elevated expression levels without any intron. Intriguingly, *SbHsf14* contains 4 introns but displayed lower expressions during stress. This infers that IME gene expression may vary depending upon the *Hsf* present in a specific species. *OsHsfA2d*, which is duplicated with *OsHsfA2c*, has two introns in place of one in the original gene *A2c* and *OsHsfB2b*/*OsHsfB2c*. This *OsHsfB2b*/*OsHsfB2c* has 2 introns and exhibited more expression during heat stress and considerably higher expression in almost all the other abiotic stresses and during seed development [62]. In *S. bicolor*, regional duplicated gene pair *SbHsf02*/*SbHsf08* has no introns instead of 2 in the original gene *SbHsf02*, *Hsf08* expressed abundantly in all the tissues and during all stress treatments. On the other hand, *SbHsf10*/*SbHsf11* has 2 introns, but not expressed during all stresses.

Class A HSFs have been characterized in more detail than class B and C HSFs in plants. In *Arabidopsis*, expression of *HsfA2* was high among the class A *HSFs* under high temperature and light stresses [8]. In rice, the expression of all *OsHsfA2* genes increased by heat stress except for *A2b*, which is actually a duplicated gene with *A2e* [62]. In *Sorghum*, 5 members of *HsfA2* genes have been noticed in contrast to 6 in rice, and are also highly induced during drought, salt, heat, and cold stresses. *HsfB1* is absent in *Sorghum* and *Oryza* but present in *Arabidopsis*. Though *HsfB1* is heat inducible, its overexpression did not lead to thermotolerance in *Arabidopsis* [14, 61]. On the other hand, in tomato, *HsfB1* is a transcription co-activator functioning along with *HsfA1* and hypothesized as a heat shock induced factor essential for maintenance and restoration of house keeping gene transcription during stress [67]. *OsHsfB2a*, *B2b* and *B2c* were induced by heat stress but expressed in developing seeds. In *Sorghum*, *Hsf B2* was not induced under any stress but observed in panicles. Double knock-out mutants for *AtHsfB1* and *B2* displayed normal fertility and thermotolerance as compared with single knock-out mutants in *Arabidopsis* [17]. In *S. bicolor*, *B3* was highly expressed in panicle and early seedling stage during droght but not in rice. On the other hand, *B4* and class C *Hsfs* are moderately induced under all stress conditions. Thus, several differences exist among different classes of *Hsfs* between water loving rice and relatively drought tolerant *S. bicolor*.

In conclusion, 25 *SbHsfs* genes were identified in the genome of *S. bicolor*. Such a systematic analysis of *Hsfs* help us in finding out the functions of *Hsf* signaling pathways associated with different abiotic stress conditions and also growth and development. The diverse expression patterns suggest that these genes may perform different physiological functions depending on the type of tissue and its needs. Some *SbHsfs* were constitutively expressed, while others exhibited a distinct expression pattern in different tissues and under diverse abiotic stress treatments, implying that *SbHsfs* genes have functional diversity. This study provides the first step towards the future studies of Hsf protein functions and enhancing drought or thermotolerance stress and also the association of *SbHsf* genes under diverse environmental conditions.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

MN is thankful to the UGC, New Delhi, for providing fellowship. PSR is thankful to the Department of Science and Technology, Govt. of India, New Delhi, for research funding through the INSPIRE Faculty Award Grant No. IFA-11LSPA-06. PBK is thankful to the CSIR, New Delhi, for providing Emeritus Fellowship. DMR is thankful to the UGC, New Delhi, for financial support through UGC-MRP.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's web site along with the published article.

REFERENCES

[1] Wang, W.; Vinocur Altman, A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta*, **2003**, *218*(1), 1-14.

[2] Feder, M.E.; Hofmann, G.E. Heat-shock proteins, molecular chaperones, and stress response: evolutionary and ecological physiology. *Annu. Rev. Physiol*., **1999**, *61*, 243-282.

[3] Scharf, K.D.; Heider, H.; Hohfeld, I.; Lyck, R.; Schmidt, E.; Nover, L. The tomato Hsf system: HsfA2 needs interaction with HsfA1 for efficient nuclear import and may be localized in cytoplasmic heat stress granules. *Mol. Cell Biol*., **1998a**, *18*(4), 2240-2251.

[4] Bharti, K.; Schmidt, E.; Lyck, R.; Bublak, D.; Scharf, K.D. Isolation and characterization of HsfA3 a new heat stress transcription factor of *Lycopersicon peruvianum. Plant J.*, **2000**, *22*(4), 355-365.

[5] Mishra, S.K.; Tripp, J.; Winkelhaus, S.; Tschiersch, B.; Theres, K.; Nover, L.; Scharf, K.D. In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermo tolerance in tomato. *Genes Dev*., **2002**, *16*(12), 1555-1567.

[6] Charng, Y.Y.; Liu, H.C.; Liu, N.Y.; Chi, W.T.; Wang, C.N.; Chang, S.H.; Wang, T.T. A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermo tolerance in *Arabidopsis. Plant Physiol*., **2007**, *143*(1), 251-262.

[7] Ogawa, D.; Kazuo, Y.; Takumi, N. High-level over expression of the *Arabidopsis* HsfA2 gene confers not only increased themo tolerance but also salt/osmotic stress tolerance and enhanced callus growth. *J. Exp. Bot*., **2007**, *58*(12), 3373-3383.

[8] Nishizawa, A.; Yabuta, Y.; Yoshida, E.; Maruta, T.; Yoshimura, K.; Shigeoka, S. *Arabidopsis* heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. *Plant J.*, **2006**, *48*(4), 535-547.

[9] Jung, H.S.; Crisp, P.A.; Estavillo, G.M.; Cole, B.; Hong, F.; Mockler, T.C.; Pogson, B.J.; Chory, J. Subset of heat- shock transcription factors required for the early response of *Arabidopsis* to excess light. *Pr. Nat. Acad. Sci.*, **2013**, *110*(35), 14474-14479.

[10] Swindell, W.R.; Huebner, M.; Weber, A.P. Transcriptional profiling of *Arabidopsis* heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. *BMC Genomics*, **2007**, *8*, 125.

[11] Yokotani, N.; Ichikawa, T.; Kondou, Y.; Matsui, M.; Hirochika, H.; Iwabuchi, M.; Oda, K. Expression of rice heat stress transcription factor OsHsfA2e enhances tolerance to environmental stresses in transgenic Arabidopsis. *Planta*, **2008**, *227*(5), 957-967.

[12] Yoshida, T.; Sakuma, Y.; Todaka, D.; Maruyama, K.; Qin, F.; Mizoi, J. Functional analysis of an Arabidopsis heat-shock transcription factor HsfA3 in the transcriptional cascade downstream of the DREB2A stress- regulatory system. *Biochem. Biophys. Res. Commun*., **2008**, *368*(3), 515-521.

[13] Li, M.Y.; Wang, F.; Jiang, Q.; Li, R.; Ma, J.; Xiong, A.S. Genome-wide analysis of the distribution of AP2/ERF transcription factors reveals duplication and elucidates their potential function in Chinese cabbage (*Brassica rapa* ssp. pekinensis). *Plant Molecular Biology Reports*, **2013**, *31*(4), 1002-1011.

[14] Prandl, R.; Hinderhofer, K.; Eggers-Schumacher, G.; Schoffl, F. HSF3, a new heat shock factor from *Arabidopsis thaliana*, derepresses the heat shock response and confers thermo tolerance when over expressed in transgenic plants. *Mol. Gen. Genet*., **1998**, *258*(3), 269-278.

[15] Nover, L.; Bharti, K.; Doring, P.; Mishra, S.K.; Ganguli, A.; Scharf, K.D. *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? *Cell Stress Chaperones*, **2001**, *6*(3), 177-189.

[16] Shim, D.; Hwang, J.U.; Lee, J.; Lee, S.; Choi, Y.; An, G.; Martinoia, E.; Lee, Y. Orthologs of the class A4 heat shock transcription factor HsfA4a confer cadmium tolerance in wheat and rice. *Plant Cell*, **2009**, *21*(12), 4031-4043.

[17] Kumar, M.; Busch, W.; Birke, H.; Kemmerling, B.; Nurnberger, T.; Schoffl, F. Heat shock factors HsfB1 and HsfB2b are involved in the regulation of Pdf1.2 expression of Pdf1.2 expression and pathogen resistance in *Arabidopsis. Molecular Plant*, **2009**, *2*(1), 152-165.

[18] Kotak, S.; Vierling, E.; Baumlein, H.; Von Koskull-Doring, P. A novel transcriptional cascade regulating expression of heat stress proteins during seed development of Arabidopsis. *The Plant Cell*, **2007**, *19*, 182-195.

[19] Prieto-Dapena, P.; Castano, R.; Almoguera, C.; Jordano, J. The ectopic overexpression of a seed-specific transcription factor, HaHSFA9, confers tolerance to severe dehydration in vegetative organs. *The Plant J.*, **2008**, *54*(6), 1004-1014.

[20] Pirkkala, L.; Nykanen, P.; Sistonen, L. Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. *FASEB J.*, **2001**, *15*(7), 1118-1131.

[21] Huang, Y.; Li, M.Y.; Wang, F.; Xu, Z.S.; Huang, W.; Wang, G.L.; Ma, J.; Xiong, A.S. Heat shock factors in carrot: genome-wide identification, classification and expression profiles response to abiotic stress. *Mol. Biol. Reports*, **2014**, DOI 10.1007/ s11033-014-3826-x.

[22] Wang, F.; Dong, Q.; Jiang, H.; Zhu, S.; Chen, B.; Xiang, Y. Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula. Mol. Biol. Reports,* **2012***, 39*(2), 1877-1886.

[23] Chung, E.; Kim, K.M.; Lee, J.H. Genome-wide analysis and molecular characterization of heat shock transcription factor family in *Glycine max. J. Genet. Genomics*, **2013**, *40*(3), 127-135.

[24] Song, X.; Liu, G.; Duan, W.; Liu, T.; Huang, Z.; Ren, J.; Li, Y.; Hou, X. Genome-wide identification, classification and expression analysis of the heat shock transcription factor family in Chinese cabbage. *Mol. Genet. Genomics*, **2014**, *289*(4), 541-551.

[25] Guo, J.; Wu, J.; Ji, Q.; Wang, C.; Luo, L.; Yuan, Y.; Wang, Y.; Wang, J. Genome-wide analysis of heat shock transcription factor families in rice and *Arabidopsis*. *J. Genet. Genomics*, **2008**, *35*(2), 105-118.

[26] Letunic, I.; Copley, R.R.; Schmidt, S.; Ciccarelli, F.D.; Doerks, T.; Schultz, J.; Ponting, C.P.; Bork, P. SMART 4.0: towards genomic data integration. *Nucleic Acids Res*., **2004**, *32D*, 142-144.

[27] Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; Thompson, J.D.; Gibson, T.J.; Higgins, D.G. Clustal W and Clustal X version 2.0. *Bioinformatics*, **2007**, *23*(21), 2947-2948.

[28] Horton, P.; Keun-Joon Park.; Takeshi Obayashi.; Naoya Fujita.; Hajime Harada, C.J.; Adams-Collier,; Kenta Nekai. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*., **2007**, *35*, W585-W587.

[29] Moller, S.; Croning, M.D.R.; Apweller, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, **2001**, *17*(7), 646-653.

[30] Guo, A.Y.; Zhu, Q.H.; Chen, X.; Luo, J.C.; Chuan, Y.I. GSDS a gene structure display server. *Yi Chuan*, **2007**, *29*(8), 1023-1026.

[31] Nguyen Ba, A.N.; Pogoutse, A.; Provart, N.; Moses, A.M. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, **2009**, *10*, 202.

[32] Hawkins, J.; Davis, L.; Boden, M. Predicting nuclear localization. *J. Proteome Res.*, **2007**, *6*, 1402-1409.

[33] la Cour, T.; Klemer, L.; Molgaard, A.; Gupta, R.; Skriver, K.; Brunak, S. Analysis and prediction of leucine- rich nuclear export signals. *Protein Eng. Des. Sel.*, **2004**, *17*(6), 527-536.

[34] Bailey, T.L.; Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, **1998**, *14*(1), 48-54.

[35] Higo, K.; Ugawa, Y.; Iwamoto, M.; Korenaga, T. Plant *cis*-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.*, **1999**, *27*(1), 297-300.

[36] Lescot, M.P.; Dehais, G.; Thijs, K.; Marchal, Y.; Moreau, Y.; Van de Peer, P.; Rombauts, R.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*., **2002**, *30*(1), 325-327.

[37] Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA 5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance and Maximum Parsimony Methods. *Mol. Biol. Evol.*, **2011**, *28*(10), 2731-2739.

[38] Gu, Z.; Cavalcanti, A.; Chen, F.C.; Bouman, P.; Li, W.H. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol. Biol. Evol.*, **2002**, *19*(3), 256-262.

[39] Yang, S.; Zhang, X.; Yue, J.X.; Tian, D.; Chen, J.Q. Recent duplications dominate NBS-encoding gene expansion in two woody

species. *Mol. Genet. Genomics*, **2008**, *280*(3), 187-198.

[40] Pfaffl, W.M.; Horgan1, W.G.; Dempfle Leo. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.*, **2002**, *30*(9), 900.

[41] Reddy, P.S.; Kishor, P.B.K; Seiler, C.; Kuhlmann, M.; Eschen-Lippold, L.; Lee, J.; Reddy, M.K.; Sreenivasulu, N. Unraveling regulation of the small heat shock proteins by the heat shock factor HvHsfB2c in barley: Its implications in drought stress response and seed development. *PLoS ONE*, **2014**, *9*(3), e89125.

[42] Hahn, A.; Bublak, D.; Schleiff, E.; Scharf, K.D. Crosstalk between Hsp90 and Hsp70 chaperones and heat stress transcription factors in tomato. *The Plant Cell*, **2011**, *23*(2), 741–755.

[43] Giorno, F.; Guerriero, G.; Baric, S.; Mariani, C. Heat shock transcriptional factors in *Malus domestica*: identification, classification and expression analysis. *BMC Genomics*, **2012**, *13*, 639.

[44] Ali, M.A.; Niaz, S.; Abbas, A.; Sabir, W.; Jabran, K. Genetic diversity and assessment of drought tolerant *Sorghum* landraces based on morpho-physiological traits at different growth stages. *Plant Omics J.*, **2009**, *2*(5), 214-227.

[45] Durbin, R.S.E.; Anders, K.; Graeme, M. Biological sequence analysis: Probabilistic models of proteins and nucleic acids. *Cambridge University Press*, **1998**.

[46] McGinnis, S.; Madden, T.L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res*., **2004**, *32W*, 20−25.

[47] Lin, Y.X.; HaiYang Jiang.; Zhang-Xin Chu,; Xiu-Li Tang,; Su-Wen Zhu,; Bei-Jiu Cheng. Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. *BMC Genomics*, **2011**, *12*, 76.

[48] Mittal, D.; Chakrabarti, S.; Sarkar, A.; Singh, A.; Grover, A. Heat shock factor gene family in rice: genomic organization and transcript expression profiling in response to high temperature, low temperature and oxidative stresses. *Plant Physiol. Biochem.*, **2009**, *47*(9), 785-795.

[49] Sarkar, N.K.; Kim, Y.K.; Grover, A. Rice sHsp genes: genomic organization and expression profiling under stress and development. *BMC Genomics*, **2009**, *10*, 393.

[50] Chauhan, H.; Khurana, N.; Nijhavan, A.; Khurana, J.P.; Khurana, P. The wheat chloroplastic small heat shock protein (*sHSP26*) is involved in seed maturation and germination and imparts tolerance to heat stress. *Plant Cell Environ.*, **2012**, *35*(11), 1912-1931.

[51] Bate, N.; Twell, D. Functional architecture of a late pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. *Plant Mol. Biol.*, **1998**, *37*(5), 859-869.

[52] Flicchkin, S.A.; Leonard, J.M.; Monteros, A.; Liu, P.P.; Nonogaki, H. A novel endo-beta mannanase gene in tomato *LeMAN5* is associated with anther and pollen development. *Plant Physiol.*, **2004**, *134*(3), 1080-1087.

[53] Plesch, G.; Ehrhardt, T.; Mueller-Roeber, B. Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression. *Plant J.*, **2001**, *28*(4), 455-64.

[54] Chen, W.; Provart, N.J.; Glazebrook, J.; Katagiri, F. Chang, H.S.; Eulgem, T.; Mauch, F.; Luan, S.; Zou, G.; Whitham, S.A.; Budworth, P.R.; Tao, Y.; Xie. Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell*, **2002**, *14*(3), 559-574.

[55] Nishiuchi, T.; Shinshi, H.; Suzuki, K. Rapid and transient activation of transcription of the ERF3 gene by wounding in tobacco leaves: Possible involvement of NtWRKYs and autorepression. *J. Biol. Chem.*, **2004**, *279*(53), 55355-55361.

[56] Chen, C.; Chen, Z. Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen- induced *Arabidopsis* transcription factor. *Plant Physiol.*, **2002**, *129*(2), 706-716.

[57] Wendel, J.F. Genome evolution in polyploids. *Plant Mol. Biol*., **2000**, *42*, 225-249.

[58] Jiao, Y.; Wickett, N.J.; Ayyampalayam, S.; Chanderbali, A.S.; Landherr, L.; Ralph, P.E.; Tomsho, P.; Hu, Y.; Liang, H.; Soltis, P.S.; Soltis, D.E.; Clifton, S.W.; Schlarbaum, S.E.; Schuster, S.C.; Ma, H.; Leebens-Mack, J.; dePamphilis, C.W. Ancestral polyploidy in seed plants and angiosperms. *Nature*, **2011**, *473*(7345), 97-100.

[59] Ma, J.; Wang, F.; Li, M.Y.; Jiang, Q.; Tan, G.F.; Tan, G.F.; Xiong, A.S. Genome-wide analysis of the NAC transcription factor family in Chinese cabbage to elucidate responses to temperature stress. *Scientia Horticultrae*, **2014**, *165*, 82-90.

[60] Yamaguchi-Shinozaki, K.; Shinozaki, K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.*, **2006**, *57*, 781-803.

[61] Busch, W,; Wunderlich, M,; Schoffl, F. Identification of novel heat shock factor-dependent genes and Bio chemical pathways in *Arabidopsis Thaliana. Plant J*., **2005**, *41*(1), 1-14.

[62] Chauhan, H,; Khurana, N,; Agarwal, P,; Khurana, P. Heat shock factors in rice (*Oryza sativa* L.): genome wide expression analysis during reproductive development and abiotic stress. *Mol. Genet. Genomics*., **2011**, *286*(2), 171-187.

[63] Callis, J,; Fromm, M,; Walbot, V. Introns increase gene expression in cultured maize cells. *Genes Dev*., **1987**, *1*(10), 1183–1200.

[64] Clancy, M,; Hannah, L.C. Splicing of maize Sh1first intron is essential for enhancement of gene expression, and a T-rich motif increases expression without affecting splicing. *Plant Physiol.*, **2002**, *130*(2), 918–929.

[65] Rose, A.B.; Last, R.L. Introns act post-transcriptionally to increase expression of the *Arabidopsis thaliana* tryptophan pathway gene PAT1. *Plant J*., **1997**, *11*(3), 455–464.

[66] LeHir, H,; Nott, A,; Moore, M.J. How introns influence and enhance gene expression? *Trends Biochem. Sci*., **2003**, *28*(4), 215-220.

[67] Bharti, K,; Von Koskull-Doring, P,; Bharti, S,; Kumar, P,; Tintschl-Korbitzer, A,; Treuter, E,; Nover, L. Tomato heat stress transcription factor HsfB1 represents a novel type of general transcription co activator with a histone-like motif interacting with the plant CREB binding protein ortholog HAC1. *Plant Cell*., **2004**, *16*(6), 1521-1535.

Received on: January 28, 2015 **Revised on: March 07, 2015** **Accepted on: March 12, 2015**