Acta Physiologiae Plantarum (2019) 41:66
https://doi.org/10.1007/s11738-019-2855-4

ORIGINAL ARTICLE

Transcriptomes of Indian barnyard millet and barnyardgrass reveal putative genes involved in drought adaptation and micronutrient accumulation

Murukarthick Jayakodi[1] · Manimekalai Madheswaran[2] · Karthikeyan Adhimoolam[3] · Sampath Perumal[4] · Dhasarathan Manickam[3] · Thangaraj Kandasamy[2] · Tae-Jin Yang[1] · Senthil Natesan[3,5]

Received: 14 April 2018 / Revised: 28 March 2019 / Accepted: 1 April 2019 / Published online: 8 April 2019

Abstract

Indian barnyard millet (*Echinochloa frumentacea*) is a rich source of dietary fiber, minerals and protein. The lack of genetic resources has slowed the discovery of genes involved in its nutrient accumulation and climate resilience. Here, we present the first transcriptomes of *E. frumentacea* [97,065 transcripts, including 65,276 protein-coding transcripts, over 90% of which have been functionally annotated, and 31,789 long noncoding RNA (lncRNA) transcripts] and its wild relative *E. crus-galli* (93,725 transcripts, including 68,480 protein-coding transcripts, 89% of which have been annotated). Comparative transcriptome analysis identified 4159 protein-coding and 2258 lncRNA transcripts in Indian barnyard millet that showed either up- or down-regulated expression when compared with *E. crus-galli*, and 3489 protein-coding transcripts unique to Indian barnyard millet were also detected. Additional analysis revealed that photosynthesis is likely crucial in the drought adaptation of Indian barnyard millet. We then identified possible genes regulation responsible for drought tolerance and Fe and Zn accumulation. Moreover, based on the simple sequence repeat (SSR)-containing sequence, 30 SSR primer pairs were arbitrarily selected, synthesized and used to screen the 30 *E. frumentacea accessions*. Of these, 10 SSR primers were polymorphic. Collectively, our results enhance the knowledge of micronutrient accumulation and drought tolerance in Indian barnyard millet, as well as of the genetic diversity of *Echinochloa* species.

Keywords Drought · Simple sequence repeats · Indian barnyard millet · Micronutrients

Introduction

Barnyard millet (*Echinochloa* spp.) is the oldest domesticated millet and mainly growing in the semi-arid tropics of Asia and Africa. The genus *Echinochloa* belongs to Poaceae family and it includes 35 species (Mabberley 1997), including the major weed species of *E. crus-galli* (barnyard grass)

Communicated by A.Chandra.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11738-019-2855-4) contains supplementary material, which is available to authorized users.

✉ Tae-Jin Yang
tjyang@snu.ac.kr

✉ Senthil Natesan
senthil_natesan@tnau.ac.in

[1] Department of Plant Science, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea

[2] Department of Plant Breeding and Genetics, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai, Tamil Nadu 625104, India

[3] Department of Biotechnology, Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai, Tamil Nadu 625104, India

[4] Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, SK S7N 0X2, Canada

[5] Department of Plant Molecular Biology and Bioinformatics, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu 641003, India

and *E. colona* (jungle-rice) and domesticated millets of *E. frumentacea* (Indian barnyard millet) and *E. esculenta* (Japanese barnyard millet) (Perumal et al. 2016). *E. crus-galli* has been distributed around the world as a weed, more plentifully in warm temperate regions than in the tropics, whereas *E. colona* is distributed as a common weed in tropical Asia, America and Africa. *E. esculenta* is domesticated from *E. crus-galli* and mostly grown in Japan, Korea and northern part of China. *E. frumentacea* is domesticated from *E. colona* and cultivated in India and several parts of Africa. In India, *E. frumentacea* is an important dry land crop, mainly growing in Himalayan region in the north and Deccan plateau in the south. It is one of the important minor millets growing fast and cultivated over a wide array of environmental conditions and poor soils (Sood et al. 2015). Recently, the crop has gained renewed interest as a health-food due to its rich nutritional profile such as carbohydrate (65%), dietary fiber (9.8%) and protein content (11.1%) that are linked to various health benefits (Hadimani and Malleshi 1993; Veena et al. 2005). Moreover, Kumari and Thayumanavan (1997) reported *E. frumentacea* is capable of reducing blood glucose and lipid levels than the other minor millets.

At present, micronutrient malnutrition is one of the major issues in India and has peaked in the present century (Kotecha 2008). Iron (Fe) and zinc (Zn) deficiencies are major risk factors in the global burden of diseases. It is important that both Fe and Zn have to be considered together because their combined effect is higher than when taken individually. Anemia caused by lack of iron affects more than 1.6 billion people worldwide or approximately 25% of the world population (Benoist et al. 2008). In addition to their importance in the human diet, Fe and Zn ions are also essential microelements for plant cellular functions (Nishizawa 2005), including chlorophyll biosynthesis, photosynthesis, and respiration. Breeding staple grain crop varieties with high Fe and Zn content is a successful strategy to help address micronutrient malnutrition (Pfeiffer et al. 2007). Indian barnyard millet is rich in micronutrients such as Fe and Zn and considered a potential food crop of the future, there are almost no genetic or genomic resources available and no comprehensive studies have been conducted on it to date. In recent years, advances in next-generation sequencing (NGS) technologies provide excellent chances for creating the genomic resources and unveiling important molecular mechanisms controlling such specific biological processes.

In this study, we report the transcriptome of Indian barnyard millet variety CO (KV) 2 leaves. The transcriptome data were de novo assembled and analyzed independently as protein-coding and putative long noncoding RNA (lncRNA) transcripts. Specifically, the objectives of this study were to (1) characterize the transcriptome of *E. frumentacea*; (2) identification of key genes responsible to drought tolerance and high Fe and Zn accumulation; (3) study the frequency and distribution of SSRs and develop SSR markers and (4) analyze the level of polymorphism in 30 diverse Indian barnyard millet genotypes.

Materials and methods

Plant material and RNA extraction

CO (KV) 2, a popular high-yielding Indian barnyard millet variety in southern India, was obtained from Millet Breeding Station, Tamil Nadu Agricultural University, Coimbatore, India, and used for transcriptome sequencing. In addition to high yield, CO (KV) 2 has high Fe and Zn contents and is drought tolerant and non-lodging (Nirmalakumari et al. 2009). Leaf samples at the fourth leaf stage were used for sequencing. The leaf samples were stored at – 80 °C in a freezer for later use. Total RNA from the leaf samples was isolated using a Plant RNeasy mini kit (QIAGEN, Germany) and/or Hybrid-R kit (GeneAll, Korea) following the manufacturers' instructions and its quality and quantity were checked using a Bioanalyzer (Agilent Technologies, USA). Approximately 2 µg of total RNA was used for RNA-seq library construction.

Illumina sequencing and data processing

RNA-seq libraries were successfully constructed using an Illumina TruSeq RNA Sample Preparation Kit according to the manufacturer's instructions. Further, libraries were sequenced by Illumina NextSeq 500 platform with paired-end (PE) reads of 150 bp at Lab Genomics Co. (Pankyo, Korea). Raw reads generated for *E. crus-galli* were downloaded from the NCBI SRA with the accessions SRR1287659, SRR1287657, and SRR1287658. The raw reads obtained from PE sequencing were processed with three major steps to get good-quality reads. First, reads containing bacterial contaminants were omitted through read mapping onto the available bacterial genome sequences by BWA (Li and Durbin 2009). Then, ribosomal (rRNA) reads and PCR duplicates were filtered using SortMeRNA (Kopylova et al. 2012) and FastUniq (Xu et al. 2012), respectively. Finally, quality control was carried out to filter low-quality reads using NGS QC Toolkit (Patel and Jain 2012) (v2.3.3).

De novo assembly and functional annotation

De novo assembly was done using Trinity (Grabherr et al. 2011) (trinityrnaseq-2.1.0) to develop contigs/transcripts. From the total transcripts, the lncRNA transcripts were separated using our pipeline (Fig. S1) to select only protein-coding mRNA transcripts for downstream analysis. The selected protein-coding transcripts were annotated using BLASTX

with *E*-value cut-off of 1E-05 against the protein datasets of foxtail millet (*S. italica*: https://foxtailmillet.genomics.org.cn/page/species/download.jsp), rice (*O. sativa*: ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/) and sorghum (*S. bicolor*: https://www.plantgdb.org/SbGDB/). Gene ontology (GO) terms were obtained based on annotation against the NCBI Nr proteins using Blast2GO program (Conesa et al. 2005). KEGG pathway annotation by KAAS (Moriya et al. 2007) was used to predict transcripts associated with various metabolic pathways.

Expression profiling and GO enrichment analysis

Filtered high-quality RNA-seq reads were mapped to protein-coding and lncRNA transcripts separately to calculate FPKM (Fragments per kilobase per million) using RSEM (Li and Dewey 2011) (v1.2.4). Bioconductor package edger (Robinson et al. 2010) was used to identify differentially expressed transcripts between Indian barnyard millet and wild *E. crus-galli*. Transcripts exhibiting over twofold changes with a significant false discovery rate (FDR) of 0.001 were considered as differentially expressed. Indian barnyard millet-specific expression was determined by the condition that FPKM value > 1 in Indian barnyard millet samples and FPKM < 0 in *E. crus-galli* samples. GO enrichment analysis was conducted on the specifically and differentially expressed transcripts by Fisher's exact test with multiple testing correction of FRD with cut-off 0.05.

Validation of differentially expressed genes through qRT-PCR analysis

Following differential expression analysis, transcripts showing upregulated expression in *E. frumentacea* and downregulated expression in *E. crus-galli* were selected. Annotation details of those transcripts were searched and different families of genes related to drought tolerance and Fe and Zn accumulation were chosen for validation through quantitative real-time PCR (qRT-PCR). To verify the transcripts involved in drought tolerance, drought stress was imposed on 20-day-old plants of *E. frumentacea* and *E. crus-galli* by withholding water for 5 days, while control plants were watered on alternate days and leaves were independently collected for three repeats. On the other hand, to verify the transcripts involved in Fe and Zn accumulation, leaves from *E. frumentacea* and *E. crus-galli* were independently collected at the 4-leaf stage for three repeats. All the collected leaves frozen in liquid nitrogen immediately, and stored at − 80 °C. Total RNA was extracted using a total RNA isolation kit according to the manufacture instructions (Roche, Berlin and Germany). After measuring the RNA quality and quantity, it was successfully converted in to cDNA using a Transcriptor First Strand cDNA Synthesis Kit (Roche), according to the manufacturer's instructions. Primer pairs were designed from selected transcript sequences using Primer Premier 6.0. Actin was used as an internal control and quantitative real-time PCR (qRT-PCR) analysis was conducted using a Light cycler 480 Real-Time PCR System (Roche) following to the manufacture instructions. The qRT-PCR mixtures contained 2.5 µL (approximately 50 ng) of first-strand cDNA, 1.0 µL of gene-specific primers (10.0 mM), 10.0 µL of FastStart Essential DNA Green Master mix (Roche) and 6.5 µL of ddH_20. PCR conditions were as follows: 95 °C for 10 s, followed by 35 cycles at 95 °C for 20 s, 60 °C for 15 s, 72 °C for 15 s and then melting for 95 °C for 5 s, 60 °C for 1 s and 97 °C for continuous and cooling for 40 °C for 30 s. Expression levels were quantified using the $\Delta\Delta^{Ct}$ method.

SSR marker discovery and validation

Simple sequence repeat (SSR) locus detection was performed using the Phobos plugin tandem repeat finder available in the Geneious software program (https://www.ruhr-uni-bochum.de/ecoevo/cm/cm_phobos.htm). SSRs were defined as di-, tri-, tetra-, penta-, and hexa-nucleotides. Primer premier 6.0 (PREMIER Biosoft International, Palo Alto, California USA) was used to design PCR primers. In total, 300 pairs of primers were designed (Table S3). The criteria for designing primers from the SSR flanking sequences were primer length range of 18–25 bases with an optimum of 22 bases, PCR product size range of 100–200 bp, annealing temperature between 50–60 °C with 55 °C as the optimum melting temperature, and a GC content of 40–60% with an optimum of 50%. A total of 30 Indian barnyard millet accessions, including cultivars and landraces, were selected for polymorphism investigation with the SSRs. The young leaves collected and were ground in liquid nitrogen and genomic DNA was isolated by CTAB method (Doyle 1987). From the 300 possible primer pairs, 30 primers were arbitrarily selected, synthesized and validated by polymerase chain reaction (PCR). PCR and agarose gel (3%) electrophoresis following the method used by Manimekalai et al. (2018).

Results

De novo assembly of the *E. frumentacea* transcriptome

We obtained 373,803,040 paired-end (PE) reads with read length of 150 bp via transcriptome sequencing. All raw reads generated in this study were deposited into the NCBI SRA (accession number SRR5860804). By filtering raw reads, high-quality sequencing reads were selected and used for

further assembly and expression analysis. This yielded a total of 125 million high-quality reads, which we assembled to obtain a total of 97,065 transcripts. We used our pipeline (Fig. S1) to filter out 31,789 putative lncRNA transcripts from among the total transcripts, which left 65,276 putative protein-coding transcripts with average and N50 sizes of 1440 bp and 1864 bp, respectively (Table 1).

Functional annotation of genes

A BLASTX-based sequence similarity search against the closely related protein datasets of foxtail millet, sorghum and rice to assign functional descriptions to each protein-coding transcript revealed 92.9%, 93.2% and 93.7% of *E. frumentacea* transcripts, respectively, to be significantly similar to proteins in foxtail millet, rice and sorghum (Table 1). We also assigned GO terms to the protein-coding transcripts: 44,166 (68%) of the transcripts were assigned at least one GO term. Of the assigned terms, 42,335 (64%), 38,136 (58%), and 37,427 (57%), respectively, fell into the categories of cellular component, biological process and molecular function. A large number of transcripts were grouped into the cellular component category, in which transcripts were enriched for GO terms cell, cell part and intracellular part (Fig. 1). The KEGG pathway database can facilitate a systematic understanding of the network of molecular interactions among genes/proteins. Such pathway-based analysis enables further understanding of biological functions. A

Table 1 Details of Indian barnyard millet de novo transcriptome assembly and functional annotation statistics

De novo transcriptome assembly	*E. frumentacea*	*E. crus-galli*[a]
Total number of transcripts	97,065	93,725
No. of filtered putative coding transcripts	65,276	68,480
Total size of transcriptomes	94 Mbp	108 Mbp
Maximum sequence length (bp)	15,596	19,453
Average sequence length (bp)	1440	1580
N50 length (bp)	1864	1917
Functional annotations	% of coding transcripts	% of coding transcripts
Foxtail millet (*S. italic*)	92.9	89.5
Rice (*O. sativa*)	93.2	89.1
Sorghum (*S. bicolor*)	93.7	89.9

[a]Raw data were retrieved from Nah et al. (2015)

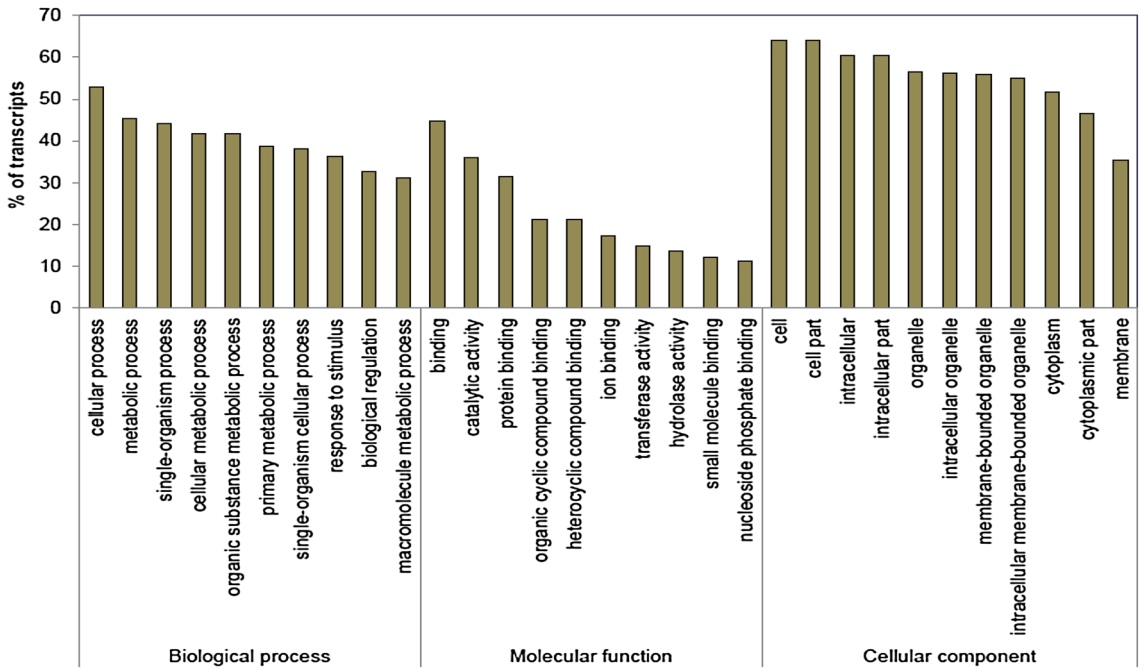

Fig. 1 Functional annotation of Indian barnyard millet (*E. frumentacea)* assembled transcripts based on gene ontology (GO) terms

total of 5,975 of the *E. frumentacea* transcripts were placed into the five top KEGG Orthology (KO) categories: organismal systems, cellular processes, environmental information processing, genetic information processing and metabolism (Fig. 2).

Comparative transcriptomes between *E. frumentacea* and *E. crus-galli*

We retrieved the raw sequencing reads from leaf samples (EC-SNU1, EC-SNU2, EC-SNU3) of *E. crus-galli* generated in Nah et al. (2015) as well as further reads deposited in the NCBI database. These 482,816,422 raw reads were processed to obtain 353,574,244 high-quality reads. These numbers are similar to those we generated for *E. frumentacea.* We performed de novo assembly and annotation on the *E. crus-galli* data, which yielded 68,480 protein-coding transcripts with average length of 1580 bp (Table 1). Approximately 89% of the protein-coding transcripts were annotated using BLASTX similarity searches against foxtail millet, rice and sorghum. Then, differential expression analysis was performed between *E. frumentacea* and *E. crus-galli* using the protein-coding and lncRNA transcripts in *E. frumentacea* as reference and mapping the *E. crus-galli* reads onto *E. frumentacea* separately. From this analysis, we found that 4,159 protein-coding (Fig. 3a) and 2258 long noncoding transcripts (Fig. 3b) showed differential expression between Indian barnyard millet and the wild weed *E. crus-galli*. In addition, GO enrichment analysis of the differentially expressed transcripts showed that they enriched for transcripts assigned to photosynthesis and light harvesting-related processes (Table S1). Further, we found that approximately 5% of transcripts were expressed only in *E. frumentacea* (Fig. 4). GO enrichment analysis of these Indian barnyard millet-specific transcripts showed that many of them were involved in photosynthetic electron transport, external stimulus and cytochrome complex assembly (Table S2).

Differentially expressed genes involved in drought tolerance and Fe and Zn accumulation

Comparative transcriptome data of *E. frumentacea* and *E. crus-galli* revealed that transcripts belonging to different families of genes involved in drought tolerance and higher Fe and Zn accumulation. The transcripts involved in biological processes such as high photosynthetic activity, including chlorophyll binding, light harvesting in photosystem I and II and photosynthetic electron transport response to drought tolerance were identified. To confirm our results, transcripts belonging to different families of genes encoding photosystem II P680 reaction center D1 protein, photo receptor super family, photosystem II CP43 chlorophyll apoprotein, photosystem I P700 chlorophyll a apoprotein A1, photosystem II PsbM protein, photosystem I subunit PsaO and photosystem II 22-kDa protein were selected and tested by qRT-PCR analysis under drought stress condition. All of the tested genes showed higher expression in *E. frumentacea* than *E. crus-galli* (Fig. 5), consistent with the results found with comparative transcriptome data. We

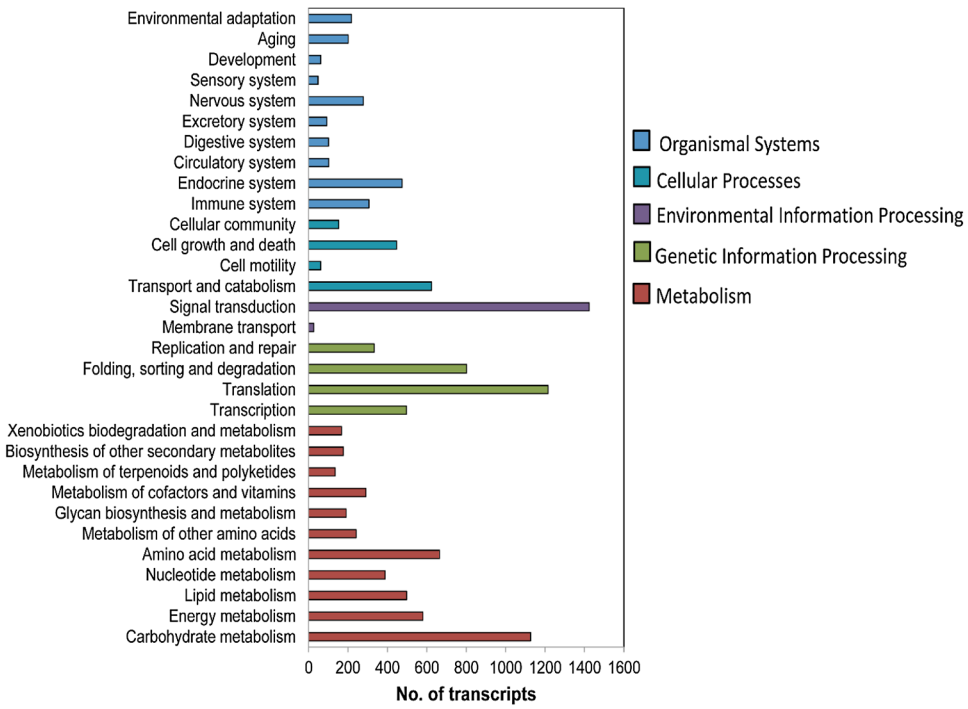

Fig. 2 KEGG pathway annotation by KAAS. Numbers of transcripts annotated are grouped into five different categories

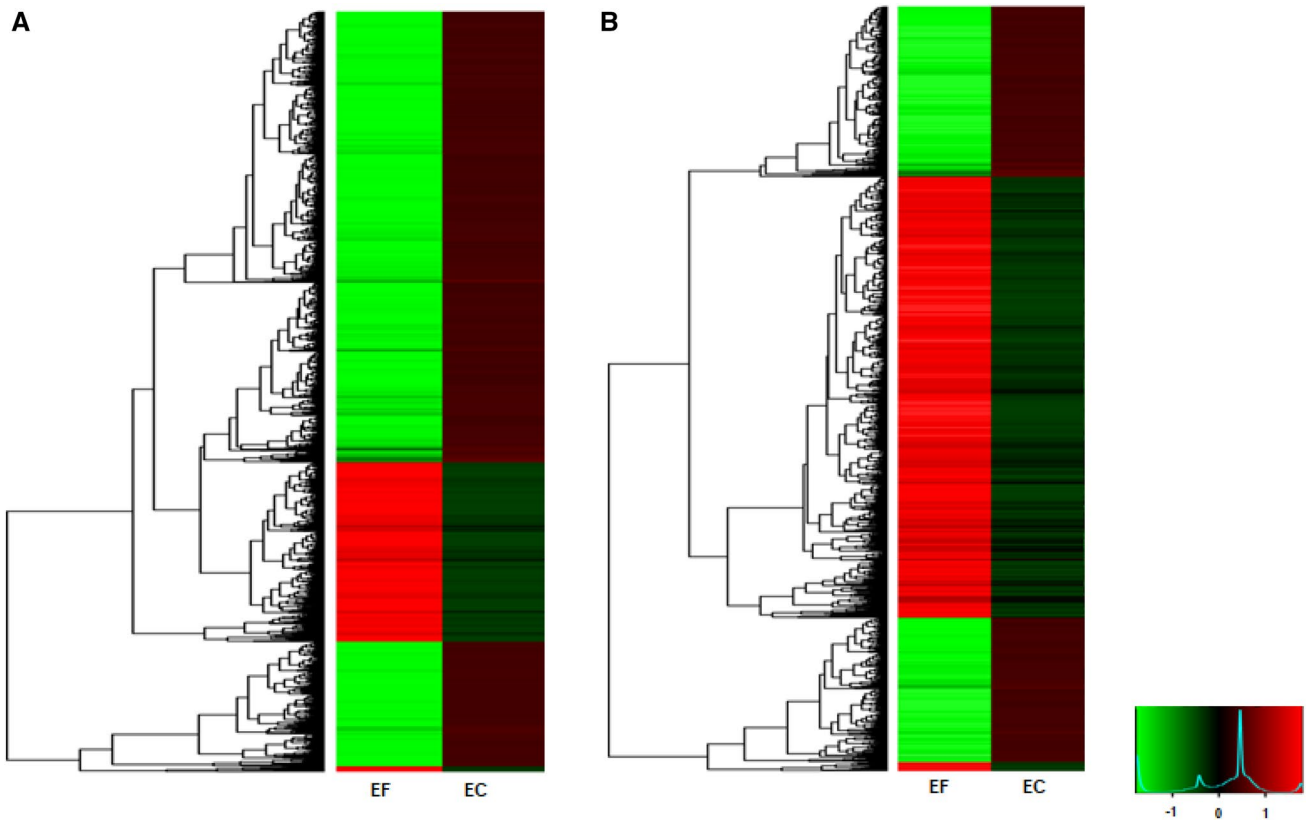

Fig. 3 Differential expression (DE) analysis between *E. frumentacea* and *E. crus-galli*. **a** Heat map of FPKM values of protein-coding transcripts over twofold differentially expressed between the domesticated Indian barnyard millet *E. frumentacea* and the wild weed *E. crus-galli*. **b** Heatmap of differentially expressed lncRNA transcripts based on FPKM

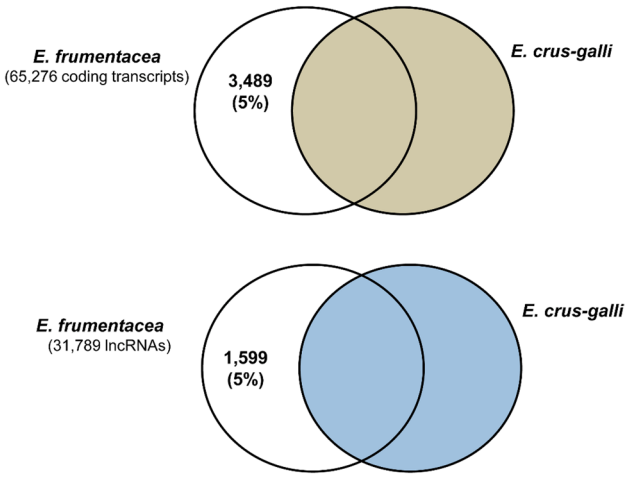

Fig. 4 *E. frumentacea*-specific coding and lncRNA transcripts (clear circles) revealed by comparison of the leaf transcriptome to transcripts from *E. crus-galli* (shaded circles)

also further identified the Fe and Zn accumulation responsible transcripts encoding vacuolar protein sorting-associated protein (VAP), helix loop helix (HLH) protein, heavy metal domain (HMD) protein, ubiquitin, uncharacterized iron-regulated protein (IRP), β glucosidase, serine/threonine protein kinases (STK/TPKs), NAC domain-containing protein and cytochrome 450. The changes in expression of Fe and Zn accumulation responsible genes determined by comparative transcriptome data were confirmed by qRT-PCR analysis between *E. frumentacea* and *E. crus-galli* (Fig. 5).

Frequency and distribution of SSR loci in the Indian barnyard millet transcriptome

Ten thousand eight hundred eighty-one SSR loci were discovered in 65,276 transcript sequences of Indian barnyard millet. Of these, the di-nucleotides were the most abundant repeats (49.02%), followed by tri- (47.74%), tetra- (2.35%), penta- (0.67%) and hexa-repeat (0.21%) repeats (Fig. 6). Among the di-repeats, the AG and AT types were the most frequent, accounting for 41.73% and 25.13% of the di-repeat SSRs, respectively. The most common type of tri-repeat was CCG (23.25%) of the repeats, followed by AGC (15.09%) and AGG (12.83%), respectively. Among the longer repeat types, 5-bp motifs appeared to be the most common, accounting for 58.56% of the longer repeats, followed by

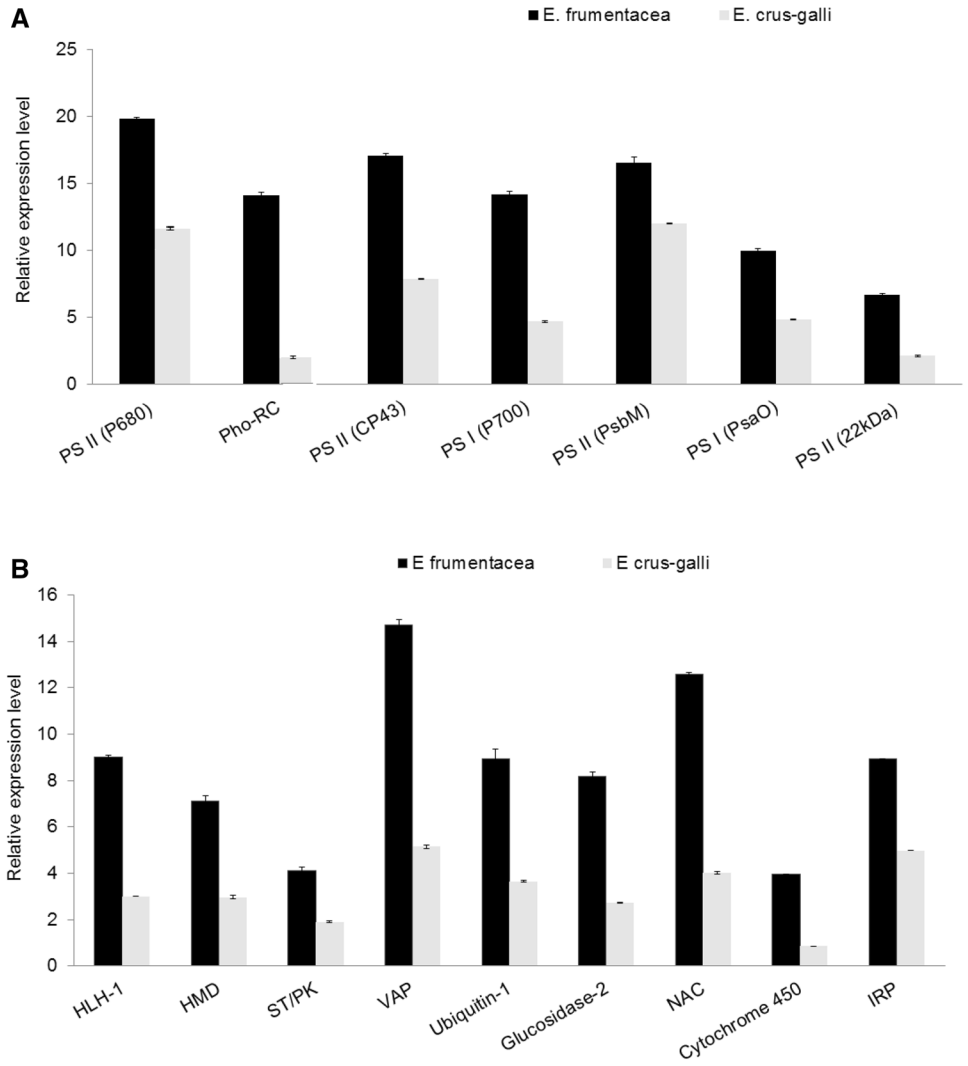

Fig. 5 Validation of differential expression of genes responsible to drought tolerance (**a**) and Fe and Zn accumulation (**b**). The bars represent the values from qRT-PCR analysis of putative genes annotated as encoding. Photosystem II P680 reaction center D1 protein [PS II (P680)], photo receptor super family (Pho-RC), photosystem II CP43 chlorophyll apoprotein [PS II (CP43)], photosystem I P700 chlorophyll a apoprotein A1 [PS I (P700)], photosystem II PsbM protein [PS 11 (PsbM)], photosystem I subunit PsaO [PS 1 (PsaO)] and photosystem II 22-kDa protein [PS II (22-kDa)], helix loop helix-1 (HLH-1), heavy metal domain (HMD), serine/threonine protein kinases (ST/PK), vacuolar protein sorting-associated protein (VAP), ubiquitin-1, glucosidase-2, NAC domain-containing protein (NAC), cytochrome 450, and uncharacterized iron-regulated protein (IRP)

6-, 7-, 8-, 9-, 11- and 10-bp motifs accounting for 17.87%, 14.32%, 3.66%, 2.20%, 1.31% and 0.25% of the longer repeats, respectively. Frequencies of different SSRs and percentages of different SSR motifs are presented in Fig. 6.

Development and applicability of the EST-SSR markers in the Indian barnyard millet genotypes

SSR primer pairs were designed for 300 of the 10,881 SSR loci identified, including 121 (40.33%) dinucleotide repeats, 122 (40.66%) trinucleotide repeats, 35 (11.66%) tetranucleotide repeats, 20 (6.66%) pentanucleotide repeats, and 2 (0.66%) hexanucleotide repeats. Of those 300 SSR primer pairs, 30 were randomly selected, synthesized and tested for stable amplification as well as polymorphism in Indian barnyard millet germplasm. Out of the tested SSR primers, DNA amplification was obtained with 28 primers (90%), and 24 primers (70%) yielded PCR products of the expected fragment size (100–300 bp). We used the 24 pairs that showed stable amplification to explore the polymorphism in 30 Indian barnyard millet accessions. Ten EST-SSR primers: BMESSR 2, BMESSR 8, BMESSR 16, BMESSR 22, BMESSR 26, BMESSR 30, BMESSR 33, BMESSR 34 and BMESSR 39, showed polymorphism in the 30 Indian barnyard millet germplasms (Fig. S2), while the remaining 14 were monomorphic. It revealed that newly developed SSR primers had good amplification efficiency and were useful to discover the polymorphism in Indian barnyard millet accessions.

Discussion

Transcriptomes of *Echinochloa* species

Transcriptome sequencing is a potential strategy for identifying transcript sequences and for discovering novel genes associated with diseases, important secondary metabolic

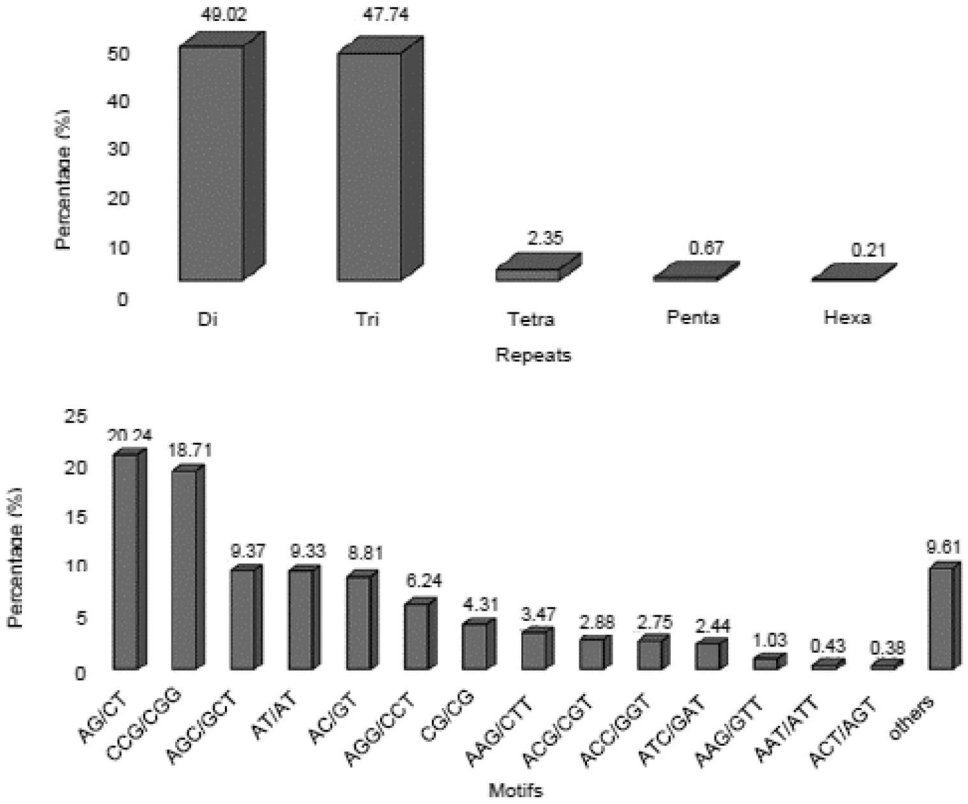

Fig. 6 Distribution and frequency of EST-SSR markers in Indian barnyard millet. **a** Distribution of SSRs by type. **b** Frequencies of different SSRs and SSR motifs

pathways (Jayakodi et al. 2015) and developing molecular markers (Han et al. 2013; Vatanparast et al. 2016). This approach has been successfully applied in many important crops, including sesame (Wei et al. 2011), chickpea (Garg et al. 2011), finger millet (Kumar et al. 2015), foxtail millet (Jo et al. 2016), peanut (Zhang et al. 2012) and maize (Yue et al. 2016). Although the genus *Echinochloa* contains a number of noxious weeds as well as two species of cereal, up until now transcriptome resources have been available only for *E. crus-galli* (Yang et al. 2013; Nah et al. 2015). In this study, we generated the first transcriptome resource for *E. frumentacea*, a domesticated species in the genus *Echinochloa*. Our results showed the transcriptome of *E. frumentacea* to be very similar to that of *E. crus-galli*. In addition, we assembled new transcripts for *E. crus-galli,* providing an unbiased resource for comparative transcriptome analysis and genetic diversity studies. The prevalence of lncRNAs in transcriptome sequencing is common, due to their mRNA-like structure, but they should be analyzed separately from protein-coding transcripts for a better view of the cellular processes likely to be active in the tissue under study. We systematically screened the putative lncRNAs we identified, and although some mis-assembled transcripts lacking open reading frames (ORFs) were identified as lncRNAs, the over 90% success in functional annotation in Indian barnyard millet clearly showed the accuracy of our assembly, as well as our exclusive selection of coding mRNAs. This study provides rich resources for protein-coding and regulatory noncoding transcripts that can be used to accelerate functional genomic research and applications in Indian barnyard millet.

Support for *E. frumentacea* as a potential source for drought-resistance genes

Rapid growth, drought tolerance and high micronutrient contents make Indian barnyard millet a valuable food crop for arid and semi-arid regions or famine areas (Sood et al. 2015). Its wild relative *E. crus-galli* has widespread habitats, from dry land to flooded paddy rice fields (Yamasue et al. 1989a, b; Nah et al. 2015), but shows only intermediate drought resistance between lowland and upland weeds (Yamasue et al. 1989a). The mechanism of Indian barnyard millet's adaptation to drought has not been well understood. Previous studies have shown that photosynthesis and cell growth are two of the primary processes affected by drought (Osakabe et al. 2014). Zivcak et al. (2013) demonstrated the role of alternative electron sinks and cyclic electron flow in photoprotection of PSII and PSI in drought stress in wheat. Comparative transcriptome data by Zhang et al. (2016) suggested that photosynthesis-related genes such as those encoding *ZF-HD_dimer* and Bzip-1 contributed to the drought tolerance in upland rice. Although Indian barnyard millet and *E. crus-galli* both

possess C_4 photosynthetic machinery, we found that the differentially expressed and Indian barnyard millet-specific transcripts were enriched for biological processes related to high photosynthetic activity, including light harvesting in photosystem I and II, photosynthetic electron transport and chlorophyll binding. This was confirmed by our qRT-PCR results (Fig. 5), exhibiting higher gene expression in Indian barnyard millet compared to *E. crus-galli* under drought stress condition. Based on these findings, we hypothesize that Indian barnyard millet might survive under drought stress using drought escape mechanisms involving high photosynthetic capacity along with rapid growth, which is in agreement with one of the drought-adaptive mechanisms in millet proposed by Tadele (2016).

Fe and Zn accumulation in *E. frumentacea*

Minor millets are nutritionally comparable or even superior to major cereals such as rice and wheat (Muthamilarasan et al. 2016). However, no reports describing genes/QTLs for nutritional traits in millets have been available. In this study, we documented higher Fe and Zn levels in Indian barnyard millet than in *E. crus-galli* (Table 2). We identified nine candidate genes responsible for this high Fe and Zn accumulation in Indian barnyard millet, including genes encoding VAP and NAC responsible for high levels of Fe and Zn in chickpea (Upadhyaya et al. 2016) and wheat (Uauy et al. 2006), ubiquitin (associated with seed Zn concentration in chickpea), HLH (Colangelo and Guerinot 2004), β glucosidase, ST/PK and cytochrome 450 reported to improve zinc content in cabbage (Li et al. 2014) and maize (Jin et al. 2015) and HMD related to Fe/Zn accumulation in rice (Chandel et al. 2011). The higher expression of these genes in Indian barnyard millet compared to *E. crus-galli* was confirmed by qRT-PCR analysis, further supporting their possible involvement in the Fe and Zn accumulation characteristic of Indian barnyard millet. These genes may be useful for increasing the Fe and Zn accumulation in commercial cultivars. Additional transcriptome sequencing at different stages of grain development and an integrated framework that combines genomics, transcriptomics and metabolomics of Indian barnyard millet would efficiently provide deep insights into nutrient accumulation and help as a model system to unveil the molecular basis for micronutrient accumulation.

Table 2 Fe and Zn content of *E. frumentacea* and *E. crus-galli*

Species	Fe (mg/100g)	Zn (mg/100g)
E. frumentacea	15.39	4.28
E. crus-galli	11.84	3.42

SSR identification and characterization

SSR markers are cost effective and excellent molecular marker system with the features of highly reproducible and polymorphic, abundant and being co-dominant. It is also successfully used in the linkage map construction (Varshney et al. 2007), diversity assessment of germplasm and identification of molecular markers for MAS (Kalia et al. 2011; Miah et al. 2013; Pandey et al. 2013). Numerous studies have explained that the simple method to identify SSR loci is through transcriptome via NGS (Obidiegwu et al. 2014; Gimode et al. 2016). In this study, the di-nucleotides were the most abundant repeat types (49.02%), followed by tri- (47.74%), tetra- (2.35%), penta- (0.67%) and hexa-repeat (0.21%) motifs. These results are in agreement with the findings of Zhang et al (2014) in foxtail millet and Mondal and Ganie (2014) in rice. Among the di-nucleotide repeats, AG was the most frequent motif, which is agreement with previous reports in foxtail millet (Zhang et al. 2014) and sweet potato (Patel and Jain 2012). In trinucleotide repeats, CCG was most abundant, constant with previous reports on maize, rice, sorghum and wheat (Grabherr et al. 2011), barley (Conesa et al. 2005), sweet potato (Patel and Jain 2012) adzuki bean (Moriya et al. 2007) and foxtail millet (Li and Dewey 2011). Twenty-four of 30 primer pair's successfully yielded amplicons at the expected sizes, and ten primer pairs were identified as polymorphic. It shows that newly developed SSR primers had good amplification efficiency and were potential to analyze the genetic diversity in barnyard millet accessions. The 10,881 SSR loci discovered in this study will offer a wealth of markers for further genetic studies in barnyard millet.

Conclusion

Indian barnyard millet is a versatile cereal with nutritious properties and climate-change resilient characteristics. This work represents is the first attempt to assemble and characterize the transcriptome of *E. frumentacea* and to compare it with a wild relative. We have identified key genes regulating a response in Fe and Zn accumulation as well as provided evidence pointing to a role for photosynthetic machinery in conferring drought stress. The SSRs developed in this study offer a marker resources and useful for construction of linkage map, analysis of genetic diversity, comparative genomics and other genetic studies in *E. frumentacea.* Our results and the resources generated in this study will facilitate discovery of new genes and further EST-SSR markers and thereby accelerate both genetic and functional genomic research in *E. frumentacea*.

Author contribution statement SN and TJY conceived and designed the study. SP and MM prepared samples and RNA-Seq libraries. MJ did all bioinformatics analysis. MM and KA performed qRT-PCR experiments. MM, DM, KA, and TK, designed and analyzed SSR markers between millet genotypes. MJ, MM, KA, TJY, and SN wrote the manuscript. All authors have read and approved the final manuscript.

Acknowledgements All the authors wish to acknowledge National Agricultural Development Programme (NADP)/Rashtriya Krishi Vikas Yojana (RKVY)—Tamil Nadu Govt, for providing financial support. Thanks to Centre of Innovation (CI), Agricultural College and Research Institute, Tamil Nadu Agricultural University, Madurai, for providing instrumentation facilities.

Compliance with ethical standards

Conflict of interest The authors have declared that no competing, or conflicts of interest exist.

References

Benoist BD, McLean E, Egli I, Cogswell M (eds) (2008) Worldwide prevalence of anaemia 1993–2005: WHO global database on anaemia. World Health Organization, Geneva

Chandel G, Samuel P, Dubey M, Meena R (2011) In silico expression analysis of QTL specific candidate genes for grain micronutrient (Fe/Zn) content using ESTs and MPSS signature analysis in rice (*Oryza sativa* L.). J Plant Genet Transgenics 2:11–22

Colangelo EP, Guerinot ML (2004) The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. Plant Cell 16:3400–3412

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatic 21:3674–3676

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Garg R, Patel RK, Tyagi AK, Jain M (2011) De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res 18:53–63

Gimode D, Odeny DA, de Villiers EP, Wanyonyi S, Dida MM, Mneney EE, Muchugi A, Machuka J, de Villiers SM (2016) Identification of SNP and SSR markers in finger millet using next generation sequencing technologies. PloS One 11:e0159437

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652

Hadimani NA, Malleshi NG (1993) Studies on milling, physico-chemical properties, nutrient composition and dietary fibre content of millets. J Food Sci Technol 30:17–20

Han XJ, Wang YD, Chen YC, Lin LY, Wu QK (2013) Transcriptome sequencing and expression analysis of terpenoid biosynthesis genes in *Litsea cubeba*. PloS one 8:e76890

Jayakodi M, Lee SC, Lee YS, Park HS, Kim NH, Jang W, Lee HO, Joh HJ, Yan TJ (2015) Comprehensive analysis of *Panax ginseng* root transcriptomes. BMC Plant Biol 15:138

Jin T, Chen J, Zhu L, Zhao Y, Guo J, Huang Y (2015) Comparative mapping combined with homology-based cloning of the rice genome reveals candidate genes for grain zinc and iron concentration in maize. BMC Genet 16:17

Jo Y, Lian S, Cho JK, Choi H, Kim SM, Kim SL, Lee BC, Cho WK (2016) De novo transcriptome assembly of *Setaria italica* variety Taejin. Genom Data 8:121–122

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan A (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309–334

Kopylova E, Noé L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217

Kotecha PV (2008) Micronutrient malnutrition in India: let us say "no" to it now. Indian J Community Med 33:9–10

Kumar A, Gaur VS, Goel A, Gupta AK (2015) De novo assembly and characterization of developing spikes transcriptome of finger millet (*Eleusine coracana*): a minor crop having nutraceutical properties. Plant Mol Biol Rep 33:905–922

Kumari SK, Thayumanavan B (1997) Comparative study of resistant starch from minor millets on intestinal responses, blood glucose, serum cholesterol and triglycerides in rats. J Food Sci Technol 75:296–302

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760

Li J, Liu B, Cheng F, Wang X, Aarts MG, Wu J (2014) Expression profiling reveals functionally redundant multiple-copy genes related to zinc, iron and cadmium responses in *Brassica rapa*. New Phyto 203:182–194

Mabberley DJ (1997) The plant book: a portable dictionary of the vascular plants. Cambridge University Press, Cambridge, p 858

Manimekalai M, Dhasarathan M, Karthikeyan A, Murukarthick J, Renganathan VG, Thangaraj K, Vellaikumar S, Vanniarajan C, Senthil N (2018) Genetic diversity in the barnyard millet (*Echinochloa frumentacea*) germplasms revealed by morphological traits and simple sequence repeat markers. Curr Plant Biol 14:71–78

Miah G, Rafii MY, Ismail MR, Puteh AB, Rahim HA, Islam KHN, Latif MA (2013) A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. Int J Mol Sci 14:22499–22528

Mondal TK, Ganie SA (2014) Identification and characterization of salt responsive miRNA-SSR markers in rice (*Oryza sativa*). Gene 535:204–209

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

Muthamilarasan M, Dhaka A, Yadav R, Prasad M (2016) Exploration of millet models for developing nutrient rich graminaceous crops. Plant Sci 242:89–97

Nah G, Im JH, Kim JW, Park HR, Yook MJ, Yang TJ, Fischer AJ, Kim DS (2015) Uncovering the differential molecular basis of adaptive diversity in three *Echinochloa* leaf transcriptomes. PLoS One 10(8):e0134419

Nirmalakumari A, Subramanian A, Sumathi V, Senthil N, Kumaravadivel N, John A, Mohanasundaram K, Muthiah AR, Raveendran TS, Raguchander T, Manoharan S (2009) A high yielding kudiraivali variety CO (KV) 2. Madras Agric J 96:319–321

Nishizawa NK (2005) The uptake and translocation of minerals in rice plants. In: Heong KL, Hardy B (eds) Toriyama K. IRRI, Philippines, pp 90–93

Obidiegwu ON, Parzies H, Obidiegwu JE (2014) Development and genotyping potentials of EST-SSRs in finger millet (*E. coracana* (L.) Gaertn.). Int J Genet Genom 2:42–46

Osakabe Y, Osakabe K, Shinozaki K, Tran LSP (2014) Response of plants to water stress. Front Plant Sci 5:86

Pandey G, Misra G, Kumari K, Gupta S, Parida SK, Chattopadhyay D, Prasad M (2013) Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [*Setaria italica* (L.)]. DNA Res 20:197–207

Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PloS One 7:e30619

Perumal S, Jayakodi M, Kim DS, Yang TJ, Natesan S (2016) The complete chloroplast genome sequence of Indian barnyard millet, *Echinochloa frumentacea* (Poaceae). Mitochondrial DNA Part B 1:79–80

Pfeiffer WH, McClafferty B (2007) HarvestPlus: breeding crops for better nutrition. Crop Sci 47:88–105

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140

Sood S, Khulbe RK, Gupta AK, Agrawal PK, Upadhyaya HD, Bhatt JC (2015) Barnyard millet—a potential food and feed crop of future. Plant Breed 134:135–147

Tadele Z (2016) Abiotic and biotic stress in plants—recent advances and future perspectives. In: AK Shanker, C Shanker (ed). InTech, London pp 768.

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky JA (2006) NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:1298–1301

Upadhyaya HD, Bajaj D, Das S, Kumar V, Gowda CLL, Sharma S, Tyagi AK, Parida SK (2016) Genetic dissection of seed-iron and zinc concentrations in chickpea. Sci Rep 6:20450

Varshney RK, Marcel TC, Ramsay L, Russell J, Röder MS, Stein N, Waugh R, Langridge P, Niks RE, Graner A (2007) A high density barley microsatellite consensus map with 775 SSR loci. Theor Appl Genet 114:1091–1103

Vatanparast M, Shetty P, Chopra R, Doyle JJ, Sathyanarayana N, Egan AN (2016) Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). Sci Rep 6:29070

Veena B, Chimmad BV, Naik RK, Shantakumar G (2005) Physicochemical and nutritional studies in barnyard millet. Karnataka J Agril Sci 18:101–105

Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X (2011) Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genom 12:451

Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S (2012) FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS One 7:e52249

Yamasue Y, Asai Y, Ueki K, Kusanagi T (1989) Anaerobic seed germination for the habitat segregation in *Echinochloa* weeds. Jpn J Breed 39:159–168

Yamasue Y, Nakamura A, Ueki K, Kusanagi T (1989) Drought resistance for the habitat segregation in *Echinochloa* weeds. Jpn J Breed 39:337–343

Yang X, Yu XY, Li YF (2013) De novo assembly and characterization of the barnyardgrass (*Echinochloa crus-galli*) transcriptome using next-generation pyrosequencing. PloS One 8:e69168

Yue R, Lu C, Qi J, Han X, Yan S, Guo S, Liu L, Fu X, Chen N, Yin H, Chi H, Tie S (2016) Transcriptome analysis of cadmium-treated roots in maize (*Zea mays* L.). Front Plant Sci 7:1298

Zhang J, Liang S, Duan J, Wang J, Chen S, Cheng Z, Zhang Q, Liang X, Li Y (2012) De novo assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). BMC Genom 13:90

Zhang S, Tang C, Zhao Q, Li J, Yang L, Qie L, Fan X, Li L, Zhang N, Zhao M, Liu X, Chai Y, Zhang X, Wang H, Li Y, Li W, Zhi H, Jia G, Diao X (2014) Development of highly polymorphic simple sequence repeat markers using genome-wide microsatellite variant analysis in Foxtail millet [*Setaria italica* (L.) P. Beauv.]. BMC Genom 15:78

Zhang ZF, Li YY, Xiao BZ (2016) Comparative transcriptome analysis highlights the crucial roles of photosynthetic system in drought stress adaptation in upland rice. Sci Rep 6:19349

Zivcak M, Brestic M, Balatova Z, Drevenakova P, Olsovska K, Kalaji HM, Yang X, Allakhverdiev SI (2013) Photosynthetic electron transport and specific photoprotective responses in wheat leaves under drought stress. Photosynth Res 117:529–546