Mol Breeding (2018) 38:95
https://doi.org/10.1007/s11032-018-0852-8

The first Illumina-based de novo transcriptome analysis and molecular marker development in Napier grass (*Pennisetum purpureum*)

Sifan Zhou · Chengran Wang · Taylor P. Frazier · Haidong Yan · Peilin Chen · Zhihong Chen · Linkai Huang · Xinquan Zhang · Yan Peng · Xiao Ma · Yanhong Yan

Received: 30 June 2017 / Accepted: 21 June 2018

Abstract *Pennisetum purpureum* belongs to the *Pennisetum* Rich genus in the family Poaceae. It is widely grown in subtropical and tropical regions as one of the most economically important cereal crops. Despite its importance, there is limited genomic data available for *P. purpureum*, which restricts genetic and breeding studies in this species. In the present study, the transcriptome of *P. purpureum* was assembled de novo and used to characterize two important *P. purpureum* cultivars: *P. purpureum* Schumab cv. Purple and *P. purpureum* cv. Mott. After assembly, a total of 197,466 unigenes were obtained for 'Purple' and 'Mott' and 103,454 of these unigenes were successfully annotated. From 'Purple' and 'Mott,' 214,648 SNPs and 21,213 EST-SSRs were identified in 40,259 unigenes and 18,587 unigenes respectively. Moreover, 50 EST-SSR primers and 6 SNP primers were designed to validate the identified markers. The transcriptomic data of present study from the two *P. purpureum* cultivars provides an abundant amount of available genomic information for *Pennisetum*. In addition, the identified SNPs and EST-SSRs will facilitate genetic and molecular studies within the *Pennisetum* genus.

Keywords Markers development · *Pennisetum purpureum* · Transcriptome sequencing

Introduction

The *Pennisetum* Rich genus belongs to the family Poaceae. There are approximately 140 species within this genus that grow under various environmental conditions all over the world (Brunken 1977). Few species are economic crop as pearl millet (*P. glaucum*), biofertilizer as Kikuyu grass (*P. clandestinum*), and ornamental plants as crimson fountaingrass (*P. setaceum*) and feathertop grass (*P. villosum*) (Fulkerson et al. 2008; Sujatha et al. 1989). In comparison to other tropical cereal crops, members of the *Pennisetum* Rich genus are highly tolerant to abiotic stresses like heat and drought. Therefore, these species can grow in the semi-arid and tropical regions of Asia and Africa, where they are considered as important staple foods (Khairwal et al. 2007; Rai et al. 2009). Few studies, however, have been performed to understand the genetic mechanisms underlying important phenotypic characteristics of members of this genus.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11032-018-0852-8) contains supplementary material, which is available to authorized users.

S. Zhou · C. Wang · H. Yan · P. Chen · L. Huang (✉) · X. Zhang · Y. Peng · X. Ma · Y. Yan
Department of Grassland Science, Faculty of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China
e-mail: huanglinkai@sicau.edu.cn

T. P. Frazier
Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996, USA

Z. Chen
National Animal Husbandry Service, Ministry of Agriculture, Beijing 100026, China

Published online: 04 July 2018

P. purpureum, also known as Napier grass or Elephant grass, is a tetraploid ($2n = 4x = 28$) perennial forage crop that is indigenous to sub-Saharan Africa (Lowe et al. 2003). It is one of the main livestock fodder crops in East Africa and it accounts for the majority of feed for the cut-and-carry zero-grazing dairy systems in that region (Wanjala et al. 2013). *P. purpureum* is an ideal fodder crop because it regenerates rapidly, produces high biomass, and is extremely palatable (Lowe et al. 2003). It also has desirable traits such as high drought and stress resistance, as well as high photosynthetic and water-retaining properties (Anderson et al. 2008). Additionally, this grass is a promising material for the production of biofuels such as charcoal, alcohol, and methane (Anderson et al. 2008; Jakob et al. 2009; Lee et al. 2010; Morais et al. 2009; Strezov et al. 2008). Despite the importance of *P. purpureum*, this species has received considerably less research attention in comparison to other cereal and energy crops.

Previous genetic studies on *P. purpureum* had focused on estimating genetic diversity by molecular markers, constructing DNA fingerprints and determining genetic relationships (Bhandari et al. 2006; Harris et al. 2010; Kawube et al. 2015). Sousa Azevedo et al. (2012) used microsatellite markers derived from *P. glaucum* to study the genetic diversity of *P. purpureum* and analyzed the applicability of these markers in cross-species amplifications. Alternatively, Kawube et al. (2015) studied the genetic diversity of *P. purpureum* using SSR markers derived from maize, pearl millet, rice, and sorghum. In that study, 23 of the SSR markers generated 339 alleles, and the unique alleles could be exploited for the genetic improvement of the farmer preferred *P. purpureum* species. Since the genome of *P. purpureum* has not been sequenced, species-specific EST-SSR markers have not yet been developed. There remains a lack of transcriptomic and genomic information for *P. purpureum*, which has significantly limited genetic and breeding studies in this species. For example, there are only 591 nucleotide sequences available for *P. purpureum* on the National Center for Biotechnology Information (NCBI) website. Therefore, more detailed molecular and genomic resources are needed in order to fully characterize the genetic diversity of *P. purpureum* species.

High-throughput sequencing technologies, such as Solexa and SOLiD, have become valuable tools for studying different areas of the plant research (Metzker 2010). These sequencing platforms had revolutionized genomics, epigenomics, and transcriptomics studies by allowing massively parallel sequencing efforts at a relatively low cost. Additionally, high-throughput sequencing technologies had opened the door for exploring the effects of environmental factors on gene expression in a wide range of organisms that currently lack a reference genome (Ockendon et al. 2016). To date, high-throughput sequencing has been increasingly and successfully used in numerous plants including *Dendrobium officinale* (Xu et al. 2017), sea cucumber (*Apostichopus japonicus*) (Zhou et al. 2014), and maize (Huang et al. 2016a). A large number of EST-SSR and SNP markers have been developed based on high-throughput sequencing of transcriptome-derived sequences and had been utilized in diverse species such as blueberry (McCallum et al. 2016), sabaigrass (Zou et al. 2013), and pear (Wu et al. 2014a). The data generated from these experiments had allowed for powerful inferences about natural biological processes such as adaptation, colonization, gene flow, and divergence (Bragg et al. 2015).

Although *P. purpureum* members are economically and ecologically important forage crops, next-generation sequencing has not been performed for this species and a substantial lack of available genomic information has limited researches of *P. purpureum*, and few EST databases are currently available for them. In the present study, the Illumina HiSeq™ 4000 platform was used to perform transcriptome analyses of two *P. purpureum* cultivars: *P. purpureum* Schumab cv. Purple and *P. purpureum* cv. Mott. The objectives of this study were to (1) provide transcriptomic information for these two *P. purpureum* genotypes and (2) assemble unigenes and develop SNP and EST-SSR markers according to the transcriptome sequencing.

Material and methods

Plant material and RNA isolation

P. purpureum Schumab cv. Purple and *P. purpureum* cv. Mott plants were planted in the field at the Chongzhou farm of Sichuan Agricultural University (30° 37′ N 103° 40′; Chongzhou City, Sichuan Province, China) on June 8, 2015. To extend genetic background and obtain more expressed genes, different genotype and different stages were collected for study. Two young leaves (Purple) and two mature leaves (Light purple) were

collected from two plants of 'Purple' on June 25, 2016. Simultaneously, two young leaves were collected from two plants of 'Mott.' Leaf samples were harvested and immediately frozen in liquid nitrogen. The six samples of two *P. purpureum* genotypes were then sent to Novogene Technologies Co., Ltd. (Beijing, China) for transcriptome analysis.

Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) according to its specifications. The quality of the isolated RNA was assessed with both 1% agarose gels and a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA) (Zhang et al. 2016b). In addition, the total RNA concentration was measured using a Qubit® RNA Assay Kit on a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) (Pei et al. 2016). Finally, the RNA integrity number (RIN) of the samples was determined with an RNA Nano 6000 Assay Kit on an Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA)(Zhang et al. 2016b).

cDNA library construction and Illumina sequencing

In total, six cDNA libraries were created using a NEBNext® UltraTM RNA Library Prep Kit for Illumina® (New England Biolabs, MA, USA) (Qiao et al. 2016). The mRNAs of each sample were purified and enriched using poly-T oligo-attached magnetic beads. Next, the mRNAs were fragmented using divalent cations under high temperature with the NEBNext First Strand Synthesis Reaction Buffer (5X) (Zhang et al. 2016a). First-strand cDNA synthesis was carried out by the M-MuLV reverse transcriptase (RNase H⁻) enzyme and random hexamer primers. Second-strand cDNA was subsequently synthesized in a buffer with dNTPs, DNA polymerase I, and RNase H. Blunt ends were produced by removing/filling the remaining overhangs with an exonuclease/polymerase treatment. After adenylation of the 3′ ends of the DNA fragments, NEBNext adaptors were ligated to the 5′ ends in preparation for hybridization. Library fragments were purified with an AMPure XP system (Beckman Coulter, MA, USA) and cDNA fragments of 150–200 bp in length were preferentially selected (Jiang et al. 2016). Next, 3 μL of USER Enzyme (New England Biolabs, MA, USA) was mixed with the selected cDNA and the mixture incubated at 37 °C for 15 min, followed by 5 min at 95 °C. After that, the PCR was disposed by universal PCR primers, an Index (X) Primer, and the Phusion High-Fidelity DNA polymerase (New England Biolabs, MA, USA) (Yang et al. 2015; Yue et al. 2015). The final products were purified using the AMPure XP system and their quality was analyzed with the Agilent Bioanalyzer 2100 system (Zhang et al. 2015a). For Illumina sequencing, a cBot Cluster Generation System was utilized to cluster the index-coded samples via the TruSeq PE Cluster Kit v3-cBot-HS (Illumina, CA, USA) (Liu et al. 2014). After cluster generation, the Illumina Hiseq platform was used for the library preparations and paired-end reads were generated for each sample (NCBI SRA: SRP100008).

De novo transcriptome assembly and annotation

All raw sequencing reads from the two genotypes were collected for transcriptome assembly. They were first processed through in-house perl scripts (Novogene Technologies Co., Ltd., Beijing, China) and clean reads were obtained by removing those reads that contained adapter and poly-N sequences (Zhao et al. 2014). During this step, low-quality reads were also removed and the Q20, Q30, GC content, and sequence duplication levels of the clean reads were calculated. All of the downstream analyses were based on the resulting high-quality clean reads (Li et al. 2015). Next, the clean reads were assembled into unigene sequences using Trinity software (http://trinityrnaseq.sourceforge.net/) with the min_K-mer_cov set to 2 and all other parameters set as default (Grabherr et al. 2011).

BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) searches (e-value < 1E-5) were used to annotate the assembled unigenes (Altschul et al. 1997). The sequences were aligned against the following databases: National Center for Biotechnology Information non-redundant protein (Nr), Protein family (Pfam), Swiss-Prot protein, Gene Ontology (GO), the Kyoto Encyclopedia of Genes and Genomes (KEGG), and Eukaryotic Orthologous Groups of proteins (KOG). Furthermore, the NCBI non-redundant nucleotide sequence (Nt) database was also searched. Functions were assigned to the unigenes based off of alignment data from the seven databases in the following order: Nr, GO, KEGG, Swiss-Prot, Pfam, KOG, and Nt (Moriya et al. 2007; Punta et al. 2011). The coding regions (CDSs) were predicted by the best BLAST alignments, and the transcripts were transformed into their protein sequences based on the standard genetic codes. Afterwards, transcript open reading frames (ORFs) were extracted. Finally, the software ESTScan (Iseli et al. 1999) was used

to predict the ORFs for the unigenes that did not align to any of the above databases.

SNP loci detection and marker validation

Prior to calling single nucleotide polymorphisms (SNPs) in 'Purple' and 'Mott,' picard-tools v1.41 (http://broadinstittute.github.io/picard) and samtools v0.1.18 (http://samtools.sourceforge.net) were employed to sort the reads, remove duplicates, and merge the bam alignment results of each sample (Huang et al. 2015). GATK3 software (http://www.broadinstitute.org/gatk/) was then used to perform SNP calling. For SNPs, variants were detected using the GATK standard filter method with the following parameters: cluster = 3, Window Size = 35, QD < 2.0, FS > 60.0, MQ < 40.0, SOR > 4.0, MQ Rank Sum < − 12.5, Read Pos Rank Sum < − 8.0, and DP < 10 (Choi et al. 2015). SNP marker validation was performed on 16 *Pennisetum* materials (Table S1) using six randomly selected primers. Briefly, 100 mg of fresh leaf tissue was harvested for each accession and DNA was extracted using a genomic DNA extraction kit (Tiangen Biotech, China). PCR amplifications were performed in 25 μl total reaction volumes that contained 1 μL (10 pmol/μL) each of the forward and reverse primers, 2.5 μL (20 ng/μL) of DNA template, 12.5 μL 2X Reaction Mix (Tiangen Biotech, China), 0.5 μL Pfu polymerase (Tiangen Biotech, China), and 7.5 μL ddH_2O. The PCRs were performed with a 5-min pre-denaturation at 94 °C, followed by 40 cycles of 30-s denaturation at 94 °C, 45-s annealing at 52–60 °C and 1-min extension at 72 °C, and a final 7-min extension at 72 °C. PCR products were electrophoresed on 1.5% agarose gels at 100 V for 30 min to separate the DNA fragments. After visualization of the bands under UV light, the sizeable PCR products were extracted from the gel and sent to the Beijing Genomics Institute for sequencing.

EST-SSR loci detection and marker validation

EST-SSR markers within the transcriptome were identified using MISA (http://pgrc.ipk-gatersleben.de/misa/misa.html). The unigene sequences were searched for perfect mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of ten, six, five, five, five, and five repeats, respectively (Singh et al. 2013). The EST-SSR primers were designed using Primer 3 (http://primer3.sourceforge.net/releases.php) based on the results from the MISA software (Lv et al. 2014). The primer design parameters were set as follows: the primer length ranged between 18 and 27 nt with 21 nt as optimum, the PCR products were 100–300 bp in size, the PCR annealing temperature ranged from 52 to 62 °C with the optimal temperature at 55 °C, and the primer GC content ranged from 40 to 60% with 50% as optimum.

The EST-SSR markers were validated on 17 *Pennisetum* materials (Table S2) using 50 randomly selected primers (Table S3). The method of DNA extraction was as above. For PCR amplification of the EST-SSR loci, reactions were performed in a total volume of 15 μL and contained 0.6 μL (10 pmol/μL) of each primer, 1.5 μL (20 ng/μL) genomic DNA, 7.5 μL 2X Reaction Mix (Tiangen, Biotech, China), 0.3 μL of Golden DNA Polymerase (Tiangen Biotech, China), and 4.5 μL distilled water. The PCRs were carried out with the following protocol: 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 54–56 °C for 45 s, 72 °C for 1 min, and a final extension at 72 °C for 7 min. The amplification products were then separated on a 6% polyacrylamide gel, stained with $AgNO_3$ solution, and photographed for scoring (Huang et al. 2014). Based on a molecular DNA marker (50 bp ladder, Tiangen, China), strong clear allelic bands with the same mobility were manually scored as either present (1) or absent (0) (Guo et al. 2016). The total number of polymorphic bands and percentage of polymorphic bands was calculated with Excel. Differences in the numbers within these three categories revealed the amount of variation among the accessions and species (Williams et al. 1990). In addition, the polymorphism information content (PIC) per EST-SSR locus was estimated using the formula described by Nei (1973). Genetic analyses, including observed number of alleles, effective number of alleles, Nei's gene diversity (H), and Shannon's information index (I), were performed for each population using POPGENE v.1.3.2 (Yeh 1997) with a model for dominant markers and individuals.

Results

Sequence analysis and assembly

In present study, the Illumina HiSeq™ 4000 platform was used to sequence the transcriptome of *P. purpureum*. High-throughput RNA sequencing

generated more than 52,610,350 raw reads (ranging from 150 to 200 bp) for each of the six *P. purpureum* cDNA libraries. After data filtering and removing adapter sequences for each sample, we obtained a total of 50,756,372 high-quality reads (~ 7.61 Gb of clean data). The GC content, Q20, and Q30% of the clean data were over 55.39, 97.62, and 93.75%, respectively, for each of the six samples (Table S4). These results indicated that the clean data reads were of high quality and they were reliable for subsequent analyses.

Using Trinity, 284,875 de novo assembled transcripts were obtained with an average length of 781 bp and an N50 value of 1354 bp (Table S5). The large transcriptome size of *P. purpureum* might be due to the tetraploid nature of the genome ($2n = 4x = 28$ chromosomes) (Campos et al. 2009). The estimated DNA size of *P. purpureum* is 2836 Mb (Taylor and Vasil 1987), and the assembly generated in this study covered 7.84%. Additionally, these transcripts were clustered into 197,466 unigenes with a mean length of 586 bp and an N50 value of 833 bp (Table S6). Among the 197,466 unigenes, the two most predominant sizes were those that ranged in length from 200 to 500 bp (137,131 unigenes or 69.45%) and those that ranged in length from 500 to 1000 bp (34,101 unigenes or 17.27%). This was followed by 17,020 unigenes (8.62%) that ranged in length from 1000 to 2000 bp and finally by 9214 unigenes (4.67%) that were more than 2000 bp in length. The sizable proportion of short unigenes (69.45%) could be attributed to the lack of an available reference genome sequence for *P. purpureum*. Moreover, the abundance of these short unigenes could be the result of fragmented transcripts or insufficient sequencing depth. This phenomenon has also been previously observed in other non-model plants (Jiang et al. 2015; Wu et al. 2014b).

Sequence annotation

The current study performed functional annotation, as well as classification, for the obtained unigenes in order to predict their structures, potential functions, and biological processes. Of the 197,466 total unigenes, 103,454 (52.39%) were successfully annotated in at least one database and 12,195 (6.17%) were annotated in all databases. Among the annotated unigenes, 72,485 (36.7%) had hits in the Nr database, 74,148 (37.54%) in Nt, 25,683 (13%) in KO, and 25,683 (13%) in KEGG. Besides, a search for homology against the Swiss-Prot database produced 54,026 hits (27.35%). This could be attributed to the smaller number of proteins in this more reliable protein bank. Similarly, a homology search against the Pfam database returned 55,245 hits (27.97%). A large number of unigenes were assigned to KOG classifications (28,665, 14.51%) and GO categories (56,999, 28.86%), indicating that the transcriptome data reflected the extensive biological functions of *P. purpureum* transcripts. However, not all of the unigenes matched known proteins in these databases, suggesting that some genes may be unique to *P. purpureum*. The remaining unigenes (47.61%) did not align with any known genes (Table S7).

The present study predicted 176,887 CDSs in total, and 80,112 of which aligned to the Nr and Swiss-Prot databases. Approximately 45.29% of the unigenes were able to uniquely match to known proteins in these databases, suggesting that a portion of the remaining genes may be unique to *P. purpureum*. Interestingly, the majority of CDSs that aligned (35,189, 43.92%) were less than 500 bp in length (Fig. S1). The remaining 96,775 CDSs were predicted using ESTScan. Not surprisingly, the number of CDSs under 500 bp long also accounted for the largest number in this group (67,796, 70.01%; Fig. S2).

Gene Ontology is a worldwide classification system for gene function, and it comprehensively categorizes the properties of genes into groups such as "biological processes," "cellular components," and "molecular function" (Tang et al. 2014). Based on sequence homology, 56,999 unigenes (28.86%) were classified into three main GO categories and 56 sub-categories. A large number of unigenes that were annotated in the "biological processes" category fell under the sub-categories "Cellular process," "Metabolic process," and "Single-organism process" (Fig. 1). In the "cellular components" group, the "Cell" was the most represented sub-category followed by "Cell part," "Organelle," and "Macromolecular complex." Among ten different molecular function categories, "Binding" and "Catalytic activity" were the two most frequent classes for the *P. purpureum* unigenes.

KOG analysis is centered on the clustering of orthologous groups for eukaryotic complete genomes (Wang et al. 2015). Based on the KOG database, 28,665 unigenes were classified into 26 functional categories, and the "General function prediction only" cluster was found to be the largest group (4995 unigenes, 17.43%). The next largest group was "Post-translational

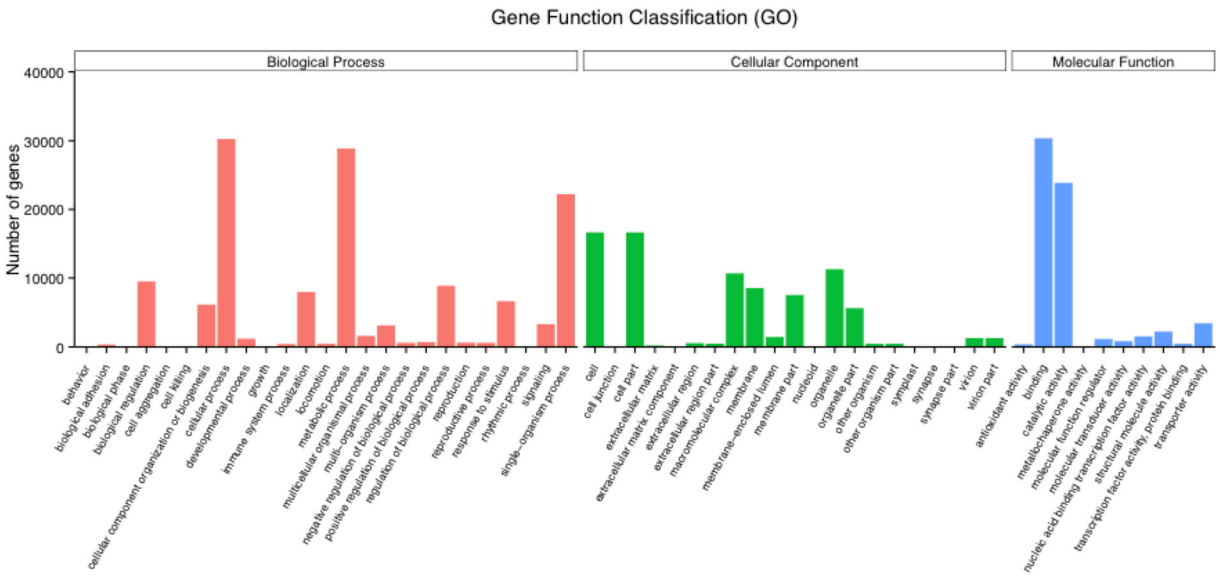

Fig. 1 Histogram representation of Gene Ontology (GO) annotation for *P. purpureum*. The genes were assigned to three main categories: biological process, molecular function, and cellular component

modification, protein turnover, chaperones" (3909 unigenes, 13.64%) followed by the "Translation, ribosomal structure and biogenesis" group (2964 unigenes, 10.34%). Alternatively, "Extracellular structures" (104 unigenes, 0.36%), "Cell motility" (24 unigenes, 0.08%), and "Unnamed protein" (1 unigene, 0.003%) groups contained significantly fewer unigenes than the above three groups (Fig. 2).

KEGG analysis is a way to analyze gene products during metabolism and determine their putative functions in cellular processes. A total of 25,683 unigenes were found and assigned to 19 biological pathways that fell under five larger groups (Cellular Processes, Genetic Information Processing, Environmental Information Processing, Metabolism and Organismal Systems). Of these 19 pathways, the six major ones included

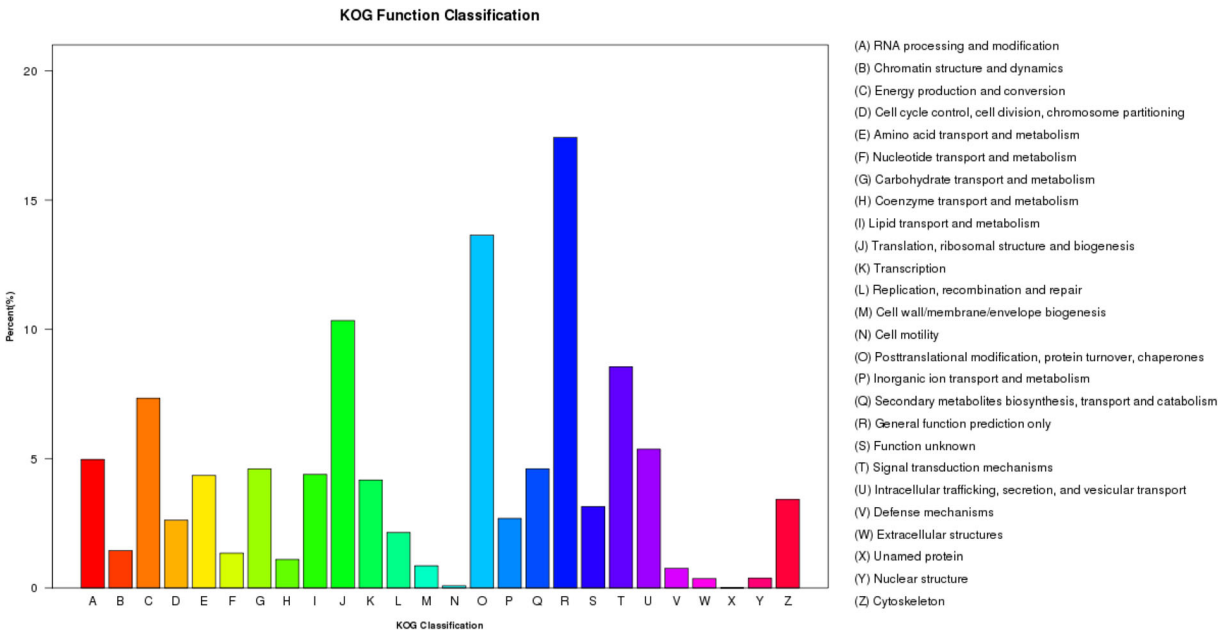

Fig. 2 EuKaryotic Ortholog Groups (KOG) classification of *P. purpureum* unigenes

"Translation" (3135, 12.21%), "Carbohydrate metabolism" (2778, 10.82%), "Overview" (2210, 8.60%), "Folding, sorting and degradation" (2094, 8.15%), "Amino acid metabolism" (1670, 6.50%), and "Energy metabolism" (1537, 5.98%) (Fig. 3).

SNP loci detection and marker validation

A total of 214,648 high-quality SNPs were detected in 40,259 unigenes of the two *P. purpureum* genotypes 'Purple' and 'Mott.' As shown in Table 1, the putative SNPs of these two genotypes included 71,242 (33.19%) transitions, of which 15.23% were A/G and 17.96% were C/T. The high percentage of transitions found in this research was consistent with other plant species (Wang et al. 2017; Wei et al. 2014). Coulondre et al. (1978) suggested that the high frequency of transitions within genomes might reflect the high levels of C/T mutations after methylation. Interestingly, the number of transitions was approximately 2.08 times higher than the number of transversions (34,232) for the two *P. purpureum* genotypes. The exact proportions of the four possible transversions were 3.74% A/C, 4.08% A/T, 4.39% C/G, and 3.74% G/T.

Further analysis of the putative SNPs revealed that 38,411 (17.89%) of them were distributed in CDS sequences. These SNPs might be associated with important economic and agronomic traits and thus could be used for genetic diversity analysis, association mapping

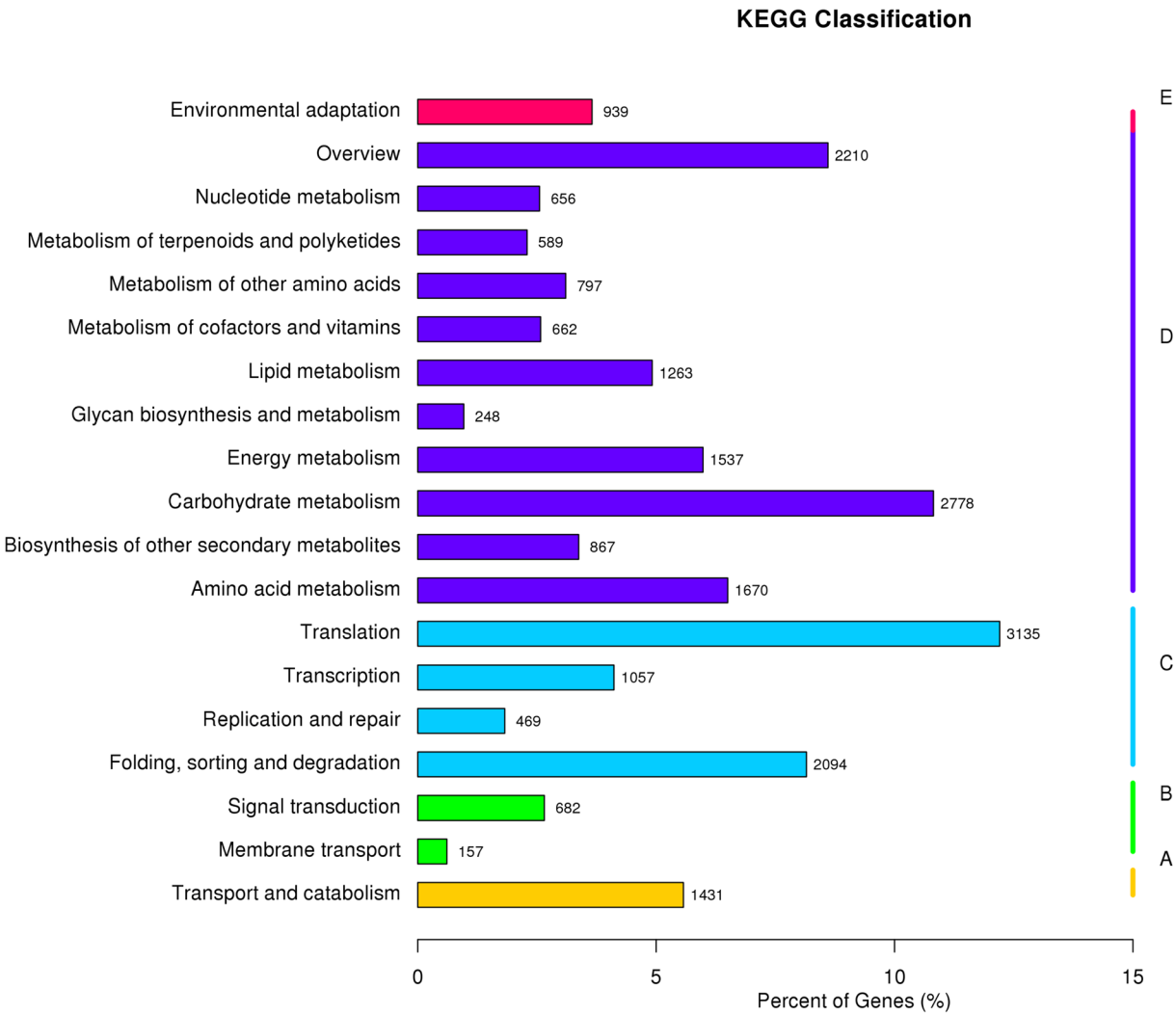

Fig. 3 Kyoto Encyclopedia of Genes and Genomes (KEGG) classification of *P. purpureum* unigenes. A, cellular process; B, environmental information processing; C, genetic information processing; D, metabolism; E, organismal systems

analysis, genetic map construction, and marker-assisted selection. Moreover, they could potentially explain the phenotypic differences between *P. purpureum* species. Among the 38,411 SNPs in the CDS regions, 39.50% of them were non-synonymous mutations, which could result in amino acid changes. These special SNPs are important for future association analyses, and they are promising candidates for studying phenotypic differences of *Pennisetum* germplasms.

Six primer pairs were chosen to validate 21 of the putative SNP loci in the present study. Sixteen different *Pennisetum* clonal materials were used to determine the feasibility and efficiency of these molecular markers for *P. purpureum* and other *Pennisetum* species. Three of the primer pairs were successfully able to discriminate the alleles, while two pairs produced no PCR products and one pair produced a single amplicon. In total, 16 SNP loci were amplified, five of which were putative SNP loci that had been discovered by SNP detection. These SNP loci were found to be polymorphic and one to two SNP loci were covered per primer pair with each pair spanning an average of 1.67 SNP loci (Table S8).

EST-SSR loci detection and marker validation

With recent progress in sequencing technologies, high-throughput sequencing platforms have enabled researchers to efficiently generate large number of EST-SSR markers (Soren et al. 2015; Sureshkumar et al. 2014). A total of 21,213 EST-SSRs were identified in 18,587 (9.41%) unigenes and these repeats ranged in lengths from 1 to 6 bp. Mono- and tri-nucleotide repeats were the most predominant and accounted for 50.04 and 33.26% of the overall EST-SSRs, respectively. Of the 21,213 EST-SSRs, the percentages of di-, tetra-, penta-, and hexa-nucleotide repeats were 15.36, 1.15, 0.14, and 0.05%, respectively (Table 2). Interestingly, 770 of the EST-SSRs contained two or more different repeats, resulting in a type of compound repeat.

The repeating unit of each type of EST-SSR varied from five to > 10. Five tandem repeats (24.01%) was the most predominant for each of the EST-SSRs examined followed by ten tandem repeats (23.33%) and six tandem repeats (15.10%) (Table S9). In addition, the repetition types and amount of di- and tri-nucleotides were analyzed. Of the tri-nucleotide repeats, CCG/GGC (3210, 45.50%), AGC/TCG (1011, 14.33%), and AGG/TCC (874, 12.39%) were the most prevalent (Table S10). This finding was in accordance with other reports in Lily (Du et al. 2015) and wheat (Kantety et al. 2002). With respect to di-nucleotide repeat motifs, the AC/TG motif had the largest frequency of 49.39% (740). This number is similar to results obtained for *Neottopteris nidus* (Jia et al. 2016).

To validate the EST-SSRs identified in the present study, more than one primer pair was designed for

Table 1 Summary of putative SNPs identified from *P. purpureum* Schumab cv. Purple and *P. purpureum* cv. Mott

SNP information	Counts
Transversion	
A/C	8031
A/T	8758
C/G	9425
G/T	8018
Transition	
A/G	32,686
C/T	38,556
Total SNPs	214,648
Number of unigenes containing SNPs	40,259
Number of annotated unigenes containing SNPs	11,005
Number of SNPs in CDS	122,370
Number of SNPs in non-CDS	92,278
Number of non-synonymous SNPs	46,468
Number of synonymous SNPs	75,902

Table 2 Summary of simple sequence repeats (EST-SSRs) identified from *P. purpureum* Schumab cv. Purple and *P. purpureum* cv. Mott

EST-SSR information	Number
Total number of sequences examined	197,466
Total size of examined sequences (bp)	115,813,152
Total number of identified EST-SSRs	21,213
Number of sequences containing EST-SSRs	18,587
Number of sequences containing more than one EST-SSR	2211
Number of EST-SSRs present in compound formation	770
Mono-nucleotide repeats	10,616
Di-nucleotide repeats	3258
Tri-nucleotide repeats	7055
Tetra-nucleotide repeats	243
Penta-nucleotide repeats	30
Hexa-nucleotide repeats	11

12,646 (61.86%) of the 21,213 EST-SSRs. A total of 50 of these EST-SSR primer pairs were randomly selected for validation. Seventeen different *Pennisetum* materials (seven *P. purpureum* species and ten other species) with different genetic backgrounds were tested to see the compatibility of these primers across *P. purpureum* species and other *Pennisetum* species (Table S11). Of the 50 EST-SSR primers tested, 20 (40.00%) of them correctly amplified PCR products with rich polymorphisms that contained di-, tri-, and tetra-nucleotide motifs. The failure of the other 30 primer pairs to amplify could be attributed to assembly errors in the cDNA contigs or the existence of null alleles. Alternatively, large unforeseen introns and primer pairs designed across splice sites would have also impeded PCR (Dutta et al. 2011). As seen in Table S11, the 20 successful primer pairs obtained 128 alleles, of which 121 (95.53%) were polymorphic, and the numbers of PCR products ranged from four to ten with an average of 6.4. The relatively small number of alleles detected with only these 20 primer pairs indicates the high efficiency of the EST-SSR markers identified in this study, as compared to other molecular markers such as microsatellites and SSRs (Kawube et al. 2015; Sousa Azevedo et al. 2012). The average polymorphism information content, Nei's gene diversity and Shannon's information index of diversity of the EST-SSRs were 0.3347, 0.3689, and 0.5405, respectively. These results revealed that the EST-SSR markers identified for *P. purpureum* in this study have a high level of polymorphism, and that they are valuable for genetic studies of *Pennisetum* species. Interestingly, the ten *Pennisetum* species that did not belong to *P. purpureum* also exhibited a relatively high level of amplification and polymorphism. The EST-SSR results were used to estimate the genetic distances of *P. purpureum*, as well as the *Pennisetum* materials, and taxonomically cluster the accessions accordingly. From these results, three major groups were identified (Fig. S3). All of the *P. purpureum* species, one *P. americanum* accession and one *P. americanum* × *P. purpureum* genotype, were grouped together in group 1, indicating that these species are genetically similar. Group 2, however, was comprised of three *P. americanum* × *P.purpureum* species that have presumably genetically differentiated from the one accession in group 1. Finally, one *P. polystachyon* accession, one *P. americanum* accession, and one *P. americanum* × *P. purpureum* genotype clustered into group 3.

Discussion

High-throughput sequencing is a promising and effective tool to obtain genomic and transcriptomic data for non-model organisms and non-sequenced genomes. Currently, RNA-Seq has been applied to several members of the *Pennisetum* genus including *P. glaucum*, *P. polystachion*, and *P. alopecuroides* (Choudhary and Padaria 2015; Sahu et al. 2012; Sarah et al. 2017); however, this study is the first report of RNA sequencing and de novo transcriptome analysis for *P. purpureum* which will provide numerous genetic information for future research and allow a very clear and extensive description of this species.

Despite the low number of unigenes annotated for *P. purpureum* in this study (52.39%), the percentage is consistent with previous studies on *Pueraria lobata* (Wang et al. 2015), *Hemarthria* (Huang et al. 2016c), and *Dactylis glomerata* L. (Huang et al. 2015). The seemingly low number of annotated unigenes might be attributed to the large amount (69.45%) of short-length (< 500 nt) unigenes, or the limited amount of publicly available EST sequences and genomic information for *P. purpureum*. Functional annotation of the *P. purpureum* unigenes and subsequent analyses provided some insight into the molecular mechanisms governing physiological processes in this species. GO analysis of the annotated unigenes found that "Binding" and "Catalytic activity" were the two most frequent classes. This phenomenon has been found in de novo transcriptome analyses of *Camellia sinensis* (Wu et al. 2014b). Of particular interest, 609 unigenes were annotated as related to reproductive processes, which may provide valuable information for further studies of reproduction in *P. purpureum*. Additionally, response to stimulus (6671 unigenes) was another abundant biological process term, and cellular response to stimulus (4382 unigenes) was the most highly represented child GO term. KEGG analysis predicted 8077 unigenes to be involved in a variety of metabolic precesses such as carbohydrate metabolism, amino acid metabolism, and energy metabolism. These unigenes were enriched for proteins that maintain the essential functions of *P. purpureum*, which was consistent with the transcriptome analyses of *Kappaphycus alvarezii* (Zhang et al. 2015b) and *Gentiana macrophylla* (Hua et al. 2014). These findings may provide useful information for research about gene expression and improving biological and physiological pathways.

As SNPs are the most abundant DNA variations in plant genomes, they can be readily utilized in crop breeding because of their high efficiency (Rafalski 2002; Riahi et al. 2013). Compared with other molecular markers, SNPs are highly convenient for comparing genomic and transcriptomic sequences (Hayashi et al. 2004; Salem et al. 2012). Presently, SNPs have been used in many species to assess genetic diversity, create genetic maps, and aid in association mapping and marker-assisted selection (MAS) breeding (Eckert et al. 2009; Li et al. 2009a; Li et al. 2009b; Xu et al. 2011). With recent improvements in next-generation sequencing platforms, the production of large-scale genomic and transcriptomic data makes it easy to identify large numbers of high-quality SNPs. For the SNPs detection, the number of transition SNPs was found to be higher than the number of transversion SNPs, which was consistent with previous reports in radish (Wang et al. 2017) and chickpea (Gaur et al. 2012). These numbers could indicate a low level of genetic divergence in *P. purpureum* genomes (Barchi et al. 2011) and might reflect high levels of C/T mutations after methylation (Coulondre et al. 1978). Alternatively, the ratio of synonymous to non-synonymous mutation was 1.63, compared with sesame (Wei et al. 2014), spruce (Pavy et al. 2006), and Arabidopsis (Schmid et al. 2003). Aside from five putative SNPs, 11 dubious SNPs were found by three primer pairs. These might have been detected due to an abundance of experimental materials. In addition, SNPs with minor allele frequencies (MAFs) lower than 0.1 may not be true SNPs, such as 1348 of c100029_g1, 1412 of c100029_g1 and 430 of c100188_g1. False SNPs were observed earlier in several SNP discovery studies based on EST sequences of some crop plants (Jhanwar et al. 2012; Varshney et al. 2008), indicating that there is room for future improvement of the methodology. The validation of *Pennisetum* SNPs in present study suggests that high-throughput sequencing and SNP discovery is a reliable method for identifying molecular markers for *Pennisetum* species. Therefore, the current sequence data and list of SNPs will greatly enrich the amount of the useful genetic marker resources available for *Pennisetum* and these resources can be used for future studies in crop development.

Prior to the present research, most molecular marker studies of *Pennisetum* were done by SRAP, RAPD, and AFLP analyses (Babu et al. 2009; Harris et al. 2010; Xie et al. 2009), which limited the number of available markers and slowed genetic research progress. Recently, an alternative molecular marker has been utilized on *Pennisetum* for assessing genetic diversity and creating genetic maps (Kawube et al. 2015; Moumouni et al. 2015; Rajaram et al. 2013). EST-SSRs are ubiquitous in transcriptomes and as such, they have ideal molecular marker features including being co-dominant, typically locus-specific, highly polymorphic, and often times having a high level of cross-species transferability (Kaur et al. 2012). Additionally, EST-SSRs are widely used as powerful molecular markers for gene mapping, genotyping, and gene diagnosis due to their different putative functions and high polymorphism (Li et al. 2004). The unigenes of *P. purpureum* obtained in this study provided a large number of EST-SSRs that can ultimately be applied to genetic research and molecular breeding. Since mono-nucleotide repeats may result from sequencing and genotyping errors (Gilles et al. 2011), they were excluded from subsequent analyses. The abundance of tri-nucleotide repeats was consistent with the similar studies of other species (Jhanwar et al. 2012; Toledo-Silva et al. 2013; Yates et al. 2014). The EST-SSR markers developed in this study were used to examine the taxonomic relationship between 17 different *Pennisetum* accessions. *P. purpureum*, *P. americanum*, and *P. americanum* × *P. purpureum* were not found to be distinctly separate, which corroborates the findings of Yao et al. (2013) and Xie and Lu (2005). In addition, these results demonstrate the close relationship of *P. purpureum*, *P. americanum*, and their hybrids. The EST-SSR and SNP markers identified in this work will aid future GWAS studies (Wang et al. 2012; Zeng et al. 2017), trait association analyses (Lakew et al. 2013; Ukoskit et al. 2018), and marker-assisted selection practices (Gupta et al. 2012; Ha et al. 2007) in both *P. purpureum* and the *Pennisetum* genus.

Conclusion

Previous studies have supported the applications of RNA-seq technologies in providing genomic data for non-model organisms (Egan et al. 2012; Ekblom and Galindo 2011; Huang et al. 2016b; Takayama et al. 2011). In this study, we assembled the first transcriptome of *P. purpureum*. The excavation of unigenes, along with their annotation, provides useful information about *P. purpureum* genes and their putative functions. In addition, this work identified a large number of SNP

and EST-SSR molecular markers that will be helpful in future molecular biology, molecular breeding, physiology, and biochemistry studies of *Pennisetum* species, and they will aid in the understanding of molecular inheritance and genetics in *P. purpureum*.

Acknowledgements We thank Tomas Hasing Rodriguez who helped polish this manuscript.

Authors' contributions L.K.H. and X.Q.Z. designed research studies; S.F.Z., C.R.W., and H.D.Y. conducted experiments, acquired data, and analyzed data; L.K.H., X.Q.Z., Y.P., X.M., and Y.H.Y. provided the experimental equipment; S.F.Z. and P.L.C. wrote the manuscript; T.P.F. and Z.H.C. revised the manuscript.Funding informationThe authors gratefully acknowledge the financial support from the National Project on Sci-Tec Foundation Resources Survey (2017FY100602), Sichuan Province Breeding Research Grant (2016NZ0098-11), and Modern Agricultural Industry System Sichuan Forage Innovation Team.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Data archiving statement Raw Illumina reads were deposited in NCBI SRA: SRP100008.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Anderson WF, Dien BS, Brandon SK, Peterson JD (2008) Assessment of bermudagrass and bunch grasses as feedstock for conversion to ethanol. Appl Biochem Biotechnol 145:13–21

Babu C, Sundaramoorthi J, Vijayakumar G, Ram SG (2009) Analysis of genetic diversity in napier grass (*Pennisetum purpureum* Schum) as detected by RAPD and ISSR markers. J Plant Biochem Biotechnol 18:181–187

Barchi L, Lanteri S, Portis E, Acquadro A, Valè G, Toppino L, Rotino GL (2011) Identification of SNP and SSR markers in eggplant using RAD tag sequencing. BMC Genomics 12:304

Bhandari AP, Sukanya D, Ramesh C (2006) Application of isozyme data in fingerprinting Napier grass (*Pennisetum purpureum* Schum.) for germplasm management. Genet Resour Crop Evol 53:253–264

Bragg JG, Supple MA, Andrew RL, Borevitz JO (2015) Genomic variation across landscapes: insights and applications. New Phytol 207:953–967

Brunken JN (1977) A systematic study of *Pennisetum* sect. *Pennisetum* (Gramineae). Am J Bot:161–176

Campos J et al (2009) In vitro induction of hexaploid plants from triploid hybrids of *Pennisetum purpureum* and *Pennisetum glaucum*. Plant Breed 128:101–104

Choi JY, Bubnell JE, Aquadro CF (2015) Population genomics of infectious and integrated Wolbachia pipientis genomes in *Drosophila ananassae*. Genome Biol Evol 7:2362–2382

Choudhary M, Padaria JC (2015) Transcriptional profiling in pearl millet (*Pennisetum glaucum* LR Br.) for identification of differentially expressed drought responsive genes. Physiol Mol Biol Plants 21:187–196

Coulondre C, Miller JH, Farabaugh PJ, Gilbert W (1978) Molecular basis of base substitution hotspots in *Escherichia coli*. Nature 274:775

Du F et al (2015) De novo assembled transcriptome analysis and SSR marker development of a mixture of six tissues from *Lilium* Oriental hybrid 'Sorbonne'. Plant Mol Biol Report 33:281–293

Dutta S et al (2011) Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. BMC Plant Biol 11(17)

Eckert AJ, Pande B, Ersoz ES, Wright MH, Rashbrook VK, Nicolet CM, Neale DB (2009) High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). Tree Genet Genomes 5:225–234

Egan AN, Schlueter J, Spooner DM (2012) Applications of next-generation sequencing in plant biology. Am J Bot 99:175–185

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107(1)

Fulkerson W, Horadagoda A, Neal J, Barchia I, Nandra K (2008) Nutritive value of forage species grown in the warm temperate climate of Australia for dairy cows: herbs and grain crops. Livest Sci 114:75–83

Gaur R et al (2012) High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). DNA Res 19:357–373

Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin J-F (2011) Accuracy and quality assessment of 454 GS-FLX titanium pyrosequencing. BMC Genomics 12:245

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

Guo Z-H et al (2016) SSRs transferability and genetic diversity of three allogamous ryegrass species. C R Biol 339:60–67

Gupta D et al (2012) Integration of EST-SSR markers of *Medicago truncatula* into intraspecific linkage map of lentil and identification of QTL conferring resistance to ascochyta blight at seedling and pod stages. Mol Breed 30:429–439

Ha B-K, Hussey RS, Boerma HR (2007) Development of SNP assays for marker-assisted selection of two southern root-knot nematode resistance QTL in soybean. Crop Sci 47:S-73–S-82

Harris K, Anderson W, Malik R (2010) Genetic relationships among napiergrass (*Pennisetum purpureum* Schum.) nursery accessions using AFLP markers. Plant Genet Resour 8:63–70

Hayashi K, Hashimoto N, Daigen M, Ashikawa I (2004) Development of PCR-based SNP markers for rice blast

resistance genes at the Piz locus. Theor Appl Genet 108: 1212–1220

Hua W, Zheng P, He Y, Cui L, Kong W, Wang Z (2014) An insight into the genes involved in secoiridoid biosynthesis in *Gentiana macrophylla* by RNA-seq. Mol Biol Rep 41: 4817–4825

Huang X et al (2014) Genetic diversity of *Hemarthria altissima* and its related species by EST-SSR and SCoT markers. Biochem Syst Ecol 57:338–344

Huang L et al (2015) Identifying differentially expressed genes under heat stress and developing molecular markers in orchardgrass (*Dactylis glomerata* L.) through transcriptome analysis. Mol Ecol Resour 15:1497–1509

Huang J, Gao Y, Jia H, Zhang Z (2016a) Characterization of the teosinte transcriptome reveals adaptive sequence divergence during maize domestication. Mol Ecol Resour 16:1465–1477

Huang J et al (2016b) De novo sequencing and characterization of seed transcriptome of the tree legume *Millettia pinnata* for gene discovery and SSR marker development. Mol Breed 36(75)

Huang X et al (2016c) De novo transcriptome analysis and molecular marker development of two *Hemarthria* species. Front Plant Sci 7:496

Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. In: ISMB, pp 138–148

Jakob K, Zhou F, Paterson AH (2009) Genetic improvement of C4 grasses as cellulosic biofuel feedstocks. In Vitro Cell Dev Biol: Plant 45:291–305

Jhanwar S, Priya P, Garg R, Parida SK, Tyagi AK, Jain M (2012) Transcriptome sequencing of wild chickpea as a rich resource for marker development. Plant Biotechnol J 10:690–702

Jia X, Deng Y, Sun X, Liang L, Su J (2016) De novo assembly of the transcriptome of *Neottopteris nidus* using Illumina paired-end sequencing and development of EST-SSR markers. Mol Breed 36:94

Jiang Q, Wang F, Tan H-W, Li M-Y, Xu Z-S, Tan G-F, Xiong A-S (2015) De novo transcriptome assembly, gene annotation, marker development, and miRNA potential target genes validation under abiotic stresses in *Oenanthe javanica*. Mol Gen Genomics 290:671–683

Jiang F, Chen X-p, Hu W-s, Zheng S-q (2016) Identification of differentially expressed genes implicated in peel color (red and green) of Dimocarpusconfinis. SpringerPlus 5:1088

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Kaur S et al (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genomics 13:104

Kawube G, Alicai T, Wanjala B, Njahira M, Awalla J, Skilton R (2015) Genetic diversity in Napier grass (*Pennisetum purpureum*) assessed by SSR markers. J Agric Sci 7:147

Khairwal I, Rai K, Diwakar B, Sharma Y, Rajpurohit B, Nirwan B, Bhattacharjee R (2007) Pearl millet crop management and seed production manual. International Crops Research Institute for the Semi-Arid Tropics, Patancheru

Lakew B, Henry RJ, Ceccarelli S, Grando S, Eglinton J, Baum M (2013) Genetic analysis and phenotypic associations for drought tolerance in *Hordeum spontaneum* introgression lines using SSR and SNP markers. Euphytica 189:9–29

Lee M-K, Tsai W-T, Tsai Y-L, Lin S-H (2010) Pyrolysis of napier grass in an induction-heating reactor. J Anal Appl Pyrolysis 88:110–116

Li Y-C, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007

Li H et al (2009a) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Li Y-H et al (2009b) Development of SNP markers and haplotype analysis of the candidate gene for rhg1, which confers resistance to soybean cyst nematode in soybean. Mol Breed 24: 63–76

Li H, Yao W, Fu Y, Li S, Guo Q (2015) De novo assembly and discovery of genes that are involved in drought tolerance in Tibetan *Sophora moorcroftiana*. PLoS One 10:e111054

Liu B, Zhang Y, Zhang W (2014) RNA-Seq-based analysis of cold shock response in Thermoanaerobacter tengcongensis, a bacterium harboring a single cold shock protein encoding gene. PLoS One 9:e93289

Lowe A, Thorpe W, Teale A, Hanson J (2003) Characterisation of germplasm accessions of Napier grass (*Pennisetum purpureum* and *P. purpureum*× *P. glaucum* hybrids) and comparison with farm clones using RAPD. Genet Resour Crop Evol 50:121–132

Lv J, Liu P, Gao B, Wang Y, Wang Z, Chen P, Li J (2014) Transcriptome analysis of the *Portunus trituberculatus*: de novo assembly, growth-related gene identification and marker discovery. PLoS One 9:e94055

McCallum S et al (2016) Construction of a SNP and SSR linkage map in autotetraploid blueberry using genotyping by sequencing. Mol Breed 36(41)

Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11(31)

Morais RF, Souza BJ, Leite JM, Soares LHB, Alves BJR, Boddey RM, Urquiaga S (2009) Elephant grass genotypes for bioenergy production by direct biomass combustion. Pesq Agrop Brasileira 44:133–140

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

Moumouni K, Kountche B, Jean M, Hash C, Vigouroux Y, Haussmann B, Belzile F (2015) Construction of a genetic map for pearl millet, *Pennisetum glaucum* (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. Mol Breed 35(5)

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci 70:3321–3323

Ockendon NF et al (2016) Optimization of next-generation sequencing transcriptome annotation for species lacking sequenced genomes. Mol Ecol Resour 16:446–458

Pavy N, Parsons LS, Paule C, MacKay J, Bousquet J (2006) Automated SNP detection from a large collection of white spruce expressed sequences: contributing factors and approaches for the categorization of SNPs. BMC Genomics 7:174

Pei M, Niu J, Li C, Cao F, Quan S (2016) Identification and expression analysis of genes related to calyx persistence in Korla fragrant pear. BMC Genomics 17:132

Punta M et al (2011) The Pfam protein families database. Nucleic Acids Res 40:D290–D301

Qiao Q et al (2016) Comparative transcriptomics of strawberries (*Fragaria* spp.) provides insights into evolutionary patterns. Front Plant Sci 7:1839

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Rai K, Gupta S, Ranjana B, Kulkarni V, Singh A, Rao A (2009) Morphological characteristics of ICRISAT-bred pearl millet hybrid seed parents. Andhra Pradesh, India, pp 4

Rajaram V et al (2013) Pearl millet [*Pennisetum glaucum* (L.) R. Br.] consensus linkage map constructed using four RIL mapping populations and newly developed EST-SSRs. BMC Genomics 14:159

Riahi L et al (2013) Characterization of single nucleotide polymorphism in Tunisian grapevine genome and their potential for population genetics and evolutionary studies. Genet Resour Crop Evol 60:1139–1151

Sahu PP, Gupta S, Malaviya D, Roy AK, Kaushal P, Prasad M (2012) Transcriptome analysis of differentially expressed genes during embryo sac development in apomeiotic non-parthenogenetic interspecific hybrid of *Pennisetum glaucum*. Mol Biotechnol 51:262–271

Salem M et al (2012) RNA-Seq identifies SNP markers for growth traits in rainbow trout. PLoS One 7:e36264

Sarah G et al (2017) A large set of 26 new reference transcriptomes dedicated to comparative population genomics in crops and wild relatives. Mol Ecol Resour 17:565–580

Schmid KJ, Sörensen TR, Stracke R, Törjék O, Altmann T, Mitchell-Olds T, Weisshaar B (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana. Genome Res 13:1250–1257

Singh RK et al (2013) Development, cross-species/genera transferability of novel EST-SSR markers and their utility in revealing population structure and genetic diversity in sugarcane. Gene 524:309–329

Soren KR et al (2015) EST-SSR analysis provides insights about genetic relatedness, population structure and gene flow in grass pea (*Lathyrus sativus*). Plant Breed 134:338–344

Sousa Azevedo AL, Costa PP, Machado JC, Machado MA, Pereira AV, José da Silva Lédo F (2012) Cross species amplification of *Pennisetum glaucum* microsatellite markers in Pennisetum purpureum and genetic diversity of Napier grass accessions. Crop Sci 52:1776–1785

Strezov V, Evans TJ, Hayman C (2008) Thermal conversion of elephant grass (*Pennisetum Purpureum* Schum) to bio-gas, bio-oil and charcoal. Bioresour Technol 99:8394–8399

Sujatha D, Manga V, Rao M, Murty J (1989) Meiotic studies in some species of *Pennisetum* (L.) rich. (Poaceae). Cytologia 54:641–652

Sureshkumar S et al (2014) Marker-assisted introgression of lpa2 locus responsible for low-phytic acid trait into an elite tropical maize inbred (*Zea mays* L.). Plant Breed 133:566–578

Takayama K, López PS, König C, Kohl G, Novak J, Stuessy TF (2011) A simple and cost-effective approach for microsatellite isolation in non-model plant species using small-scale 454 pyrosequencing. Taxon 60:1442–1449

Tang X, Xiao Y, Lv T, Wang F, Zhu Q, Zheng T, Yang J (2014) High-throughput sequencing and de novo assembly of the *Isatis indigotica* transcriptome. PLoS One 9:e102963

Taylor M, Vasil I (1987) Analysis of DNA size, content and cell cycle in leaves of Napier grass (*Pennisetum purpureum* Schum.). Theor Appl Genet 74:681–686

Toledo-Silva G, Cardoso-Silva CB, Jank L, Souza AP (2013) De novo transcriptome assembly for the tropical grass *Panicum maximum* Jacq. PLoS One 8:e70781

Ukoskit K, Posudsavang G, Pongsiripat N, Chatwachirawong P, Klomsa-ard P, Poomipant P, Tragoonrung S (2018) Detection and validation of EST-SSR markers associated with sugar-related traits in sugarcane using linkage and association mapping. Genomics. https://doi.org/10.1016/j.ygeno.2018.03.019

Varshney R et al (2008) Identification and validation of a core set of informative genic SSR and SNP markers for assaying functional diversity in barley. Mol Breed 22:1–13

Wang M, Yan J, Zhao J, Song W, Zhang X, Xiao Y, Zheng Y (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Sci 196:125–131

Wang X, Li S, Li J, Li C, Zhang Y (2015) De novo transcriptome sequencing in Pueraria lobata to identify putative genes involved in isoflavones biosynthesis. Plant Cell Rep 34:733–743

Wang Y et al (2017) Development of SNP markers based on transcriptome sequences and their application in germplasm identification in radish (*Raphanus sativus* L.). Mol Breed 37(26)

Wanjala BW et al (2013) Genetic diversity in Napier grass (*Pennisetum purpureum*) cultivars: implications for breeding and conservation. AoB Plants 5

Wei L et al (2014) Development of SNP and InDel markers via de novo transcriptome assembly in *Sesamum indicum* L. Mol Breed 34:2205–2217

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Wu J et al (2014a) High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. J Exp Bot 65:5771–5781

Wu Z-J, Li X-H, Liu Z-W, Xu Z-S, Zhuang J (2014b) De novo assembly and transcriptome characterization: novel insights into catechins biosynthesis in *Camellia sinensis*. BMC Plant Biol 14:277

Xie X-M, Lu X-L (2005) Analysis of genetic relationships of cultivars in *Pennisetum* by RAPD markers. Acta Pratacultural Science 2

Xie X-M, Zhou F, Zhang X-Q, Zhang J-M (2009) Genetic variability and relationship between MT-1 elephant grass and closely related cultivars assessed by SRAP markers. J Genet 88:281–290

Xu P et al (2011) A SNP and SSR based genetic map of asparagus bean (Vigna. unguiculata ssp. sesquipedialis) and comparison with the broader species. PLoS One 6:e15952

Xu M, Liu X, Wang J-W, Teng S-Y, Shi J-Q, Li Y-Y, Huang M-R (2017) Transcriptome sequencing and development of novel genic SSR markers for *Dendrobium officinale*. Mol Breed 37(18)

Yang X, Hang X, Tan J, Yang H (2015) Differences in acid tolerance between *Bifidobacterium breve* BB8 and its acid-resistant derivative B. breve BB8dpH, revealed by RNA-sequencing and physiological analysis. Anaerobe 33:76–84

Yao YF, Hong JJ, Zeng RQ (2013) SRAP analysis on genetic diversity of *Pennisetum*. J Gansu Agric Univ 4:108–109

Yates SA et al (2014) De novo assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. BMC Genomics 15:453

Yeh F (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg J Bot 129:157

Yue X, Nie Q, Xiao G, Liu B (2015) Transcriptome analysis of shell color-related genes in the clam *Meretrix meretrix*. Mar Biotechnol 17:364–374

Zeng A et al (2017) Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. Mol Breed 37(30)

Zhang Y, Cheng Y, Ya H, Han J, Zheng L (2015a) Identification of heat shock proteins via transcriptome profiling of tree peony leaf exposed to high temperature. Genet Mol Res 14:8431–8442

Zhang Z, Pang T, Li Q, Zhang L, Li L, Liu J (2015b) Transcriptome sequencing and characterization for *Kappaphycus alvarezii*. Eur J Phycol 50:400–407

Zhang W, Guo Y, Li J, Huang L, Kazitsa EG, Wu H (2016a) Transcriptome analysis reveals the genetic basis underlying the seasonal development of keratinized nuptial spines in *Leptobrachium boringii*. BMC Genomics 17:978

Zhang Y, Tao S, Yuan C, Liu Y, Wang Z (2016b) Non-monotonic dose–response effect of bisphenol A on rare minnow *Gobiocypris rarus* ovarian development. Chemosphere 144: 304–311

Zhao W et al (2014) RNA-Seq-based transcriptome profiling of early nitrogen deficiency response in cucumber seedlings provides new insight into the putative nitrogen regulatory network. Plant Cell Physiol 56:455–467

Zhou Z et al (2014) Transcriptome sequencing of sea cucumber (*Apostichopus japonicus*) and the identification of gene-associated markers. Mol Ecol Resour 14:127–138

Zou D, Chen X, Zou D (2013) Sequencing, de novo assembly, annotation and SSR and SNP detection of sabaigrass (*Eulaliopsis binata*) transcriptome. Genomics 102:57–62