



***In silico* studies predict role of *PgUCP1* from *Pennisetum glaucum* in heat stress tolerance**

ALBERT MAIBAM¹, HARINDER VISHWAKARMA² and JASDEEP CHATRATH PADARIA³

ICAR-National Institute for Plant Biotechnology, New Delhi 110 012, India

Received: 21 February 2019; Accepted: 23 July 2019

ABSTRACT

Heat stress adversely affects crop plants leading to high yield losses. To protect themselves, plants respond by expressing large number of genes. This includes reported/known genes as well as hypothetical or uncharacterized genes. Genes for uncharacterized or hypothetical proteins form a major proportion of data generated by different functional genomic approaches. It is quite important to assign function to these stress responsive uncharacterized genes for better understanding of stress responsive molecular mechanisms. In the present study, full length coding sequence of a gene for an uncharacterized protein1 PgUCP1 (624 bp) was cloned from pearl millet genotype 841-B (ICMB841) at National Institute for Plant Biotechnology, New Delhi in year 2017–18. The gene PgUCP1 (Accession number MK33595) was identified in the heat responsive transcriptome data generated in leaf tissue of *P. glaucum* plants grown in National Phytotron Facility, IARI, New Delhi. The CDS was successfully isolated and cloned in pGEM-T easy vector. The predicted 3-Dimensional structure of PgUCP1 showed that it is able to interact with ligands [AMP (Adenosine monophosphate), ADP (Adenosine diphosphate), ATP (Adenosine triphosphate)] depicting presence of active site residues. Phylogenetic analysis showed PgUCP1 to be closely related to zinc finger protein of *Setaria italica*. The predicted transcript in this study clearly indicated its role in providing heat stress tolerance. Further, the role of identified transcript can be validated in model plant system under abiotic stress conditions. The gene may be a potent prospective resource for development of abiotic stress tolerant crops.

Key words: 3-Dimensional structure, Heat stress, *Pennisetum glaucum*, Phylogenetic, Uncharacterized protein

Pennisetum glaucum (L.) R. Br. (pearl millet) has the inherent ability to grow and withstand harsh environmental conditions such as low moisture stress and high temperature stress (Maibam and Padaria 2015). Temperature is one of the physical parameter that has a vital role in maintaining proper growth and development of plant. Maximum growth and development of plant occurs at optimum temperature and its value varies based on plant species (Wahid *et al.* 2007). Many genes associated with heat stress response have been identified, which includes structural genes such as heat shock proteins (*Hsps*) (Craig *et al.* 1993), ascorbate peroxidase (*APX*) (Padaria *et al.* 2014), Ca-dependent protein kinases (*CDPKs*) (Hasanuzzaman *et al.* 2013) and regulatory genes such as heat shock factors (*Hsfs*) (Liu *et al.* 2011), *WRKY* transcription factor (Dang *et al.* 2013) etc. Functional genomics has helped in deciphering a number of heat stress responsive genes but the annotated data have a large proportion of uncharacterized genes.

Sequence attributes, expression or even localization are some parameters supporting evidence that most of the uncharacterized genes are bonafide genes. In most of genomic scale studies, characterized genes are given more emphasis as compared to uncharacterized genes. Regardless of the success of high-throughput techniques in molecular biology, many biological researchers find it difficult to annotate uncharacterized proteins (Kolker *et al.* 2014), as a result limited studies have targeted uncharacterized genes. In the present study based on the fold changes and *p*-value adjust parameters, an uncharacterized gene, was selected from heat stress response transcriptome data of *Pennisetum glaucum* cytoplasmic male sterile line ICMB841 (P1537587) (Singh *et al.* 1990) and named as Uncharacterized Protein 1 (*PgUCP1*). *PgUCP1* was cloned and further *in silico* analysis was performed to add new information and enable us to predict the functions for the uncharacterized gene. The *in silico* based studies predicted the role of *PgUCP1* in heat stress tolerance. Further experimental evidence of this gene may be studied by developing transgenics tolerant to abiotic stresses.

Present address: ¹PhD Scholar (maibamalbert@gmail.com) ICAR-IARI; ²Research Associate (harinder.v@gmail.com), ³Principal Scientist (jasdeep_kaur64@yahoo.co.in) ICAR-National Institute for Plant Biotechnology.

MATERIALS AND METHODS

Plant materials, growth conditions and heat stress

treatment: Seeds of pearl millet genotype 841-B (ICMB841) were collected from Division of Genetics, Indian Agricultural Research Institute (IARI), New Delhi, India. Seeds were surface sterilized and sown in plastic pots (10 inches) filled with vermiculite and grown under glasshouse condition, at National Phytotron facility, IARI, New Delhi in the year 2016–17. At flowering stage (55 days after sowing), the plants were subjected to a heat stress treatment (temperature: 42°C for 6 h) in a growth chamber. For heat stress treatment, the temperature was gradually increased (1°C per 10 min) till 42°C was reached. Leaf samples were collected after 6 h of heat stress, quickly frozen in liquid nitrogen to be stored at -80°C for further use. Total RNA was isolated from the flag leaf of pooled samples using Trizol method (Chomczynski *et al.* 1987) and purified using MN (Macherey-Nagel, Germany) kit. cDNA was prepared from 5 µg of total RNA using SuperScript™ III first strand cDNA synthesis kit (Invitrogen, USA).

Construct preparation and transformation in *E. coli* DH5a cells: The CDS (Coding DNA Sequence) of gene *PgUCP1* (Accession no. MK335955) was amplified from total cDNA using gene specific primers. The forward (*PgUCP1* F: CGGGGTACCCCGATGATGTGCATGCAATGCCT) and reverse primers (*PgUCP1* R: CCGGAATTCCGGTTAGTTTGATGTAGAACACTCTACCGTA) were designed, having the *KpnI* and *EcoRI* restriction sites respectively for downstream cloning experiment. *PgUCP1* gene was amplified in a three step PCR programme (initial denaturation at 94°C for 3 min, then 32 cycles of 94°C for 30 s, 58°C for 30 s, 72°C for 60 s). The obtained specific amplicon was observed on 1% agarose gel electrophoresis and subsequently extracted from the gel using QIAquick Gel Extraction Kit (Qiagen, US). The *PgUCP1* amplicon was ligated to vector pGEM®-T Easy (Insert:Vector:: 5:1) using Rapid DNA Ligation Kit (Thermo Scientific™, USA). The ligated product was transformed in *E. coli* DH5a using heat shock method and the recombinant cells were plated in Luria–Bertani (LB) agar plate having ampicillin (100 mg/l), isopropyl-β-D thiogalactoside (IPTG) (1 mM) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (20 mg/l) for blue/white selection. The confirmation of the recombinant clones was carried out by colony PCR. Plasmid was isolated from the positive clones using GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich, USA) and restriction digestion carried to confirm the presence of desirable insert. The recombinant plasmid having *PgUCP1* gene was sequenced in year 2017-18.

In silico studies of *PgUCP1*: Using ProtParam tool (<http://web.expasy.org/cgi-bin/protparam>) and Expert Protein Analysis System (ExPASy) (Artimo *et al.* 2012) different protein parameters were estimated. Multiple alignment of *PgUCP1* with similar predicted proteins of other related species was carried out using Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar 2004). MEGA7 (<https://www.megasoftware.net>) (Kumar *et al.* 2016) software was employed to construct

phylogenetic tree with Neighbour-joining statistical method for studying evolutionary relationship with similar predicted proteins of other plant species. The boot strap value was set as 1500 and the evolutionary distance was calculated using p-distance model.

Secondary structure prediction, modelling and Ramachandran plot: PSIPRED software tool (<http://bioinf.cs.ucl.ac.uk/psipred/>) was employed to predict α-helices, β-strand, coils and amino acid percentage composition (McGuffin *et al.* 2000). Phyre2 software tool (<http://www.sbg.bio.ic.ac.uk/phyre2>) was used to predict the level of disorder in the secondary structure of protein and to normalize QMEAN4 score and 3D protein structure prediction. To identify ligand binding site and to predict active site of *PgUCP1*, the generated model was tied up to a ligand using 3D Ligand Site prediction server (<http://www.sbg.bio.ic.ac.uk/3dligandsite>). Ramachandran plot was used to examine the quality of conformations of the obtained 3D structure in terms of amino acid residues percentage in favourable regions. Using Verify 3D (<http://servicesn.mbi.ucla.edu/Verify3D/>) and ERRAT software (<http://servicesn.mbi.ucla.edu/ERRAT/>) tool, Ramachandran plot was assessed.

Subcellular location, phosphorylation site and kinase specific prediction: CELLO v.2.5 (<http://cello.life.nctu.edu>) software was used to predict the subcellular location of *PgUCP1*. For prediction of the phosphorylation site and kinase specific prediction, Netphos 3.1a server (<http://cbs.dtu.dk/services/Netphos/>) was used. Using TMHMM server (<http://www.cbs.dtu.dk/services/TMHMM/>) transmembrane helix region was estimated. To find out the N-linked and O-linked glycosylation site respectively in *PgUCP1*, NetNGly 1.0 server (<http://www.cbs.dtu.dk/services/NetNGly/>) and YinOYang 1.2 server (<http://www.cbs.dtu.dk/services/YinOYang/>) were used.

RESULTS AND DISCUSSION

Amplification and cloning of *PgUCP1* gene: Total RNA isolated from flag leaf of plants subjected to heat stress, was checked for quality by electrophoresis on 1%

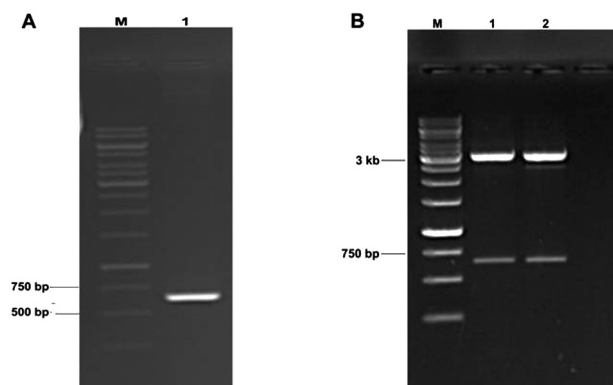


Fig 1 (A) PCR amplification of full length *PgUCP1* CDS. (B) Restriction digestion confirmation of isolated plasmid. M is 1 kb DNA marker

Table 1 Predicted subcellular location of PgUCP1 using CELLO v.2.5 server

Location	Score
Extracellular	3.352
Nucleus	0.553
Plasma membrane	0.459
Mitochondrial	0.305
Cytoplasmic	0.170
Chloroplast	0.061
Lysosomal	0.040
Vacuole	0.017
Endoplasmic reticulum	0.011
Golgi	0.011
Cytoskeletal	0.010
Peroxisomal	0.010

agarose gel. CDS synthesized from the isolated RNA was used as a template for amplification of *PgUCP1* (624bp) CDS (GenBank accession number MK335955). The obtained amplicon of 624 bp was observed on 1 % agarose gel electrophoresis. The *PgUCP1* amplicon was ligated with pGEM-T easy vector and transformed in *E. coli* DH5 α with a transformation efficiency of 4.3×10^4 colony forming unit (cfu)/ μ g. Putative recombinant clones (white colonies) were confirmed by colony-PCR and an amplicon of ~ 750 bp comparable to the size for *PgUCP1* CDS was observed on 1% agarose gel electrophoresis. Colony PCR positive clones were again confirmed for the presence of the candidate gene by double digestion using *KpnI* and *EcoRI* restriction endonucleases and successful excision of inserted *PgUCP1* CDS was observed. The positive clone was finally reconfirmed with sequencing using T7 & SP6 primer pairs (Fig 1).

In silico based analysis of PgUCP1: PgUCP1 was found to have cysteine (14%) and tryptophan (0.5%) as the highest and lowest amino acid residue respectively. PgUCP1 was observed to have nearly neutral pI (7.34) which might be due to almost equal number of charged amino acid residue i.e. total number of negatively charged residues (aspartic acid and glutamic acid) were 19 while total number of positively charged residues (arginine and lysine) were 20. The instability index and GRAVY (Grand average of hydropathicity) values were 46.73 and -0.122 respectively which suggests its unstable and hydrophilic nature. The molecular weight of PgUCP1 was predicted to be 23.603 kDa having an isoelectric point of 7.34, with atomic composition of $C_{1009}H_{1549}N_{281}O_{294}S_{40}$. Predicted subcellular localization of PgUCP1 shows extracellular location having the highest score of 3.352, followed by nucleus and plasma membrane which suggests its highly soluble and hydrophilic nature (Table 1). No site of transmembrane helix as well as no signal peptide site was observed in the PgUCP1. This suggests that PgUCP1 might not being targeted to any particular organelle. Post-translational modifications such

as phosphorylation of serine, threonine and tyrosine amino acid residues, play important role in proper functioning of protein. Sites for 17 different kinases {ATM (Ataxia-telangiectasia mutated), CKI (Casein kinase I), CKII, (Casein kinase II) CaM-II (Ca²⁺/calmodulin-dependent kinase II), DNAPK (DNA dependent protein kinase), EGFR (Epidermal growth factor receptor), GSK3 (Glycogen synthase kinase 3), INSR (Insulin receptor tyrosine kinase), PKA (Protein kinase A), PKB (Protein kinase B), PKC (Protein kinase C), PKG (Protein kinase G), RSK (Receptor Tyr kinase), SRC (Sarcoma family kinases), cdc2 (Cell division cycle protein kinase), cdk5 (Cyclin Dependent Kinase 5) and p38MAPK Mitogen-activated protein kinase (<http://www.cbs.dtu.dk/services/NetPhos/>)} and 10 phosphorylating sites were observed to be present in the PgUCP1. These phosphorylated sites of PgUCP1 might have a role in regulating response to abiotic stresses (Ogiso *et al.* 2004). Moreover, phosphorylation also influenced others cellular activities like splicing, protein-protein interaction (Zhang *et al.* 2009) etc. PgUCP1 was found to have five potential

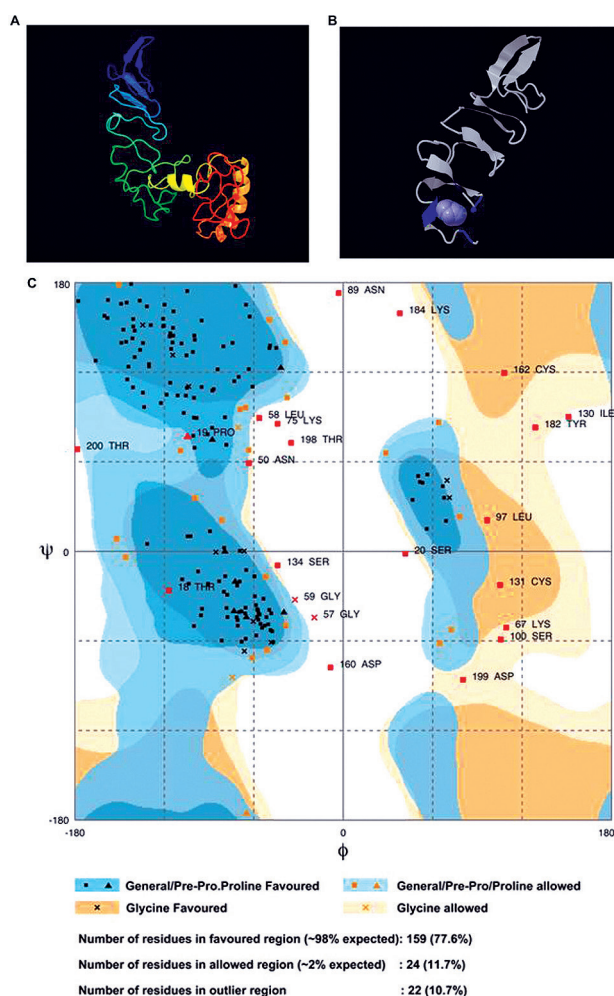


Fig 2 Three dimensional structure prediction of PgUCP1. (A) Modelling of PgUCP1 based on template- c2k2cA. (B) PgUCP1 docking to a predicted ligand (coloured portion shows active site interaction with ligand). (C) Ramachandran plot showing the favourable region of PgUCP1 protein

Table 2 List of PgUCP1 related orthologous genes (retrieved from NCBI database).

Plant	Family	Accession no	Sequence length CDS (bp)	Deduced amino acid	% Sequence similarity (nucleotide)	% Sequence similarity (amino acid)
<i>Setaria italica</i>	Poaceae	XM_004961260.4	3672	1223	97	99
<i>Sorghum bicolor</i>	Poaceae	XM_002440153.2	3648	1215	94	97
<i>Zea mays</i>	Poaceae	XM_008651747.2	3618	1205	93	97
<i>Oryza brachyantha</i>	Poaceae	XM_006654685.2	3114	1037	88	92
<i>Oryza sativa japonica</i>	Poaceae	XM_015783612.1	3081	1026	87	92
<i>Brachypodium distachyon</i>	Poaceae	XM_003567899.4	3705	1234	86	89
<i>Hordeum vulgare</i>	Poaceae	AK363968.1	3705	1234	85	88
<i>Aegilops tauschii</i>	Poaceae	XM_020291568.1	3714	1237	85	88
<i>Gossypium arboreum</i>	Malvaceae	XM_017755782.1	3705	1234	80	80

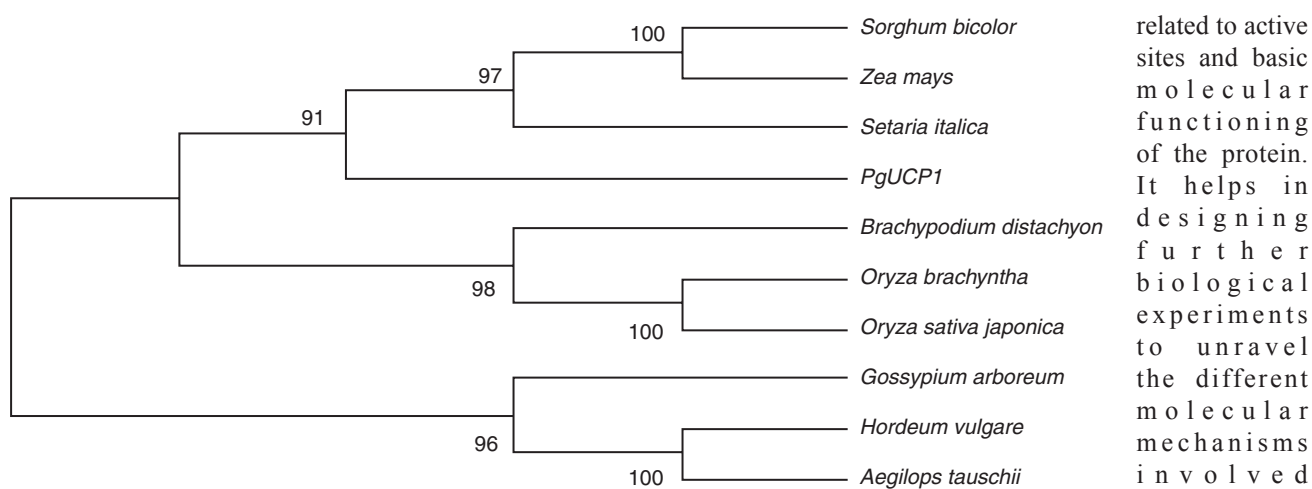


Fig 3 Phylogenetic analysis of PgUCP1 with other plant species homologs

O-linked glycosylation site at position 18, 100, 204, 205 and 206. O-linked glycosylation of intracellular proteins play an important role in response to oxidative and others abiotic stresses (Chen *et al.* 2018). Secondary and 3D structure prediction of protein provides information related to active sites and basic molecular functioning of the protein (Kolahi *et al.* 2018). Secondary structure of PgUCP1 showed presence of α -helices (15%), β -strands (24%) and coils. The protein was observed to have disorderness of 15%. PgUCP1 was 96% modelled at >90% confidence, the best match of PgUCP1 was with solution nuclear magnetic resonance (nmr) structure of n-terminal domain of human (p53-induced RING-H2 protein) pirh2.2 northeast structural genomics consortium (nesg) target ht2a (Protein data bank (PDB) id: c2k2cA) (42% identical with 100% confidence, PDB info-signalling molecule, resolution-3.90 Å, Model dimensions (Å)- X:41.347, Y:71.386, Z:35.943) (Fig 2A). PgUCP1 was further docked to a molecule (predicted ligand) as shown in Fig 2B. Active site of PgUCP1 consists of Cys65, Lys67, Cys68, Met76, Ile77 and Cys81 residues. These amino acids of PgUCP1 proposed to be involved in interacting with ligands [(AMP (Adenosine monophosphate), ADP (Adenosine diphosphate), ATP (Adenosine triphosphate)]. Secondary and 3D structure prediction of protein provides information

related to active sites and basic molecular functioning of the protein. It helps in designing further biological experiments to unravel the different molecular mechanisms involved during cellular processes. Ramachandran

plot analysis showed 77.6% amino acid residues were in the favoured region, 11.7% amino acid residues were in the allowed region while 10.7% amino acid residues were in the outlier region (Fig 2C).

Phylogenetic analysis: BLAST (Altschul *et al.* 1990) analyses of the PgUCP1 showed high similarity percentage with *Setaria italica* zinc finger protein, available in NCBI database (Accession no: XM_004961260.4) (Table 2). Phylogenetic analysis revealed that the evolutionary distance of PgUCP1 from pearl millet genotype 841-B of the present study with that of other plant species. PgUCP1 was found to be closely related to gene for zinc finger protein *Setaria italica*, *Sorghum bicolor* and *Zea mays* (Fig 3).

Zinc finger proteins have been found to be involved in binding specific sequence of DNA and protein-protein interaction (Mackay and Crossley 1998). Zinc finger domains are actively involved in plant growth, development and regulating resistance mechanisms under various stress conditions (Gourcilleau *et al.* 2011, Kodaria *et al.* 2011, Parthasarathy and Murthy 2000). PgUCP1 similarity with gene for Zinc finger protein, does indicate that this gene is involved in stress tolerance mechanism in pearl millet. Upregulation of the expression of PgUCP1 in response to heat stress in the tolerant genotype of pearl millet, further

validates the role of *PgUCP1* in molecular mechanism of heat stress tolerance.

In this study, *PgUCP1* full length CDS was cloned and its function was predicted based on *in silico* analysis. Secondary and 3-Dimension structure of uncharacterized protein PgUCP1 was predicted and validated by Ramachandran plot. The best match of PgUCP1 was with zinc finger protein of *Setaria itilica*, zinc finger domains are actively involved in growth and development especially at the time of stress conditions in plants. PgUCP1 also showed high homology with RING finger and CHY zinc finger domain containing 1 (RCHY1) of *Homo sapiens*, having a role in ubiquitination and proteasomal degradation. In addition, we performed phylogenetic analysis to give evolutionary relationship of PgUCP1 with uncharacterized protein of other plant species. The present study indicates a role of *PgUCP1* in abiotic stress tolerance in pearl millet. However, further biological experimentation shall decipher the involvement of *PgUCP1* in different abiotic stress tolerance and metabolic pathway.

REFERENCES

- Altschul S F, Gish W, Miller W, Myers E W and Lipman D J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403–10.
- Artimo P, Jonnalagedda M, Arnold K *et al.* 2012. ExpASy: SIB bioinformatics resource portal. *Nucleic Acids Research*. 40(Web Server issue):W597-603.
- Chen P H, Chi J T and Boyce M. 2018. Functional crosstalk among oxidative stress and O-GlcNAc signaling pathways. *Glycobiology* **28**(8): 556–64.
- Chomczynski P and Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162**(1): 156–9.
- Craig E A, Gambill B D and Nelson R J. 1993. Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiological Reviews* **57**(2): 402–14.
- Dang F F, Wang Y N, Yu L. 2013. CaWRKY40, a WRKY protein of pepper, plays an important role in the regulation of tolerance to heat stress and resistance to *Ralstonia solanacearum* infection. *Plant, Cell and Environment* **36**(4): 757–74.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**(5): 1792–7.
- Gourcilleau D, Lenne C, Armenise C, Moulia B, Julien J, *et al.* 2011. Phylogenetic study of plant Q-type C2H2 zinc finger proteins and expression analysis of poplar genes in response to osmotic, cold and mechanical stresses. *DNA Research* **18**: 77–92.
- Hasanuzzaman M, Nahar K, Alam M M, Roychowdhury R and Fujita M. 2013. Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *International Journal of Molecular Science* **14**(5): 9643–84.
- Kodaira K, Qin F, Tran L P, Maruyama K, Kidokoro S. 2011. Arabidopsis Cys2/His2 zinc-finger proteins AZF1 and AZF2 negatively regulate abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions. *Plant Physiology* **157**: 742–56.
- Kolahi M, Yazdi M, Goldson-Barnaby A and Tabandeh MR. 2018. *In silico* prediction, phylogenetic and bioinformatic analysis of *SoPCS* gene, survey of its protein characterization and gene expression in response to cadmium in *Saccharum officinarum*. *Ecotoxicology and Environmental Safety* **163**: 7–18.
- Kolker E, Picone A F and Galperin M Y. 2005. Global profiling of *Shewanella oneidensis* MR-1: Expression of hypothetical genes and improved functional annotations. *Proceedings of the National Academy of Sciences* **102**(6): 2099–104.
- Kumar S, Stecher G and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* **33**(7): 1870–4.
- Liu H C, Liao H T and Charng Y Y. 2011. The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in Arabidopsis. *Plant Cell and Environment* **34**(5): 738–51.
- Maibam A and Padaria J C. 2015. Pearl millet : A Genetic resource for abiotic tolerant transgenics. *Biotech Today* **5**(1): 21–4.
- Mackay JP and Crossley M. 1998. Zinc fingers are sticking together. *Trends in Biochemical Sciences* **23**: 1–4.
- McGuffin L J, Bryson K and Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* **16**(4): 404–5.
- Ogiso H, Kagi N, Matsumoto E. 2004. Phosphorylation analysis of 90 kDa heat shock protein within the cytosolic arylhydrocarbon receptor complex. *Biochemistry* **43**(49): 15510-15519.
- Padaria J C, Vishwakarma H, Biswas K, Jasrotia R S and Singh G P. 2014. Molecular cloning and *in silico* characterization of high temperature stress responsive *pAPX* gene isolated from heat tolerant Indian wheat cv. Raj 3765. *BMC Research Notes* **7**(1):713.
- Parthasarathy S and Murthy M R. 2000. Protein thermal stability: Insights from atomic displacement parameters (B values). *Protein Engineering* **13**: 9–13.
- Singh SD, Singh P, Rai KN and Andrews DJ. 1990. Registration of ICMA 841 and ICMB 841. *Crop Science* **36**(6): 1378.
- Wahid A, Gelani S, Ashraf M and Foolad M R. 2007. Heat tolerance in plants: An overview. *Environmental and Experimental Botany* **61**(3):199–223.
- Zhang X N, Mount S M. 2009. Two alternatively spliced isoforms of the Arabidopsis SR45 protein have distinct roles during normal plant development. *Plant Physiology* **150**(3): 1450–8.