

RESEARCH ARTICLE

Open Access



# Transcriptome analysis of heat stress and drought stress in pearl millet based on Pacbio full-length transcriptome sequencing

Min Sun<sup>1†</sup>, Dejun Huang<sup>2†</sup>, Ailing Zhang<sup>1†</sup>, Imran Khan<sup>1</sup>, Haidong Yan<sup>3</sup>, Xiaoshan Wang<sup>1</sup>, Xinquan Zhang<sup>1</sup>, Jian Zhang<sup>2\*</sup> and Linkai Huang<sup>1\*</sup>

## Abstract

**Background:** Heat and drought are serious threats for crop growth and development. As the sixth largest cereal crop in the world, pearl millet can not only be used for food and forage but also as a source of bioenergy. Pearl millet is highly tolerant to heat and drought. Given this, it is considered an ideal crop to study plant stress tolerance and can be used to identify heat-resistant genes.

**Results:** In this study, we used Pacbio sequencing data as a reference sequence to analyze the Illumina data of pearl millet that had been subjected to heat and drought stress for 48 h. By summarizing previous studies, we found 26,299 new genes and 63,090 new transcripts, and the number of gene annotations increased by 20.18%. We identified 2792 transcription factors and 1223 transcriptional regulators. There were 318 TFs and 149 TRs differentially expressed under heat stress, and 315 TFs and 128 TRs were differentially expressed under drought stress. We used RNA sequencing to identify 6920 genes and 6484 genes differentially expressed under heat stress and drought stress, respectively.

**Conclusions:** Through Pacbio sequencing, we have identified more new genes and new transcripts. On the other hand, comparing the differentially expressed genes under heat tolerance with the DEGs under drought stress, we found that even in the same pathway, pearl millet responds with a different protein.

**Keywords:** Pearl millet, Pacbio sequencing, Illumina sequencing, Heat stress, Drought stress

## Background

Currently, dramatic climate change, including severe heat and drought, are threatening agricultural production [1]. According to the report by the Intergovernmental Panel on Climate Change (IPCC), the global temperature is

expected to rise 0.2 °C per decade, and by 2100, it will be 1.8 to 4.0 °C higher than the current level. One previous study revealed that every 1 °C increase in the global average temperature is reducing the yield of major agricultural crops by the following amounts: wheat 6%, rice 3.2%, corn 7.4%, and soybean 3.1% [2]. The average precipitation in many mid-latitude and subtropical arid regions is expected to decrease and severely affect the crop yield. A report by the United Nations Environment Program showed that the crop yields in semi-arid and arid regions were

\* Correspondence: [zjturf@163.com](mailto:zjturf@163.com); [huanglinkai@sicau.edu.cn](mailto:huanglinkai@sicau.edu.cn)

<sup>†</sup>Min Sun, Dejun Huang and Ailing Zhang contributed equally to this work.

<sup>2</sup>Herbivorous Livestock Research Institute, Chongqing Academy of Animal Sciences, Chongqing, China

<sup>1</sup>Department of Grassland Science, Sichuan Agricultural University, Chengdu 611130, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

reduced by approximately 3.6 billion hectares (25% of the world's highlands).

Pearl millet (*Pennisetum glaucum* (L.) R. Br.) ranks 6th behind the world's most important cereal crops, such as wheat, rice, maize, sorghum and barley [3], and it is grown on an area of 31 million hectare (ICRISAT 2016) in the world. Pearl millet is an ideal plant for bioethanol production and can be used as a sustainable and alternative energy source, due to its high concentration of easily extractable fermentable sugar.

Pearl millet is highly tolerant to heat because its root branchings and flowering speed increase along with the growth of new leaves at certain high temperature levels [4]. The relative growth rate and net assimilation rate (NAR) significantly increase in pearl millet and slightly decrease in maize subjected to high temperature (38/27 °C) [5]. Previous reports distinguish pearl millet as a crop that is highly resistant toward drought. Under a certain degree of drought stress, pearl millet has a higher resistance to drought than other species of millet (finger millet, Job's tears, barnyard millet, common millet, and foxtail millet) and maintains an unchanged morphology, particularly with respect to the leaf area and shoot fresh and dry weight [6, 7]. Therefore, pearl millet is an ideal material to study the mechanism of heat and drought resistance of cereal crops. Studies on the root activity [4], seed germination rate [8] and seedling growth rate of pearl millet under heat stress [9] have previously been reported. Some researchers have cloned and investigated the HSP (sHSP [10], hsp70 [11] and HSP90 [12]) genes in pearl millet, but the presence of only a few reports do not provide sufficient information on the sequencing of pearl millet under heat stress. In addition, studies on the sequencing of pearl millet under drought stress are still limited [13, 14].

The genome of pearl millet was reported in 2017 [15], but short reads cannot be mapped to the genes due to incomplete genome annotations. Single-molecule real-time sequencing technology also known as the next-generation sequencing, such as PacBio sequencing, allows the direct production of full-length transcripts, making it ideal for transcript recovery and isoform detection of well-sequenced and/or incomplete genomic sequences [16, 17], but its disadvantage is low throughput [18]. Second-generation sequencing using the Illumina platform is an effective way to quantify the gene expression and high-quality reads. However, due to the short length of the reads produced by the second generation of sequencing, computational assembly is required [18, 19]. To avoid these problems, we combined the two sequencing techniques to study the similarities and differences of the molecular mechanism of the response of pearl millet under heat stress and drought stress. First, we used Illumina sequencing data to correct the raw data of the full-length

transcriptome, and second, we used the corrected full-length transcriptome data as a reference sequence to analyze the short sequencing data.

In this study, we obtained the following results: a. We identified 63,090 new transcripts and 26,299 new genes; b. Compared with previous studies, the annotation rate of pearl millet genes increased by 20.18%; c. The heat shock protein under HSPs still function after 48 h of heat stress; d. The plants can still be regulated by the abscisic acid (ABA) pathway under 48 h drought stress; e. Even with the same mechanism of response, plants tend to select different protein species when they face different stresses. This study demonstrated the changes of pearl millet under 48 h heat stress and drought stress at the molecular level, which provided a new concept to study the tolerance of pearl millet. In addition, the new genes that were discovered provide more resources for breeding projects.

## Results

### A total of 132,488 non-redundant transcripts and 64,878 genes were identified by full-field transcriptome modeling

Acquisition of the full-length transcriptome data was based on the third-generation sequencing platform of PacBio Sequel. By filtering the raw data, we removed the connector and original offline data that was less than 50 bp in length to obtain 6,842,837 Subreads based on 17.89 G (Table 1). The average length of the Subreads was 2615 bp. A total of 354,139 circular consensus (CCS) were obtained through conditional screening (full passes of 1 and quality of 0.80). Finally, 306,369 full length non-chimeric reads (Flnc) were obtained with complete 5'-primers, 3'-primers and poly-A tails in which an average of the Flnc read length was 2897.

The third-generation sequencing technology, represented by PacBio, has the advantage of long read lengths, but the technique has a high single base error rate. To reduce the high error rate, the Illumina data were used for correction. After correcting the data, 132,588 consensus, 3302 N50, and 2006 N90 were obtained. CD-HIT software removed redundant and similar sequences that resulted in 132,488 non-redundant transcripts and 64,878 genes.

### Annotation of 62,436 (96.24%) genes

To obtain a comprehensive genetic annotation, we analyzed the total of the 64,878 genes using NCBI nucleotide sequences (Nt) by BLASTN (E-value  $10^{-10}$ ), NCBI non-redundant protein sequence (Nr), Protein family (Pfam), eukaryotic Ortholog Groups (KOG), Swiss-prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and BLASTx (E-value  $10^{-10}$ ) databases. The Venn diagram demonstrated that 27,190

**Table 1** Summary of PacBio transcripts and comparisons with genomes reported in 2017

	Pacbio		Varshney et al., 2017
	Before correct	After correct	
Subreads base(G)	17.89		
Subreads number	6,842,837		
Average subreads length	2615		
N50(subreads)	2787		
CCS	354,139		
5'-primer	332,357		
3'-primer	332,288		
Poly-A	326,534		
Full length	306,369		
Flnc	303,627		
Average flnc read length	2897		
Consensus reads	132,488		
Total_nucleotides	409,482,590	410,858,545	
Total_number	132,488	132,488	
Mean_length	3091	3102	1014.71
Min_length	200	201	
Max_length	14,226	14,277	
N50(consensus)	3288	3302	
N90(consensus)	1999	2006	
Number of transcripts		132,488	69,398
Number of Genes		64,878	38,579
Number of genes annotated		62,436(96.24%)	29,344 (76.06%)
Number of genes unannotated		2442(3.76%)	9235 (23.94%)

genes were simultaneously annotated in all the databases (Nr, Nt, Pfam, GO, KEGG, KOG, and Swiss-prot) (Fig. 1a). There were 62,436 genes that were annotated in at least one database of which 60,870 (93.82%) genes were annotated by the NR database. Based on the alignment of sequence homology, the result showed that 48,226 (74.33%) sequences were found against *Setaria italica*; 2471 (3.81%) sequences had significant hits for *Sorghum bicolor*, followed by *Zea mays* (2422, 3.73%), *Dichanthelium oligosanthes* (2218, 3.42%) and *Oryza sativa* (1100, 1.70%). A total of 13.01% of the sequences were homologous to other species (Fig. 1b).

A total of 40,159 (61.90%) genes were annotated as GO terms. The results of the GO enrichment analysis showed that the genes were primarily enriched in metabolic process, cellular process, single-organism process, cell, cell part, membrane, binding, catalytic activity, transporter activity in biological process (BP), molecular function (MF), and cellular component (CC) (Fig. 1c).

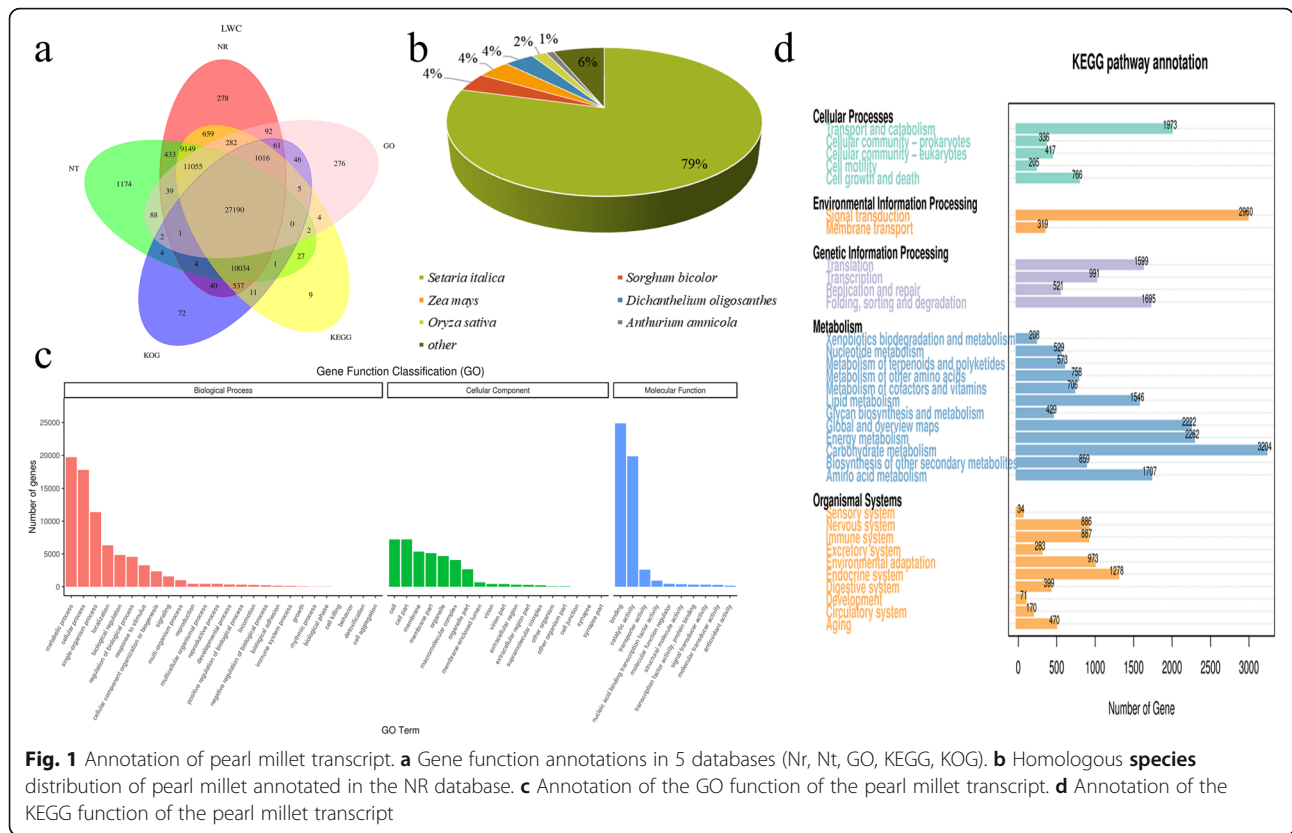
A total of 59,981 (92.45%) genes were annotated in the KEGG database (Fig. 1d). The KEGG pathway analysis revealed that 3204 (4.94%) genes were clustered in the carbohydrate metabolism pathway, 2960 (4.56%) genes

in the signal transduction pathway, and 2262 (3.49%) in energy metabolism.

We annotated 39,024 (60.15%) genes in the KOG database. A total of 7332 (11.30%) genes were annotated in general function prediction only, while 5256 (8.10%) were in signal transduction mechanisms and 3997 (6.16%) in post-translational modification, protein turnover and chaperones.

#### A total of 2792 TFs, 1223 TRs, 694 lncRNA, and 1124 alternative splicing events were identified in pearl millet

By predicting 132,488 non-redundant transcripts using iTAK software, 2792 genes were predicted to be TFs, and 1223 genes were predicted to be TRs (Fig. 2a). A total of 272 transcripts were predicted to be MYB-related TFs. MYB-associated transcription factors are important telomere-binding proteins that have the effect of maintaining the integrity of the chromosomal structure and regulating gene transcription, and 235 transcripts were predicted as other TFs. A total of 196 transcription factors were considered C3H transcription factors, which are involved in abiotic stress responses in plants [20]. There are 190 transcripts that were anchored

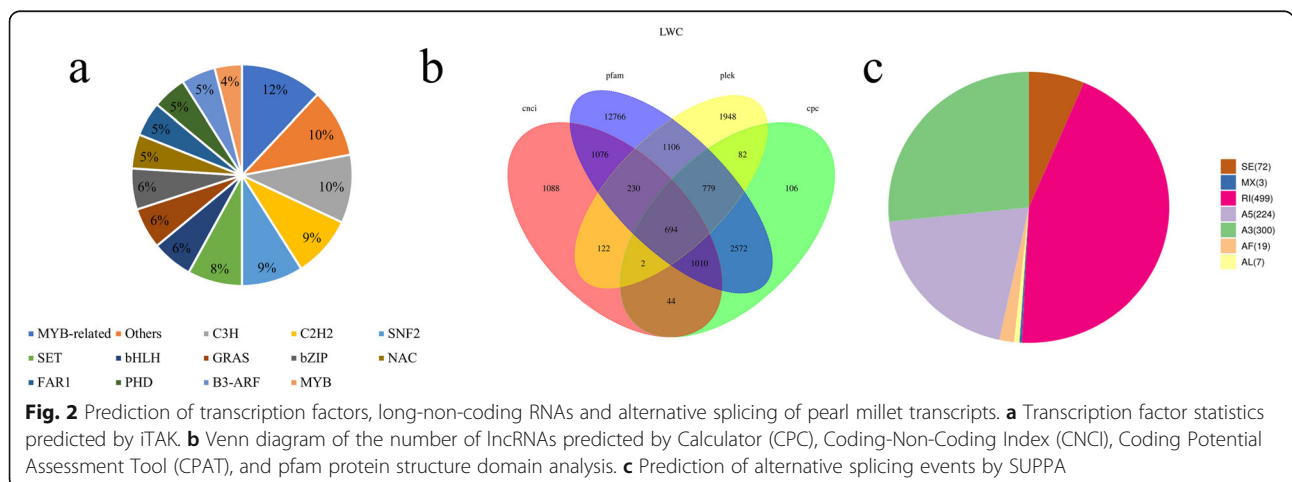


as C2H2 TFs, primarily related to plant growth and development and the response to environmental stresses [21].

In addition, CNCI, CPC, Pfam-scan and PLEK tools were used to predict long non-coding RNA (lncRNA) (Fig. 2b). Based on known annotations, 4266 transcripts were detected as long non-coding proteins through CNCI. Consistent with the NCBI database, 5389 transcripts were predicted as long non-coding proteins by CPC (e-value  $1e^{-10}$ ). Translation and identification of

transcripts via Pfam-scan (default parameters of  $-E$  0.001  $--domE$  0.001) showed that 20,233 transcripts were long non-coding proteins. There were 4963 non-coding transcripts belonging to long protein by screening with PLEK (default parameters of  $-minlength$  200), while 694 transcripts were simultaneously screened by CNCI, CPC, Pfam-scan and PLEK.

A total of 1124 alternative splicing events were identified by SUPPA (Fig. 2c). There were 499 genes belonging



to the Retained Intron Type, 300 genes as Alternative 3' Splice Sites, and 224 genes were related to Alternative 5' Splice Sites.

### Photosynthetic proteins may be involved in resistance to heat stress

There were a total of 6920 differentially expressed genes (DEGs), of which 3555 and 3365 genes were up- and down-regulated, respectively.

GO enrichment analysis displayed that “transmembrane transport” (114 up-regulation and 252 down-regulation), “transport” (308 up-regulation and 444 down-regulation) and “establishment of localization” (308 up-regulation and 444 down-regulation) were the most enriched terms of the biological process. “photosystem II oxygen evolving complex” (5 up-regulation and 7 down-regulation) and “photosystem II” (5 up-regulation and 10 down-regulation) were the most enriched terms of the cellular component (Additional file 2). “transporter activity” (141 up-regulation and 248 down-regulation) and “monooxygenase activity” (104 up-regulation and 10 down-regulation) were the most enriched terms of the cellular component. We found that “RNA splicing”, “anatomical structure morphogenesis”, “regulation of translation”, “spliceosomal complex”, “translation regulator activity”, “zinc ion binding” and “nucleic acid binding” were the most enriched GO terms among the differentially expressed up-regulated genes. “single-organism localization”, “single-organism transport”, “transmembrane transport”, “intrinsic component of membrane”, “integral component of membrane”, “membrane part”, “transmembrane transporter activity”, “transporter activity”

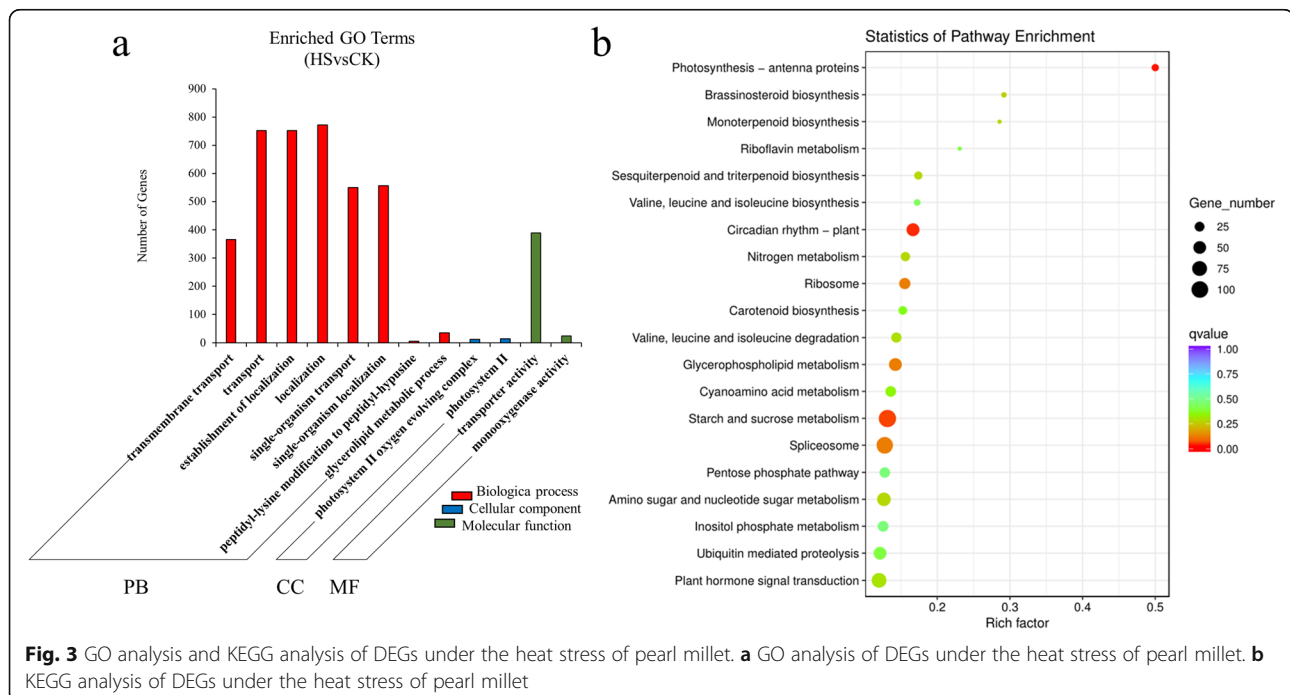
and “diacylglycerol O-acyltransferase activity” were the most enriched GO terms among the differentially expressed down-regulated genes (Fig. 3a).

The KEGG enrichment analysis of 6920 DEGs showed that “Photosynthesis - antenna proteins”, “Circadian rhythm-plant” and “Starch and sucrose metabolism” were the most enriched pathways. A KEGG enrichment analysis of 3555 up-regulated DEGs revealed that “Spliceosome”, “Valine, leucine and isoleucine degradation” and “Valine, leucine and isoleucine degradation” were the most enriched terms, and a KEGG enrichment analysis of 3365 down-regulated differentially expressed genes indicated that “Photosynthesis - antenna proteins”, “Glycerophospholipid metabolism” and “Circadian rhythm - plant” were the most enriched pathways (Fig. 3b).

### Glycerophospholipid metabolic pathway may play an important role in the drought tolerance of pearl millet

By comparing the gene expression levels under drought treatment with the control conditions, a total of 6484 DEGs ( $P < 0.05$ ) were shown to be up- or down-regulated between samples collected at 48 h. There were 3041 up-regulated and 3443 down-regulated DEGs.

GO enrichment analysis displayed that “single-organism process” (964 up-regulation and 959 down-regulation), “single-organism metabolic process” (593 up-regulation and 552 down-regulation), “oxidation-reduction process” (289 up-regulation and 303 down-regulation) “photosystem II oxygen evolving complex” (0 up-regulation and 3 down-regulation), “photosystem II” (0 up-regulation and 21 down-regulation), “thylakoid membrane” (0 up-regulation and 18 down-



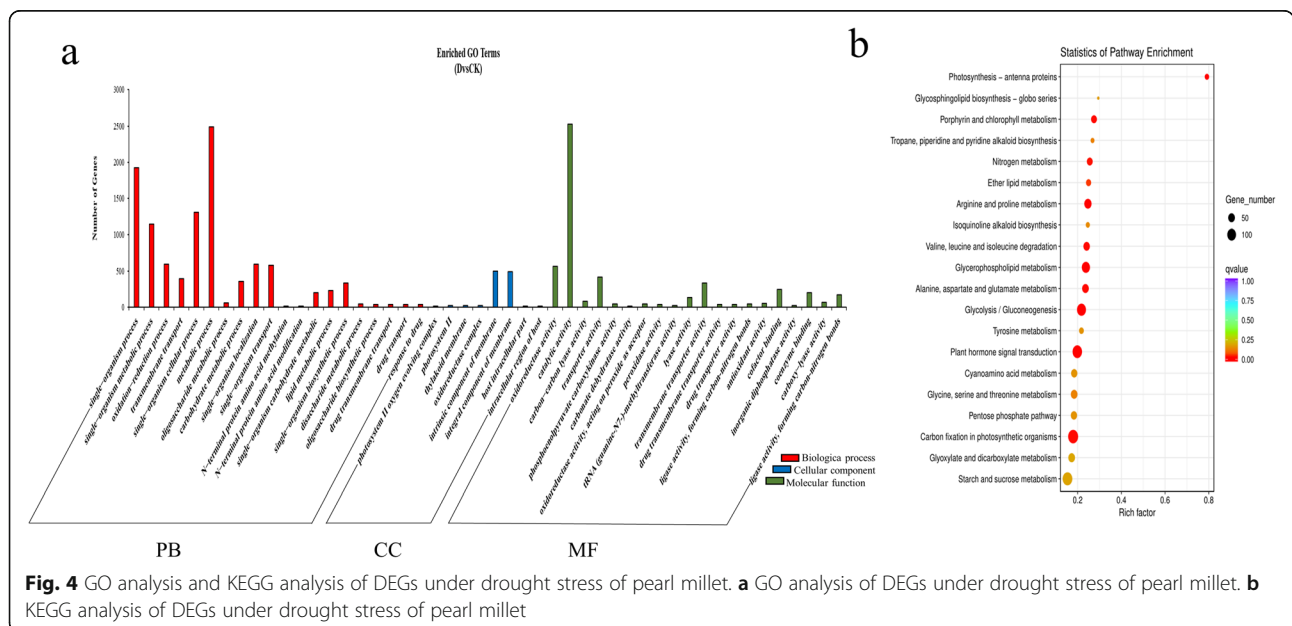
regulation), “oxidoreductase activity” (284 up-regulation and 276 down-regulation), “catalytic activity” (1254 up-regulation and 3165 down-regulation) and “carbon-carbon lyase activity” (81 up-regulation and 17 down-regulation) were the most enriched terms. The results showed that the up-regulated DEGs were primarily enriched in “single-organism process”, “single-organism metabolic process”, “oligosaccharide metabolic process”, “tRNA (guanine-N7-)-methyltransferase activity”, “catalytic activity” and “tRNA (guanine) methyltransferase activity” terms. However, the down-regulated DEGs were primarily enriched in “single-organism process”, “metabolic process”, “protein import into nucleus”, “photosystem II”, “photosystem II oxygen evolving complex”, “thylakoid membrane”, “carbon-carbon lyase activity”, “phosphoenolpyruvate carboxykinase activity” and “lyase activity” (Fig. 4a).

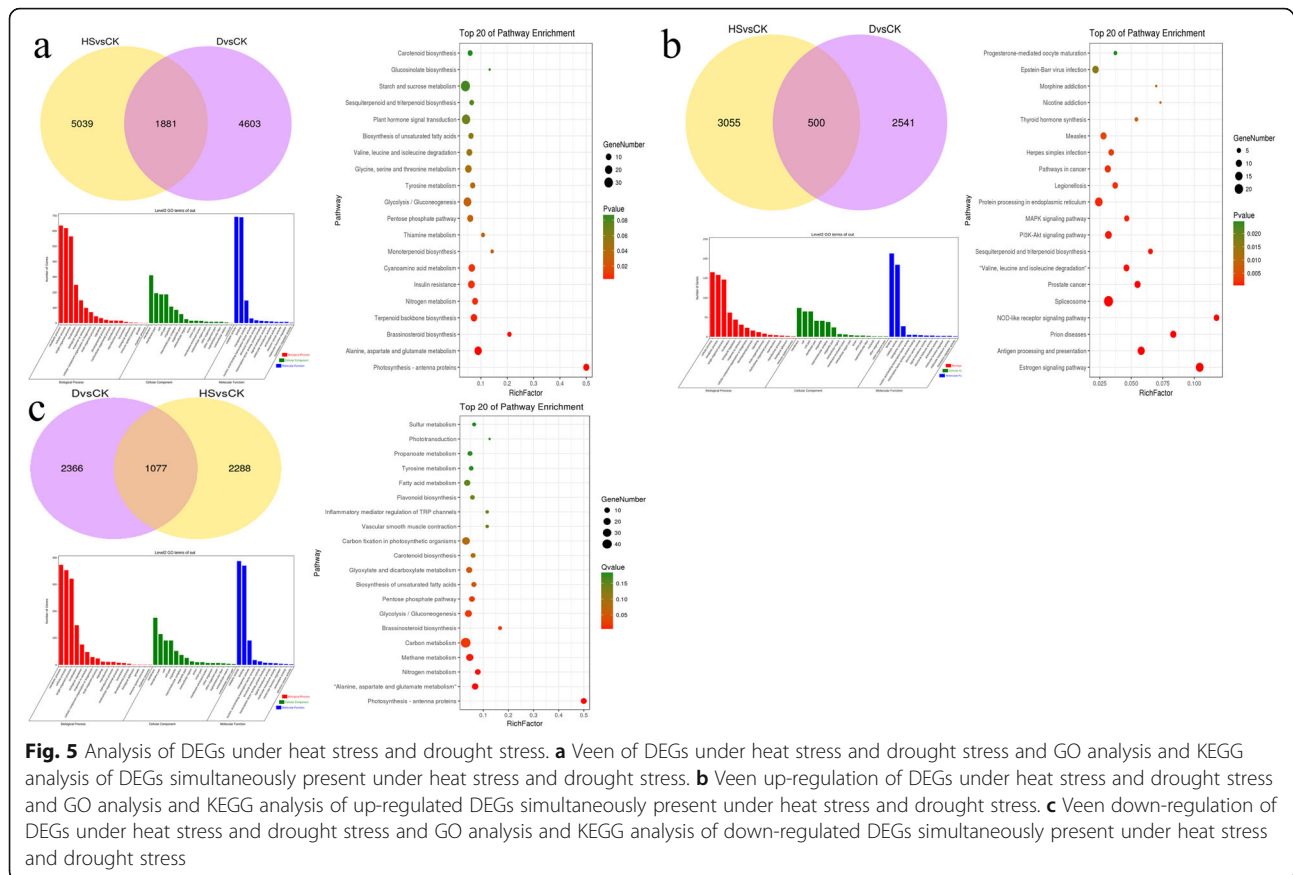
The results of the KEGG enrichment analysis indicated that the DEGs were primarily enriched in “Photosynthesis - antenna proteins”, “Glycerophospholipid metabolism” (Additional file 3) and “Glycolysis/Gluconeogenesis” pathway. The up-regulated DEGs were primarily enriched in “Glycerophospholipid metabolism”, “Valine, leucine and isoleucine degradation” and “Arginine and proline metabolism” pathway. The down-regulated DEGs were primarily enriched in “Carbon fixation in photosynthetic organisms”, “Photosynthesis - antenna proteins” and “Porphyrin and chlorophyll metabolism” pathways (Fig. 5b).

**A total of 1881 genes were involved in both heat stress and drought stress responses**

A total of 1881 DEGs were simultaneously found in both heat stress and drought stress (Fig. 5a). GO enrichment analysis results show that the most enriched GO terms

were “metabolic process”, “cellular process”, “membrane”, “membrane part”, “catalytic activity”, and “binding” (Fig. 5a). The KEGG enrichment analysis showed that “Photosynthesis - antenna proteins”, “Alanine, aspartate and glutamate metabolism”, “Brassinosteroid biosynthesis”, and “Terpenoid backbone biosynthesis” were the most enriched pathways (Fig. 5a). Among the 1881 DEGs, 500 genes were up-regulated under both heat stress and drought stress (Fig. 5b). Through GO enrichment analysis, it was found that the 500 DEGs were primarily enriched in “cellular process”, “metabolic process”, “membrane”, “cell”, “binding”, and “catalytic activity” GO terms (Fig. 5b). A KEGG enrichment analysis found that the main enrichments were “Estrogen signaling pathway”, “Antigen processing and presentation”, and “NOD-like receptor signaling pathway” pathway (Fig. 5b). A total of 1077 DEGs were simultaneously down-regulated under heat stress and drought stress (Fig. 5c). GO enrichment analysis showed that the main enrichment was in “metabolic process”, “cellular process”, “membrane”, “membrane part”, “catalytic activity”, and “binding” GO terms (Fig. 5c). A KEGG enrichment analysis found that the main enrichment was in “photosynthesis - antenna protein”, “alanine, aspartic acid and glutamate metabolism”, and “nitrogen metabolism” (Fig. 5c). A total of 122 DEGs were up-regulated under heat stress but down-regulated under drought stress. GO enrichment analysis of the 122 DEGs showed that “metabolic process”, “cellular process”, “membrane”, “membrane part”, “binding”, “catalytic activity” were the most enriched GO terms (Fig. 6a). The results of the KEGG enrichment analysis indicated that 122 differentially expressed genes were primarily enriched in the





“Thiamine metabolism”, “Cyanoamino acid metabolism”, “Starch and sucrose metabolism”, and “Phenylpropanoid biosynthesis” pathways (Fig. 6b). A total of 182 DEGs were down-regulated under heat stress but were up-regulated under drought stress. GO enrichment analysis of the 182 DEGs revealed that the main enrichment was at the “single-organism process”, “metabolic process”, “membrane”, “membrane part”, “catalytic activity”, and “binding” GO terms (Fig. 6c). The results of a KEGG enrichment analysis indicate that the main enrichment is in the “Glycerophospholipid metabolism”, “Monoterpenoid biosynthesis”, and “Phosphonate and phosphinate metabolism” pathways (Fig. 6d).

There were 5039 and 4603 DEGs specific to heat stress and drought stress, respectively (Fig. 5a). A total of 5039 DEGs analysed by GO enrichment found that these genes were primarily enriched in “metabolic process”, “cellular process”, “membrane”, “cell”, “binding”, and “catalytic activity” (Fig. 7a). A KEGG enrichment analysis showed that 5039 genes were primarily enriched in the “p53 signaling pathway”, “Circadian rhythm - plant”, and “Ribosome” and “Ubiquitin mediated proteolysis” pathways (Fig. 7b). The 4603 DEGs were only found in drought stress enriched in “metabolic process”, “cellular process”, “membrane”, “cell”, “catalytic activity”, and

“binding” GO terms (Fig. 7c). A KEGG enrichment analysis showed that 4603 DEGs were primarily enriched in the “Glycerophospholipid metabolism”, “Arginine and proline metabolism”, “Glycolysis/Gluconeogenesis”, and “Biosynthesis of amino acids” pathways (Fig. 7d).

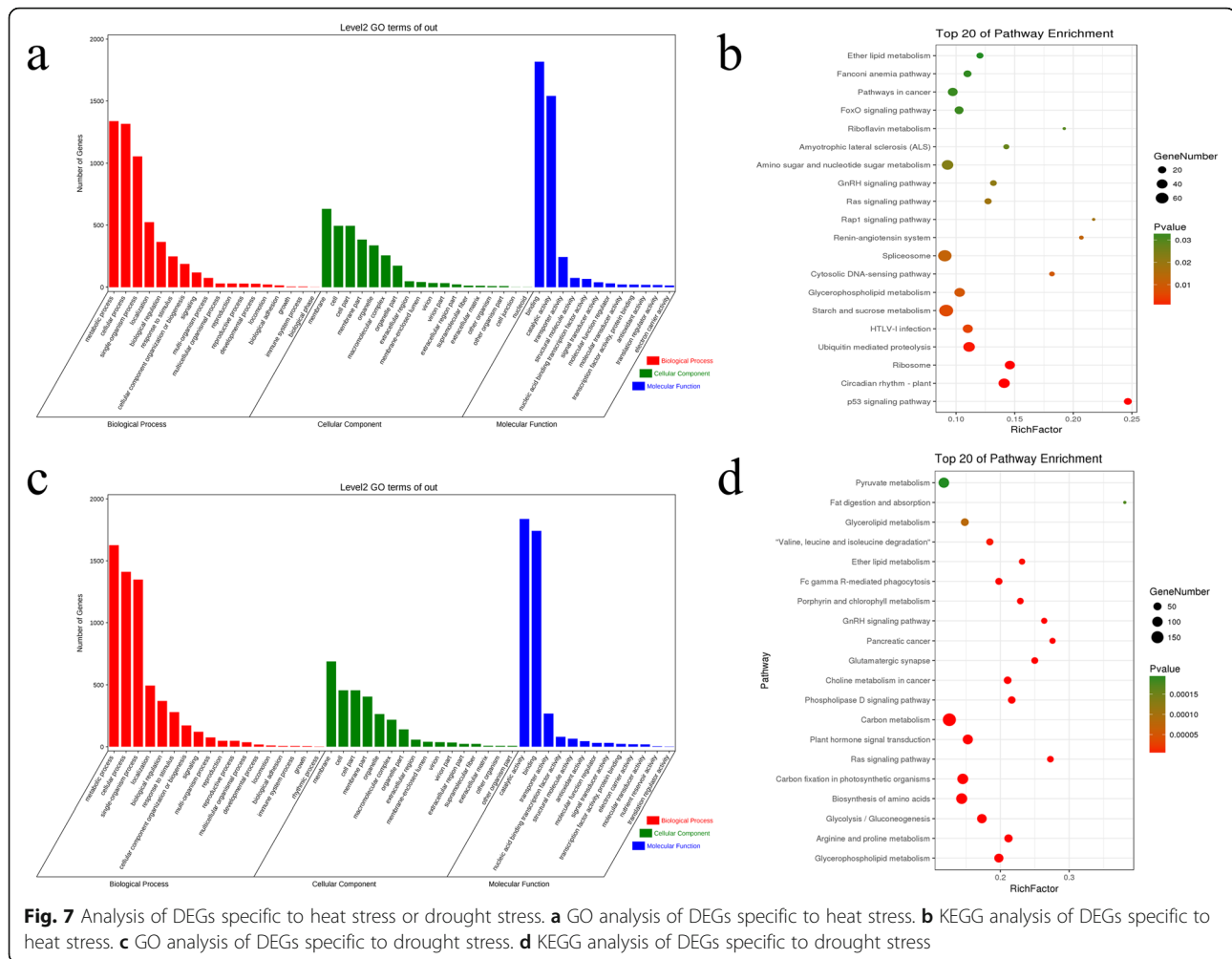
## Discussion

**A total of 63,090 new transcripts, 26,299 new genes were identified, and annotations increased by 20.18%**

Pearl millet is a versatile grain that can be used as grain, hay and green fodder. It is distributed in arid and semi-arid regions, such as India and sub-Saharan Africa [22]. In recent years, there have been many reports on the sequencing of corn [23, 24], rice [25] and wheat [26], but only a few that examined pearl millet. Some researchers constructed a genetic linkage map of 640 cM using GBS sequencing (genotyping-by-sequencing) to identify the genes associated with Striga and other agronomic traits [27]. In April 2016, researchers performed an RNA-Seq analysis of pearl millet planted in two climate gradients to explore their adaptability to climate [28]. In May 2016, researchers constructed the highest density genetic linkage map using GBS sequencing. The total average density of SNP markers in the map was 23.23 SNP/cM, and the total length was 716.7 cM [29]. In June 2016,







regarding RNA sequencing exists. We performed RNA sequencing on heat-stressed pearl millet and discovered DEGs associated with reactive oxygen species (ROS) produced by exposure to heat stress. As a starting signal of plants, ROS prompts local tissues to respond to local abiotic stress stimulation, but the expression of the ROS genes are temporary [32, 33]. We identified nine DEGs associated with ROS production, two DEGs encoding amine oxidases (AOC), and the remaining seven DEGs encoding polyamine oxidases (PAO). With the exception of two genes that were up-regulated, the others were down-regulated (Additional file 4). These results indicated that a variety of genes involved in ROS production were inhibited after 48 h of heat stress. This is consistent with previous studies of the expression of genes involved in ROS production that have not been continuously expressed under stress [14]. ROS scavenging enzymes play an important role in protecting plants from temperature stress and are essential for the detoxification of ROS produced under stress conditions [34]. We found that 10 DEGs encoded ROS scavenging enzymes,

five of which were up-regulated (Additional file 4), while the other five genes were down-regulated. These data suggest that when ROS is reduced, ROS clearance enzymes will decrease, and this effect is not sustained. The accumulation of heat shock proteins (HSP) under the control of heat stress transcription factors (HSFs) plays an important role in the heat stress response (HSR) and subsequently produces heat tolerance in plants and other organisms [35, 36]. In this study, we found that 30 DEGs were associated with HSPs; three DEGs encoded HSP100; 14 DEGs encoded HSP90; 12 DEGs encoded HSP70, and one DEG encoded sHSP, and all were up-regulated DEGs (Additional file 4). We hypothesize that the heat shock proteins still function even if the plants are subjected to sustained heat stress.

#### Expression of typical drought-related genes in pearl millet

Under drought stress, the ABA content of plants increases, which regulates the opening and closing of pores, reduces water loss and maintains the moisture

content of plants [37, 38]. Under drought stress, we found five DEGs encoding 9-cis-epoxycarotenoid dioxygenases (NCED), 1 DEG encoding zeaxanthin epoxidase (ABA1), and two DEGs encoding short-chain dehydrogenases/reductases (ABA2), four of which were up-regulated DEGs, and the other four were down-regulated DEGs (Additional file 5). We also discovered a DEG encoding PYL8, which is a receptor for ABA. The gene was down-regulated under drought stress (Additional file 5). This indicates that under 48 h of drought stress, the ABA pathway still functions in the drought stress response. Previous studies have found that slow anion channel correlation (SLAC) regulates the opening and closing of stomata, which is necessary for stomatal closure [39]. After 48 h of drought stress, a SLAC-encoding DEG was identified, which indicates that even after prolonged drought stress, pearl millet reduces water evaporation by regulating stomatal closure. (Additional file 5). The *Asr* gene family is classified as a new group of late embryogenic abundant (LEA) genes and is involved in adaptation to drought stress [40]. Under stress condition, the expression of the *Asr* gene was up-regulated. We found two *Asr* genes with different levels of expression, which were up-regulated under drought stress and are consistent with previous reports. (Additional file 5).

#### **In the same molecular mechanism, the pearl millet selects different proteins to respond to different stresses**

The results showed that 5039 DEGs and 4603 DEGs were regulated under heat stress and drought stress, respectively, indicating that plants have different molecular mechanisms for heat and drought stress [41]. In addition, even if the same mechanism exists under both types of stresses, the types of proteins involved, the number of proteins, and the choice of related genes are different, such as ROS clearance mechanisms. Under drought stress, we found 14 DEGs associated with ROS scavenging enzymes, two DEGs encoding SOD, 11 DEGs encoding APX, and one DEG encoding GPX, six of which were up-regulated DEGs, while the other eight DEGs were down-regulated. The gene encoding CAT was significantly expressed 48 h after heat stress, but the expression of the gene was not significant under drought stress. The number of DEGs encoding APX proteins was higher under drought stress than heat stress. The number and type of ROS scavenging enzymes expressed under drought stress were different than those expressed from heat stress, indicating that ROS scavenging enzymes are specific to the type of stress (Table 2). A total of 1881 DEGs were collectively expressed under heat stress and drought stress, indicating that these 1881 DEGs play important roles under both types of stresses.

Among the 1881 DEGs, 1577 DEGs with the same pattern of expression (both up-regulated or down-regulated under heat stress and drought stress) are essential to combat heat stress and drought stress, such as HSPs. HSPs are essential molecular chaperones in eukaryotic cells that play an important role in the folding and activation of proteins involved in signal transduction and regulation of the cell cycle. Under drought stress, we found that 26 DEGs encode HSP, of which 11 DEGs encode HSP70, 14 encode HSP90, and one encodes sHSP, all of which up-regulate DEGs (Table 3). Among them, 12 DEGs encoding HSP existed in both heat stress and drought stress. There were 304 DEGs with different expression patterns under heat stress and drought stress, indicating that different stresses have different levels of regulation of the same gene.

#### **Conclusions**

We combined two sequencing technologies to study the similarities and differences of the molecular mechanism of millet under heat and drought stress. In this study, 63,090 new transcripts and 26,299 new genes were identified, and the functional annotation of genes was improved by 20.18%. Under heat and drought stress, 6920 and 6484 genes were differentially expressed, respectively, and 1881 differentially expressed genes were present in both types of stresses. This information provides additional resources to identify and unearth important genes in pearl millet. In addition, we found that under the same mechanism of response, plants have different protein choices when faced with different types of stresses. This lays the foundation to study the heat and drought resistance mechanism of pearl millet.

#### **Methods**

##### **Plant materials, cultivation and treatment**

The pearl millet variety Tifleaf 3 used in this study was provided by Sichuan Agricultural University and is preserved in the Department of Grassland Science, College of Animal Science and Technology, Sichuan Agricultural University, Ya'an, Sichuan Province, China. Pearl millet seeds were grown in pots (10\*15 cm) containing quartz sand and placed in a growth chamber. The pots were exposed to 14 h of light at 26 °C and 10 h of darkness at 22 °C for 13 days with 50% Hoagland's nutrient solution (1 mM MgSO<sub>4</sub>, 1 mM KH<sub>2</sub>PO<sub>4</sub>, 1 mM NH<sub>4</sub>NO<sub>3</sub>, 0.5 mM CaCl<sub>2</sub>, 0.1 mM FeNa-EDTA, 25 mM NaCl, 0.1 mM H<sub>3</sub>BO<sub>3</sub>, 0.1 mM Na<sub>2</sub>SiO<sub>3</sub>, 1.5 μM CuSO<sub>4</sub>, 50 μM KCl, 10 μM MnSO<sub>4</sub>, 0.075 μM Na<sub>2</sub>MoO<sub>4</sub> and 2 μM ZnSO<sub>4</sub>). Thirteen-day-old plants were divided into three groups (control, heat treatment group and drought treatment group). The culture conditions for the control group were 14 h of light at 26 °C and 10 h of darkness at 22 °C with 50% of Hoagland's nutrient solution. The plants in

**Table 2** Expression of gene encoding peroxide scavenging enzyme under heat stress and drought stress

annotation	Heat stress		drought stress	
	Gene	log <sub>2</sub> (HS/CK)	Gene	log <sub>2</sub> (DS/CK)
SOD	<i>i0_LQ_LWC_c2218/f1p0/833</i>	1.1734	<i>i2_HQ_LWC_c7659/f3p1/2700</i>	1.8944
			<i>i1_LQ_LWC_c42429/f1p0/1061</i>	1.5801
GPX	<i>i0_HQ_LWC_c63/f5p0/847</i>	-1.3626	<i>i1_LQ_LWC_c42213/f1p2/1035</i>	-1.4874
CAT	<i>i1_LQ_LWC_c20347/f1p0/1948</i>	-4.4651		
	<i>i2_HQ_LWC_c41068/f2p7/2070</i>	1.6112		
	<i>i1_LQ_LWC_c12593/f1p0/1872</i>	-3.7907		
	<i>i2_HQ_LWC_c4824/f3p1/2226</i>	-4.5763		
APX	<i>i1_LQ_LWC_c18498/f1p3/1627</i>	2.9239		
	<i>i1_LQ_LWC_c38392/f1p0/1178</i>	-1.3366		
	<i>i3_LQ_LWC_c37944/f1p0/3280</i>	1.2566		
	<i>i1_LQ_LWC_c22741/f1p1/1544</i>	1.7925	<i>i1_LQ_LWC_c22741/f1p1/1544</i>	-2.7765
			<i>i1_HQ_LWC_c28271/f15p0/1733</i>	-1.3777
			<i>i1_HQ_LWC_c40231/f3p0/1461</i>	3.1134
			<i>i1_LQ_LWC_c13218/f1p0/1219</i>	1.4302
			<i>i1_LQ_LWC_c13821/f1p5/1427</i>	-1.2859
			<i>i1_LQ_LWC_c18173/f1p0/1818</i>	-2.3155
			<i>i1_LQ_LWC_c18294/f1p0/1406</i>	-1.5224
		<i>i1_LQ_LWC_c26053/f1p6/1719</i>	-2.2266	
		<i>i1_LQ_LWC_c36589/f1p0/1714</i>	-1.9296	
		<i>i1_LQ_LWC_c38632/f1p0/1092</i>	1.4572	
		<i>i2_LQ_LWC_c65698/f1p0/2629</i>	2.4454	

Note: HS heat treatment, CK control, DS drought treatment

the heat treatment group were exposed to 14 h of light at 40 °C and 10 h of darkness at 35 °C with 50% of Hoagland's nutrient solution. The plants in the drought treatment group were exposed to 14 h of light at 26 °C and 10 h of darkness at 22 °C with 20% PEG (polyethylene glycol 6000) dissolved in 50% of Hoagland's nutrient solution. All the treatment time is 48 h (2 days).

#### RNA preparation for Iso-Seq

Leaf and root samples of the treated plants were collected separately at the three-leaf and five-leaf stages. When the pearl millet was in the heading stage, the spikelet, leaves, stems and roots were collected. While at the flowering stage, the spikelet, leaves, stems and roots were collected, and these samples were immediately stored at -80 °C. The RNA was extracted using the RNeasy Plant Mini Kit [42], and the RNA quality was analysed using RNA gel electrophoresis. Equal amounts of RNA from 12 samples (1 µg per sample) were pooled together to form total RNA, and then the SMRT library was prepared using 3 µg total RNA.

#### PacBio Iso-Seq library preparation and sequencing

The Iso-Seq library was prepared using the Isoform Sequencing protocol (Iso-Seq) with the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin Size Selection System protocol as described by Pacific Biosciences (P/N100-377-100-05 and P/N100-377-100-04) with some modifications. First, 3 µl of total RNA solution was added to deionized water containing a single primer and incubated at 72 °C for 3 min (3' SMART primer IIA from the Clontech SMARTer kit 5'-AAG-CAGTGGTATCAACGCAGAGTACTNN-3'). Next, the SMARTER II A oligonucleotide (from the Clontech SMARTer kit 5'-AAGCAGTGGTATCAACGCAGAGTACXXXX-3'), 5X first-strand buffer, DTT, dNTP mix, RNase inhibitor and SMARTScribe reverse transcriptase were added to the mixture and incubated at 72 °C for 1 h. Finally, the reaction was terminated at 70 °C. After 23 PCR cycles, the length of the PCR product was screened by the BluePippin Size Selection System and divided into 1-2 Kb, 2-3 Kb and 3-10 Kb fragments. After size screening, the cDNA was subjected to 12 cycles of PCR reactions. The PCR amplification

**Table 3** Expression of gene encoding heat shock protein under heat stress and drought stress

Annotation	Heat stress		drought stress		
	Gene	log <sub>2</sub> (HS/CK)	Gene	log <sub>2</sub> (DS/CK)	
sHSP	<i>i0_LQ_LWC_c967/f1p0/765</i>	4.9898	<i>i0_LQ_LWC_c967/f1p0/765</i>	10.152	
HSP 70			<i>i1_LQ_LWC_c32521/f1p1/1976</i>		
			<i>i1_LQ_LWC_c22890/f1p0/1926</i>		
			<i>i2_LQ_LWC_c53787/f1p5/2383</i>	5.0347	
			<i>i2_LQ_LWC_c86046/f1p8/3023</i>	1.3889	
			<i>i3_HQ_LWC_c31973/f15p0/3104</i>	1.0617	
			<i>i2_HQ_LWC_c43630/f6p12/2432</i>	6.901	
		<i>i2_LQ_LWC_c104760/f1p1/2446</i>	1.1635	<i>i2_LQ_LWC_c104760/f1p1/2446</i>	1.7534
		<i>i2_LQ_LWC_c3333/f1p2/2529</i>	4.6887	<i>i2_LQ_LWC_c3333/f1p2/2529</i>	5.7184
		<i>i2_LQ_LWC_c81928/f1p8/2468</i>	3.0743	<i>i2_LQ_LWC_c84968/f1p1/2090</i>	1.5487
		<i>i2_LQ_LWC_c84968/f1p1/2090</i>	1.2513	<i>i3_LQ_LWC_c21708/f1p2/3195</i>	1.3274
		<i>i3_LQ_LWC_c21708/f1p2/3195</i>	1.2128	<i>i3_LQ_LWC_c27490/f1p0/3150</i>	6.7401
		<i>i3_LQ_LWC_c27490/f1p0/3150</i>	4.1929		
		<i>i2_HQ_LWC_c127192/f35p1/2517</i>	1.2971		
		<i>i2_HQ_LWC_c55656/f2p12/2212</i>	1.9817		
		<i>i2_HQ_LWC_c59895/f32p11/2345</i>	1.6303		
		<i>i2_HQ_LWC_c69799/f2p1/2652</i>	2.5412		
		<i>i2_LQ_LWC_c22856/f1p14/2305</i>	2.0376		
HSP 90			<i>i2_HQ_LWC_c127285/f2p25/2794</i>	1.4443	
			<i>i2_HQ_LWC_c34808/f2p10/2596</i>	1.4052	
			<i>i2_HQ_LWC_c97463/f19p4/2142</i>	1.4033	
			<i>i2_LQ_LWC_c104439/f1p16/2591</i>	1.9891	
			<i>i2_LQ_LWC_c3878/f1p2/2975</i>	1.7125	
			<i>i2_LQ_LWC_c64587/f1p17/2400</i>	2.3499	
			<i>i2_LQ_LWC_c8803/f1p15/2771</i>	2.3145	
			<i>i1_LQ_LWC_c13864/f1p0/1351</i>	1.279	
			<i>i2_LQ_LWC_c25038/f1p7/2643</i>	2.0169	
			<i>i2_HQ_LWC_c49563/f2p1/2825</i>	3.2962	
			<i>i2_HQ_LWC_c50337/f2p2/2836</i>	1.6639	
			<i>i2_LQ_LWC_c106886/f1p1/2786</i>	2.7368	
			<i>i2_LQ_LWC_c90259/f1p14/2485</i>	2.9776	
			<i>i4_LQ_LWC_c19908/f1p0/5045</i>	3.1358	
			<i>i2_LQ_LWC_c121080/f1p8/2511</i>		
			<i>i3_LQ_LWC_c19538/f1p2/3645</i>		
			<i>i2_LQ_LWC_c11597/f1p0/2377</i>		
		<i>i2_LQ_LWC_c126787/f141p12/2519</i>			
		<i>i2_LQ_LWC_c33264/f1p2/2930</i>			
		<i>i2_LQ_LWC_c33469/f1p1/2516</i>			
HSP100			<i>i2_LQ_LWC_c102911/f1p4/3026</i>	2.213	
			<i>i2_LQ_LWC_c12834/f1p3/2671</i>	1.5904	
			<i>i2_LQ_LWC_c88848/f1p1/2115</i>	3.2613	

Note: HS heat treatment, CK control, DS drought treatment

products were used to construct the SMRTbell Template libraries using the Iso-Seq protocol. The libraries were prepared for sequencing by annealing a sequencing primer (component of the SMRTbell Template Prep Kit 1.0) and binding polymerase to the primer-annealed template.

#### Iso-Seq data analysis

Processing of raw data was conducted by SMRTlink 5.1 software (<https://www.pacb.com/videos/tutorial-minor-variant-analysis-smrt-link-v5-0-0/>). The parameter settings were min\_length 50, max\_drop\_fraction 0.8, no\_polish TRUE, min\_zscore -9999.0, min\_passes 2, min\_predicted\_accuracy 0.8, max\_length 15,000, generating a cyclic consensus sequence from the sub read BAM file. Pbcclassify.py, ignorepolyA false, minSeqLength 200 were used to divide the samples into full length and non-full length reads. Non-full length and full-length FASTA files were produced and fed into the cluster step, which performs isoform-level clustering (ICE), followed by final Arrow polishing, hq\_quiver\_min\_accuracy 0.99, bin\_by\_primer false, bin\_size\_kb 1, qv\_trim\_5p 100, qv\_trim\_3p 30. Since the frequency of nucleotide indels and mismatches in the Iso-Seq reads were much higher than those in the shorter high-throughput sequencing, LoRDEC software was used to correct additional nucleotide errors in consensus reads based on Illumina RNA-Seq data. The redundancy in the corrected readings was obtained using software CD-HIT (-c 0.95 -T 6 -G 0 -aL 0.00 -aS 0.99) to obtain the final transcript for subsequent analysis.

#### Gene functional annotation

We used the following five databases to align all predicted protein coding sequences, NCBI non-redundant protein (NR, cutoff E value  $\leq 1^{e-5}$ ), NCBI non-redundant nucleotide (NT, E value  $\leq 1^{e-5}$ ), Swiss-Prot protein (<http://www.expasy.org/sprot/>, cutoff E value  $\leq 1^{e-5}$ ), KOG (<http://www.ncbi.nlm.nih.gov/COG/>, cut-off E value  $\leq 1^{e-3}$ ) [43], protein family (Pfam: <http://pfam.sanger.ac.uk/>, cutoff E-value  $\leq 0.01$ ), KEGG (<http://www.genome.jp/kegg>, cut-off E value  $\leq 1^{e-5}$ ) pathways [44]. Utilize Blast2GO (<http://www.blast2go.com>) program for GO annotation (<http://www.geneontology.org>) based on NR annotation (cutoff E-value  $\leq 1^{e-10}$ ).

#### CDS prediction

Protein coding sequences from cDNA were identified by the ANGEL pipeline (a long-read implementation of ANGLE). We used closely related species to ensure that the protein sequences were ANGEL-trained and then performed an ANGEL prediction for a given sequence.

#### TF analysis and Lnc-RNA analysis

Transcription factors were predicted by iTAK software.

We predicted lncRNA using four software: a) The CNCI (Coding-Non-Coding-Index, parameters as default parameters) was an effective software to distinguish the protein-encoding and non-coding sequences by profiles adjoining nucleotide triplets without relying on known annotations. b) The CPC (Coding Potential Calculator) was used to assess the ORF extent and quality of transcripts and search the sequenced base eukaryote protein database from NCBI (e value " $1^{e-10}$ ") to identify coding transcripts and non-coding transcripts. c) All the transcripts were translated, and then Pfam Scan (-E 0.001 --domE 0.001) was used to determine if they have a domain of a known protein family. d) Predicting Lnc-RNA with PLEK SVM classifier (-minlength 200). The PLEK SVM classifier identifies transcripts of encoded proteins by optimizing the K-mer approach, which eliminates the need for reference genomes and annotations. All of the four software programs described above identified non-coding transcripts, which were identified as Lnc-RNA.

#### RNA preparation for RNA-Seq

Three treatments were conducted simultaneously. After 48 h of treatment, we randomly selected the leaves of 16 seedlings and collected them in a cryogenic vials and immediately stored at  $-80^{\circ}\text{C}$  (Additional file 1). Three biological replicates were set for each treatment. The RNA was extracted from the samples using the RNeasy Plant Mini Kit, and the RNA quality was checked using RNA gel electrophoresis.

#### RNA-Seq library preparation and sequencing

The purity of RNA was detected using a NanoDrop spectrophotometer (California, USA), and the concentration of RNA was determined by a Qubit RNA assay kit in a Qubit 2.0 fluorometer system (California, USA). The library was constructed using the NEBNext® UltraTM Directional RNA Library Prep Kit for Illumina® (California, USA). The NEBNext® Poly (A) mRNA Magnetic Isolation Module was used to enrich the mRNA, and Fragmentation Buffer was added to break the mRNA into short segments. A strand of cDNA was synthesized with random hexamer primers. The second strand of cDNA was synthesized by adding buffer, dNTPs and DNA polymerase I. The double strand cDNA was purified by AMPure XP beads. The purified cDNA was repaired at the end; a tail was added and sequenced, and the fragment size was selected by AMPure XP beads. Finally, the final cDNA library was obtained by PCR enrichment. Qubit 2.0 was used for preliminary quantification; Agilent 2100 was used to detect the inserted fragments of the library, and finally Illumina

Hi-Seq 2000 was used for sequencing. A total of 9 RNA-Seq libraries were established.

### Quantification of the gene expression levels

Identify the gene expression level of each sample by RSEM [45]. The clean data generated by Illumina sequencing were mapped to SMRT sequencing data, and the read count of each gene was obtained from the mapping results [46]. The read count value of each gene was converted to the FPKM value (Fragments per Kilobase Million), and genes with FPKM > 0.3 were selected for analysis.

### Identification and function analysis of DEGs

Differential expression analysis was performed using the DESeq R package (1.10.1) [47] to identify DEGs between the heat-stressed and control samples and between drought-stressed and control samples. DESeq is a statistical program that determines the differential expression in digital gene expression data using a model-based negative binomial distribution. The P rate was adjusted by the p.adjust function to control the error rate. The genes with an adjusted P-value < obtained by DESeq software were considered to be differentially expressed, and the significance of DEGs determined by the absolute value of  $\log_2(\text{Group1}/\text{Group2}) \geq 1$  as the threshold.

We use the GOseq R package to perform GO enrichment analysis on DEGs. The software package is based on Wallenius non-central hypergeometric distribution, and estimates and adjusts the preference of DEGs length. Finally, the KEGG enrichment analysis of DEGs was carried out by KOBAS software [48].

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12870-020-02530-0>.

**Additional file 1.** Pearl millet after 48 h of heat treatment, drought treatment and control treatment

**Additional file 2.** Photosynthesis-related gene expression under heat stress

**Additional file 3.** Expression of Glycerophospholipid metabolic related genes during drought stress

**Additional File 4.** DEGs expression under the heat stress of pearl millet

**Additional File 5.** DEGs expression under the drought stress of pearl millet

### Abbreviations

DEG: Differentially expressed gene; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genome; HSP: heat shock proteins; ROS: Reactive oxygen species; AOC: Amine oxidases; PAO: Polyamine oxidases; HSFS: Heat stress transcription factors; NCED: 9-cis-epoxycarotenoid dioxygenases; ABA1: Zeaxanthin epoxidase; ABA2: Short-chain dehydrogenases/reductases; SLAC: Slow anion channel correlation; SOD: Superoxide Dismutase; CAT: Catalase; apx: ascorbate peroxidase; LEA: Ate embryogenic abundant

### Acknowledgements

Not applicable.

### Authors' contributions

MS, DH and AZ participated in material processing and transcriptome data analysis. IK, HY, XW and XZ provided bioinformatics support. XZ revised the paper. JZ, LH designed the research, provided supervision and contributed to the manuscript. All authors read and approved the final manuscript.

### Funding

This work was supported by the Chongqing performance incentive guide special project (cstc2019xjl80010), the Sichuan Province Research grant (2016NYZ0036), the Modern Agro-industry Technology Research System (CARS-34) and Modern Agricultural Industry System Sichuan Forage Innovation Team (SCCXTD-2020-16).

### Availability of data and materials

The datasets supporting the conclusions of this article are included within the article (and its additional files). Sequencing database for pearl millet could download from NCBI under the accession number SRR11816223, and the data will be shared on reasonable request of the corresponding author.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declared that they have no conflicts of interest to this work.

### Author details

<sup>1</sup>Department of Grassland Science, Sichuan Agricultural University, Chengdu 6111130, China. <sup>2</sup>Herbivorous Livestock Research Institute, Chongqing Academy of Animal Sciences, Chongqing, China. <sup>3</sup>Department of Horticulture, Virginia Tech, Blacksburg, VA 24061, USA.

Received: 17 January 2020 Accepted: 26 June 2020

Published online: 08 July 2020

### References

1. Akter S. Identification of common and unique stress responsive genes of *Arabidopsis thaliana* under different abiotic stress through RNA-Seq meta-analysis. Virginia Tech; 2018.
2. Zhao C, Liu B, Piao S, Wang X, Lobell DB, Huang Y, Huang M, Yao Y, Bassu S, Ciais P. Temperature increase reduces global yields of major crops in four independent estimates. *Proc Natl Acad Sci*. 2017;114(35):9326–31.
3. Singh R, Singh D, Tyagi P. Effect of Azotobacter, farmyard manure and nitrogen fertilization on productivity of pearl millet hybrids (*Pennisetum glaucum* (L) r. br) in semi-arid tropical environment. *Arch Agron Soil Sci*. 2003;49(1):21–4.
4. Pearson CJ. Thermal adaptation of *Pennisetum*: seedling development. *Funct Plant Biol*. 1975;2(3):413–24.
5. Ashraf M, Hafeez M. Thermotolerance of pearl millet and maize at early growth stages: growth and nutrient relations. *Biol Plant*. 2004;48(1):81–6.
6. Kusaka M, Ohta M, Fujimura T. Contribution of inorganic components to osmotic adjustment and leaf folding for drought tolerance in pearl millet. *Physiol Plant*. 2010;125(4):474–89.
7. Walter ZL, Morio I. Deep root water uptake ability and water use efficiency of pearl millet in comparison to other millet species. *Plant Prod Sci*. 2005; 8(4):454–60.
8. Ong CK, Monteith JL. Response of pearl millet to light and temperature. *Field Crop Res*. 1985;11(2–3):141–60.
9. Launders TE. The effects of early season soil temperatures on emergence of summer crops on the north-western plains of NSW. *Aust J Exp Agric*. 1971; 11(48):39–44.
10. Nitnavare RB, Yeshvekar RK, Sharma KK, Vadez V, Reddy MK, Reddy PS. Molecular cloning, characterization and expression analysis of a heat shock protein 10 (Hsp10) from *Pennisetum glaucum* (L), a C4 cereal plant from the semi-arid tropics. *Mol Biol Rep*. 2016;43(8):861–70.
11. Reddy PS, Mallikarjuna G, Kaul T, Chakradhar T, Mishra RN, Sopory SK, Reddy MK. Molecular cloning and characterization of gene encoding for cytoplasmic Hsc70 from *Pennisetum glaucum* may play a protective role against abiotic stresses. *Mol Genet Genomics*. 2010;283(3):243–54.

12. Reddy PS, Thirulogachandar V, Vaishnavi CS, Aakrati A, Sopory SK, Reddy MK. Molecular characterization and expression of a gene encoding cytosolic Hsp90 from *Pennisetum glaucum* and its role in abiotic stress adaptation. *Gene*. 2011;474(1):29–38.
13. Bidingger F, Mahalakshmi V, Rao GDP. Assessment of drought resistance in pearl millet (*Pennisetum americanum* (L.) Leeke). II. Estimation of genotype response to stress. *Aust J Agric Res*. 1987;38(1):49–59.
14. Choudhary M, Jayanand PJC. Transcriptional profiling in pearl millet (*Pennisetum glaucum* L.R. Br.) for identification of differentially expressed drought responsive genes. *Physiol Mol Biol Plants*. 2015;21(2):187.
15. Varshney RK, Shi C, Thudi M, Mariac C, Wallace J, Peng Q, He Z, Zhao Y, Wang X, Rathore A. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nat Biotechnol*. 2017;35(10):969–76.
16. Abdelghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Benhur A, Reddy ASN. A survey of the sorghum transcriptome using single-molecule long reads. *Nat Commun*. 2016;7:11706.
17. Wang M, Wang P, Liang F, Ye Z, Li J, Shen C, Pei L, Wang F, Hu J, Tu L. A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytol*. 2017;217(1):163.
18. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*. 2015;13(5):278–89.
19. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016;17(6):333–51.
20. Jiang AL, Xu ZS, Zhao GY, Cui XY, Chen M, Li LC, Ma YZ. Genome-wide analysis of the C3H zinc finger transcription factor family and drought responses of members in *Aegilops tauschii*. *Plant Mol Biol Report*. 2014;32(6):1241–56.
21. Kielbowicz-Matuk A. Involvement of plant C2H2-type zinc finger transcription factors in stress responses. *Plant Sci*. 2012;185–186(4):78–85.
22. Anuradha N, Satyavathi CT, Chellapilla B, Thirunavukkarasu N, Sankar SM, Singh SP, Meena MC, Singhal T, Srivastava RK. Deciphering Genomic Regions for High Grain Iron and Zinc Content Using Association Mapping in Pearl Millet. *Front Plant Sci*. 2017;8:412.
23. Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X. Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat Genet*. 2018;50(9):1289.
24. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun*. 2016;7:11708.
25. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*. 2018;557(7703).
26. Wang X, Chen S, Shi X, Liu D, Ma C. Hybrid sequencing reveals insight into heat sensing and signaling of bread wheat. *Plant J*. 2019.
27. Moumouni KH, Kountche BA, Jean M, Hash CT, Vigouroux Y, BIG H, Belzile F. Construction of a genetic map for pearl millet, *Pennisetum glaucum* (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. *Mol Breed*. 2015;35(1):5.
28. Berthouly-Salazar C, Thuillet AC, Rhoné B, Mariac C, Ousseini IS, Couderc M, Tenaillon MI, Vigouroux Y. Genome scan reveals selection acting on genes linked to stress response in wild pearl millet. *Mol Ecol*. 2016;25(21):5500–12.
29. Punnuri SM, Wallace JG, Knoll JE, Hyma KE, Mitchell SE, Buckler ES, Varshney RK, Singh BP. Development of a High-Density Linkage Map and Tagging Leaf Spot Resistance in Pearl Millet Using Genotyping-by-Sequencing Markers. *Plant Genome*. 9(2).
30. Kulkarni KS, Zala HN, Bosamia TC, Shukla YM, Kumar S, Fougat RS, Patel MS, Narayanan S, Joshi CG. De novo Transcriptome Sequencing to Dissect Candidate Genes Associated with Pearl Millet-Downy Mildew (*Sclerospora graminicola*Sacc.) Interaction. *Front Plant Sci*. 2016;7(252):847.
31. Burgarella C, Cubry P, Kane NA, Varshney RK, Mariac C, Liu X, Shi C, Thudi M, Couderc M, Xu X. A western Sahara Centre of domestication inferred from pearl millet genomes. *Nat Ecol Evol*. 2018.
32. Aaron B, Ron M, Nobuhiro S. ROS as key players in plant stress signalling. *J Exp Bot*. 2014;65(5):1229–40.
33. Wang L, Ma KB, Lu ZG, Ren SX, Jin B. Differential physiological, transcriptomic and metabolomic responses of arabidopsis leaves under prolonged warming and heat shock. *BMC Plant Biol*. 2020;20:86.
34. Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K. Transcriptional regulatory network of plant heat stress response. *Trends Plant Sci*. 2017;22(1):53–65.
35. Bokszzanin KL, Fragkostefanakis S. Perspectives on deciphering mechanisms underlying plant heat stress response and thermotolerance. *Front Plant Sci*. 2013;4(315):315.
36. Saidi Y, Finka A, Goloubinoff P. Heat perception and signalling in plants: a tortuous path to thermotolerance. *New Phytol*. 2011;190(3):556–65.
37. McAdam SA, Brodribb TJ, Banks JA, Hedrich R, Atallah NM, Cai C, Geringer MA, Lind C, Nichols DS, Stachowski K. Abscisic acid controlled sex before transpiration in vascular plants. *Proc Natl Acad Sci*. 2016;113(45):12862–7.
38. Mcadam SAM, Brodribb TJ. Fern and lycophyte guard cells do not respond to endogenous abscisic acid. *Plant Cell*. 2012;24(4):1510–21.
39. Triin V, Hannes K, Yong-Fei W, Noriyuki N, Wai-Yin C, Gabriel V, Airi LK, Mikael B, Heino M, Radhika D. SLAC1 is required for plant guard cell S-type anion channel function in stomatal signalling. *Nature*. 2008;452(7186):487–91.
40. Magwanga RO, Lu P, Kirungu JN, Lu H, Wang X, Cai X, Zhou Z, Zhang Z, Salih H, Wang K. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. *BMC Genet*. 2018;19(1):6.
41. Zandalinas SI, Mittler R, Balfagón D, Arbona V, Gómez-Cadenas A. Plant adaptations to the combination of drought and high temperatures. *Physiol Plant*. 2017;162(1).
42. Huang LK, Yan HD, Zhao XX, Zhang XQ, Wang J, Frazier T, Yin G, Huang X, Yan DF, Zang WJ, et al. Identifying differentially expressed genes under heat stress and developing molecular markers in orchardgrass (*Dactylis glomerata*L.) through transcriptome analysis. *Mol Ecol Resour*. 2015;15(6):1497–509.
43. Huang L, Feng G, Yan H, Zhang Z, Bushman BS, Wang J, Bombarely A, Li M, Yang Z, Nie G. Genome assembly provides insights into the genome evolution and flowering regulation of orchardgrass. *Plant Biotechnol J*. 2019.
44. Zhou S, Chen J, Lai Y, Yin G, Chen P, Pennerman KK, Yan H, Wu B, Zhang H, Yi X. Integrative analysis of metabolome and transcriptome reveals anthocyanins biosynthesis regulation in grass species *Pennisetum purpureum*. *Ind Crop Prod*. 2019;138:111470.
45. Bo L, Victor R, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*. 2010;26(4):493–500.
46. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12(1):1–16.
47. Likun W, Zhixing F, Xi W, Xiaowo W, Xuegong Z. DESeq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*. 2010;26(1):136–8.
48. Mingfeng D, Guijie D. Analysis of SSR loci of functional gene linked to drought resistance based on Transcriptome sequences in *Pinus massoniana* under drought stress. *For Res*. 2018.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

