SMRT Sequencing of a Full-length Transcriptome Reveals Transcript Variants Involved in C18 Unsaturated Fatty Acid Biosynthesis and Metabolism Pathways at Chilling Temperature in Pennisetum giganteum

Qingyuan Li
Wuhan Academy of Agricultural Sciences https://orcid.org/0000-0002-8579-3390
Conglin Xiang
Wuhan academy of agricultural sciences; Huazhong agricultural university
Lin Xu
Wuhan academy of agricultural sciences
Jinghua Cui
Wuhan academy of agricultural sciences; Huazhong agricultural university
Shao Fu
Wuhan academy of agricultural sciences
Baolin Chen
Wuhan academy of agricultural sciences
Shoukun Yang
Wuhan Academy of Agricultural Sciences
Pan Wang
Wuhan academy of agricultural sciences; Huazhong agricultural university
Yanfeng Xie
Wuhan academy of agricultural sciences
Ming Wei
Wuhan academy of agricultural sciences
Zhanchang Wang (wzc_ttt@sina.com)

Research article

Keywords: Pennisetum giganteum, full-length transcriptome, alternative splicing, C18 unsaturated fatty acids, chilling temperature

Posted Date: June 21st, 2019

DOI: https://doi.org/10.21203/rs.2.10522/v1

Version of Record: A version of this preprint was published on January 16th, 2020. See the published version at https://doi.org/10.1186/s12864-019-6441-3.

Abstract

Background Pennisetum giganteum, an abundant, fast-growing perennial C4 grass that belongs to the genus Pennisetum, family Poaceae, has been developed as a source of biomass for mushroom cultivation and production, as a source of forage for cattle and sheep, and as a tool to remedy soil erosion. However, having a chilling-sensitive nature, P. giganteum seedlings need to be protected while overwintering in most temperate climate regions. Results To elucidate the cold stress responses of P. giganteum, we carried out comprehensive full-length transcriptomes from leaf and root tissues under room temperature (RT) and chilling temperature (CT) using PacBio Iso-Seq long reads. We identified 196,124 and 140,766 full-length consensus transcripts in the RT and CT samples, respectively. We then systematically performed functional annotation, transcription factor identification, long non-coding RNAs (lncRNAs) prediction, and simple sequence repeat (SSR) analysis of those full-length transcriptomes. Moreover, isoform analysis revealed that alternative splicing events may be induced by cold stress in P. giganteum, and transcript variants may be involved in C18 unsaturated fatty acid biosynthesis and metabolism pathways at chilling temperature in P. giganteum. Conclusions We provide the first comprehensive full-length transcriptomic resource for the abundant and fast-growing perennial grass Pennisetum giganteum. Our results provide a useful genomic resource for exploring the biological pathways involved in the cold stress responses of P. giganteum.

Background

Temperature is a major environmental factor that affects plant growth, development, productivity and distribution [1]. In agriculture, cold stress may limit production, causing preharvest and postharvest damage and resulting in qualitative and quantitative losses [2]. Plants from temperate regions can increase their freezing tolerance by being exposed to chilling and non-freezing temperatures, which is known as cold acclimation [3]. By contrast, plants of tropical and subtropical origins, such as rice, maize, and C_4 grasses, largely lack such a capacity for cold acclimation and are sensitive to chilling stress [1].

As the principal barrier between the cytoplasm and the extracellular milieu, the plasma membrane is regarded as a key site of injury during cold stress [4]. As temperatures below the optimal requirements for organisms cause membrane lipid rigidity and abnormal cellular activities, plants respond to such environmental influences by remodeling the lipid composition of their membrane [5]. As phospholipids rich in unsaturated fatty acids often have a considerably lower transition temperature compared to phospholipids containing high amounts of saturated fatty acids, plants with higher unsaturated fatty acids content in the plasma membranes usually exhibit more cold tolerance [6, 7]. The maintenance of polyunsaturated fatty acid levels in chloroplast lipids has been shown to contribute to survival at low temperatures and the normal formation of chloroplast membranes in plants under cold stress [8]. Trienoic fatty acids (TAs), such as hexadecatrienoic acid (16:3) and linolenic acid (18:3), which are considered to be the major polyunsaturated fatty acid species in plant membrane lipids, are important to ensure the correct biogenesis and maintenance of chloroplasts during plant growth under low temperatures [9]. A study on *Camellia japonica* showed that α-linolenic acid biosynthesis and metabolism pathways may play roles in plant cold responses [7].

C_4 photosynthesis is theoretically more efficient than C3 photosynthesis in light, nitrogen and water use [10]. C_4 grasses dominate most open biomes in tropical and subtropical areas, where they achieve greater biomass and higher growth rates [11]. However, in cooler environments, the peak yields of most C_4 plants are markedly reduced [12]. As a result, the present global distribution of C_4 grasses is largely limited to warmer climate regions, and strong positive relationships between C_4 grass abundance and growing season temperature have been documented at continental scales and along elevational gradients on tropical mountains across the globe [13].

Pennisetumgiganteum, an abundant, fast growing perennial C_4 grass that belongs to the genus *Pennisetum*, family Poaceae, is native to eastern and northeastern African tropical regions, such as Kenya, Eritrea and Ethiopia. This grass has been planted in more than 30 provinces in China and more than 80 countries worldwide [14]. In addition, at present, *P. giganteum* has been developed as a source of biomass for mushroom cultivation and production, a source of forage for cattle and sheep, and a tool to remedy soil erosion [15-18]. However, having a chilling-sensitive nature, *P. giganteum* seedlings need to be protected during overwintering in most temperate climate regions. Therefore, improving the chilling tolerance of *P. giganteum*will be important for livestock husbandry and earth ecology.

Third-generation sequencing platforms, such as single-molecule real-time (SMRT) sequencing from PacBio, can generate full-length cDNA sequences without assembly [19, 20]. Isoform sequencing (Iso-Seq), which is based on the SMRT sequencing platform, has been used to analyze full-length transcriptomes in various plant species [21, 22]. In this study, we used PacBio Iso-Seq to generate comprehensive full-length transcriptomes for *P. giganteum* under room temperature (RT) and chilling temperature (CT). We then systematically carried out functional annotation, transcription factor identification, long non-coding RNAs (lncRNAs) prediction, and simple sequence repeat (SSR) analysis of those full-length transcriptomes. Moreover, isoform analysis revealed the complexity of alternative splicing in *P. giganteum*, and transcript variants may be involved in C18 unsaturated fatty acid biosynthesis and metabolism pathways at chilling temperature in *P. giganteum*. In this study, we not only systematically characterized the profile of the *P. giganteum* full-length transcriptome but also provided a valuable resource for investigating the biological pathways involved in the cold response in *P. giganteum*.

Results

P. giganteum transcriptome analysis using PacBio Iso-seq

Two pooled samples (from leaf and root tissues) of RT and CT were sequenced to obtain a wide coverage of the *P. giganteum* full-length transcriptome using PacBio Iso-Seq. For the RT sample, a total of 509,371 circular consensus sequencing (CCS) reads were generated, with a total of 558,634,435 nucleotides. For the CT sample, a total of 371,590 CCS reads were generated, with a total of 442,366,361 nucleotides (Supplementary Table 1). The subreads distribution is shown in Supplementary Figure 1a.

By applying the standard Iso-Seq classification and clustering protocol on the above CCS reads, we produced 393,678 full length reads, including 382,945 full-length non-chimeric (Flnc) reads with an average length of

2,370 for the RT sample. For the CT sample, we generated 273,168 full length reads, including 263,302 Flnc reads with an average length of 2,489. Finally, 196,124 and 140,766 full-length consensus transcripts were generated in the RT and CT samples, respectively (Table 1). The Flnc reads distribution is shown in Supplementary Figure 1b.

After combining the RT and CT data, a total of 336,890 transcripts and 319,926 unigenes were generated, and the distribution of transcripts and unigenes is shown in Supplementary Table 2.

Functional annotation of the assembled transcriptome

To obtain a comprehensive functional annotation of the *P. giganteum* transcriptome, we assessed the non-redundant transcripts from RT and CT samples using a BLASTX search against the following databases: Nr (NCBI non-redundant protein sequences), Nt (NCBI non-redundant nucleotide sequences), Swiss-Prot (a manually annotated and reviewed protein sequence database), GO (Gene Ontology), COG (cluster of orthologous groups) and KEGG (Kyoto Encyclopedia of Genes and Genomes). For the RT sample, 163,748 (83.49%), 165,942 (84.61%), 124,098 (63.28%), 65,341 (33.32%), 87,244 (44.48%), and 158,795 (80.97%) unigenes returned BLAST results and showed identity with sequences in the Nr, Nt, SwissProt, GO, KOG and KEGG databases, respectively. For the CT sample, 120,529 (85.62%), 122,147 (86.77%), 99,970 (71.02%), 71,041 (50.47%), 73,324 (52.09%), 117,980 (83.81%) unigenes returned BLAST results and showed identity with sequences in the Nr, Nt, SwissProt, GO, COG and KEGG databases, respectively (Supplementary Table 3).

To determine the potential functions of unigenes, we used GO assignments to classify the predicted *P. giganteum* genes. It is interesting that unigenes identified from the GO database in the CT sample (71,041) were more than those in the RT sample (65,341), although the total unigenes in the CT sample (140,766) were less than those in the RT sample (196,124). Overall, we observed highly similar GO functional classifications in the RT and CT samples (Figure 1a). In terms of biological processes, ‘metabolic processes’ (28.26% in RT sample and 26.16% in CT sample) and ‘cellular processes’ (26.49% in RT sample and 24.67% in CT sample) were the top two GO terms in both treatments. It is notable that the percentages of unigenes in the terms ‘biological regulation’, ‘regulation of biological process’, ‘response to stimulus’ and ‘signaling’ in the CT sample were greater than that of the RT sample, which may indicate that plants were undergoing stress stimulated biological regulations in the CT sample. In the molecular function category, the transcripts were predominantly assigned to the ‘binding’ (49.47% in RT sample and 50.62% in CT sample) and ‘catalytic activities’ (37.88% in RT sample and 37.84% in CT sample) groups in both treatments. In the cellular component category, the unigenes were frequently assigned to ‘cell part’ (~20%), ‘cell’ (~20%), ‘membrane’ (~15%) in both CT and RT samples.

To identify the biological pathways in the annotated *P. giganteum* sequences, we annotated the unigenes to the reference pathways in the KEGG using KeggArray software. A total of 158,795 and 117,980 unigenes were assigned to six specific pathways, including ‘Organism Systems’, ‘Metabolism’, ‘Human disease’, ‘Genetic Information Processing’, ‘Environmental Information Processing’, and ‘Cellular Processes’ in the RT and CT samples, respectively (Figure 1b). As with the GO classification, the percentages of different classes of KEGG pathway terms were highly similar in the RT and CT samples. Nearly 40% of annotated unigenes

were classified as belonging to 'Metabolism'-related pathways, in which 'Energy metabolism', 'Carbohydrate metabolism', and 'Global and overview maps' were the top 3 pathways with the most abundant unigenes in both RT and CT samples. The difference between the two treatments with respect to KEGG annotation is the percentages of genes involved in 'Energy metabolism' and 'Carbohydrate metabolism'. There was a higher percentage of unigenes associated with 'Energy metabolism' in RT compared with CT sample (35.15% compared with 24%); however, there was a lower percentage of unigenes associated with 'Carbohydrate metabolism' in RT compared with CT sample (18.36% compared with 21.42%) (Figure 1b).

To classify the orthologous gene products, 87,244 and 73,324 unigenes were subdivided into COG classifications in the RT and CT samples, respectively (Figure 1c). The percentages of different classes of COG classifications were highly similar in the RT and CT samples. In both treatments, the cluster of 'general function prediction only' (16.19% in RT sample and 17.40% in CT sample) represented the largest group, followed by 'signal transduction mechanisms' (10.49% in RT sample and 11.02% in CT sample) and 'posttranslational modification, protein turnover, chaperones' (9.83% in the RT sample and 9.39% in the CT sample).

Taken together, the results from GO, KEGG and COG annotation and classification of assembled genes allowed us to obtain a comprehensive functional characterization for the full-length transcriptomes from CT and RT treatments of *P. giganteum*. The overall similar functional classification of transcripts in these two treatments indicates that the transcriptome at the pathway level is generally conserved, although some subtle differences can be found between the CT and RT samples.

Transcription Factors identification

A total of 4,974 and 5,170 putative TF genes were identified in the RT and CT samples, respectively (Supplementary Table 4). Notably, the number of putative TF genes in the CT sample was greater than that in the RT sample, although the total number of unigenes in the CT sample (140,766) was less than that in the RT sample (196,124). Among all TF families, the C2H2 family was the largest group in both RT and CT samples (391, 7.86% in RT sample and 400, 7.74% in CT sample). The C3H family (362, 7.28%) and the GRAS family (338, 6.80%) were followed by in the RT sample. For the CT sample, the second and third largest groups were represented by the MYB-related family (360, 6.96%) and the AP2/ERF family (334, 6.46%). Furthermore, the unigenes in the FAR1, bHLH, WRKY, bZIP, B3-ARF and HB-BELL TF families in the CT sample were more than those in the RT sample (Figure 2, Supplementary Table 4).

LncRNA prediction

LncRNAs from these PacBio Iso-Seq data sets were predicted by CNCI, Pfam, PLEK and CPC protein structure domain analysis. In total, 18,461 and 12,701 candidate lncRNAs of ≥200 bp were predicted by all four methods in the RT and CT samples, respectively (Figure 3a). The lncRNAs had a length ranging from 200 to 15,913 bp with an average length of 559 bp in the RT sample and a length ranging from 200 to 7359 bp with an average length of 491 bp in the CT sample (Supplementary Table 5). The length distribution of lncRNAs is

shown in Figure 3b, and most of them were single-isoform transcripts present in both treatments (Figure 3c). The functions of these lncRNAs need to be further characterized.

Genic-SSR identification

SSRs were highly abundant in the assembled *P. giganteum* transcriptome. In total, 39,309 potential SSRs with a minimum of five repetitions for all motifs (di- to hexa- nucleotides) were identified from 33,463 contigs, representing 12.29% of the total 319,926 unigenes in RT and CT combined transcriptome. The frequency of occurrence of SSR loci was one in every 17.6 kb of full-length unigene sequences. Among all repeat types, the length of SSRs was distributed from 12 to 140 bp with an average of 16.61 bp.

Incidences of different repeat types and frequencies for each motif were evaluated based on the repeat unit number (Table 2). SSRs existed mainly as dinucleotide repeats and trinucleotide repeats, accounting for 97.65%. Trinucleotide repeats, comprising 72.11% of the total SSRs, were the most abundant repeat unit, followed by di- (24.37%), tetra-(2.34%), penta-(0.62%) and hexa-nucleotides (0.56%). Most (97.39%) of the motifs had 5-10 repeat units, while motifs with more than 10 reiterations were rare, exhibiting a frequency of 2.61%. Within the identified SSRs, AG/CT comprised 54.13% of all dinucleotide repeat motifs and was the most common type (Figure 4a). The predominant trinucleotide repeat motifs were CCG/CGG and AGC/CTG, which accounted for 41.41% and 19.86%, respectively (Figure 4b). In tetranucleotide repeats, the most frequent motif was AATG/CATT (16.01%) followed by ACAT/ATGT (9.48%) and AAAT/ATTT (8.28%) (Figure 4c).

Gene alternative splicing detection

One advantage of PacBio Iso-Seq is its ability to describe the complexity of alternative splicing at the whole-transcriptome scale. To detect the alternative splicing event inthe *P. giganteum* transcriptome, the Coding GENome reconstruction Tool (Cogent) was used to further partition these error-corrected non-redundant transcripts into transcript families and reconstruct each family into one or several full-length unique transcript model(s) (referred to as UniTransModel). In total, 63,696 and 48,102 full-length UniTransModels were obtained for RT and CT samples, respectively. Then, transcript isoforms were identified in both samples, a total of 28.61% of full-length UniTransModels had more than one isoform in the RT sample and a slightly more percentage (33.74%) in the CT sample (Figure 5a).

We identified alternative splicing events by using the UniTransModels as our reference. In total, 6,571 and 8,088 UniTransModel-based alternative splicing events were identified in RT and CT samples, representing 10.32% of the 63,696 full-length UniTransModels for RT sample and 16.81% of the 48,102 full-length UniTransModels for CT sample, respectively. The transcript isoform numbers in both samples with alternative splicing events are shown in Figure 5b. It is notable that the percentages of UniTransModels had two isoforms and more in CT sample were all higher than that in the RT sample. These results indicated that alternative splicing events may be induced by cold stress in *P. giganteum*.

By mapping Illumina short reads to transcript models, we were able to confirm the reliability of isoform detection using our pipeline, even in the absence of a reference genome (Figure 5c). We also detected different splicing isoforms of the same UniTransModels in RT and CT samples (Figure 5c).

Gene alternative splicing involved in α-linolenic acid biosynthesis and metabolism pathways

The octadeca-carbon unsaturated fatty acids play important roles in plant cold responses. We searched our UniTransModels with alternative splicing events using BLASTX against the KEGG databases. We determined that 12 genes involved in the α-linolenic acid metabolism pathway had alternative splicing events in the RT sample and 14 in the CT sample (Supplementary Table 6). We also found that several of these genes might have different transcription isoforms in RT and CT samples. For example, the enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase gene *MFP2* had 4 isoforms in the RT sample whereas had 9 isoforms in the CT sample (Figure 6).

Interestingly, we also determined that 6 genes involved in α-linolenic acid biosynthesis pathway had alternative splicing events in the CT sample, including one very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase gene (HACD), PB.15065_8_path0, one acyl-[acyl-carrier-protein] desaturase gene (FAB2), PB.11646_0_path0, three 3-oxoacyl-[acyl-carrier protein] reductase genes (FabGs), PB.14731_0_path0, PB.18504_0_path0, PB.3134_0_path0, and one acyl-coenzyme A thioesterase gene (ACOT), PB.8528_0_path0 (Supplementary Table 6), whereas no genes were found that had alternative splicing events in such a pathway in the RT sample.

These results indicated that the alternative splicing events in α-linolenic acid biosynthesis and metabolism pathway genes may be induced by low temperature in *P. giganteum*. Our full-length Iso-Seq transcriptome can provide not only additional information for characterization of the α-linolenic acid biosynthesis and metabolism biosynthesis pathways at a deeper transcription isoform level but will also help understand other biological pathways under normal conditions and cold stress in *P. giganteum*.

Discussion

C_4 plants play a key role in world agriculture, crops such as maize and sorghum are major contributors to world food production in both developed and developing nations, and C_4 grasses are the major plant sources of bioenergy [23]. However, most C_4 plants are chilling sensitive [12]. It is important to develop an effective strategy for the genetic breeding of chilling tolerance in *P. giganteum*.

With the application of third-generation sequencing technology that can produce average read lengths of more than 10,000 bp [24], transcriptome reconstruction and annotation, particularly for species without a reference genome sequence, has improved significantly. Although several transcriptomic analyses in the genus *Pennisetum* have been reported [25, 26], transcriptomic studies in *P. giganteum* are still rare. In this study, we generated 382,945 Flnc reads with an average length of 2,370 for the RT sample and 263,302 Flnc reads with an average length of 2,489 for the CT sample (Supplementary Table 1). These large-scale and

high-quality full-length transcriptome data not only provide useful reference data but can also be used to elucidate chilling response pathways and conduct subsequent functional genomics studies in *P. giganteum* and other plants of the genus *Pennisetum*. To the best of our knowledge, this report describes the first full-length transcriptomic analysis of *P. giganteum*. Moreover, a total of 178,330 of the 196,124 unigenes (90.93%) in the RT sample and 129,746 of the 140,766 unigenes (92.17%) in the CT sample were annotated to public databases (Nr, Nt, SwissProt, GO, KOG and KEGG) for comprehensive analysis (Supplementary Table 3). Compared with previous studies in plants of the genus *Pennisetum*, which reported 103,454 annotated unigenes from 197,466 unigenes [27], the present study obtained more complete annotation information.

LncRNAs are a large and diverse class of transcribed RNA molecules with a length of more than 200 nucleotides that do not encode proteins (or lack > 100 amino acid open reading frame) and potentially play important roles in the gene regulation of eukaryotic cells [28]. Despite accumulating evidence suggesting that lncRNAs have crucial roles in gene expression control during both developmental and differentiation processes, such as in dosage compensation, genomic imprinting, cell differentiation and organogenesis [29], only a relatively small proportion of lncRNAs have been studied in detail. In this study, we identified 18,461 and 12,701 candidate lncRNAs in RT and CT samples, respectively (Figure 3). Although the functions of these lncRNAs have not been characterized yet, these still provide a useful resource for understanding the potential functions of lncRNAs in *P. giganteum*.

In eukaryotic cells, approximately 95% of genes undergo RNA transcript splicing, where most introns are removed and exons are retained, resulting in multiple alternative transcripts (isoforms) of the gene(s) [30, 31]. In metazoans, alternative splicing plays an important role in generating different protein products that function in diverse cellular processes, including cell growth, differentiation and death [32]. In this study, we identified 6,571 and 8,088 UniTransModel-based alternative splicing events in RT and CT samples (Figure 5). The high proportion of transcripts with multiple splicing isoforms suggests a high degree of transcriptome complexity in *P. giganteum*. Previous studies have suggested that alternative splicing may play important roles in plant adaptation to environmental stresses [33, 34]. A recent study showed that rapid and dynamic alternative splicing impacts the cold response in Arabidopsis [1]. Similarly, in this study, we found that the alternative splicing level in the CT sample was higher than that in the RT sample (Figure 5), indicating that alternative splicing may be one of the major drivers of transcriptome reprogramming for the chilling response in *P. giganteum*.

Although no specific receptor has been identified in plants in response to cold stress, it has been considered that the plasma membrane itself could be the primary sensor of temperature fluctuations [35]. Unsaturated fatty acids constitute the major components of membrane lipids of both prokaryotes and eukaryotic organisms [5]. Previous studies showed that fatty acid compositions were altered during plants exposed to cold temperatures [36, 37], and the level of unsaturated fatty acids is associated with cold tolerance [6, 7]. In our study, 6 genes involved in the α-linolenic acid biosynthesis pathway were identified as having alternative splicing events in the CT sample, whereas no genes were found to have alternative splicing events in the same pathway in the RT sample (Supplementary Table 6), implying that the C18 unsaturated fatty acid biosynthesis pathway might be involved in the chilling stress in *P. giganteum*. It is notable that α-linolenic

acid, one of the most important C18 unsaturated fatty acids in plants, is the substrate of jasmonic acid (JA). Genes involved in the α-linolenic acid metabolism pathway are also involved in JA biosynthesis [38, 39]. It is well known that JA and its derivatives are important regulators of plant responses to abiotic stresses [39, 40]. In the present study, we found that a number of genes involved in the α-linolenic acid metabolism pathway had alternative splicing events in both RT and CT samples, and several of these genes had different transcript isoforms in RT and CT samples (Figure 6; Supplementary Table 6). These results indicate that as a downstream pathway of α-linolenic acid biosynthesis, the α-linolenic acid metabolism pathway may also be involved in chilling responses in *P. giganteum*. However, the relationship between the alternative splicing events and transcription levels of related genes, and the actual functions of the alternative splicing isoforms will need to be determined in future studies.

Conclusions

In this study, we performed an in-depth full-length transcriptome analysis of the responses to chilling temperature in *P. giganteum*. Our results systematically characterized the transcriptome information of *P. giganteum* at RT and CT and revealed that transcript variants may be involved in C18 unsaturated fatty acid biosynthesis and metabolism pathways at chilling temperature in *P. giganteum*, further expanding our knowledge of chilling responses in C_4 grasses.

Methods

Plant materials and preparation

Mother plant of *P. giganteum* seedlings were obtained from National Engineering Research Center of JUNCAO Technology, Fujian, China. The plant material was identified by Professor Shoukun Yang, Forestry and Fruit Tree Research Institute, Wuhan Academy of Agricultural Sciences. And the germplasm (No. WH-H-JJC-2013) is deposited in the Germplasm Resources Nursery of Forestry and Fruit Tree Research Institute, Wuhan Academy of Agricultural Sciences, Wuhan, China.

In this study, *P. giganteum* seedlings were grown in growth chambers in Forestry and Fruit Tree Research Institute, Wuhan Academy of Agricultural Sciences, China, under long-day conditions (16 h light/8 h dark) under white fluorescent light at 25 °C during the day and 22 °C at night. Plants were then divided into two groups. The first group served as the RT sample, while the other group was moved and cultured at 4 °C as the CT sample.

Leaves and roots of *P. giganteum* seedlings at RT and at 2, 8 and 24 h of CT were sampled for RNA isolation. Total RNA was extracted using a Plant Total RNA Extraction Kit (BioTeke, China) following the manufacturer's instructions and then treated with RNase-free DNase I (Thermo Scientific, USA) to remove genomic DNA contamination. The quality of extracted RNAs was evaluated using an Agilent 2100 Bioanalyzer (SA Pathology, Adelaide, SA, Australia). For the RT sample, RNA isolated from leaves and roots was evenly pooled. For the CT sample, RNA isolated from leaves and roots at 2, 8 and 24 h of CT was evenly pooled.

Library preparation and sequencing

RNA samples were sent to Beijing Novogene Bioinformatics Technology Co., Ltd., where the libraries were produced and sequenced. Briefly, the Iso-Seq library was prepared according to the Isoform Sequencing protocol (Iso-Seq) using the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PN 100-092-800-03).

Data processing

Sequence data were processed using SMRT Link 4.0 software. Circular consensus sequence (CCS) was generated from subread BAM files, parameters: min_length 200, max_drop_fraction 0.8, no_polish TRUE, min_zscore -999, min_passes 1, min_predicted_accuracy 0.8, and max_length 18000. CCS. BAM files were output, which were then classified into full-length and non-full-length reads using pbclassify.py, ignorepolyA false, minSeqLength 200. Subsequently produced non-full-length and full-length fasta files were then fed into the cluster step, which performs isoform-level clustering (ICE) followed by final Arrow polishing, hq_quiver_min_accuracy 0.99, bin_by_primer false, bin_size_kb 1, qv_trim_5p 100, and qv_trim_3p 30.

Gene functional annotation

Gene function was annotated based on the following databases: Nr (NCBI non-redundant protein sequences); Nt (NCBI non-redundant nucleotide sequences); KOG/COG (Clusters of Orthologous Groups of proteins); Swiss-Prot (A manually annotated and reviewed protein sequence database); KO (KEGG Ortholog database); and GO (Gene Ontology).

Coding potential analysis and lncRNA identification

CNCI (Coding-Non-Coding-Index) profiles adjoining nucleotide triplets to effectively distinguish protein-coding and non-coding sequences independent of known annotations. We use CNCI with default parameters. CPC (Coding Potential Calculator) mainly assesses the extent and quality of the ORF in a transcript and searches the sequences with known protein sequence databases to clarify the coding and non-coding transcripts. We used the NCBI eukaryote protein database and set the e-value ‘1e-10’ in our analysis. We translated each transcript into all three possible frames and used Pfam Scan to identify the occurrence of any of the known protein family domains documented in the Pfam database. Any transcript with a Pfam hit would be excluded in the following steps. Pfam searches use default parameters of -E 0.001 -- domE 0.001. PLEK (Predictor of Long non-coding RNAs and Messenger RNAs based on an Improved k-mer Scheme) is an efficient alignment-free computational tool to distinguish lncRNAs from mRNAs in RNA-seq transcriptomes of species lacking reference genomes. PLEK is especially suitable for PacBio or 454 sequencing data and large-scale transcriptome data. We use PLEK with default parameters.

We filtered out the sequences with an ORF≥200 aa, after which the remaining sequences were characterized for potential non-coding RNAs, including small RNAs, microRNAs and lncRNAs.

SSR analysis

SSR of the transcriptome was identified using MISA. This analysis allows the identification and localization of perfect microsatellites, as well as compound microsatellites that are interrupted by a certain number of bases.

Full-length unique transcript model reconstruction

The non-redundant transcripts were processed with Coding GENome reconstruction Tool (Cogent v1.4, https://github.com/Magdoll/Cogent). In general, Cogent first creates the k-mer profile of non-redundant transcripts and calculates pairwise distances and then clusters transcripts into families based on their k-mer similarity. Each transcript family is further reconstructed into one or several unique transcript model(s) (referred to as UniTransModels) using a De Bruijn graph method [21].

Isoform identification

Error-corrected non-redundant transcripts (transcripts before Cogent reconstruction) were mapped to UniTransModels using GMAP v2014-08-04 [41]. Splicing junctions for transcripts mapped to the same UniTransModels were examined, and transcripts with the same splicing junctions were collapsed. Collapsed transcripts with different splicing junctions were identified as transcription isoforms of UniTransModels. AS events were detected with SUPPA using default settings [42].

Abbreviations

SMRT: Single Molecule Real-Time; RT: Room Temperature; CT: Chilling Temperature; Iso-Seq: Isoform Sequencing; lncRNA: long non-coding RNA; SSR: Simple Sequence Repeat; TAs: Trienoic fatty Acids; CCS: Circular Consensus Sequencing; Flnc: Full-length non-chimeric; Nr: NCBI non-redundant protein sequences; Nt: NCBI non-redundant nucleotide sequences; Go: Gene Ontology; COG: Cluster of Orthologous Groups; KEGG: Kyoto Encyclopedia of Genes and Genomes; TF: Transcription Factor; CNCI: Coding-Non-Coding-Index; PLEK: Predictor of Long non-coding RNAs and Messenger RNAs based on an Improved k-mer Scheme; CPC: Coding Potential Calculator; Cogent: Coding GENome reconstruction Tool.

Declarations

Acknowledgments

We would like to express our sincere thanks to Beijing Novogene Bioinformatics Technology Co., Ltd. for helping in the analysis of transcriptome data.

Funding

This work was supported by the Technological Projects of Wuhan (2015021388010001 2016020899010991). The funding bodies had no role in the design, collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All the data was available in present paper. The sequencing data was available at NCBI database with accession number of PRJNA544770.

Authors' Contributions

QL and ZW conceived and designed the experiments. CX and LX performed the RNA isolation experiment. QL analyzed the data and wrote the manuscript. JC, SF and PW helped to analyze the data and draft the manuscript. BC, YX and MW participated in the design of the study. SY identified the plant material. ZW coordinated the study and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Maleki M, Ghorbanpour M: Cold tolerance in plants: molecular machinery deciphered. In: *Biochemical, Physiological and Molecular Avenues for Combating Abiotic Stress Tolerance in Plants.* Elsevier; 2018: 57-71.

2. Sanghera GS, Wani SH, Hussain W, Singh N: Engineering cold stress tolerance in crop plants. *Current genomics* 2011, 12(1):30.

3. Miura K, Furumoto T: Cold signaling and cold response in plants. *International journal of molecular sciences* 2013, 14(3):5312-5337.

4. Steponkus PL: Cold acclimation and freezing injury from a perspective of the plasma membrane. *Environmental injury to plants* 1990:1-16.

5. Garba L, Shukuri Mo M, Nurbaya Os S, Noor Zalih R: Review on fatty acid desaturases and their roles in temperature acclimatisation. *Journal of Applied Sciences* 2017, 17:282-295.

6. Murata N, Ishizaki-Nishizawa O, Higashi S, Hayashi H, Tasaka Y, Nishida I: Genetically engineered alteration in the chilling sensitivity of plants. *Nature* 1992, 356(6371):710.

7. Li Q, Lei S, Du K, Li L, Pang X, Wang Z, Wei M, Fu S, Hu L, Xu L: RNA-seq based transcriptomic analysis uncovers alpha-linolenic acid and jasmonic acid biosynthesis pathways respond to cold acclimation in *Camellia japonica*. *Sci Rep* 2016, 6:36463.

8. Upchurch RG: Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnology letters* 2008, 30(6):967-977.

9. Routaboul J-M, Fischer SF: Trienoic fatty acids are required to maintain chloroplast function at low temperatures. *Plant physiology* 2000, 124(4):1697-1705.

10. Long SP, Spence AK: Toward cool C_4 crops. *Annual Review of Plant Biology* 2013, 64:701-722.

11. Atkinson R, Mockford E, Bennett C, Christin P, Spriggs E, Freckleton R, Thompson K, Rees M, Osborne C: C_4 photosynthesis boosts growth by altering physiology, allocation and size. *Nature Plants* 2016, 2.

12. Watcharamongkol T, Christin PA, Osborne CP: C_4 photosynthesis evolved in warm climates but promoted migration to cooler ones. *Ecology letters* 2018, 21(3):376-383.

13. Edwards EJ, Still CJ: Climate, phylogeny and the ecological distribution of C_4 grasses. *Ecology letters* 2008, 11(3):266-276.

14.Lin Z: Juncao Science. Beijing: National School of Administration Press; 2013.

15. Lin X, Lin Z, Lin D, Lin H, Luo H, Hu Y, Lin C, Zhu C: Effects of planting Pennisetum sp.(Giant Juncao) on soil microbial functional diversity and fertility in the barren hillside. *Acta Ecol Sin* 2014, 34(15):4304-4312.

16. Lin Z, Lin D, Su D, Lin H, Li J, Zheng D, Yu S: Effect of different salt-affected soils on biological characteristics of Pennisetum sp. *Southwest China Journal of Agricultural Sciences* 2015, 28(2):675-680.

17. Lin X, Lin Z, Lin D, Lin H, Luo H, Hu Y, Lin C, Zhu C: Preliminary study on Juncao as biomass fuel. *Journal of Fujian College of Forestry* 2013, 1.

18. Yang L, Fan J, Liu X, Lin Z, Liu B: Effects of grass-substrates on growth and nutrients of ganoderma lucidum. *Fujian Journal of Agricultural Sciences* 2017, 32(5):508-511.

19. Dong L, Liu H, Zhang J, Yang S, Kong G, Chu JS, Chen N, Wang D: Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC genomics* 2015, 16(1):1039.

20. Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS: A survey of the sorghum transcriptome using single-molecule long reads. *Nature communications* 2016, 7:11706.

21. Li J, Harata-Lee Y, Denton MD, Feng Q, Rathjen JR, Qu Z, Adelson DL: Long read reference genome-free reconstruction of a full-length transcriptome from Astragalus membranaceus reveals transcript variants involved in bioactive compound biosynthesis. *Cell discovery* 2017, 3:17031.

22. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D: Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nature communications* 2016, 7:11708.

23. Von Caemmerer S, Ghannoum O, Furbank RT: C_4 photosynthesis: 50 years of discovery and innovation. *Journal of experimental botany* 2017, 68(2):97.

24. Roberts RJ, Carneiro MO, Schatz MC: The advantages of SMRT sequencing. *Genome biology* 2013, 14(6):405.

25. Jaiswal S, Antala TJ, Mandavia M, Chopra M, Jasrotia RS, Tomar RS, Kheni J, Angadi U, Iquebal M, Golakia B: Transcriptomic signature of drought response in pearl millet (*Pennisetum glaucum* (L.) and development of web-genomic resources. *Scientific reports* 2018, 8(1):3382.

26. Dudhate A, Shinde H, Tsugama D, Liu S, Takano T: Transcriptomic analysis reveals the differentially expressed genes and pathways involved in drought tolerance in pearl millet [*Pennisetum glaucum* (L.) R. Br]. *PloS one* 2018, 13(4):e0195908.

27. Zhou S, Wang C, Frazier TP, Yan H, Chen P, Chen Z, Huang L, Zhang X, Peng Y, Ma X: The first Illumina-based de novo transcriptome analysis and molecular marker development in Napier grass (*Pennisetum purpureum*). *Molecular breeding* 2018, 38(7):95.

28. Mercer TR, Dinger ME, Mattick JS: Long non-coding RNAs: insights into functions. *Nature reviews genetics* 2009, 10(3):155.

29. Fatica A, Bozzoni I: Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews Genetics* 2014, 15(1):7.

30. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* 2008, 40(12):1413.

31. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008, 456(7221):470.

32. Chen M, Manley JL: Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature reviews Molecular cell biology* 2009, 10(11):741.

33. Filichkin S, Priest HD, Megraw M, Mockler TC: Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Current opinion in plant biology* 2015, 24:125-135.

34. Laloum T, Martín G, Duque P: Alternative splicing control of abiotic stress responses. *Trends in plant science* 2018, 23(2):140-150.

35. Moustafa K, AbuQamar S, Jarrar M, Al-Rajab AJ, Trémouillaux-Guiller J: MAPK cascades and major abiotic stresses. *Plant cell reports* 2014, 33(8):1217-1225.

36. Guo Y, Liu S, Yang Z, Tian S, Sui N: Responses of unsaturated fatty acid in membrane lipid and antioxidant enzymes to chilling stress in sweet sorghum (*Sorghum bicolor* (L.) Moench) seedling. *J Agric Sci* 2016, 8(9):71.

37. Shahandashti SSK, Amiri RM, Zeinali H, Ramezanpour SS: Change in membrane fatty acid compositions and cold-induced responses in chickpea. *Molecular biology reports* 2013, 40(2):893-903.

38. Wasternack C, Hause B: Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Annals of botany* 2013, 111(6):1021-1058.

39. Wasternack C: Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Annals of botany* 2007, 100(4):681-697.

40. Fang L, Tong J, Dong Y, Xu D, Mao J, Zhou Y: De novo RNA sequencing transcriptome of Rhododendron obtusum identified the early heat response genes involved in the transcriptional regulation of photosynthesis. *PloS one* 2017, 12(10):e0186376.

41. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ: GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. In: *Statistical Genomics.* Springer; 2016: 283-334.

42. Alamancos GP, Pagès A, Trincado JL, Bellora N, Eyras E: Leveraging transcript quantification for fast computation of alternative splicing profiles. *Rna* 2015, 21(9):1521-1531.

Tables

Table 1 Summary of consensus transcripts after Iso-Seq classification and clustering protocol.

Sample	Number of 5'-primer reads	Number of 3'-primer reads	Number of Poly-A reads	Number of full length reads	Number of flnc reads	Average flnc read length	Full-length percentage (%)	Consensus transcripts
RT	464,021	464,131	448,301	393,678	382,945	2,370	77	196,124
CT	330,088	331,916	324,607	273,168	263,302	2,489	74	140,766

Table 2 Frequencies of different SSR repeat motif types observed in *P. giganteum* transcriptome.

SSR motif	Repeat number								Percentage
	5	6	7	8	9	10	>10	Tatal	(%)
Dinucleotide	0	4,596	1,813	1,216	710	387	859	9,581	24.37
Trinucleotide	18,897	6,009	2,135	819	274	89	121	28,344	72.11
Tetranucleotide	649	134	85	26	9	2	13	918	2.34
Pentanucleotide	170	28	18	6	2	2	19	245	0.62
Hexanucleotide	127	54	17	4	2	2	15	221	0.56
Total	19,843	10,821	4,068	2,071	997	482	1,027	39,309	100.00
Percentage (%)	50.48	27.53	10.35	5.27	2.54	1.23	2.61	100.00	

Figures

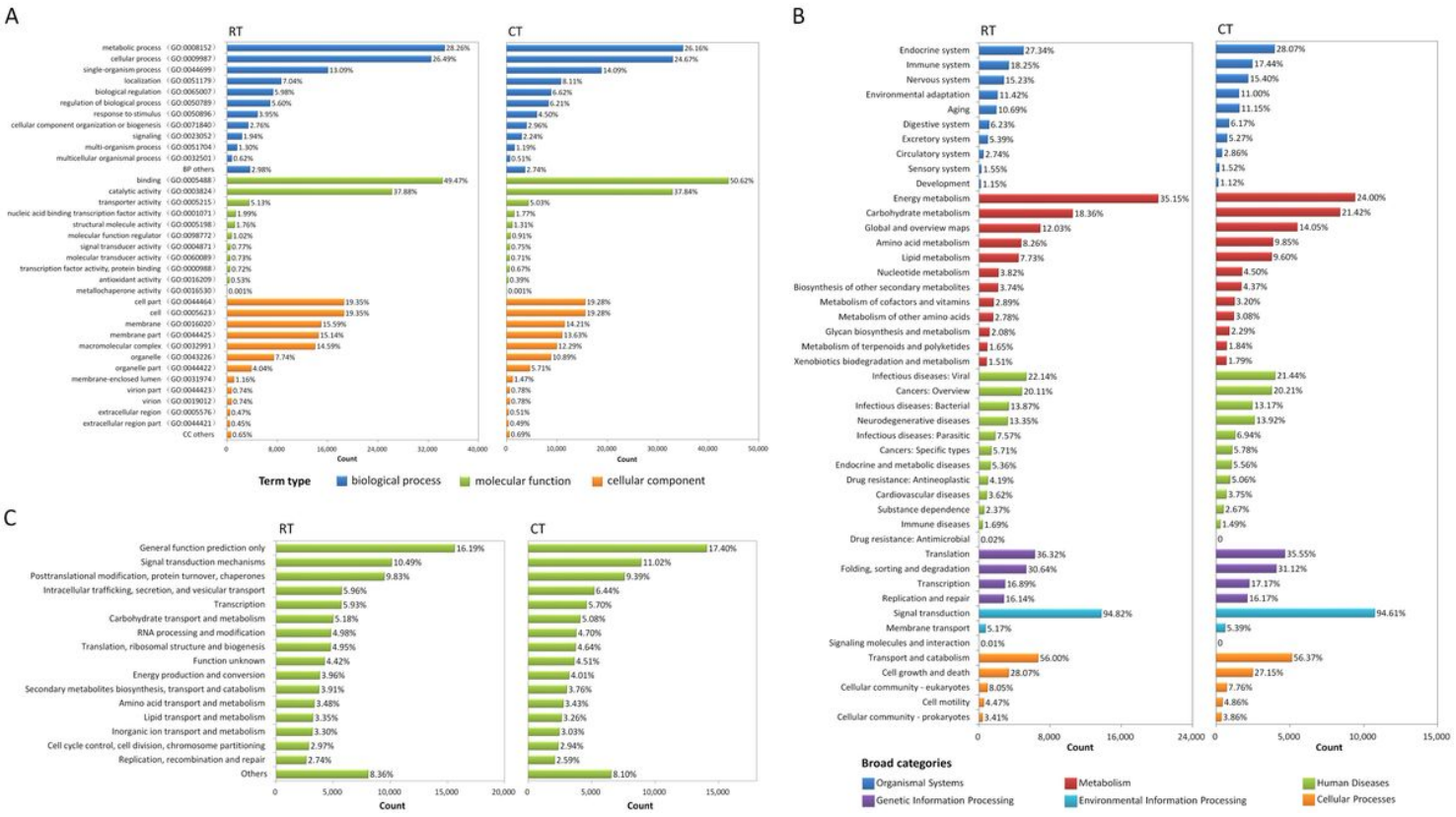

Figure 1

Function annotation and classification of P. giganteum assembled transcriptomes under RT and CT. (a) GO classification of the annotated unigenes in RT and CT samples. (b) KEGG classification of the annotated unigenes in RT and CT samples. (c) COG classification of the putative proteins in RT and CT samples.

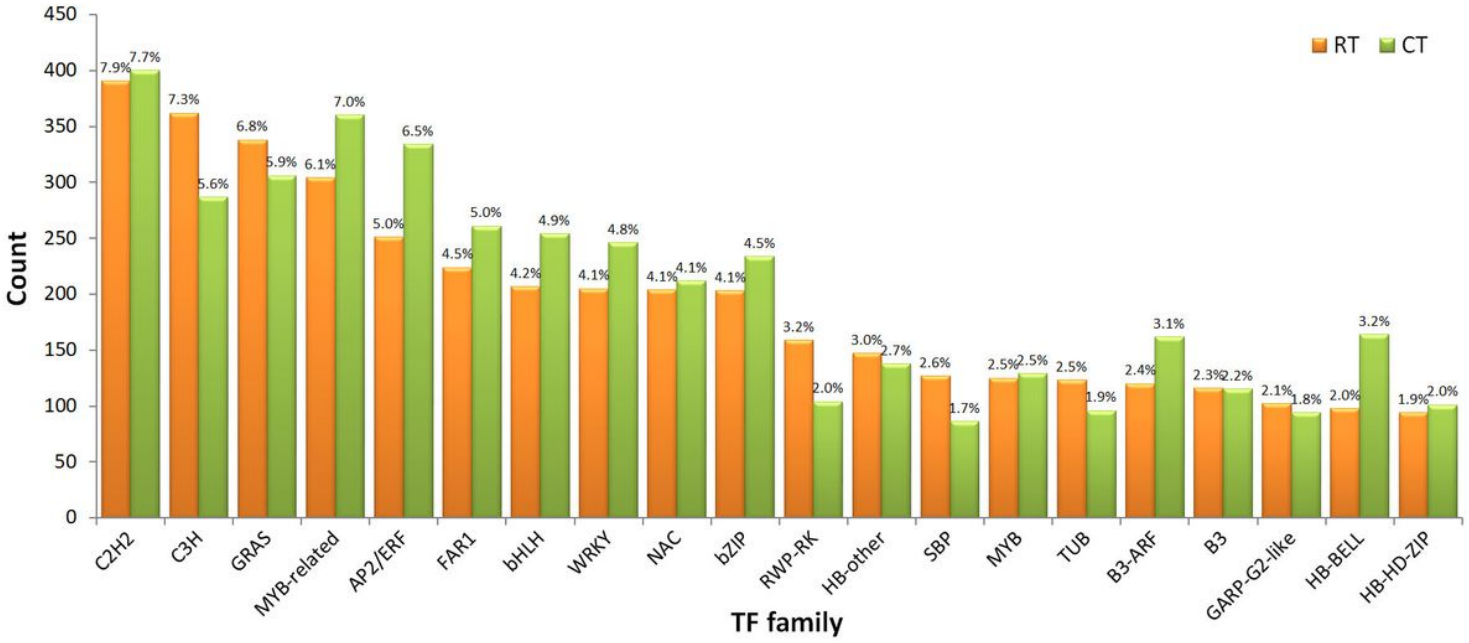

Figure 2

Putative TF gene families in P. giganteum assembled transcriptomes under RT and CT.

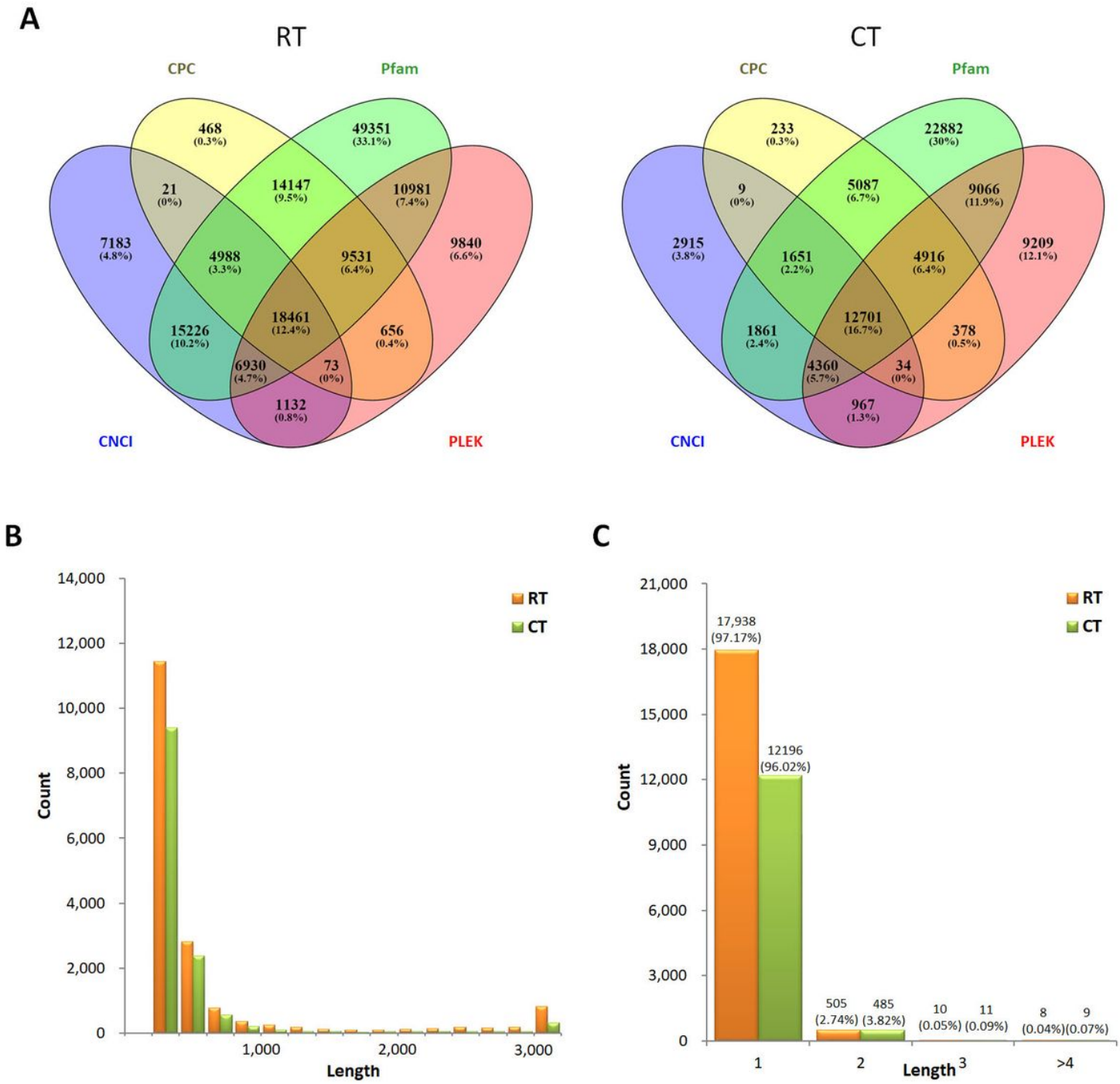

Figure 3

Identification of P. giganteum lncRNAs. (a) Venn diagram of the number of lncRNAs predicted by CNCI, CPC, Pfam and PLEK. (b) Length distribution of identified lncRNAs in RT and CT samples. (c) Distribution of isoform numbers for lncRNAs in RT and CT samples.

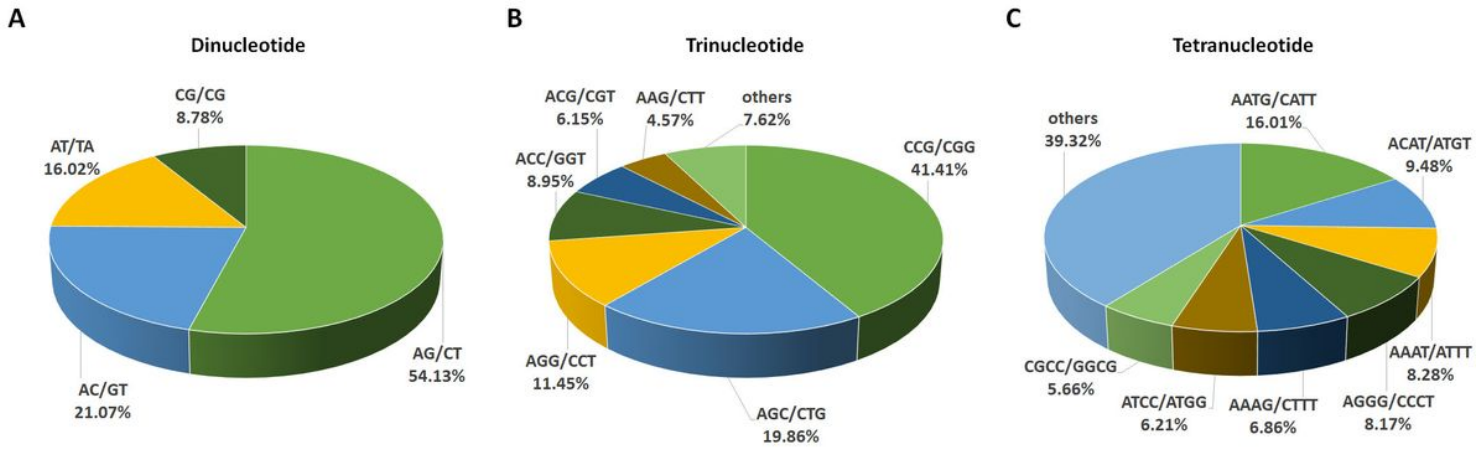

Figure 4

Percentages of different motifs among dinucleotide (a), trinucleotide (b) and tetranucleotide (c) repeats in P. giganteum RT and CT combined transcriptome.

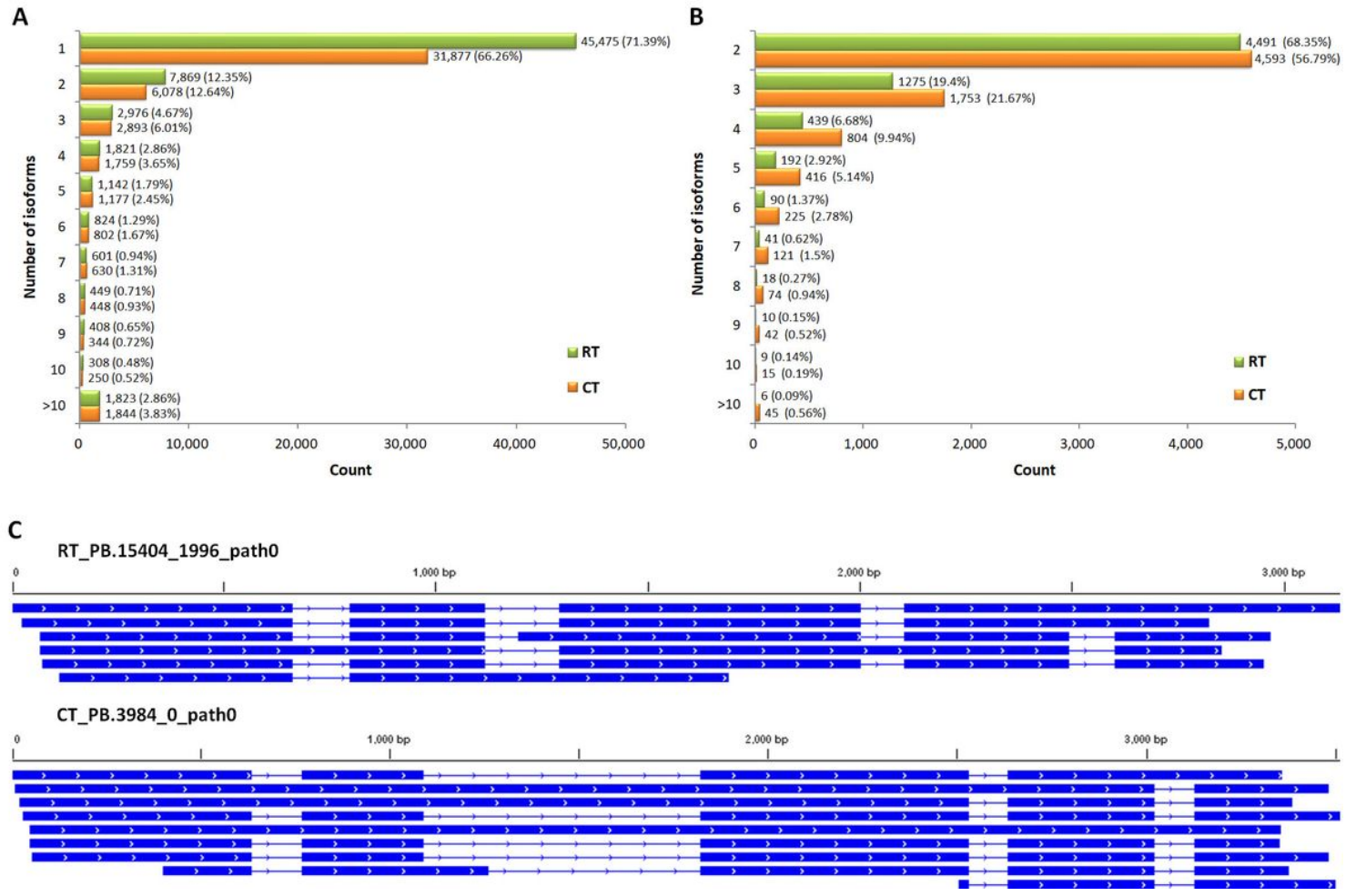

Figure 5

Isoform analysis of P. giganteum full-length transcriptomes using Iso-Seq. (a) Distribution of isoform numbers for UniTransModels in RT and CT samples. (b) Distribution of isoform numbers for UniTransModels

that have alternative splicing events in both samples. (c) Different splicing isoforms of the same UniTransModels in RT and CT samples. For each isoform, blocks in blue represent exons and lines in-between represent introns.

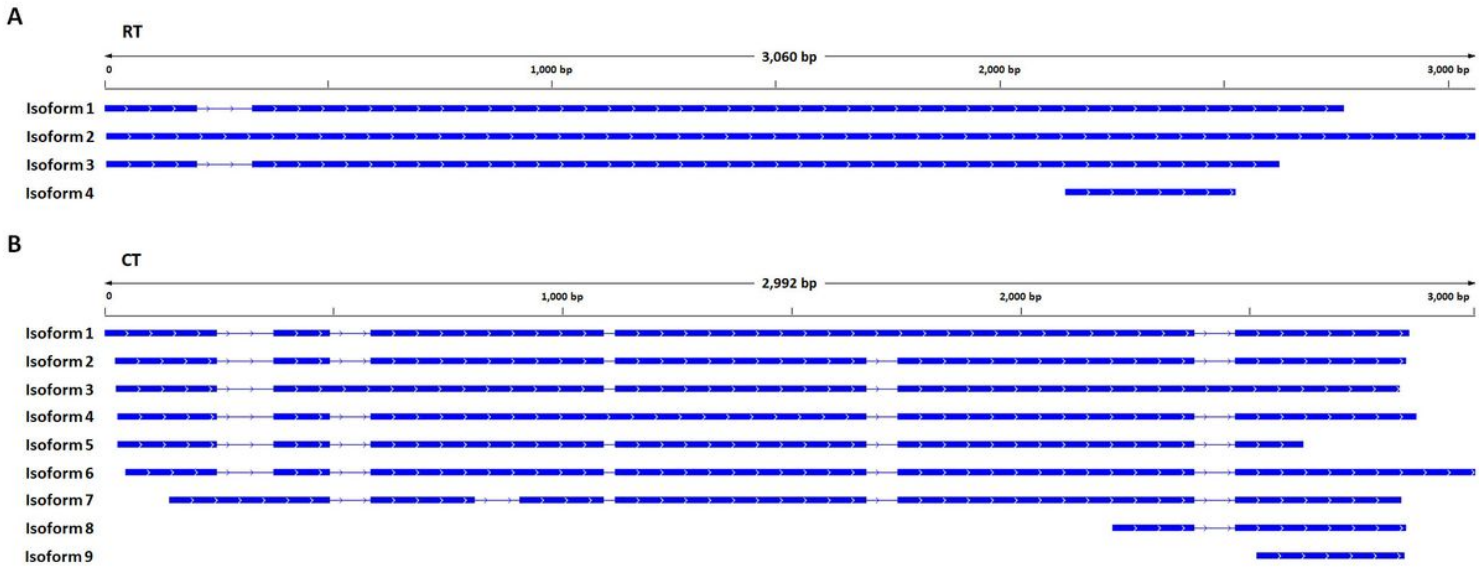

Figure 6

Different isoforms of the MFP2 gene in P. giganteum in RT (a) and CT (b) samples. For each isoform, blocks in blue represent exons and lines in- between represent introns.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.xlsx
- supplement2.xlsx
- supplement3.xlsx
- supplement4.jpg
- supplement5.xlsx
- supplement6.xlsx
- supplement7.xlsx