Journal of Genetics (2019) 98:46
https://doi.org/10.1007/s12041-019-1087-0

© Indian Academy of Sciences

RESEARCH ARTICLE

Transcriptome analysis of finger millet (*Eleusine coracana* (L.) Gaertn.) reveals unique drought responsive genes

M. S. PARVATHI[1], KARABA N. NATARAJA[1*], Y. A. NANJA REDDY[1,2], MAHANTESHA B. N. NAIKA[1,3] and M. V. CHANNABYRE GOWDA[2]

[1]*Department of Crop Physiology, University of Agricultural Sciences, and* [2]*AICRP (Small Millets), ICAR-UAS, GKVK, Bengaluru 560 065, India*
[3]*Department of Biotechnology and Crop Improvement, K.R.C. College of Horticulture, University of Horticultural Sciences, Bagalkot 591 218, India*
*For correspondence. E-mail: nataraja_karaba@yahoo.com, nnkaraba@uasbangalore.edu.

Received 27 July 2018; revised 14 December 2018; accepted 17 December 2018; published online 9 May 2019

Abstract. Finger millet (*Eleusine coracana* (L.) Gaertn.), an important C_4 species is known for its stress hardiness and nutritional significance. To identify novel drought responsive mechanisms, we generated transcriptome data from leaf tissue of finger millet, variety GPU-28, exposed to gravimetrically imposed drought stress so as to simulate field stress conditions. *De novo* assembly-based approach yielded 80,777 and 90,830 transcripts from well-irrigated (control) and drought-stressed samples, respectively. A total of 1790 transcripts were differentially expressed between the control and drought-stress treatments. Functional annotation and pathway analysis indicated activation of diverse drought-stress signalling cascade genes such as serine threonine protein phosphatase 2A (*PP2A*), calcineurin B-like interacting protein kinase31 (*CIPK31*), farnesyl pyrophosphate synthase (*FPS*), signal recognition particle receptor α (*SRPR α*) etc. The basal regulatory genes such as TATA-binding protein (TBP)-associated factors (TAFs) were found to be drought responsive, indicating that genes associated with housekeeping or basal regulatory processes are activated under drought in finger millet. A significant portion of the expressed genes was uncharacterized, belonging to the category of proteins of unknown functions (PUFs). Among the differentially expressed PUFs, we attempted to assign putative function for a few, using a novel annotation tool, Proteins of Unknown Function Annotation Server. Analysis of PUFs led to the discovery of novel drought responsive genes such as pentatricopeptide repeat proteins and tetratricopeptide repeat proteins that serve as interaction modules in multiprotein interactions. The transcriptome data generated can be utilized for comparative analysis, and functional validation of the genes identified would be useful to understand the drought adaptive mechanisms operating under field conditions in finger millet, as has been already attempted for a few candidates such as *CIPK31* and *TAF6*. Such an attempt is needed to enhance the productivity of finger millet under water-limited conditions, and/or to adopt the implicated mechanisms in other related crops.

Keywords. finger millet; drought; transcriptome; ontology; differentially expressed genes; *Eleusine coracana*.

Introduction

Finger millet (*Eleusine coracana* (L.) Gaertn.) is a drought-tolerant C_4 plant belonging to the family Poaceae, and subfamily Chloridoideae. It is an annual allotetraploid ($2n = 4x = 36$, AABB) millet existing as two subspecies *coracana* and *africana* (Hilu and De Wet 1976; Hilu 1994). India is considered the secondary centre of diversity of finger millet, which is the fourth most important millet crop, cultivated in more than 25 countries (Upadhyaya *et al.* 2007; Vetriventhan *et al.* 2015; Hittalmani *et al.* 2017). The crop is an important cereal widely grown in Africa and Asia, known for its high-nutritive value, with multiple traits associated with drought and salinity tolerance, and blast disease resistance (Shailaja and

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12041-019-1087-0) contains supplementary material, which is available to authorized users.

Thirumeni 2007; Agarwal *et al.* 2011). Using traditional breeding approaches many agronomic traits have been improved in finger millet, and with the help of omics-based approaches novel genes have also been identified from the crop (Rahman *et al.* 2014, 2016; Hittalmani *et al.* 2017). High-quality whole-genome sequence data of finger millet (cultivar PR202) with sufficient coverage and assembled via a novel multiple hybrid assembly workflow is now available (Hatakeyama *et al.* 2018). Attempts have been also made to prospect a few drought-responsive genes (Parvathi *et al.* 2013) and validate their significance in abiotic stress tolerance (Ramegowda *et al.* 2012; Babitha *et al.* 2015; Nagarjuna *et al.* 2016; Parvathi and Nataraja 2017). Since drought tolerance is governed by multiple traits and associated pathways, investigations on the mechanisms and pathways prevalent in finger millet could facilitate crop improvement programmes. With an aim to unravel stress responsive pathways in finger millet, we performed transcriptome analysis using leaf tissue of finger millet variety GPU-28 grown under well-irrigated and drought conditions.

Materials and methods

Plant material and stress treatment

***Creation of drought stress*:** Healthy plants of finger millet (*E. coracana* (L.) Gaertn., variety GPU-28) were raised in battery pots containing 18–20 kg of 2:1 garden soil and sand mixture, by adopting all timely plant production and protection measures. The water status of the pots was maintained at 100% field capacity (FC) by controlled irrigation, and drought stress was imposed on 20-day old plants by a gravimetric approach to simulate field conditions (Karaba *et al.* 2007; Parvathi and Nataraja 2017), where soil FCs of 80, 60, 40 and 20% were maintained. Gradual moisture stress was imposed by weighing the pots and the loss of water in the pots was replenished with the required amount of water to arrive at the desired FC of soil. At the end of stress period, when soil reached the required FC, the pots were maintained at that particular FC for one week and then assessed for stress damage by determining loss of cell viability. For comparison, control plants were maintained at 100% FC by regular watering. Subsequently, after 10 days of stress imposition, leaf tissues were harvested, frozen in liquid nitrogen and stored at −70°C for RNA isolation.

Drought-stress imposition and assessment of stress damage

***Estimation of soil water potential (SWP)*:** To assess the extent of drought-stress imposed, SWP was monitored using a dewpoint potentiometer (Parvathi *et al.* 2013). The soil sampling was performed at a depth of 10–15 cm from the rhizosphere of stressed and nonstressed plants and the SWPs of the different samples were recorded using the WP4 dewpoint potentiometer (Decagon Devices, USA). The SWP measurement was performed at a temperature of 25°C. The experiment was conducted in three replicates.

***Estimation of cell viability*:** The estimation of cell viability is a good measure of stress induced damage. The 2,3,5-triphenyltetrazolium chloride (TTC; HiMedia Laboratories, India) assay was performed to measure the extent of cell viability (Senthil-Kumar and Udayakumar 2006). Leaf discs from fully expanded leaves of stressed plants 1-week post stress exposure as well as from nonstressed plants were incubated in TTC solution at room temperature for 5 h while shaking. The leaf discs were then washed to remove unbound form and boiled with 5 mL of 2-methoxyethanol to dryness. The bound TTC was extracted with 5 mL of 2-methoxyethanol and absorbance was measured at 485 nm using a UV-visible spectrophotometer (UV-VIS 2450, Shimadzu, Kyoto, Japan). The reduction in cell viability over nonstressed plants was calculated using the following equation and expressed in per cent:

$$\text{Reduction in cell viability} = ((A_{485})\text{control} - (A_{485})\text{stressed})/((A_{485})\text{control}) \times 100(\%)$$

RNA extraction, cDNA library preparation and sequencing

Total RNA was extracted from frozen leaf samples of individual stress treatments (80, 60, 40, 20% FCs) along with control (100% FC) using the TRIzol reagent (Sigma-Aldrich, USA), as per the manufacturer's instructions. Total RNA concentration and purity of samples were quantified using a Nanodrop spectrophotometer and a Qubit Fluorometer (Invitrogen, USA), respectively. The two samples used for transcriptome analysis are as follows: sample A (total RNA from leaf tissues of absolute control, 100% FC) and sample B (pooled from equal quantity of total RNA from leaf tissues collected from treatments 80, 60, 40 and 20% FCs). A pooled stressed sample would ensure the capture of early as well as late stress responsive candidates. The RNA integrity of samples was checked using an Agilent Bioanalyzer chip (Agilent Technologies, USA). Library preparation was performed following the Illumina TruSeq RNA library protocol as per the manufacturer's instructions. One microgram of RNA was subjected to poly(A) purification of mRNA and purified mRNA was fragmented for 4 min at an elevated temperature (94°C) in the presence of divalent cations and reverse transcribed with Superscript III Reverse transcriptase in the presence of first strand mix. Second strand cDNA was synthesized using second strand mix. The cDNA was cleaned up using Agencourt AMPureXP SPRI beads (Beckman Coulter,

USA) and adapters were ligated to the cDNA molecules after end repair and addition of A base. SPRI bead clean-up was performed after ligation. The library was purified, and amplified using eight cycles of polymerase chain reaction (PCR) for enrichment of adapter-ligated fragments. The prepared library was quantified using a spectrophotometer and validated for quality by running an aliquot on a High Sensitivity Bioanalyzer Chip (Agilent Technologies, USA). The RNA seq libraries were sequenced using an Illumina NextSeq 500 (2× paired end (151 bp)).

Data processing, de novo assembly and functional annotation

The Illumina NextSeq 500 paired end raw reads were quality checked using FastQC 8.0. (Andrews 2010) and processed by using in-house scripts for trimming adapters and low-quality bases towards the 3′-end. *De novo* assembly of the processed read data was performed using Trinity (Grabherr *et al.* 2011) for default k-mers i.e. 25. After assembly, the contigs were rechecked for adapter sequences and none were found. All of the raw data have been deposited in the Short Read Archive (SRA) at the National Centre for Biotechnology Information (NCBI) database under the project accession number SRP057792. The obtained unique transcripts were utilized for performing BLAST-2.2.29+ (Altschul *et al.* 1990) search with a cut-off E-value of 1×10^{-5}, at NCBI. Further, transcripts were annotated using the proteins of Viridiplantae kingdom taken from the UniProt database. Gene ontology (GO) terms were assigned to the *E. coracana* assembled transcripts based on the GO terms annotated to their corresponding homologues in the UniProt database (Apweiler *et al.* 2004).

Pathway analysis

Pathway analysis was performed by KEGG Automatic Annotation Server (KAAS) (Moriya *et al.* 2007), using *Arabidopsis thaliana*, *Glycine max*, *Fragaria vesca*, *Vitis vinifera*, *Solanum lycopersicum*, *Oryza sativa japonica* and *Theobroma cacao* as reference organisms. In addition to the routine analysis, data were filtered out from the entire transcriptome data generated, with a focus on upstream basal regulators. Based on pathway annotation results, genes with GO terms having 'transcription' as molecular function was filtered out and classified.

Differential gene expression analysis

Differentially expressed gene(s) (DEGs) were estimated based on the number of reads mapped against unique locus under both control and treated conditions using the DESeq tool (Anders and Huber 2010). Data were log 2 transformed to identify the genes with significant differential expression during drought stress with a criteria of fold change, ≥ 1 log2 fold change ≤ -1 and P value < 0.05.

Functional characterization of uncharacterized proteins using PUFAS server

The genes identified as uncharacterized proteins from annotation of DEGs were analysed using the free online web server, Proteins of Unknown Function Annotation Server (PUFAS; http://caps.ncbs.res.in/pufas/; Dhanyalakshmi *et al.* 2016), in an attempt to unravel possible novel pathways linked to drought hardiness in finger millet. A few highly upregulated and downregulated DEGs which were identified in this study, annotated as uncharacterized proteins, were used as the input into the PUFAS server.

Results

Drought stress effect

To obtain a global view of the transcriptome of finger millet under drought and control conditions, pot-grown plants were exposed to a prolonged period of drought-stress simulating field conditions. There was reduction in the SWP with the increase in the intensity of stress level (figure 1). The reduced cell viability (figure 1) indicated the effect of drought stress on plant physiology.

Sequencing and de novo assembly

High-throughput Illumina NextSeq 500 paired-end sequencing technology was performed on poly(A) enriched RNAs (mRNA-Seq) from drought-stressed and control leaves, with a specific aim to prospect specific pathways

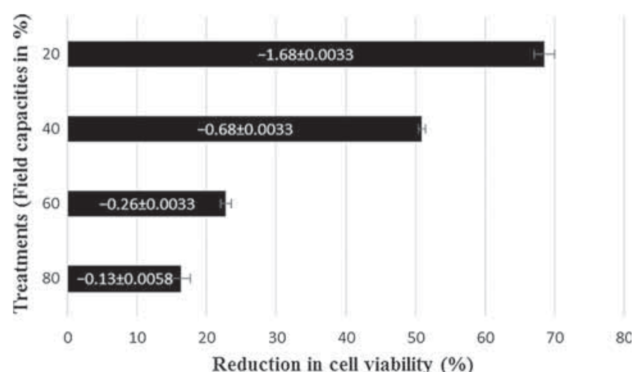

Figure 1. Assessment of stress imposition and stress damage. Analysis of reduction in cell viability of stressed leaf tissue over control and the SWPs corresponding to different drought-stress treatments. The graph bars represent mean ± standard error (SE) of three replicates. Horizontal bars indicate SE. The numerals on the bars represent SWP (Ψw in MPa ± SE).

and genes responsible for tolerance in finger millet leaf tissue experiencing field level drought stress. Total raw and processed reads were recorded for control (44,861,506 and 41,206,698) and drought-stressed (48,751,234 and 44,691,572) samples, respectively. *De novo* assembly with 25k-mer length, generated 80,777 and 90,830 transcripts from control and drought-stressed samples, respectively, with average length of 1050 and 1047 bp with N50 length of 1487 and 1495 bp, respectively (table 1). The raw sequence data have been deposited in the NCBI-SRA with accession ID SRP057792.

Annotation of assembled transcripts

NCBI-BLAST search identified 27,469 and 29,710 unique protein accessions for control and drought-treated samples, respectively, suggesting that Illumina paired-end sequencing had likely captured a substantial proportion of the drought-responsive genes of finger millet. We further annotated the unique transcripts by assigning them to GO terms. In total, 22,586 unigenes corresponding to known proteins were assigned to GO classes using 4036 functional terms (table 1 in electronic supplementary material). As shown in figure 2, a&b, the majority of the unigenes were assigned to the categories of molecular functions (25,479), followed by biological processes (15,507) and cellular components (13,212). Genes which failed to obtain a GO term, fell into distinct categories such as uncharacterized, putative and hypothetical proteins, which were used for a separate analysis in this study (table 1 in electronic supplementary material). The GO category of biological processes includes genes associated with DNA integration, transcription, carbohydrate metabolic process, translation, DNA repair, transmembrane transporter, protein-folding lipid metabolic process etc. The genes associated with integral component of membrane, nucleus, membrane, chloroplast, cytoplasm, cytosol, mitochondrion etc., were included under the category of cellular components. Genes involved in adenosine triphosphate binding, zinc ion binding, nucleic acid binding, DNA binding, serine/threonine kinase activity, metal ion binding, RNA binding, catalytic activity, sequence-specific DNA-binding transcription factor binding sites etc., were prominently represented under the category of molecular functions. This indicated that important cellular response mechanisms occurred in finger millet in response to drought stress. The top 10 groups in the biological process, cellular components and molecular function categories are shown in figure 2, a&b (table 1 in electronic supplementary material).

Pathway analysis

To identify the biological pathways that are active, all expressed genes were mapped to the reference canonical

Table 1. Summary of the finger millet (*E. coracana* L.) transcriptome by *de novo* assembly.

Sample name	Transcripts generated	Maximum transcript length	Minimum transcript length	Average transcript length	Median transcript length	Transcripts ≥ 300 bp	Transcripts > 500 bp	Transcripts > 1 kb	Transcripts > 10 kb	N50	Total nucleotides
Control	80,777	14,598	300	1050	1115	80,777	54,292	30,237	10	1487	84,815,096
Drought	90,830	13,512	300	1047	411	90,830	60,740	33,784	9	1495	95,098,860

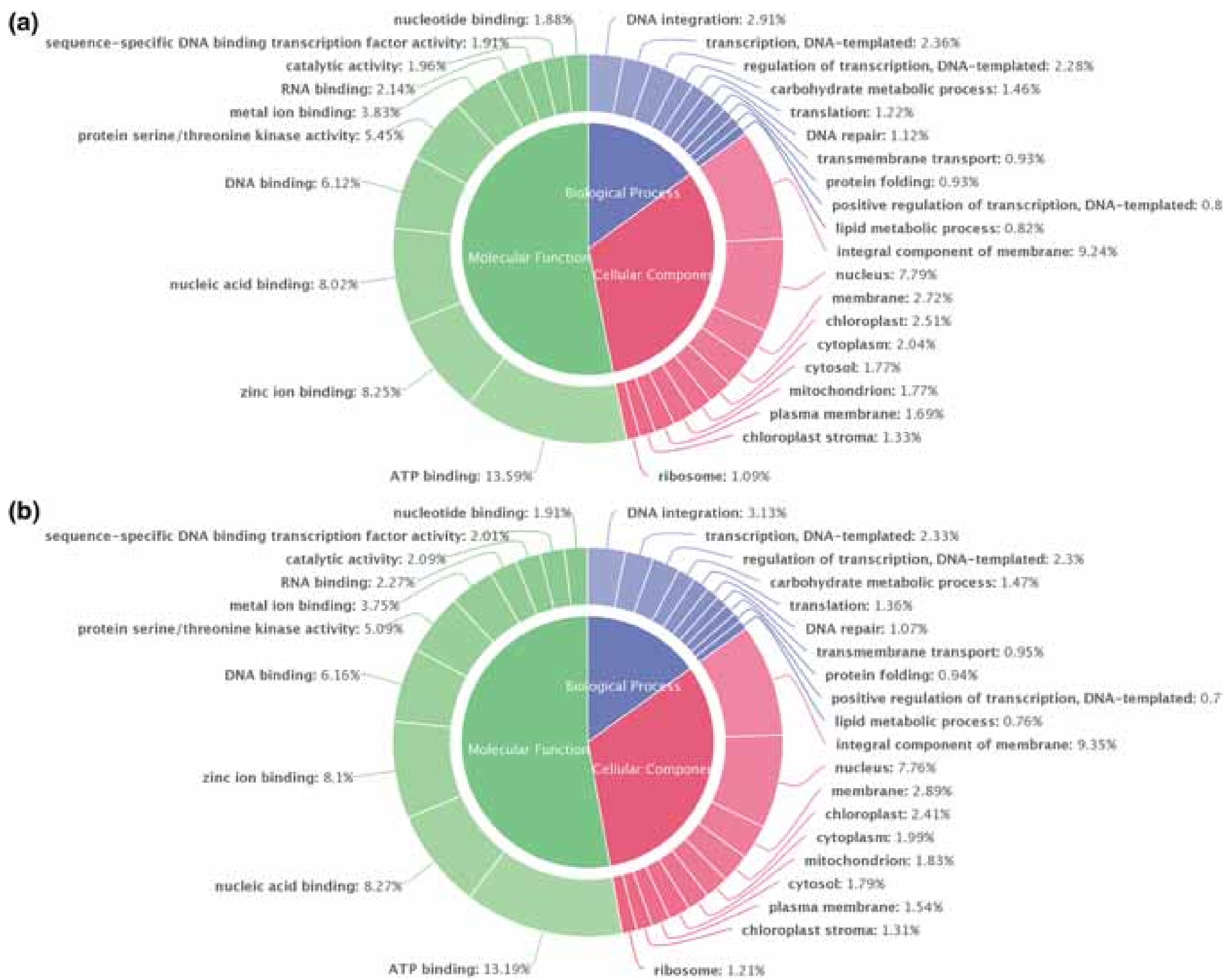

Figure 2. Representation of the selected top 10 transcript annotation classes from the transcriptome data of (a) control and (b) treated leaf tissue samples of finger millet.

pathways in the KEGG. A total of 131 unique pathways belonging to 19 clades were identified under five major KEGG categories, including metabolism, genetic information processing, environmental information processing, cellular processes and organismal systems. Among them, metabolism of cofactors and vitamins, carbohydrate metabolism, biosynthesis of other secondary metabolites, amino acid metabolism, energy metabolism, nucleotide metabolism, metabolism of terpenoids and polyketides, lipid metabolism, metabolism of other amino acids, glycan biosynthesis and metabolism, folding, sorting and degradation, transcription, translation, replication and repair, signal transduction, transport and catabolism, membrane transport and environmental adaptation emerged as the key pathways (figure 3; table 2 in electronic supplementary material). These pathways provide a valuable resource for investigating specific processes, functions and pathways in plant drought-stress research.

Identification of DEG

The DEG analysis revealed 1790 transcripts that were differentially expressed between the control and drought-stress treatments (table 3 in electronic supplementary material). The control sample showed lower number of DEGs (83 genes) when compared to the drought-stressed sample (163 genes). Genes that were contrastingly regulated were mapped on a heatmap (figure 4). Gene identities of the highly differentially expressed transcripts are shown in table 4 in electronic supplementary material. To further characterize the expression changes of DEGs in drought-treated leaves compared to control, GO analysis was also undertaken (table 4 in electronic supplementary material). The representation of DEGs during the progression of stress shows the stress responsiveness and their putative role in stress tolerance in finger millet.

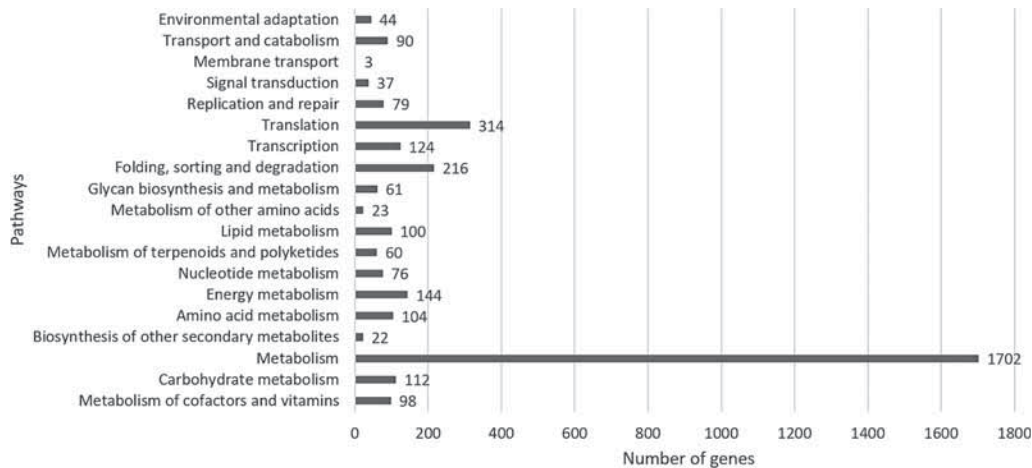

Figure 3. Pathway analysis of assembled transcripts from drought-stressed leaf tissue of finger millet.

Functional characterization of uncharacterized proteins among DEGs

The considerable representation of uncharacterized proteins in the transcriptome data led to the need for a careful analysis pertaining to their identity. We made use of an in-house web analysis tool to identify proteins of unknown functions (PUFs) based on their functional domain / protein family / fold characteristics (Dhanyalakshmi *et al.* 2016). We attempted to identify selected upregulated and downregulated DEGs that were initially annotated as uncharacterized proteins, the details of which are given in table 5 in electronic supplementary material. The reannotation of these selected uncharacterized proteins using PUFAS predicted that most of the upregulated DEGs were either regulatory in nature (such as transcription factors and kinases/phosphatases) or other functional proteins (such as pumps/translocators) (table 5 in electronic supplementary material). The downregulated DEGs were annotated to diverse functionalities such as enzymes, cell cycle related proteins, receptors etc., (table 5 in electronic supplementary material). There were genes which did not relate to any protein based on homology/domain characteristics per se but were annotated based on protein fold similarities such as Myc box-dependent-interacting protein 1 (upregulated DEG; ~4.7 log 2 fold change) and pyrrolidone-carboxylate peptidase (downregulated DEG; ~−2.7 log 2 fold change), to name a few. The detailed analysis results are given in table 5 in electronic supplementary material.

Full-length clones generated from the transcriptome information

With leads from the earlier work on prospecting candidate genes linked to drought tolerance in finger millet (Parvathi *et al.* 2013), we cloned a few identified pathway genes and other novel regulators. The genes cloned were serine threonine protein phosphatase 2A (*PP2A*), calcineurin B-like interacting protein kinase31 (*CIPK31*), farnesyl pyrophosphate synthase (*FPS*), signal recognition particle receptorα (*SRPRα*) and TBP associated factor6 (*TAF6*). The clones were confirmed by sequencing and annotation, the results of which are given in figure 1 in electronic supplementary material. The sequence information of the cloned genes was submitted to the NCBI database (GenBank accession numbers: *EcCIPK31* – KT288194, *EcTAF6-1* – KT824872, *EcPP2A* – KT824869, *EcSRPRα* – KT824870, *EcFPS* – KT824871) the details of which are given in table 2.

Identification of novel basal regulators

The transcriptome data generated was curated with a focus on upstream basal regulators, such that genes related to transcription regulation and transcript abundance were captured. Based on pathway annotation results, genes coming under the category of pathway function as ‘Transcription’ under the major group ‘Genetic Information Processing’ were filtered out and classified as shown in figure 5. Genes belonging to the categories of RNA polymerase-related genes, basal transcription factors and splicing factors were the major representatives. Among all three, there was an appreciable representation of basal transcription regulators contributing to about 18% (figure 5). The different basal transcription regulators identified are presented in table 3. The basal transcription factor representatives belonged to the TFIID family of general transcription factors (TFs) involved in preinitiation complex assembly, namely TATA-binding protein (TBP)-associated factors (TAFs). However, there were also representations from other general transcription factor families such as TFIIH-1, TFIIH-2, TFIIH-3 and TFIIH-4 (transcription initiation factor TFIIH subunits) (data not shown).

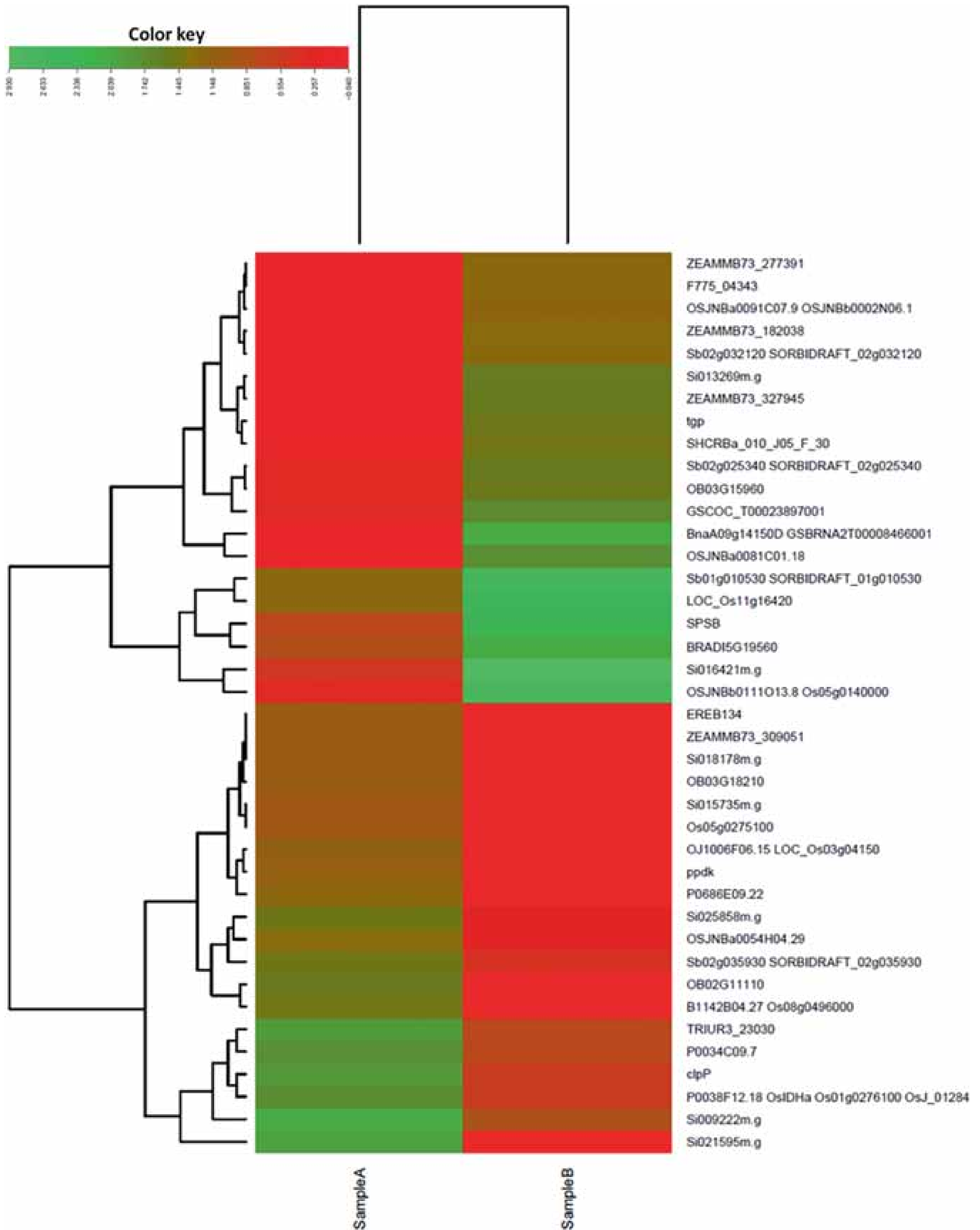

Figure 4. Heatmap showing DEGs between control and drought-stress treatments. (variance stabilized data obtained with the DESeq package was used to generate the heatmap. Green indicates higher expression and red indicates lower expression. Colour key indicates the intensity associated with normalized expression values.)

Table 2. Full-length clones generated in this study.

	Gene name	Gene notation	GenBank accession no.	CDS (bp)
1	CBL interacting protein kinase 31-like	*EcCIPK31*	KT288194	1350
2	Signal recognition particle receptor subunit alpha-like	*EcSRPRα*	KT824870	1854
3	Transcription initiation factor TFIID subunit TBP associated factor 6-1	*EcTAF6*	KT824872	1635
4	Farnesyl pyrophosphate synthase-like	*EcFPS*	KT824871	1062
5	Serine/threonine-protein phosphatase PP2A-4 catalytic subunit-like	*EcPP2A*	KT824869	942

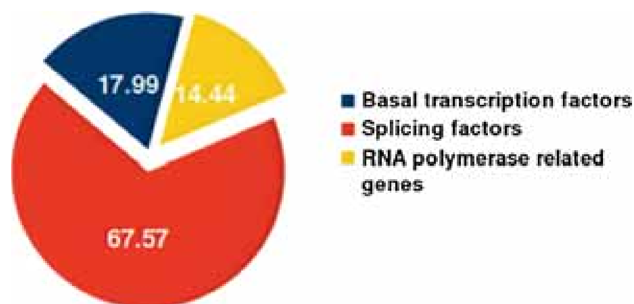

Figure 5. Depiction of representation of factors involved in transcription regulation based on pathway annotation results obtained from transcriptome data generated.

Discussion

Prospecting novel stress responsive mechanisms/genes from adapted species

Drought tolerance is critically linked to cellular stability attained by the coordinated expression of multiple genes. Cellular function is reliant on timely manoeuvre of multiple vital processes, and hence it is very critical to ensure cellular tolerance under stressful conditions. Cellular stress response pathways are controlled by a number of highly conserved signalling molecules and transcriptional regulators which eventually govern gene

Table 3. List of basal transcription regulators identified from finger millet drought-stressed leaf transcriptome.

Contig identity	Contig ID
CCNH; cyclin H	c27423_g1_i8
CDK7; cyclin-dependent kinase 7	c34484_g3_i2
ERCC2; DNA excision repair protein ERCC-2	c29351_g1_i1
TAF1; transcription initiation factor TFIID subunit 1	c37362_g1_i1
TAF2; transcription initiation factor TFIID subunit 2	c37256_g2_i2
TAF4; transcription initiation factor TFIID subunit 4	c36108_g1_i2
TAF5; transcription initiation factor TFIID subunit 5	c34887_g1_i1
TAF6; transcription initiation factor TFIID subunit 6	c34506_g1_i2
TAF7; transcription initiation factor TFIID subunit 7	c28800_g1_i1
TAF8; transcription initiation factor TFIID subunit 8	c42357_g1_i1
TAF9B; transcription initiation factor TFIID subunit 9B	c27166_g1_i1
TAF10; transcription initiation factor TFIID subunit 10	c21317_g1_i1
TAF11; transcription initiation factor TFIID subunit 11	c23054_g1_i1
TAF12; transcription initiation factor TFIID subunit 12	c33700_g1_i2
TAF13; transcription initiation factor TFIID subunit 13	c30884_g1_i3
TBP; transcription initiation factor TFIID TATA-box-binding protein	c34605_g1_i2
TFIIA1; transcription initiation factor TFIIA large subunit	c28946_g1_i1
TFIIA2; transcription initiation factor TFIIA small subunit	c30587_g3_i1
TFIIB; transcription initiation factor TFIIB	c34535_g1_i1
TFIIE1; transcription initiation factor TFIIE subunit alpha	c31589_g1_i2
TFIIE2; transcription initiation factor TFIIE subunit beta	c35227_g2_i3
TFIIF1; transcription initiation factor TFIIF subunit alpha	c33472_g1_i1
TFIIF2; transcription initiation factor TFIIF subunit beta	c22810_g1_i1
TFIIH1; transcription initiation factor TFIIH subunit 1	c34754_g1_i1
TFIIH2; transcription initiation factor TFIIH subunit 2	c29956_g1_i1
TFIIH3; transcription initiation factor TFIIH subunit 3	c32319_g2_i2
TFIIH4; transcription initiation factor TFIIH subunit 4	c33272_g1_i6
TTDA; TFIIH basal transcription factor complex TTD-A subunit	c26456_g1_i1

expression. Since many factors interact and alter acclimation response, identification of novel regulators or functional proteins to evolve tolerance strategies by genetic manipulation to develop stress-tolerant plants under field conditions would be ideal. The adapted plants might have novel proteins/regulators or integrated stress resistance pathways (Parvathi *et al.* 2013). With this outlook, there were earlier attempts to identify stress specific proteins from drought-adapted plants. However, due to the complex nature of drought-stress responses, it is evident that there are novel pathways to be unveiled. We attempted to generate resources by transcriptome analysis of finger millet (cultivar GPU-28) under drought stress.

Transcriptome analysis of finger millet variety GPU-28 under drought

RNA sequencing of finger millet (cultivar GPU-28) leaf tissue and functional annotation of transcripts indicated the activation of genes belonging to diverse categories (figure 2; table 1 in electronic supplementary material). Pathway analysis predicted 131 unique pathways belonging to 19 clades under five major KEGG categories (figure 3; table 2 in electronic supplementary material). Earlier studies also found similar patterns in the plant model system during abiotic stress (Naika *et al.* 2013; Jaiswal *et al.* 2018). Global analysis of DEGs provided a comprehensive dataset responding to drought stress in leaves (figure 4). The genes identified would serve as resources for functional analysis and crop improvement. The information can also be used to identify genic markers for crop breeding programmes. We predicted a total of 13,569 SSR markers from the transcriptome data, which are to be validated (data not shown).

Cloning of a few novel representative stress responsive signalling cascade genes

Our major focus of transcriptome analysis was to identify genes associated with drought adaptive response in finger millet. Using the leads generated, a few pathway and regulatory genes were cloned into full length from finger millet. The major genes cloned were *EcPP2A* and *EcCIPK31* involved in signalling pathways, *EcTAF6* which is a basal transcriptional regulator, *EcFPS* involved in posttranslational modification (PTM) and *EcSRPRα* involved in protein targeting.

Protein phosphatase 2A (*PP2A*) was reported to be 2-fold higher in drought-stressed finger millet leaf tissue (Parvathi *et al.* 2013). *PP2A* has been reported to be a key regulator of many metabolic processes in vascular plants and also various signalling cascades under diverse stress conditions (Pais *et al.* 2009). *PP2A* mediated regulation in complex HMG-CoA reductase (HMGR) activity under stress conditions indicates its significance (Antolin-Llovera *et al.* 2011). Calcineurin B-like protein (CBL) -interacting protein kinase31 (*CIPK31*) was found to be highly stress responsive under different abiotic stresses such as methyl viologen-induced oxidative stress, salinity and heat stresses at seedling stage as well as under whole plant level drought stress in finger millet, indicating that it is an active component of diverse stress signalling pathways (Nagarjuna *et al.* 2016). The roles of *CIPKs* in imparting abiotic stress tolerance have been well documented (Xiang *et al.* 2007; Tripathi *et al.* 2009; Zhang *et al.* 2014), but specific *CIPKs* could have distinct functionalities as evidenced by *CIPK31* identified in finger millet, which could also be involved in seed calcium accumulation. Hence, key signalling partners such as *PP2A* and *CIPKs* could be involved in signalling modules linked to drought hardiness in finger millet.

In earlier reports, the expression of *FPS* was not consistent at different stress levels in finger millet, showing very low expression at 20% FC although the expression levels were relatively high under severe stress conditions (30% FC) (Parvathi *et al.* 2013). One of the important PTM mechanisms is farnesylation, facilitated by an enzyme *in vivo* (FPS) for membrane localization (Benetka *et al.* 2006). Farnesylated proteins have been reported earlier to be involved in maintaining membrane integrity and having regulatory roles in plant growth and development and stress responses like ABA signalling (Taylor 1996; Nambara and McCourt 1999). Hence, PTM or regulation of proteins could be a key regulatory mechanism linked to drought hardiness in finger millet.

During stress acclimation response, basal transcriptional regulation is critical, which is coordinated by several cofactors, among which TAFs are the major proteins involved in preinitiation complex assembly. Among all 12 TAFs identified in finger millet, *EcTAF6* was selected for further evaluation based on its uniqueness with respect to the presence of a downstream promoter element in *Drosophila* and implications in multiple regulatory processes in yeast and human system. *EcTAF6* was initially identified from the drought stress-specific cDNA library of finger millet and further functionally validated to report for the first time the stress responsiveness of a TBP-associated factor (Parvathi and Nataraja 2017). Targeted expression analysis revealed that *TAF6* exhibited differential upregulation pattern under multiple abiotic stresses in finger millet. Virus-induced gene silencing of *NbTAF6* in *Nicotiana benthamiana* resulted in a developmentally defective and stress sensitive phenotype. These phenotypes were further justified by the expression pattern of different categories of genes in the silenced leaf tissue when compared to wild-type and mock-treated plants (Parvathi and Nataraja 2017). Hence, the role of a crucial component of TFIID transcription factor complex was elucidated which

indicates that there could be more key basal regulators that are important for stress homoeostasis.

SRPRα was also found to be upregulated under drought in finger millet in a pattern such that fold increase in transcript levels was 1.4 and 2.3 at 40 and 60% FC, respectively, displaying its stress responsiveness (Parvathi 2010). The signal recognition particle (SRP) and the SRP receptor (SR) were reported to cotranslationally deliver newly synthesized proteins from the cytosol to target membranes (Walter and Johnson 1994; Keenan *et al.* 2001). The SRP-dependent protein-targeting pathway has already been reported in chloroplast (Schuenemann *et al.* 1998). However, an interaction between the SR and translocon was evidenced to be critical during cotranslational protein translocation (Jiang *et al.* 2008). Proper localization of synthesized proteins is mediated by crucial protein-targeting machinery consisting of the SRP carrying the signal sequences and the functionality of SR could be a critical molecular realm governing stress hardiness in finger millet. Novel regulators involved in multiple biochemical regulations along a stress signalling cascade could be crucial regulatory checkpoints defining the stress hardiness of finger millet, as also evidenced by their representation in the KEGG pathway analysis. The relevance of these identified candidate pathway linked genes from finger millet under drought stress could be analysed and characterized in future, for unravelling novel signalling pathways, as has been elucidated in case of *EcCIPK31*-like (Nagarjuna *et al.* 2016) and *EcTAF6* (Parvathi and Nataraja 2017).

Identification of novel basal regulatory genes from finger millet

With special focus on proteins involved in transcriptional regulation, it was identified that basal transcriptional regulation was effected by genes related to RNA polymerase activity, splicing and basic transcription factor complexes in finger millet. Among these, basal transcription factors accounted for around 18% which included TFIIA, B, D, E, F and H family members (figure 5; table 3). The TFIID general transcription factor family includes major factors involved in transcription preinitiation complex assembly called TATA box binding protein (TBP)-associated factors (TAFs). There were 12 TAFs represented in finger millet transcriptome both in control and stress conditions, owing to their basal constitutive expression, namely *TAF1, 2, 4, 5, 6, 7, 8, 9B, 10, 11, 12, 13* as against 14 TAFs in *Arabidopsis* (Lago *et al.* 2004). Among all the TAFs, *TAF6* is the only member with a downstream promoter element as reported in *Drosophila*; thus, dissecting its role assumes significance, since it could enhance stability of transcriptional machinery. This resulted in the identification of a novel transcriptional regulator, *EcTAF6* (Parvathi and Nataraja 2017). We further probed into the stress responsive nature of 15 TAFs reported in *Arabidopsis* by *in silico* e-northern analysis under drought stress (Rakshitha 2017). Similarly, an *in silico* expression analysis of the selected TAFs was carried out under drought as well as salt stress at different time points in rice. TAFs such as *TAF4* and *TAF6* were found to be commonly upregulated in both *Arabidopsis* and rice which was also experimentally reconfirmed by reverse transcription-PCR in finger millet (Rakshitha 2017). Careful analysis of the generated stress specific transcriptome data, in conjunction with the genome information available on cultivars ML365 (Hittalmani *et al.* 2017) and PR202 (Hatakeyama *et al.* 2018) can yield significant stress-related homologues by resequencing analyses, if demanded. Copies of important stress-related genes or their homologues could be systematically analysed to reveal novel drought-related functionalities.

Functional characterization of uncharacterized DEGs

The drought stress specific transcriptome has representation of PUFs and proteins with domains of unknown functions, indicating that novel proteins still exist to be identified in finger millet that may contribute to its drought hardiness. Although it is true that this analysis yielded meaningful annotation to many unknown proteins, it is difficult to attribute their contribution to a specific defined stress responsive pathway. Nevertheless, it is significant to mention that there are seemingly interesting implications of the data generated on annotation of PUFs. For example, the pentatricopeptide repeat proteins (PPRs) which are regulatory at the RNA level (Manna 2015) and the tetratricopeptide repeat proteins (TPRs) which serve as interaction modules in multiprotein complexes (Zeytuni and Zarivach 2012) were annotated from the uncharacterized upregulated DEGs. It is interesting to note that among the annotated upregulated DEGs, a few that belong to common pathways were also represented like PPR-associated serine/threonine-protein kinase prpf4B-like involved in pre-mRNA splicing (Eckert *et al.* 2016), along with the HEAT repeat-containing protein 5B isoform and TPR-associated F-box/LRR-repeat protein 3 which are members of other important multiple repeat protein families that nature employs to promote protein–protein interaction (Zeytuni and Zarivach 2012). Hence, PPRs and TPRs could be key players involved in multiple crucial pathways related to stress response in finger millet. This analysis also revealed the apparent existence of intron retention splicing in finger millet which could be extrapolated to the dependence of splicing on prp4 kinase activity which is similar to serine/threonine-protein kinase prpf4B-like protein found in our transcriptome data (Eckert *et al.* 2016). Alternate splicing could thus be a critical post-transcriptional regulatory mechanism under drought stress, capable of evoking

drought-shock memory responses. At the same time, two downregulated DEGs, namely annotated as lipopolysaccharide (LPS)-binding protein (Thomas *et al.* 2002) and brassinosteroid insensitive 1-associated receptor kinase 1 (BAK1) (Chinchilla *et al.* 2007) were implicated to be involved in pattern recognition receptors signalling mediated by pathogen-associated molecular patterns which include LPS and FLS2 (complexes with BAK1), that mediates innate immunity (Kumar *et al.* 2011). However, it is not clear as to how the downregulation of plant defense signalling is related to an abiotic stress response scenario. Similarly, other downregulated DEGs include cyclin-dependent kinase 2/F-1-like (Levkau *et al.* 1998) and proteasomal ubiquitin receptor ADRM1 (Qiu *et al.* 2006) both of which are implicated in cell cycle progression and apoptosis in eukaryotes. A downregulation in the cell death related events in finger millet under stress could be a possible explanation for its sturdy nature under adverse conditions, although other complementary pathways are required to support such responses. Critical insights into probable novel pathways can be obtained by further analysis, and such an analysis could yield cross disciplinary comprehension of the multitude of stress responsive attributes in a drought hardy crop like finger millet.

In conclusion, we generated transcriptome data from drought-stressed finger millet leaf tissue to unravel novel stress responsive regulators/pathways/mechanisms underlying the drought hardy nature of finger millet. Identification of stress responsive basal regulators such as TAFs indicates that maintenance of housekeeping cellular activities could be a crucial mechanism under drought stress in finger millet. Basal transcriptional regulators such as *TAF6* could be involved in multiple pathways to regulate multiple stress responsive genes. Novel regulatory genes such as *PP2A*, *CIPK31*, *FPS* and *SRPRα* have been found to be associated with drought response. However, further biological significance can only be reinstated upon systematic validation approaches. An additional approach to ascertain novel functionalities to PUFs is required to identify unique pathways operating in finger millet. The information generated in this study therefore, gives an insight into the various stress responsive genes and probable novel pathway modules that they could be involved in, under stress.

Acknowledgements

We acknowledge the Department of Biotechnology, Government of India, New Delhi for funding the adhoc project Phenomics and Genomics of Ragi (file no. BT/PR5791/AGR/2/852/2012; D.O. no. BT/AB/Climate/2007 (Part II)). This project is also partially funded by the Indian Council of Agricultural Research (AICRP-Small Millets), New Delhi, India. We also thank the Department of Science and Technology, Government of India, for infrastructure support under DST-FIST programme. MS Parvathi would like to thank University Grants Commission, New Delhi, India for awarding Junior Research Fellowship (F. 17-3/2002 (SA-I)) during the course of the study.

References

Agarwal S., Singh P. K. and Singh A. 2011 Responses of some genotypes of finger millet (*Eleusine coracana* Gaertn.) for their salt tolerance. *Int. J. Curr. Res.* **3**, 45–50.

Altschul S., Gish W., Miller W., Myers E. and Lipman D. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Anders S. and Huber W. 2010 Differential expression analysis for sequence count data. *Genome Biol.* **11R**, 106.

Andrews S. 2010 FastQC: a quality control tool for high throughput sequence data. Available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Antolin-Llovera M., Leivar P., Arro M., Ferrer A., Boronat A. and Campos N. 2011 Modulation of plant HMG-CoA reductase by protein phosphatase 2A: positive and negative control at a key node of metabolism. *Plant Signal Behav.* **6**, 1127–1131.

Apweiler R., Bairoch A., Wu C. H., Barker, W. C., Boeckmann B., Ferro S. *et al.* 2004 UniProt: the Universal Protein knowledgebase. *Nucl. Acids Res.* **32**(Database issue), D115-9.

Babitha K. C., Vemanna R. S., Nataraja K. N. and Udayakumar M. 2015 Overexpression of EcbHLH57 transcription factor from *Eleusine coracana* L. in tobacco confers tolerance to salt, oxidative and drought stress. *PLoS One* **10**, e0137098.

Benetka W., Koranda M., Maurer-Stroh S., Pittner F. and Eisenhaber F. 2006 Farnesylation or geranylgeranylation? Efficient assays for testing protein prenylation in vitro and in vivo. *BMC Biochem.* **7**, 6.

Chinchilla D., Zipfel C., Robatzek S., Kemmerling B., Nurnberger T., Jones J. D. *et al.* 2007 A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**, 497–500.

Dhanyalakshmi K. H., Naika M. B. N., Sajeevan R. S., Mathew O. K., Shafi K. M., Sowdhamini R. *et al.* 2016 An approach to function annotation for proteins of unknown function (PUFs) in the transcriptome of Indian mulberry. *PLoS One* **11**, e0151323.

Eckert D., Andree N., Razanau A., Zock-Emmenthal S., Lutzelberger M., Plath S. *et al.* 2016 Prp4 kinase grants the license to splice: control of weak splice sites during spliceosome activation. *PLoS Genet.* **12**, e1005768.

Grabherr M. G., Haas B. J., Yassour M., Levin J. Z., Thompson D. A., Amit I. *et al.* 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652.

Hatakeyama M., Aluri S., Balachadran M. T., Sivarajan S. R., Patrignani A., Grüter S. *et al.* 2018 Multiple hybrid de novo genome assembly of finger millet, an orphan allotetraploid crop. *DNA Res.* **25**, 39–47.

Hilu K. 1994 Validation of the combination *Eleusine coracana* subspecies Africana (Kennedy-O'Byrne) Hilu et Dewet. *Phytologia* **76**, 410–411.

Hilu K. W. and De Wet J. 1976 Domestication of *Eleusine coracana*. *Econ. Bot.* **30**, 199–208.

Hittalmani S., Mahesh H. B., Shirke M. D., Biradar H., Uday G., Aruna Y. R. *et al.* 2017 Genome and transcriptome sequence of finger millet (*Eleusine coracana* (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. *BMC Genomics* **18**, 465.

Jaiswal S., Antala T. J., Mandavia M. K., Chopra M., Jasrotia R. S., Tomar R. S. *et al.* 2018 Transcriptomic signature of drought response in pearl millet (*Pennisetum glaucum* (L.) and development of web-genomic resources. *Sci. Rep.* **8**, 3382.

Jiang Y., Cheng Z., Mandon E. C. and Gilmore R. 2008 An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. *J. Cell Biol.* **180**, 1149–1161.

Karaba A., Greco R., Aharoni A., Trijatmiko K. R., Marsch-Martinez N., Krishnan A. *et al.* 2007 Improvement of water use efficiency in rice by expression of HARDY, an Arabidopsis drought and salt tolerance gene. *Proc. Natl. Acad. Sci. USA* **104**, 15270–15275.

Keenan R. J., Freymann D. M., Stroud R. M. and Walter P. 2001 The signal recognition particle. *Annu. Rev. Biochem.* **70**, 755–775.

Kumar H., Kawai T. and Akira S. 2011 Pathogen recognition by the innate immune system. *Int. Rev. Immunol.* **30**, 16–34.

Lago C., Clerici E., Mizzi L., Colombo L. and Kater M. M. 2004 TBP-associated factors in Arabidopsis. *Gene* **342**, 231–241.

Levkau B., Koyama H., Raines E. W., Clurman B. E., Herren B., Orth K. *et al.* 1998 Cleavage of p21Cip1/Waf1 and p27Kip1 mediates apoptosis in endothelial cells through activation of Cdk2: role of a caspase cascade. *Mol. Cell.* **1**, 553–563.

Manna S. 2015 An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* **113**, 93–99.

Moriya Y., Itoh M., Okuda S., Yoshizawa A. C. and Kanehisa M. 2007 KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185.

Nagarjuna K. N., Parvathi M. S., Sajeevan R. S., Pruthvi V., Mamrutha H. M. and Nataraja K. N. 2016 Full-length cloning and characterization of abiotic stress responsive *CIPK31*-like gene from finger millet, a drought-tolerant crop. *Curr. Sci.* **111**, 890–894.

Naika M., Shameer K. and Sowdhamini R. 2013 Comparative analyses of stress-responsive genes in Arabidopsis thaliana: insight from genomic data mining, functional enrichment, pathway analysis and phenomics. *Mol. Biosyst.* **9**, 1888–1908.

Nambara E. and McCourt P. 1999 Protein farnesylation in plants: a greasy tale. *Curr. Opin. Plant Biol.* **2**, 388–392.

Pais S. M., Tellez-Inon M. T. and Capiati D. A. 2009 Serine/threonine protein phosphatases type 2A and their roles in stress signaling. *Plant Signal Behav.* **4**, 1013–1015.

Parvathi M. S. 2010 Prospecting candidate genes for drought tolerance from finger millet (*Eleusine coracana* (L.) Gaertn). M.Sc. thesis, UAS Bengaluru, India.

Parvathi M. S. and Nataraja K. N. 2017 Discovery of stress responsive TATA-box binding protein associated Factor6 (*TAF6*) from finger millet (*Eleusine coracana* (L.) Gaertn). *J. Plant Biol.* **60**, 335–342.

Parvathi M. S., Nataraja K. N., Yashoda B. K., Ramegowda H. V., Mamrutha H. M. and Rama N. 2013 Expression analysis of stress responsive pathway genes linked to drought hardiness in an adapted crop, finger millet (*Eleusine coracana*). *J. Plant Biochem. Biotechol.* **22**, 192–201.

Qiu X. B., Ouyang S. Y., Li C. J., Miao S., Wang L. and Goldberg A. L. 2006 Hrpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, UCH37. *EMBO J.* **25**, 5742–5753.

Rahman H., Jagadeeshselvam N., Valarmathi R., Sachin B., Sasikala R., Senthil N. *et al.* 2014 Transcriptome analysis of salinity responsiveness in contrasting genotypes of finger millet (*Eleusine coracana* L.) through RNA-sequencing. *Plant Mol. Biol.* **85**, 485–503.

Rahman H., Ramanathan V., Nallathambi J., Duraialagaraja S. and Muthurajan R. 2016 Over-expression of a NAC67 transcription factor from finger millet (*Eleusine coracana* L.) confers tolerance against salinity and drought stress in rice. *BMC Biotechnol.* **16**, 35.

Rakshitha H. C. 2017 Studies on the role of TATA Box binding proteins (TBP) Associated Factors (TAFs) from finger millet in regulating abiotic stress acclimation response. M.Sc. thesis, UAS, Bengaluru, India.

Ramegowda V., Senthil-Kumar M., Nataraja K. N., Reddy M. K., Mysore K. S. and Udayakumar M. 2012 Expression of a finger millet transcription factor, EcNAC1, in tobacco confers abiotic stress-tolerance. *PLoS One* **7**, e40397.

Senthil-Kumar M. and Udayakumar M. 2006 High-throughput virus-induced gene-silencing approach to assess the functional relevance of a moisture stress-induced cDNA homologous to *Lea4*. *J. Exp. Bot.* **57**, 2291–2302.

Shailaja H. and Thirumeni S. 2007 Evaluation of salt-tolerance in finger millet (*Eleusine coracana*) genotypes at seedling stage. *Indian J. Agric. Sci.* **77**, 672–674.

Schuenemann D., Gupta S., Persello-Cartieaux F., Klimyuk V. I., Jones J. D., Nussaume L. *et al.* 1998 A novel signal recognition particle targets light-harvesting proteins to the thylakoid membranes. *Proc. Natl. Acad. Sci. USA* **95**, 10312–10316.

Taylor C. 1996 Farnesylation – a sticky business. *Plant Cell* **8**, 2151–2153.

Thomas C. J., Kapoor M., Sharma S., Bausinger H., Zyilan U., Lipsker D. *et al.* 2002 Evidence of a trimolecular complex involving LPS, LPS binding protein and soluble CD14 as an effector of LPS response. *FEBS Lett.* **531**, 184–188.

Tripathi V., Parasuraman B., Laxmi A. and Chattopadhyay D. 2009 CIPK6, a CBL-interacting protein kinase is required for development and salt tolerance in plants. *Plant J.* **58**, 778–790.

Upadhyaya H., Gowda C. and Reddy V. G. 2007 Morphological diversity in finger millet germplasm introduced from southern and Eastern Africa. *J. SAT Agric. Res.* **3**, 1–3.

Vetriventhan M., Upadhyaya H. D., Dwivedi S. L., Pattanashetti S. K. and Singh S. K. 2015 Finger and foxtail millets. In: *Genetic and genomic resources for grain cereals improvement* (ed. Mohar Singh and Hari D. Upadhyaya), pp. 291–319. Academic Press, Elsevier.

Walter P. and Johnson A. E. 1994 Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10**, 87–119.

Xiang Y., Huang Y. and Xiong L. 2007 Characterization of stress-responsive CIPK genes in rice for stress tolerance improvement. *Plant Physiol.* **144**, 1416–1428.

Zeytuni N. and Zarivach R. 2012 Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* **20**, 397–405.

Zhang H., Yang B., Liu W. Z., Li H., Wang L., Wang B. *et al.* 2014 Identification and characterization of CBL and CIPK gene families in canola (*Brassica napus* L.). *BMC Plant Biol.* **14**, 8.

Corresponding editor: MANOJ PRASAD