Transcriptome analysis of two contrasting genotypes of pearl millet to gain insight into heat stress responses

CURRENT STATUS: UNDER REVIEW

BMC Genomics  BMC Series

Albert Maibam
ICAR-National Institute for Plant Biotechnology

Sunil Nigombam
ICAR-National Institute for Plant Biotechnology

Harinder Vishwakarma
ICAR-National Institute for Plant Biotechnology

Showkat Ahmad Lone
University of Kashmir

Kishor Gaikwad
ICAR-National Institute for Plant Biotechnology

S. V. Amitha Mithra
ICAR-National Institute for Plant Biotechnology

Madan Pal Singh
Division of Plant Physiology, ICAR-Indian Agricultural Research Institute

Sumer Pal Singh
Division of Genetics, ICAR-Indian Agricultural Research Institute

Monika Dalal
ICAR-National Institute for Plant Biotechnology

Jasdeep Chatrath Padaria
NATIONAL INSTITUTE FOR PLANT BIOTECHNOLOGY

jasdeep_kaur64@yahoo.co.in  Corresponding Author
ORCiD: https://orcid.org/0000-0002-5246-2691

10.21203/rs.3.rs-23605/v1
SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
Pennisetum glaucum, heat stress, flag leaf, RNA sequencing, SSRs, SNPs
Abstract

Background

_Pennisetum glaucum_ (L.) R. Br. is mainly grown in arid and semi-arid regions. Being naturally tolerant to various adverse conditions, it is a good biological resource for deciphering the molecular basis of abiotic stresses such as heat stress in plants but limited studies have been carried out till date to this effect. Here, we performed RNA-sequencing from the leaf of two contrasting genotypes of pearl millet (841-B and PPMI-69) subjected to heat stress (42 °C for 6 h).

Results

Over 274 million high quality reads with an average length of 150 nt were generated. Assembly was carried out using trinity, obtaining 47,310 unigenes having an average length of 1254 nucleotides, N50 length of 1853 nucleotides and GC content of 53.11%. Blastx resulted in annotation of 35,628 unigenes and functional classification showed 15,950 unigenes designated to 51 Gene Ontology terms, 13,786 unigenes allocated to 23 Clusters of Orthologous Groups and 4,255 unigenes distributed into 132 functional KEGG pathways. 12,976 simple sequence repeats were identified from 10,294 unigenes for the development of functional markers. A total of 3,05,759 SNPs were observed in the transcriptome data. Out of 2,301 differentially expressed genes, 10 potential candidates genes were selected based on log2 fold change and adjusted p-value parameters for their differential gene expression by qRT-PCR.

Conclusions

The dynamic expression changes in two genotypes of _P. glaucum_ reflect transcriptome regulation of signaling pathways in heat stress response. In order to develop genetic markers, 12,976 simple sequence repeats (SSRs) were identified. The sequencing data generated in this study shall serve as an important resource for further research in the area of crop biotechnology.

Background

Global warming has a negative impact on crop production. Temperature is a key physical parameter which affects the growth and development of plant. According to the Inter-Governmental Panel on Climate Change (IPCC), there has been an average increase of 4 °C in global atmospheric temperature since late 20th century (1). Plant undergoes a number of morphological, physiological, biochemical and molecular changes during heat stress to ensure its survival (2). These changes
include, reduction in chlorophyll content, changes in membrane fluidity and protein stability, reactive oxygen species (ROS) production, changes at transcriptome level etc. In some crops occurrence of heat stress during the flowering period leads to poor grain setting (3, 4). Increased water stress due to heat throughout the growing cycle can reduce the crop yield (5). However, the impact of temperature beyond the optimum required for growth and underlying heat tolerance mechanisms are not clearly understood in many crops.

Pearl millet [*Pennisetum glaucum* (L.) R. Br.], is widely grown in African and Indian subcontinent, since prehistoric times. The crop’s main center of diversity is known to be Sahel zone of West Africa. Pearl millet is a C4 species having diploid number 2n = 14, genome size around 1.79 Gb (6) and is mostly grown under drought-prone semi-arid and arid tropics in the regions with 200 to 800 mm of annual rainfall. Pearl millet is grown in an area of about 31 mha worldwide (http://exploreit.icrisat.org/profile/Pearl%20Millet/178) with 27.83 mtonnes production accounting for 50% of global millet production. Optimum temperature required for its growth is about 30–35 °C (7). It is well known for its tolerance to extreme environmental conditions and limited genomic resources are available as compared to other crop species.

Next-generation sequencing (NGS) based technology for analysis of transcriptome is more powerful and accurate as compared to, the Sanger based EST sequencing, suppression subtractive hybridization and hybridisation based microarrays (8, 9). Moreover, over the last decade several NGS platforms including the Illumina are becoming more affordable and efficient for transcriptome sequencing. Additionally, transcriptome sequencing has emerged as an alternative core technology for discovery and understanding of genes associated with desired traits, where full genome sequencing is not economically feasible especially in case of non-model plants. RNA-sequencing (RNA-Seq) has become a benchmark tool for whole transcriptome gene expression quantification and identification of differentially expressed genes (DEGs). It provides scope for the identification of probable candidate genes involved in abiotic and biotic stress tolerance and further for the development of molecular markers (10, 11, 12).

Enormous amounts of ESTs generated from various transcriptomic studies and other genomic
sequences are available in public databases for many model plant species. However, limited research emphasis has been given to the non-model crops including pearl millet, as evidenced by the presence of only 75,493 ESTs (https://www.ncbi.nlm.nih.gov/nuccore/?term=pearl+millet+ESTs) of this crop in GenBank (13). It is now possible to assemble transcripts without reference genome via de novo assembly using trinity (14) and/or one of several other available software tools. Recently, genome wide expression profiling in various non-model plant species growing under abiotic or biotic stresses, for various biosynthetic pathways or developmental stages has been carried out using RNA-Sequencing as a tool. These plants includes *Raphanus sativus*, *(Scrophularia ningpoensis, Camellia sinensis, Lilium, Brassica rapa, Sesamum indicum, Asparagus officinalis* and *Agrostis* (15–22).

Flag leaf acts as an immediate source for panicle development during reproductive stage in plants (23), RNA-Sequencing of flag leaf subjected to heat stress during flowering stage was carried out using Illumina sequencing platform. *De novo* assembly resulted 47,310 unigenes and further, functional annotation (gene ontology, corresponding metabolic pathways) was carried out. The aim of the present study was to unravel the gene pool responsible for conferring heat tolerance to pearl millet. Being the first transcriptome report of pearl millet in response to heat stress, the data presented here will be a primary source of information for the research on genomics and functional genomics in this orphan crop.

**Results And Discussion**

**Determination of physio-biochemical properties of P. glaucum genotypes**

The MDA (Malondialdehyde) content and MSI (Membrane Stability Index) analysis as observed in *P. glaucum* genotypes 841B and PPMI69 revealed that genotype 841B was more heat tolerant as compared to genotype PPMI69 (Supplementary Figs. 1 and 2, Additional File 1).

**Illumina Sequencing And Raw Data Pre-processing**

The whole transcriptome sequencing was performed using Paired end (PE) 2 x 150 bp library on Illumina HiSeq 2500. The sequencing run produced a total raw data of 288,876,956, details are shown in Table 1. After removal of low quality sequences, ambiguous bases and adapter sequences by Trimmomatic tool (24), a total of 274,721,009 high quality clean reads, containing 39,782,593,275
nucleotides (nts) were generated having an average length of 150 nt and GC content of 57.17%. The sequencing data has been deposited to NCBI (National Centre for Biotechnology Information) Sequence Read Archive (SRA) database under the accession number SRP151237.

| Samples     | Total raw reads | Total clean reads | Total clean nucleotides (nt) | Total clean nucleotides (nt GB) | Average length of clean reads (nt) | Q30% | N % | GC % |
|-------------|-----------------|-------------------|------------------------------|--------------------------------|----------------------------------|------|-----|------|
| 841-B ctrl  | 76413068        | 72512870          | 10486909631                  | 10.49                          | 150                              | 93.04| 2.49%| 57.80%|
| 841-B 6 h   | 71818610        | 67530958          | 9724478182                   | 9.72                           | 150                              | 92.19| 2.48%| 56.72%|
| PPMI-69 ctrl| 72236940        | 69132264          | 10043345563                  | 10.04                          | 150                              | 94.14| 0.34%| 57.91%|
| PPMI-69 6 h | 68408338        | 65544917          | 9527859899                   | 9.53                           | 150                              | 94.22| 0.34%| 56.25%|

De novo assembly of pearl millet flag leaf transcriptome

Using the Trinity program based on the de Bruijn graph algorithm, we performed de novo transcriptome assemblies using their default K-mer sizes. Generating 1,47,934 contigs having mean length of 1059 nt and N50 length of 1526 nt (Table 2). In order to reduce the assembled contig numbers, CD-HIT software, was used for grouping and estimating similarity and dissimilarity of nucleotide sequences, which resulted in the number of contigs being reduced from 147,934 to 129,893 due to the removal of redundant sequences. TGICL tools were used further, 1,29,893 contigs processed into more complete and longer 1,09,001 contigs, with N50 length of 1,649 nt. In order to retain the biologically significant contigs, EvidentialGene tools was used and these contigs were assembled in a non-redundant manner as described previously. 47,310 high quality unigenes were generated, with a total length of 59,323,119 nt, a mean length 1,254 nt, N50 length of 1,853 nt and GC content 53.11% (Table 2). Assembly statistics of P. glaucum transcriptome is shown in Table 3.

The use of assembly tools (CD-HIT, TGICL and EvidentialGene) led to the improvement of N50 values, compared to raw assembly. To find the read usage in the assembly, we aligned all the 274.721 million reads to 47,310 unigenes using Bowtie 2 software tool (25), 72.62% of reads aligned to the assembled transcripts.
Table 2
Summary of pearl millet transcriptome assembly

|                    | Contigs          | Unigenes        |
|--------------------|------------------|-----------------|
| Total number       | 147934           | 47310           |
| Total length (nt)  | 156661397        | 59323119        |
| Mean Length (nt)   | 1059             | 1254            |
| N50                | 1526             | 1853            |
| Total consensus sequences | 156661397        | 59323119        |

Table 3
Assembly statistics of P. glaucum transcriptome

| Tools used       | Trinity | CD-HIT-EST | TGICL   | Evidential gene |
|------------------|---------|------------|---------|-----------------|
| Number of transcripts | 147934  | 129893     | 109001  | 47310           |
| Total size of transcripts | 156661397 | 133060357 | 123852445 | 59323119 |
| Longest transcript | 14931   | 14931      | 23213   | 23213           |
| Shortest transcript | 301     | 301        | 301     | 301             |
| Number of transcripts > 300 nt | 147934  | 129893     | 109001  | 47310           |
| Number of transcripts > 500 nt | 99692   | 86416      | 76452   | 34992           |
| Number of transcripts > 1K nt | 52013   | 43823      | 42991   | 21055           |
| Number of transcripts > 10K nt | 19      | 14         | 15      | 11              |
| N50 transcript length | 1526    | 1449       | 1649    | 1853            |

Structural, Functional annotation and classification of P. glaucum transcriptome

The transcriptome structural annotation analysis was performed using TransDecoder tool (https://github.com/TransDecoder/TransDecoder/wiki). Out of 47,310 unigenes analysed, 29,919 (63.24%) were found to be coding sequences, in which 11,893 unigenes (25.13%) were detected with ORFs (Open Reading Frame) (Table 4). Functional annotation of all assembled unigenes were compared to the NCBI non-redundant (nr) protein database (Blastx program) (26) with a cut-off value of 1.0 E-06. A total of 35,628 unigenes (75.31% of all unigenes) were annotated against nr protein database while the remaining 11,682 (24.69%) were not annotated (Table 5). Based on E-value distribution of Blastx results, 81.34% of the unigenes showed E-value < 1.0E-45 while 18.66% of unigenes had E-value in the range of 1.0E-0.6 to 1.0E-45 (Fig. 1A). 83.35% of the aligned unigenes showed more than 80% of similarity distribution (Fig. 1B). Blast top hits analysis showed that 57.07% of the annotated sequences correspond to Setaria italica, followed by Sorghum bicolor (6.61%), Cronobacter sakazakii (4.72%), Zea mays (4.33%) and Oryza sativa japonica (2.74%) (Fig. 2). The thirty top-hit species based on nr annotation are shown in Fig. 2. Based on Gene Ontology (GO), 15,950 unigenes were designated into 3 GO classes i.e., 23 biological processes, 14 cellular components and 14 molecular functions as shown in Fig. 3A. Transcriptional sequences for cellular
process, biological regulation, establishment of localization, localization, pigmentation, response to stimulus, metabolic process, among others were significantly enriched under the biological process. Within the cellular component, unigene sequences for cell, cell part, organelle, organelle part, macromolecular complex were identified as highly enriched categories. The major proportion of unigenes was assigned to binding, catalytic activity and transporter categories under molecular function. Moreover, Clusters of Orthologous Groups (COGs) analysis showed that 13,786 unigenes (29.14% of all unigenes) were allocated to 23 COGs categories (Table 4; Fig. 3B). Of these, maximum unigenes fall under the category of unknown functions (4,353), followed by a large number of unigenes falling under the categories of signal transduction mechanisms (1,776), transcription (1,105), carbohydrate metabolism and transport (1,095) and post-translational modification, protein turnover, chaperone functions (961). Minimum unigenes were observed to fall under the categories of extracellular structures (9) and cell motility (3). The assigned function of unigenes showed GO (44.77%) and COGs (29.14%) classifications, representing a broad range of cellular transcripts in pearl millet. We used PlantTFcat tool (http://plantgrn.noble.org/PlantTFcat/) and identified 3841 unigenes associated with plant transcription factors (Fig. 4A).

| Total unigene | 47310 |
|---------------|-------|
| Unigene with ORFs | 29919 |
| Unigene complete ORFs | 11893 |
| Unigene without ORFs | 5498 |

Identification of heat responsive genes involved in biological pathways during flowering

To identify potential heat responsive genes and understand their role in various biological pathways, KEGG (Kyoto Encyclopedia of Genes and Genomes) analysis was performed with a cut-off E-value of 1.0E-05. In total, 4,255 unigenes (11.94% of total unigenes) were categorised into 132 KEGG pathways (Table 5; Fig. 4A). The most represented pathways were the ones related to “biosynthesis of antibiotics” (10.86%), “purine metabolism” (6.46%), “starch and sucrose metabolism” (3.45%), “pyrimidine metabolism” (3.2%), “phenylpropanoid biosynthesis” (2.70%) and “glycolysis/gluconeogenesis” (2.63%). A total of 147 transcripts of starch and sucrose metabolism (3.45%) and 112 transcripts of “glycolysis/gluconeogenesis” (2.63%) were identified. Our study will
provide, better understanding of molecular mechanisms that are prevailing during flowering stage of pearl millet under heat stress. Interestingly, this analysis may help in specifying pathways related to synthesis and turnover of compounds, which have favourable effects in grain filling and yield.

Table 5
Summary of annotations against public available databases

| Database | Number of Unigenes | Percentage (%) |
|----------|--------------------|----------------|
| Total Unigene | 47310 | 100 |
| NR | 35628 | 75.31 |
| GO | 15950 | 44.77 |
| COG | 13786 | 29.14 |
| KEGG | 4263 | 11.97 |
| TF | 3841 | 8.119 |

Identification of simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) from P. glaucum transcriptome

Assembled transcriptome of P. glaucum was used for identification of SSRs. Based on criteria with a minimum of (5–10) repetitions of mono to hexa-nucleotide motifs, MISA software was used to search SSR markers in all unigenes (27). A total of 12,976 SSRs from 10,294 unigenes were detected, of which 2,116 unigenes had more than one SSRs (Table 6). Among all the identified SSRs, 50.88% fall under tri-nucleotide repeats, followed by mono-nucleotides (31.88%) and di-nucleotides (13.7%) but tetra-nucleotide, penta-nucleotide and hexa-nucleotide were represented only as a small fraction (Fig. 4B).

Table 6
Statistics related to SSRs obtained

| Statistics related to SSRs obtained | Value |
|-------------------------------------|-------|
| Total number of unigenes examined | 47,310 |
| Total size of examined sequences (bp) | 5,93,23,119 |
| Total number of identified SSRs | 12,976 |
| Number of unigenes containing SSRs | 10,294 |
| Number of unigenes containing more than 1 SSR | 2,116 |
| Number of SSRs present in compound formation | 986 |

SSRs are simple motifs of nucleotides (1–10 nucleotides), which may occur as tandem or interspersed repeats and are abundant within genome of prokaryotes and eukaryotes (28). Genetic variability for heat stress tolerance in P. glaucum is still unexplored. Therefore mining the SSR markers from P. glaucum would be utilized by breeders to develop heat stress tolerant crops. Several studies show that SSRs are not distributed randomly along the genome. For example in case of Arabidopsis thaliana and rice (29) and in Gossypium raimondii (30). It has been reported that occurrence of GC-rich trinucleotides SSRs were frequent in exon regions, whereas distribution of AT-rich trinucleotides SSRs were found throughout the genome (coding sequences, untranslated regions, introns and intergenic
SSRs are codominant, multi-allele genetic markers that are highly reproducible and transferable among related species. As a result, it has been the most widely used marker for genotyping and other various breeding purposes. Identification of new SSRs will provide the necessary impetus to the research community interested in genotyping, genetic mapping and genetic diversity studies in various *Pennisetum* species. A total of 3,05,759 SNPs were observed in the *P. glaucum* transcriptome data. Analysis of SNPs showed that 64.76% (5,81,800/8,98,460) of the nucleotide changes were transitions, while 35.24% (3,16,660/8,98,460) were transversions. The observed transition:transversion (Ts/Tv) ratio is 1.84. The ratio of transition to transversion is expected to be more than one due to mutational processes in plant genome. The ratio is lower than estimates from maize (2.5) and *Arabidopsis* (2.4). In our studies, an average of 1 variant was observed on every 5,116 bases. The identified SNPs may be utilized as a genomic resource for *P. glaucum* improvement by mining alleles of genes and genome assisted breeding for future genome-wide association studies.

**Differential gene expression analysis and validation of heat stress responsive genes**

FPKM (Fragments per kilobase per million fragments) unit was used to calculate the expression levels of genes in *P. glaucum* flag leaves transcriptome. Each sample reads were aligned separately to 47,310 unigenes. 75.02% reads got aligned (out of 72,512,870) for genotype 841-B control, 68.25% reads got aligned (out of 67,530,958) for genotype 841-B 6 h, 72.84% reads got aligned (out of 69,132,264) for genotype PPMI-69 control and 74.21% reads got aligned (out of 65,544,917) for the genotype PPMI-69 6 h. After filtering based on adjusted *p*-value, less than 0.005 and (-1 < log2 fold change > 1) there were 850 differentially expressed genes identified between 841-B control and 841-B 6 h, of which 494 genes were up-regulated and 356 genes were down-regulated. Among 1,934 differentially expressed genes identified between PPMI-69 control and PPMI-69 6 h, 964 genes were up-regulated and 970 genes were down-regulated (Fig. 5).

The hierarchical cluster analysis of two different genotypes with heat stress treatment shows that the differential gene expression pattern of the transcriptome readily differentiates *P. glaucum* genotype PPMI-69 6 h from others (841-B control, 841-B 6 h and PPMI-69 control) (Fig. 5), possibly indicating
the variation in differential gene expression occurring between genotypes in response to heat stress. The maximum number of differentially expressed genes (1934 genes) was observed between PPMI-69 6 h and PPMI-69 control (Table 7). Differential gene expression patterns across the treatments and comparison of differential expression patterns between the two different genotypes, were analysed by cluster analysis with hierarchical clustering method (Fig. 5). These differentially expressed genes between samples 841-B 6 h and PPMI-69 6 h are grouped into two clusters, representing that genes involved in thermotolerance have different level of differential gene expression. This suggests that 841-B 6 h has better ability to maintain homeostasis during heat stress as compared to PPMI-69 6 h. A comparative analysis between differential expressed genes in 841-B and PPMI-69 genotypes was performed to identify common differentially expressed genes in both the genotypes under heat stress and those that are unique to each genotype of the *P. glaucum*. A total of 2,301 genes were differentially expressed, 20.99% (483 genes) of which were shared common by both genotypes. 15.95% (367 genes) of differentially expressed genes were unique to 841-B genotypes and 63.06% (1451 genes) of differentially expressed genes were unique to PPMI-69 (Fig. 5). However, the differential expression of genes between the two genotypes cannot be attributed exclusively to the treatments alone, as the two genotypes have inherently different levels heat stress tolerance and therefore could have different gene expression profile. The DEGs were also visualized by volcano plots (Fig. 6)

| Comparisons                        | Up-Regulated | Down-Regulated | Total |
|------------------------------------|--------------|----------------|-------|
| 841-B Control vs PPMI _69 Control | 323          | 355            | 678   |
| 841-B 6 h vs 841-B Control         | 494          | 356            | 850   |
| 841-B 6 h vs PPMI-69 Control       | 813          | 508            | 1321  |
| 841-B 6 h vs PPMI-69 6 h           | 748          | 422            | 1170  |
| PPMI-69 6 h vs 841-B Control       | 970          | 936            | 1906  |
| PPMI-69 6 h vs PPMI-69 Control     | 964          | 970            | 1934  |

In order to investigate the temporal expression patterns by qRT-PCR analysis, 10 target genes were selected based on the fold changes and adjusted p-values. The selected genes displayed varying patterns in response to different durations of heat stress (Fig. 7). Validation of selected 10 genes (7 known and 3 uncharacterized) by qRT-PCR deciphered their significant role in heat stress
management in *P. glaucum*. These genes might play an essential role in amelioration of heat stress in *P. glaucum*.

**DnaJ**

Heat stress induces changes in protein conformation and increase improper folding of native proteins. As a result, accumulation of many heat shock proteins is triggered to counter balance the negative effect of heat stress in plants (35). Among many heat shock proteins, DnaJ or Hsp40 is known to play significant role in plant development, signal transduction and abiotic stresses response, either by itself or in association with Hsp70 partner (36, 37). *PgDnaJ* expression show significantly upregulation, about 15 folds in 841-B in response to 30 min heat stress, and 16 folds in PPMI-69. However, this result indicates the important role of *DnaJ* in maintaining cellular protein homeostasis during heat stress (42 °C).

**GST**

Plasma membrane acts as the primary sensor of the cell when the cell is subjected to heat stress (38). Consequently, membrane properties undergo a number of changes in its composition, ion concentration and ion channels in response to heat stress (39, 40). Peroxidation of lipid membrane is one of the most important change that occurs in the cell in response to various stresses (41). Malondialdehyde (MDA) content is a direct indicator of lipid peroxidation. These by-product of lipid peroxidation such as electrophiles or xenobiotics are detoxified by reactive oxygen species (ROS)-scavenging enzymes such glutathione S-transferases (GST), superoxide dismutases, catalases and ascorbate peroxidases (APX). The expression of *PgGST* in our study was significantly upregulated (7 folds) in response to heat stress in 841-B but significantly downregulated in PPMI-69 indicating its role in the heat stress tolerance pathway. This result correlates with malondialdehyde (MDA) content (Fig. S2), as MDA content was observed to be high in PPMI-69 as compared with 841-B.

**NAC67**

NAC transcription factors (TFs), (NAM, No apical meristem; ATAF, Arabidopsis transcription activation factor and CUC, Cup-shaped cotyledon) play important role in plant growth and development and in regulating response to abiotic or biotic stresses (42, 43). Among these NAC genes, *NAC67* has a role
in imparting tolerance to multiple abiotic stress such as drought, salt and cold stresses. \textit{NAC67} has been reported to be involved in conferring tolerance to abiotic stresses in rice (44). However, so far there has been no report indicating its role in heat stress tolerance. In this study, \textit{PgNAC67} expression was found to be significantly upregulated (27 folds) in 841-B in response to 30 min and 6 folds in PPMI-69 in response to 6 h of heat stress. It shows that \textit{PgNAC67} plays a role in heat stress response in \textit{P. glaucum}.

**TIL**

Temperature-induced lipocalins (TILs), a plasma membrane protein have an important role in basal and acquired thermotolerance in plant. TILs alleviate the heat induced lipid peroxidation in membrane. \textit{PgTIL} expression was observed to be significantly upregulated (9 folds) in response to 6 h of heat stress in 841-B but significantly downregulated in PPMI-69. This might the reason why 841-B is able to maintain low malondialdehyde (MDA) content (Fig. S2) as compared to PPMI-69.

**EXP**

\textit{PgEXP} expression shows 5 folds significant upregulation in 841-B in response to 6 h of heat stress. Association of expansin genes and heat stress tolerance in some plants has been reported (45, 46). Over-expression of the \textit{EXP1} gene exhibits low electrolyte leakage, decrease in membrane lipid peroxidation but higher chlorophyll content, net photosynthetic rate, relative water content, activity of antioxidant enzyme in transgenic plants (47).

**Hd1**

Heading 1 (\textit{Hd1}) and early heading 1 (\textit{Ehd1}) are mainly known for the regulation of flower development and flowering, leading to either induction or suppression corresponding to the particular photoperiod (48). Environmental factors such as day length, abiotic and biotic stresses regulate the expression of these genes. Earlier studies show that inhibition of early heading 1 (\textit{Ehd1}), in response to drought stress delays flowering in rice (49). In our studies, \textit{PgHd1} expression shows significant downregulation in both genotypes in response to 30 min and 6 h heat stress.

**LTP**

\textit{PgLTP} expression shows significant (19 folds) upregulation in response to 6 h of heat stress in PPMI-69
but significantly downregulated in 841-B indicating involvement of high activity with regard to transfer of lipid molecules in cell. This shows active regulation of membrane fluidity in PPMI-69 in response to heat stress. Lipid Transfer Protein (LTP) are reported to be involved in growth and development, response to abiotic and biotic stresses but their functions remain unclear (50-53). Moreover, LTPs has ability to facilitate the transfer of phospholipids between membranes in vitro (54).

**Uncharacterized genes**

Among differentially expressed contigs, uncharacterised ORFs share maximum proportion in our study. Uncharacterised genes with a predicted protein domain associated with zinc fingers (ZnF), ribonuclease (RN) and chaperone were validated for their expression during heat stress. Some of the uncharacterised genes were expressed uniquely either in 841-B or PPMI-69. For example, uncharacterised gene *PgUCP1*, with predicted zinc fingers (ZnF) domain is significantly upregulated, (34 folds) in 841-B and (10 folds) in PPMI-69 in response to heat stress. ZnF is known for involvement in multiple stress response, but their exact molecular mechanism and their interaction is yet to be deciphered (55, 56). On contrary, *PgUCP2* uncharacterised gene with predicted ribonuclease (RN) domain is significantly downregulated in 841-B but significantly upregulated (29 folds) in PPMI-69 in response to heat stress. It has been reported that loss of ribonuclease function in *Arabidopsis* enhances heat stress tolerance (57). However, uncharacterised gene *PgUCP3*, with predicted chaperone domain show considerable upregulation (4 folds) in 841-B and is downregulated in PPMI-69 in response to heat stress. These uncharacterised genes could likely represent important genes involved in imparting variation in thermotolerance among different genotypes. Further, detailed investigation of these uncharacterised genes is required for understanding the role in response to stress.

**Conclusion**

This study investigated the transcriptome profile of pearl millet flag leaves in response to heat stress. In this study, high quality 47,310 unigenes were generated and annotated. This EST data will provide the foundation for research on gene expression, genomics, and functional genomics in pearl millet improvement program. Further, the SSRs obtained in this study shall facilitate the research on
genotyping, and diversity studies of this important crop. The candidate genes whose expression patterns were validated by qRT-PCR shall serve as important resource for their effective utilization in the development of transgenic crops tolerant to heat stress.

Methods

**Plant materials, heat stress treatment, RNA isolation and cDNA library construction**

Seeds of two contrasting genotypes 841-B and PPMI-6925 were collected from Division of Genetics, Indian Agricultural Research Institute (ICAR-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 (temperature 32±2°C, relative humidity 70-80%, under day length of 12 h). Ten seeds per genotype were grown with one seedling per pot. At flowering stage (55 days after sowing), heat stress was conducted in a growth chamber at a temperature of 42°C, relative humidity of 70-80% and normal light conditions for different time intervals (30 min and 6 h). Plants grown at 32-34°C under normal light conditions in glasshouse served as control. Different plant samples used in the study were given independent Identity Number (Supplementary Table 1, Additional File 1). For RNA extraction, one flag leaf per plant was collected from each plant sample respectively in biological triplicates and immediately frozen in liquid nitrogen before storing at -80°C. RNA isolation was carried out in three biological replicates using TRIzol reagent (ThermoFisher Scientific, USA) and purified using NucleoTrap mRNA mini kit (Macherey-Nagel, Germany). DNA contamination was removed using TURBO DNase (Ambion, USA) according to the manufacturer’s protocol. The RNA quality was assessed using the 2100 Bioanalyzer (Agilent Technology, USA). The isolated total RNA from control and heat stressed (30 min and 6h pooled together) were used for RNA library construction. The isolated total RNA from control and heat stressed (30 min and 6h) of both genotypes (841-B and PPMI-69) (three biological replicates of each) were pooled together as four independent RNA samples. One µg of the total pooled RNA from each sample was used to purify poly-A containing mRNA molecules using Oligotex mRNA mini kit (Qiagen, Germany) as described by the manufacturer’s protocol. Four independent RNA-seq libraries were constructed using TruSeq® Stranded mRNA Library Prep Kit (Illumina, USA) according to the manufacturer’s instructions.
The RNA libraries thus constructed were sequenced using Illumina Hiseq platform.

**Determination of physio-biochemical characteristics of plants**

Malondialdehyde (MDA) content was determined according to Heath and Packer (58). 0.5 g of leaf tissue were taken, and homogenised in 10% trichloroacetic acid (TCA) and 0.65% thiobarbituric acid (TBA). These are incubated at 95°C for 30 min, allowed to cool down to room temperature, and centrifuged at 10,000 g for 10 min. The absorbance was measured using Shimadzu UV-vis Spectrophotometer (UV-1800) at 532 nm and 600 nm. The MDA equivalent was calculated in nmole/g fresh weight as MDA= [(A532-A600)]/155000]*100. Membrane stability index (MSI) was measured as per Blum and Ebercon (59). Leaf samples were washed with double distilled water (DDW) to remove surface contamination and 10 leaf discs were taken in sealed vials containing 10 ml of DDW separately, followed by incubation at 4°C for 24 h, the electrical conductivity (EC1) was recorded. For electrical conductivity (EC2), the samples were autoclaved at 120°C for 20 min, and allowed to cool down to room temperature. The membrane stability index was calculated as per the equation: MSI (%)= 1-(EC1/EC2)* 100.

**De novo assembly of flag leaves transcriptome**

The next generation sequencing run for whole transcriptome was performed using Paired end (PE) 2x150 bp library on Illumina HiSeq 2500. Using FastQC tools (60), quality check was performed. Trimmomatic was used for pre-processing of the raw reads to eliminate adapter sequences and poor quality reads. Trinity program was used for de novo assembly with default parameters (14, 61). Cluster Database at High Identity with Tolerance (CD-HIT) program (62) was run to remove the similar short sequences based on 90% alignment coverage to longer sequence and produces a set of 'nonredundant' (nr) representative sequences and eliminating short redundant sequences. Sequences were clustered using TGICL tools (63) with default parameters to produce longer, more complete consensus sequences. Gene construction was carried out using EvidentialGene tools with default parameters, to retain the biologically significant transcripts.

**Annotation of transcriptome and identification of SSRs and SNPs**
The transcriptome structural annotation was performed using TransDecoder tools. The functional annotation was performed using BLAST+ tools, with BLASTx using a translated nucleotide query (unigenes). Gene Ontology mapping was performed using Blast2GO (64), to specify all the annotated unigenes to various categories such as biological processes, molecular functions and cellular components. Pathway mapping of unigenes was performed using KEGG database (65). The unigene sequences were aligned to the Clusters of Orthologous Groups (COGs database) (66) to predict and classify proteins. PlantTFcat online tool (http://plantgrn.noble.org/PlantTFcat/) was used to identify transcription factor in the generated data. SSRs were identified using MiraSATellite identification tool MISA. The Microsatellite search module (MISA) is available online for public (http://pgrc.ipk-gatersleben.de/misa/). The SNPs were identified by using GATK best practice pipeline Version 4.1.2.0 (https://software.broadinstitute.org/gatk/best-practices/), the cleaned reads were mapped against the Transfuse.fasta file using BWA aligner (http://bio-bwa.sourceforge.net/). The alignment was performed in default mode. Picard tool was used co-ordinate, sort and remove duplicates from aligned bam files. The GATK tool was used for processing the alignments and variant calling. SplitNCigarReads and HaplotypeCaller from GATK tools were used for reassigning mapping qualities and variant calling.

**Expression analysis**

Fragments per kilobase of transcript per million mapped reads (FPKM) unit was used to calculate the expression level of unigenes. Read count for each unigenes were calculated and then converted to FPKM using formula (Read count * 10^9)/(Sum of read count * Length). Differential gene expression was determined using DEGSeq (67), a R package. The significant differentially expressed unigenes were filtered based on adjusted p-value < 0.005 and log ratio > 1 and -1 between the samples. Heatmaps of the significant genes were generated with the heatmap package (68), in R package. Using R package pvclust (69), hierarchical clustering was performed with 1000 bootstrap replications.

**Quantitative RT-PCR**

Total RNA was extracted from the treated and control flag leaves as described above to study the differential expression patterns under heat stress (42°C for 30 min and 6 h) of the few selected genes. A total of 10 genes (PgDnaJ, PgGST, PgNAC67, PgTIL, PgEXP, PgHd1, PgLTP, PgUCP1, PgUCP2,
PgUCP3) involved in heat stress response were selected for differential expression analysis, from the generated pearl millet transcriptome data based on log2 fold change ≥ 2 (for upregulated transcripts) and adjusted p-values. cDNA was synthesised for selected 10 DEGs with samples of flag leaves originally used for RNA-seq, using SuperScript® III First-Strand Synthesis System (Invitrogen, USA). qRT-PCR was performed using LightCycler® 480 System (Roche, Switzerland) and KAPA SYBR® FAST qPCR Kits (Kapa Biosystems, USA) was used as reaction components. Gene specific primers were designed using PrimerQuest Tool (Integrated DNA Technologies (IDT), USA) (Table 8). PCR programme was set as: 94°C for 5 min and 40 cycles of 94°C for 10 sec, 60°C for 10 sec, and 72°C for 10 sec. PgActin was used as internal control to normalize all the data. 2−ΔΔCt method was used to calculate relative fold change expression (70). Significance level was calculated using two-tailed unpaired t-test.

| Gene   | Sequence (5’ to 3’) |  |
|--------|---------------------|---|
| PgDnaJ | F                   | AATCCATGTTGGACGGGATG |
|        | R                   | GCTCTTGCGCTGATGTATGA |
| PgGST  | F                   | GACTACGAGTACAAGTCGGTAAAT |
|        | R                   | GAGCGGATGTGCAGGATATT |
| PgNAC67| F                   | TCTCGATTCAGGGAACCAAAATAA |
|        | R                   | CTACTACCAGAACCAGAAACAA |
| PgTIL  | F                   | AATACGCGCACCACCTAC |
|        | R                   | GAGGTTAGAAGGCTGAGTAATG |
| PgEXP  | F                   | GAGGAGGAGAGGGAAGAGG |
|        | R                   | GTAGCCAGATGCTAGAAGT |
| PgHd1  | F                   | ACTTGTGCACTTGAAAGATA |
|        | R                   | ATGCCCTGTCCACAAACAA |
| PgLTP  | F                   | CTACCTACGTAACCTGAGT |
|        | R                   | ACGTGCCATACACATAGA |
| PgUCP1 | F                   | TCCAGTGACAGGAGTCAAG |
|        | R                   | CGGAGGCGCCTGGTTTT |
| PgUCP2 | F                   | GGAGATCCTTCTTACCCCTT |
|        | R                   | GATCCGAGCACACACATA |
| PgUCP3 | F                   | TACGAGATCAACCACAAAC |
|        | R                   | CAGGAGCGTACCAGGAAAG |
| PgActin| F                   | CCCAAGGCCAATAGAGAGA |
|        | R                   | CACTGCGTACAAGGAAAGA |

Table 8
List of primers used for qRT-PCR analysis

Abbreviations
nt: nucleotides, h: hours, BLAST: Basic Local Alignment Search Tool, SNPs: Simple sequence repeat, mm: milli metre, ESTs: Expressed Sequence Tags

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable
**Availability of data and materials:** The transcriptome data is available on NCBI SRA database

**Competing interests:** The authors declare that they have no competing interests

**Funding:** Support by ICAR (Indian Council of Agricultural Research) project on National Innovations in Climate Resilient Agriculture (NICRA) for providing necessary equipments and facilities to carryout qRT-PCR analysis and bioinformatic studies is duly acknowledged.

**Authors contributions:** AM: Performed the experiments, Data curation, writing-original draft preparation, HV: writing-review & editing, Visualization; SAL: Formal analysis, writing-review & editing; KG: writing-review & editing, SN, SVAM, MPS, SPS and MD: Formal analysis, JCP: Conceptualization, Resources, Project administration, writing-review & editing, Funding acquisition.

**Acknowledgements:** Authors are grateful to the Director, ICAR-National Institute for Plant Biotechnology, New Delhi, India for providing necessary facilities. Director, ICAR-IARI is thanked for facilitating the use of National Phytotron Facility. Award of fellowship to the first author and second author by and Department of Biotechnology, Government of India and by PG School-IARI respectively are duly acknowledged. Thanks are also due to Dr. Tara Satyavathi Chellapilla, Division of Genetics, IARI, New Delhi for providing the seed material.

**References**

1. Porter JR, Xie L, Challinor AJ, Cochrane K, Howden SM, Iqbal MM, Lobell DB, et al. Food security and food production systems. In: Climate Change: Impacts, Adaptation, and Vulnerability. Part A: Global and sectoral aspects. Contribution of working Group II to the fifth assessment report of the Intergovernmental panel on climate change. 2014;485-533.

2. Wahid A, Gelani S, Ashraf M, Foolad MR. Heat tolerance in plants: An overview. Environ Exp Bot. 2007;61:199–223.

3. Iqbal MM, Goheer MA, Khan MA. Climate-change aspersions on food security of Pakistan. Sci Vis. 2009;15:15-24.

4. Moriondo M, Giannakopoulos C, Bindi M. Climate change impact assessment: The role
of climate extremes in crop yield simulation. Clim Change. 2011;104:679–701.

5. Lobell DB, Hammer GL, McLean G, Messina CD. The critical role of extreme heat for maize production in the United States. Nat Clim Change. 2013;3:497–501.

6. Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, et al. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat Biotechnol. 2017;35:10.

7. Maibam A, Padaria JC. Pearl millet: A Genetic resource for abiotic tolerant transgenics. Biotech Today. 2015;5:21–4.

8. Agarwal A, et al. Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. BMC Genom. 2010;11:383.

9. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. Comparison with gene expression arrays RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008. doi:10.1101/gr.079558.108.

10. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63.

11. Strickler SR, Bombarely A, Mueller LA. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. Am J Bot. 2012;99:257–66.

12. Garber M, Grabherr MG, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods. 2011;8:469–77.

13. James D, et al. Development and characterization of a high temperature stress responsive subtractive cDNA library in pearl millet Pennisetum glaucum (L.) R.Br. Indian J Exp Biol. 2015;53:543–50.

14. Haas BJ, et al. De novo transcript sequence reconstruction from RNA-Seq: reference
generation and analysis with Trinity. Nat Protoc. 2014;8:1-43.

15. Nie S, et al. De novo transcriptome analysis in radish (Raphanus sativus L.) and identification of critical genes involved in bolting and flowering. BMC Genomics. 2016;17:389.

16. Pan Y, Chen DX, Song XH, Zhang X, Li LY. Transcriptome characterization for Scrophularia ningpoensis based on high-throughput sequencing technology and related genes for synthesis of terpenoid compounds. China J Chinese Mater Medica. 2017;42:2460-6.

17. Guo F, Guo Y, Wang P, Wang Y, Ni D. Transcriptional profiling of catechins biosynthesis genes during tea plant leaf development. Planta. 2017. doi:10.1007/s00425-017-2760-2.

18. Hu Z, et al. Transcriptome sequencing analysis reveals a difference in monoterpane biosynthesis between scented Lilium ‘Siberia’ and unscented Lilium ‘Novano’. Front Plant Sci. 2017;8:1351.

19. Greenham K, et al. Temporal network analysis identifies early physiological and transcriptomic indicators of mild drought in Brassica rapa. eLife. 2017;6.

20. Dossa K, et al. Transcriptomic, biochemical and physio-anatomical investigations shed more light on responses to drought stress in two contrasting sesame genotypes. Sci Rep. 2017;7:8755.

21. Li SF, et al. Comparative transcriptome analysis reveals differentially expressed genes associated with sex expression in garden asparagus (Asparagus officinalis). BMC Plant Biol. 2017;17:143.

22. Ma Y, Shukla V, Merewitz EB. Transcriptome analysis of creeping bentgrass exposed to drought stress and polyamine treatment. PLoS One. 2017;12:e0175848.

23. Wang X, et al. Pre-anthesis high-temperature acclimation alleviates damage to the
flag leaf caused by post-anthesis heat stress in wheat. J Plant Physiol. 2011;168:585–93.

24. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

27. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet. 2003;106:411–22.

28. Vieira MLC, Santini L, Diniz AL, Munhoz CF. Microsatellite markers: What they mean and why they are so useful. Genet Mol Biol. 2016;39:312–28.

29. Lawson MJ, Zhang L. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biol. 2006;7:R14.

30. Zou C, Lu C, Zhang Y, Song G. Distribution and characterization of simple sequence repeats in *Gossypium raimondii* genome. Bioinformation. 2012;8:801–6.

31. Temnykh S, et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 2001;11:1441–52.

32. Mason AS. in Methods in molecular biology. (Clifton NJ). 2015;1245:77–89.

33. Senthilvel S, et al. Development and mapping of Simple Sequence Repeat markers for pearl millet from data mining of Expressed Sequence Tags. BMC Plant Biol. 2008;8:119.

34. Nigombam S, Mahajan R, Vishvakarma H, Mithra A, Pal SS, Padaria JC. Microsatellite
markers associated with heat stress tolerance in *Pennisetum glaucum* (L.) R. Br. Res. J. Biotech. 2019 (Accepted).

35. Schlesinger MJ. Heat shock proteins. J Biol Chem. 1990;265:12111-4.

36. Yang Y, et al. The Arabidopsis chaperone J3 regulates the plasma membrane H+-ATPase through interaction with the PKS5 kinase. Plant Cell. 2010;22:1313-32.

37. Shen L, Kang YGG, Liu L, Yu H. The J-Domain Protein J3 mediates the integration of flowering signals in *Arabidopsis*. Plant Cell. 2011;23:499-514.

38. Murata N, Los DA. Membrane Fluidity and Temperature Perception. Plant Physiol. 1997;115:875-9.

39. Saidi Y, et al. Membrane lipid composition affects plant heat sensing and modulates *Ca(2+)*-dependent heat shock response. Plant Signal Behav. 2010;5:1530-3.

40. Prasad TK. Mechanisms of chilling-induced oxidative stress injury and tolerance in developing maize seedlings: changes in antioxidant system, oxidation of proteins and lipids, and protease activities. Plant J. 1996;10:1017-26.

41. Thompson JE, Froese CD, Madey E, Smith MD, Hong Y. Lipid metabolism during plant senescence. Prog Lipid Res. 1998;37:119-41.

42. Souer E, van Houwelingen A, Kloos D, Mol J, Koes R. The No Apical Meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. Cell. 1996;85:159-70.

43. Nuruzzaman M, Sharoni AM, Kikuchi S. Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. Front Microbiol. 2013;4:248.

44. Rahman H, et al. Over-expression of a NAC 67 transcription factor from finger millet (*Eleusine coracana* L.) confers tolerance against salinity and drought stress in rice. BMC Biotechnol. 2016;16:35.
45. Xu J, Tian J, Belanger FC, Huang B. Identification and characterization of an expansin gene AsEXP1 associated with heat tolerance in C3 Agrostis grass species. J Exp Bot. 2007;58:3789–96.

46. Kuluev B, et al. Expression profiles and hormonal regulation of tobacco expansin genes and their involvement in abiotic stress response. J Plant Physiol. 2016;206:1–12.

47. Xu Q, Xu X, Shi Y, Xu J, Huang B. Transgenic tobacco plants overexpressing a grass PpEXP1 gene exhibit enhanced tolerance to heat stress. PLoS One. 2014;9:e100792.

48. Endo-Higashi N, Izawa T. Flowering time genes Heading date 1 and Early heading date 1 together control panicle development in rice. Plant Cell Physiol. 2011;52:1083–94.

49. Cho L, Yoon J, An G. The control of flowering time by environmental factors. Plant J. 2017;172:708–19.

50. Debono A, et al. Arabidopsis LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface. Plant Cell. 2009;21:1230–8.

51. Pan Y, et al. A non-specific Setaria italica Lipid Transfer Protein gene plays a critical role under abiotic stress. Front Plant Sci. 2016;7:1752.

52. Guo C, Ge X, Ma H. The rice OsDIL gene plays a role in drought tolerance at vegetative and reproductive stages. Plant Mol Biol. 2013;82:239–53.

53. Yu K, et al. A feedback regulatory loop between G3P and Lipid Transfer Proteins DIR1 and AZI1 mediates azelaic-acid-induced systemic immunity. Cell Rep. 2013;3:1266–78.

54. Kader JC. Lipid-transfer proteins in plants. Annu Rev Plant Physiol Plant Mol Biol. 1996;47:627–54.
55. Bogamuwa S, Jang JC. Plant Tandem CCCH Zinc finger proteins Interact with ABA, drought, and stress response regulators in processing-bodies and stress granules. PLoS One. 2016;11:e0151574.

56. Vij S, Tyagi AK. A20/AN1 zinc-finger domain-containing proteins in plants and animals represent common elements in stress response. Funct Integr Genomics. 2008;8:301-7.

57. Nguyen AH, et al. Loss of Arabidopsis 5′-3′ exoribonuclease AtXRN4 function enhances heat stress tolerance of plants subjected to severe heat stress. Plant Cell Physiol. 2015;56:1762-72.

58. Heath RL, Packer L. Photoperoxidation in isolated chloroplasts. I. Kinetics and stoichiometry of fatty acid peroxidation. Arch Biochem Biophys. 1968;125:189-98.

59. Blum A, Ebercon A. Cell membrane stability as a measure of drought and heat tolerance in wheat. Crop Sci. 1981;21:43-7.

60. Andrews S. FastQC: A quality control tool for high throughput sequence data. 2014. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

61. Grabherr MG. et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644-52.

62. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658-9.

63. Pertea G, et al. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics. 2003;19:651-2.

64. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics. 2008. doi:10.1155/2008/619832.

65. Ogata H, et al. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999;27:29-34.
66. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28:33–6.

67. Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 2010;26:136–8.

68. Perry M. Heatmaps. Flexible heatmaps for functional genomics and sequence features. R package version 1.4.0. 2018.

69. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;22:1540–2.

70. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 2001;25:402–8.

Supplemental File

Supplementary Table 1. Identification number given to samples used in the study.

Supplementary Fig. 1. Membrane stability Index (%) (MSI) under control and heat stress condition imposed at flowering stage in two contrasting genotypes (841B and PPMI69). Two tailed unpaired t-test was used to calculate $p$ value, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Supplementary Fig. 2. Malondialdehyde content (nmole/g fresh weight) (MDA) under control and heat stress condition imposed at flowering stage in two contrasting genotypes (841B and PPMI69). Two tailed unpaired t-test was used to calculate $p$ value, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Figures

Figure 1

Characteristics of sequence similarity of unigenes against NR database (a) E-value distribution of BLAST hits for each unigenes, (b) Similarity distribution of BLAST hits of each unigene
Figure 2

Distribution of the top BLAST hits in different species
Gene annotations (A) GO (Gene Ontology) classification of the transcriptome, (B) COGs (Clusters of Orthologous Groups) classification
Figure 4

(a) Summary of annotation of all unigenes, (b) Distribution of different SSRs
Figure 5

(a) Heatmaps of the significant differentially expressed genes with hierarchical clustering, (b)-Venn diagram of differential expressed genes under heat stress conditions in 841-B and PPMI-69 genotypes
log2 Fold change

Figure 6
Volcano plots displaying differential expressed genes, (a) In genotype 841-B, (b) In genotype PPMI-69. Red color represents upregulated genes, green color indicates downregulated genes while non significant genes are shown as black dots

Figure 7
Real time PCR validation of 10 target genes in A) 841-B and B) PPMI-69 genotypes. Two tailed unpaired t-test was used to calculate p value, *p < 0.05, **p < 0.01, ***p < 0.001

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
Supplementaryinformation.pdf