Combination of Single-Molecule Long-Read Sequencing and Illumina Sequencing Revealed the Mechanism of Anthocyanins Accumulation in an Ornamental Grass, Pennisetum Setaceum 'Rubrum'

Lingyun Liu (liulingyun01@163.com)
Beijing Forestry University https://orcid.org/0000-0003-0972-6003
Ke Teng
Beijing Academy of Agriculture and Forestry Sciences
Xifeng Fan
Beijing Academy of Agriculture and Forestry Sciences
Hui Zhang
Beijing Academy of Agriculture and Forestry Sciences
Chao Han
Beijing Academy of Agriculture and Forestry Sciences
Juying Wu
Beijing Academy of Agriculture and Forestry Sciences
Zhihui Chang
Beijing Academy of Agriculture and Forestry Sciences

Research Article

Keywords: Pennisetum setaceum 'Rubrum', Single-molecule long-read sequencing, Next generation sequencing, Alternative splicing events, Anthocyanin accumulation

Posted Date: October 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-922257/v1

Combination of single-molecule long-read sequencing and Illumina sequencing revealed the mechanism of anthocyanins accumulation in an ornamental grass, *Pennisetum setaceum* 'Rubrum'

Ling-yun Liu[1,2]*, Ke Teng[1], Xi-feng Fan[1], Hui Zhang[1], Chao Han[1], Ju-ying Wu[1]*, Zhi-hui Chang

1. Beijing Research and Development Center for Grass and Environment, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, P.R. China; tengke.123@163.com (K.T.); fanxifengcau@163.com (X.-F.F.); peakzeal@163.com (H.Z.); han_chao1988@126.com (C.H.); wujuying@grass-env.com (J.-Y.W.);

2. College of Grassland Science, Beijing Forestry University, Beijing 100083, China; liulingyun01@163.com (L.-Y.L.); changzh@bjfu.edu.cn (Z.-H.C.);

***Corresponding author:**

Ling-yun Liu (liulingyun01@163.com); Ju-ying Wu (wujuying@grass-env.com)

Abstract

Key message **Combination of single-molecule long-read sequencing and Illumina sequencing shed new light on the transcriptome and revealed the anthocyanins accumulation mechanism of *Pennisetum setaceum* 'Rubrum'.**

Abstract *Pennisetum setaceum* 'Rubrum' is an ornamental herb with purple leaves, and it is widely used in the construction of landscaping. However, the current next generation sequencing (NGS) transcriptome information is not satisfactory mainly because of the enormous difficulty in obtaining full-length transcripts. What's more, the molecular mechanisms of anthocyanin accumulation have not been thoroughly studied. In this study, we used PacBio full-length transcriptome sequencing combined with NGS sequencing technology to conduct transcriptome analysis on leaves showing different colors at different stages to clarify the molecular mechanism involved in the color change of *P. setaceum* 'Rubrum'. A total of 280,413 full-length non-chimeric reads (FLNC) sequences were obtained based on single-molecule long-read sequencing technology. We obtained 140,633 high quality (HQ) transcripts and 2,683 low quality (LQ) transcripts and identified 5,352 alternative splicing (AS). In addition, a total of 93,066 ORFs, including 57,457 full open links and 2,910 lncRNA sequences were screened out. Furthermore, a total of 10,795 differentially expressed genes were identified. Gene ontology (GO) cluster and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis revealed the underlying mechanism of anthocyanin accumulation. In this study, to our best knowledge, we provided the full-length transcriptome information of *P. setaceum* 'Rubrum' for the first time. The underlying mechanism of anthocyanin accumulation in *P. setaceum* 'Rubrum' was further discussed based on the newly generated transcriptome data. The information will not only facilitate the gene function studies but also pave the way for future breeding projects of *Pennisetum setaceum*.

Key words

Pennisetum setaceum 'Rubrum', Single-molecule long-read sequencing, Next generation sequencing, Alternative splicing events, Anthocyanin accumulation

Abbreviations

AS, Alternative splicing events; CCS, Circular Consensus; CDS, Putative coding sequences; DEGs, Different expressed genes; FLNC, Full-length non-chimeric reads; FPKM, Fragments Per Kilobase of transcript Per Million Fragments; GO, Gene ontology; HQ, High quality isoforms; ICE, Iterative isoform-clustering program; KEGG, Kyoto Encyclopedia of Genes and Genomes; LncRNA, long non-coding RNA; LQ, Low quality isoforms; NFL, Non-full-length; NGS, Next generation sequencing; ORF, Open Reading Frame; ROI, Reads of insert; ROI, Reads of insert; SMRT, The PacBio single-molecule long-read sequencing technology; TFs, Transcription factors

Introduction

Pennisetum setaceum is one of the most widely used ornamental grasses in the world which is originated in Africa (Beckwith et al., 2004;Zhu et al., 2020) and *P. setaceum* 'Rubrum' are cultivars that have been selected for their ornamental purple foliage. It is not only inherited the excellent resistance and unique form, but also has been further enhanced in ornamental color. However, studies on the *P. setaceum* focused on new variety breeding and physiological and ecological aspects, but ignored the supplement of transcriptome information(Goergen and Daehler, 2002;Ma Gonzalez-Rodriguez et al., 2010;Bella and D'Urso, 2012). Therefore, the theoretical basis is urgently needed to study *P. setaceum*. The mechanism of heat tolerance and rapid seed germination of pearl millet (*Pennisetum glaucum* (L.) R. Br.) has revealed by combining phenotype, physiology and transcriptome (Sun et al., 2021;Wu et al., 2021). As well as Zhu et al. (2020) performed NGS transcriptome sequencing of *P. setaceum* 'Rubrum'. However, comprehensive data needs to be available caused by the sequence reading defects of NGS.

The PacBio single-molecule long-read sequencing technology (SMRT) provides complete transcript sequence information while avoiding assembly errors. The readings are longer and more accurate (Workman et al., 2018;Chao et al., 2019). For species with no reference or bad reference genome, full-length transcriptome can obtain full-length sequence of transcripts and improve the information of genome annotation. It has been demonstrated in many plants, such as *Trifolium pratense* (Chao et al., 2018), *Allium sativum* (Chen et al., 2018) and *Cynodon dactylon* (Bing et al., 2018). Error correction is performed using NGS, and unique third-generation information such alternative splicing events (AS) can be found(Wang et al., 2017a;Wang et al., 2017c). Full-length transcriptome sequencing can be used to

explain various molecular mechanisms of plant growth and development, such as the senescence mechanism of *Glycine Max* seed (Fleming et al., 2018), the molecular mechanism of interaction between *Alternanthera philoxeroides* and host plants and ecosystems (Jia et al., 2018). Similarly, the application of this method to the molecular mechanism of anthocyanin accumulation will be more conducive to the amplification of genetic information of *P. setaceum*.

Anthocyanins are water-soluble pigments and chromatic substances that determine the color of plant leaves, flowers, and fruits and they are phenylalanine derived flavonoids (Yokozawa et al., 1998;Jian et al., 2019;Xu et al., 2020). They have a wide range of colors, from orange/red to purple/blue (Tanaka et al., 2008;He and Giusti, 2010). At present, 635 anthocyanins have been identified, covering fruits, vegetables, flowers and so on (Kyoungwon et al., 2016). Purple and dark colors caused by delphinidin, petunidin, and malvidin, while bright-red-colors based on cyanidin and pelargonidin mostly (Jaakola, 2013). Anthocyanins have strong light absorption ability, which can protect plants to a certain extent (Hoch et al., 2001). At the same time, the colors displayed give new vitality to plants and greatly improve the ornamental value of plants.

Among them, the main secondary metabolites related to the mechanism of color change are flavonoids (Wang et al., 2017b). Early steps in anthocyanin biosynthesis include the catalytic activity of chalcone synthase (CHS), chalcone isomerase (CHI), flavonoid 3-hydroxylase (F3H), and flavonoid 3’-hydroxylase. Later stages of the anthocyanin synthesis pathway include dihydroflavonol 4-reductase, anthocyanidin synthase (ANS), and anthocyanidin 3-glycosyltransferase. Three most widely studied transcription factors in anthocyanin synthesis in plants are MYB, bHLH and WD40 (Lepiniec et al., 2015;Lai et al., 2016;Shi et al., 2020). In recent years, although

several researches have been conducted on the regulation of plant color by anthocyanins and the improvement of anthocyanin content in fruits and vegetables, insights into the regulation of leaf color of grass species have remained future objectives (Harborne and Self, 1987;Fossen and Andersen, 1998;Fossen et al., 2001). In fruits, 158 flavonoids have been identified as an important factor of 'Tailihong' jujube fruit skins at different developmental stages, among which cyanidin-3-O-rutinoside and peonidin-3,5-O-diglucoside were the primary anthocyanins (Shi et al., 2020). In vegetables, anthocyanidin 3-O-glucosyltransferase effected anthocyanin biosynthesis of purple potato (*Solanum tuberosum*) at the transcriptional level (Hu et al., 2011). In flowers, blue-colored gentian plants flowers changed to a new cultivar named 'Polarno White' due to mutation of *GTMYB3* (Nakatsuka et al., 2008). In grasses, anthocyanins have been identified in *Festuca rubra* and *Panicum melinis*, 11 anthocyanins were identified by 23 Poaceae grass species and anthocyanins acylated with one or two malonic acid moieties dominated the anthocyanin profiles of all the species in the subfamilies Pooideae and Panicoideae (Fossen et al., 2002).

In this study, we aim at obtaining full-length transcriptome sequencing information of *P. setaceum* 'Rubrum' at first time, meanwhile, predicting AS candidate events and long non-coding RNA (lncRNA). In addition, we focused on differential expression of anthocyanins in green, lilac and purple *P. setaceum* 'Rubrum' using NGS sequencing technology. The results provided new insights into the full-length transcriptome information of *P. setaceum* 'Rubrum' to explain the molecular mechanism of anthocyanin accumulation, and provide a basis for improving its horticultural and ornamental quality.

Materials and Methods

Plant Materials and RNA samples preparation

In 21 July 2020, we collected green, lilac and purple three kinds of leaves of different branches on a single plant as experimental materials in the breeding base of Beijing Academy of Agriculture and Forestry Sciences, China. (Figure. 1). We also collected roots, flowers, stalks as materials of total RNA libraries. After quick-freezing in the liquid nitrogen, the collected leaves were stored at -80 °C for subsequent experiments. Using a plant RNA kit (OMEGA, Georgia, USA, No. R6827–01) to extract total RNA. RNA samples quantity and integrity were tested in Drop ND-1000 spectrophotometer (NanoDrop Technologies, Delaware, USA) and 2100 Bioanalyzer (Agilent Technologies, California, USA) (Zhang et al., 2016), respectively.

NGST and SMRT sequencing

Nine RNA-seq libraries (three different colors of green, lilac and purple × 3 replicates) were sequenced and constructed, and the quality of the prepared RNA samples was evaluated. After the quality control, eukaryotic mRNA was enriched with magnetic beads with Oligo (dT) and mRNA was randomly broken by Fragmentation Buffer. The first cDNA strand was synthesized with random hexamers using mRNA as template. The second cDNA strand was synthesized by adding Buffer, dNTPs, RNase H and DNA polymerase I. The purified double-stranded cDNA ends were repaired, A tail was added and sequencers were connected. Then AMPure XP beads were used to select segment sizes. Finally, cDNA libraries were obtained by PCR enrichment. NGS (Next generation sequencing) was performed at Illumina HiSeq4000 (San Diego, CA, USA). For SMRT sequencing, total RNA (equally mixed with RNAs of roots, flowers, stalks and three different color leaves) were used to construct cDNA library based on the SMARTer ™ PCR cDNA short Kit. The cDNA library was built with the size of 1-6kb, the full-length cDNA was amplified by PCR, and then the full-length cDNA

was repaired at the end and connected to the SMRT dumbbell connector. Finally, exonuclease digestion was performed to obtain the sequencing library. The library building was generated at Biomarker Technology Co. (Biomarker, Beijing, China).

Quality filtering and error correction

The CCS sequences were extracted from the original sequence according to the condition full passes> =3 and the sequence accuracy is greater than 0.9. Among them, the sequence containing the correct 5' primers, the 3' primers and the poly(A) tail were full-length sequences, and the rest were non-full-length sequences. The insertion sequences of CCS were obtained by removing cDNA primer sequences and poly(A) sequences. The sequences were divided into full-length sequences and non-full-length sequences, chimeric sequences and non-chimeric sequences according to the differences of primers at both ends of CCS sequences. Used the IsoSeq module in SMRTLink software to cluster the similar sequences (i.e., multiple copies of the same transcript) of the FLNC into a cluster, and each cluster will get a Consensus isoform. The consistent sequences in each cluster were further corrected to obtain HQ with accuracy greater than 99% and LQ, respectively. CD-HIT software (Godzik, 2006) was used to combine the sequences with high similarity and remove the redundant sequences in the transcript.

Alternative splicing and prediction of CDS and LncRNA

The software BLAST (Altschul et al., 1997)was used to perform pairwise alignment of all de-redundancy transcriptome sequences for the prediction of AS candidate events. Briefly, there are three kinds of identification methods. The length of both sequences is greater than 1000bp and have two high-scoring Segment (HSPs), variable splicing Gap are greater than 100bp in comparison and at least 100bp

distance from the 3' and 5' end, allowed 5bp overlap of all variable transcripts. Using TransDecoder software to predict the Coding Sequence (CDS) of transcriptional Sequence and its corresponding amino acid Sequence based on the Open Reading Frame (ORF), log-likelihood Score, alignment between amino acid sequence and protein domain sequence in Pfam database. lncRNA is another significant component of the transcriptome. Four methods including CPC analysis, CNCI analysis, Pfam protein domain analysis and CPAT analysis were used to screen the coding potential of the transcripts.

Differential Gene Expression Analysis and Prediction of Gene Function

BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) software (version 2.2.26) was used to compare the obtained non-redundant transcripts with GO (www.geneontology.org) and KEGG (Minoru et al., 2004) databases to obtain annotated information of the transcripts. The Fragments Per Kilobase of Transcript Per Million Fragments (FPKM) method was used to detect whether the number of mapped Reads and the length of transcripts in the sample had reached the gene expression level (Dewey and Bo, 2011). Deseq2 (Robinson et al., 2010) was used for differential expression analysis between sample groups to obtain the transcripts of differential expression between the two biological conditions. In this study, the differences between green leaf and red leaf, green leaf and semi-red leaf, semi-red leaf and whole red leaf were analyzed, and the functional annotation of the differentially expressed transcripts was conducted in the database. The identified differentially expressed genes (DEGs)were mapped to each term of KEGG metabolic pathway, and the optimal comparison results were selected to determine the functions of the genes in the metabolic network and signaling pathway.

cDNA synthesis and quantitative reverse transcription (qRT-PCR) analysis

The cDNA reverse transcription kit was used to reverse transcribe the extracted total RNA into cDNA, and the product was purified by cycle-pure kit (OMEGA, Georgia, USA, No. D6492-01). The up-regulated and down-regulated genes of differentially expressed genes were randomly selected, and primers were designed by Primer Premier 5. qRT-PCR was run in 7500 rapid quantitative PCR system (Applied Biosystems) using SYBR Green PCR Master mix (TAKARA). Actin was used as the reference gene, and the expression was normalized by $2^{-\Delta\Delta CT}$ method. Three biological replicates were used for all gene expression analysis, and the primers used for gene expression analysis were listed in Table S2.

Photosynthetic Data Assays

According to the instructions of CIRAS-3 (Hasha Scientific Instruments Limitied), a portable photosynthetic apparatus, to determine the photosynthetic rate of *P. setaceum* 'Rubrum' cultivated in the open air. The time was fixed at 10 a.m. and the penultimate or penultimate fourth fully expanded leaf of a single plant was selected for measurement. The rectangle (18 mm × 25 mm) was set as the size of the leaf chamber window, the light source was LED, red light, the light intensity was 100 $\mu mol \cdot m^{-2} \cdot s^{-1}$, and the reference air humidity was 80%~100%. The concentration of CO_2 was the same as the concentration of it in the air. Net photosynthetic rate (Pn, $\mu mol \cdot m^{-2} \cdot s^{-1}$), intercellular CO_2 concentration (Ci, $\mu mol \cdot mol^{-1}$), stomatal conductance (Gs, $mmol^{-2} \cdot s^{-1}$), transpiration rate (Tr, $mmol^{-2} \cdot s^{-1}$) were measured at the same time. The determination of processing repeated 10 times, average was calculated as the final result.

Anthocyanin Measurements

Clean the leaves and set 3 replicates for each color. Cut off the vein part of the leaf, weigh 0.1g leaf tissue, cut it into pieces and grind it. The powdered sample were placed in a 50ml centrifuge tube, then added 10ml 1% HCI-methanol solution. The mixture was placed in incubator at a constant temperature for 5h, and shocked it every 1h. With 1% HCI-methanol solution as the control, the absorbance values at 530nm and 657nm were measured by ultraviolet spectrophotometer (TU-1810). The measurement was repeated for 5 times, and the average value of each 2 absorbance values at the same wavelength was used for subsequent calculation. The formula of anthocyanin content was $\Delta A = (A530-0.25A657)/g$.

Chlorophyll Content Measurements

Chlorophyll content was measured following the method of Teng et al. (2016) (Teng et al., 2016). In brief, weighed 0.05~0.08g of fresh leaf tissues, recorded the mass, and then cut them into pieces and placed them in a 10ml centrifuge tube. Add 8ml 95% ethanol and let it stand for 24h to 48h in the dark. Using ultraviolet spectrophotometer (TU-1810) to measure the absorbance values at 665nm, 649nm and 470nm based on 95% ethanol as a contrast. The calculation formula of chlorophyll a and b is C_a (mg/L) = $13.95A_{665}$-$6.88A_{649}$, C_b (mg/L) = $24.96A_{649}$-$7.32A_{665}$, Total chlorophyll concentration C_t (mg/L) = C_a+C_b, Carotenoid concentration C_{x+c}=(1000*A470-2.05*Ca-114.8*Cb)/245, Photosynthetic pigment concentration $C_{pho} = C_a+C_b+C_{x+c}$ (LICHTENTHALER et al., 1983). The sum of the chlorophyll a and chlorophyll b is the concentration of total chlorophyll. Finally, the content of chlorophyll in plant tissues can be further calculated according to the following formula, Chlorophyll content (mg/g) = [concentration of chlorophyll × volume of extraction liquid]/ fresh weight of the sample.

Flavonoid Content determination

Anthocyanin composition in green, lilac and purple leaves were collected based on UPLC-MS/MS following the method described in Jia et al. (2021) (Jia et al., 2021a;Jia et al., 2021b). In brief, green, lavender, and purple leaves are ground into a fine powder with liquid nitrogen. 0.1g of each sample was weighed and added into 4mL of 70% methanol aqueous solution containing 0.1% formic acid. The samples were vorced for 1 min and subjected to ultrasonic treatment at 40 Hz 20°C for 30 min. After centrifugation, the supernatant was analyzed by a 30-A UPLC high performance liquid chromatography-mass spectrometry (Shimadzu Corporation, Japan). The mobile phases A and B were 1% formic acid solution containing 5% methanol, 1% formic acid - methanol solution, respectively, and the elution was carried out according to the properties of anthocyanin monomers. The determination indexes included Delphinidin, Delphindin 3-O-B-D-glucoside, Delphinidin-3.5-diglucoside, Cyanin, Cyanidin, Pelargonidin, Pelargonin, Peonidin, Peonidin 3-O-glcucoside, Malvidin, Malvin, Malvidin 3-galactoside, Kuromanin, Callistephin, Petunidin, Petunidin 3-O-B-D-chloride, Delphindin 3-O-rutinoside, Peonidin-3.5-di-O-glucoside.

Results

PacBio Analysis of of *Pennisetum setaceum* 'Rubrum'

In this study, leaves of different colors (green leaf, lilac leaf and purple leaf) were collected (Figure 1). In order to obtain comprehensive reads, a total of six samples of roots, stems, flowers and leaves of three different colors were selected for mixed samples. PacBio long read sequencing was performed. The cDNA library size was 1-6kb and 316,274 circular consensus (CCS) read was obtained by full-length transcriptome sequencing (Table 1; Figure 2a). In order to improve the utilization of sequencing data, full passes were selected based on the condition that the total pass

number is ≥3 to generate reads of inserts (ROIs) consensus sequences. The higher the full passes, the higher the sequence accuracy. The distribution of full passes of each cDNA database was shown in Figure 2b. ROIs consensus sequences were divided into FLNC and NFL by detecting whether the CCS sequences contained the correct 5' primers, 3' primers and poly(A) tails. FLNC read length distribution After removing cDNA primers from CCS sequences and insertion sequences from poly(A) sequences, a total of 280,413 FLNC sequences were obtained, accounting for 88.66% of the total CCS sequences (Figure 2c). The similar sequences in the FLNC sequences were clustered into a cluster, and a consensus isoform was obtained for each cluster depended on the IsoSeq module in SMRTLink software. Consensus isoform sequence length varies due to different cDNA lengths (Figure. 2d). A total of 143,505 consensus isoform were obtained. By further correcting the consensus isoform in each cluster, 140,633 HQ transcripts and 2,683 LQ transcripts were obtained, respectively.

Full-length transcriptome sequencing reads were longer, but the sequence accuracy was lower than NGS short reads. In order to improve the accuracy of the data, the NGS based on nine samples of three replicates of green leaves, lilac leaves and purple leaves were sequenced to correct the LQ consensus isoform. We aligned the NGS of short reads to the full-length transcriptome sequence to be corrected, and replaced the bases in the full-length position with the bases with the most counts in each position by alignment. CD-HIT software was used to combine the sequences with high similarity and remove the redundant sequences in the transcript. Finally, 97,450 non-redundant transcripts were obtained.

Alternative splicing analysis and prediction of CDS and LncRNA

The de-redundant transcripts were compared with BLAST. A total of 5,352 AS events was found (Table S1). TransDecoder software is used to identify reliable potential Coding sequences (CDS) from transcription sequences based on ORF length, log-likelihood Score, alignment between amino acid Sequence and protein domain Sequence in Pfam database. A total of 93,066 ORFs were obtained, including 57,457 complete ORFs. The length distribution (Figure S1) showed that the largest number is between 100-200 bases, accounting for 20.94%. Those whose length is less than 1,000 bases account for 97.55% of the total.

Next, four methods including cpc analysis, cnci analysis, Pfam protein domain analysis and CPAT analysis were used to screen the coding potential of the transcripts, and 2,910 lncRNA sequences were screened (Figure S2). The target genes of lncRNA were predicted by complementary pairing with mRNA bases. Among them, 905 lncRNAs successfully predicted target genes (Table S2).

Transcription factor prediction and transcription function annotation

In order to find potential proteins that can bind to the upstream of the gene to regulate gene expression, iTAK software was used to predict the transcription factors in the transcripts of *P. setaceum* 'Rubrum'. The results showed that the most common type of transcription factor was RLK-PELLE_DLSY, and BHLH and MYB, which were related to anthocyanin synthesis, were also included (Figure S3).

Transcriptome expression quantification and Gene function annotation

With full-length transcriptome sequencing as reference, the second-generation data were compared and quantified using RSEM software. The number of mapped Reads and the length of transcripts were normalized using Fragments Per Kilobase of

transcript Per Million Fragments (FPKM) as an indicator to measure the expression level (Figure S4-a, b). The overall distribution of 9 sample transcripts showed that the whole transcriptome expression level was high and DEGs analysis can be performed. The Pearson Correlation Coefficient (r) of the expression correlation of the repeated samples was between 0.884-1, indicating that the study had good reproducibility Figure S4C).

By searching NR, Swissprot, GO, COG, KOG, Pfam, and KEGG databases, 97,450 non-redundant transcripts were functionally annotated, with a total of 89,758 transcripts (92.11%) annotated (Table 2). GO database is an international standardized Gene functional classification system, which provides a set of dynamically updated standard vocabulary to comprehensively describe the functional properties of genes and Gene products in an organism. The annotated 73,264 genes were classified using the GO database, which were later classified into three functional categories: biological processes, cellular components, and molecular functions. In the category of cell components, these sequences are further divided into 15 categories, of which the most representative subcategory is cell and cell part. In the molecular function category, these unigenes are divided into 14 categories, among which the largest subcategory is binding, and the second is catalytic activity. Among the 21 biological processes, metabolic process, cellular process, single-organism process, and biological regulation are the most important components (Figure 3). The results of NR alignment showed that 64,016 unigenes had the highest homology with millet, accounting for 71.61% of the total. *Setaria italic*， *Zea mays* and *Sorghum bicolor* accounted for 9.16%, 4.43% and 3.99% of the total, respectively (Figure 4).

DEGs Analysis

Transcriptome sequencing of 9 samples was completed, and Q30 reached 85%, indicating that the reads have high quality and could be used to analysis DEGs. A total of 10,795 non-redundant DEGs were detected by transcriptome sequencing. 2,562 DEGs were up-regulated and 3,931 DEGs were down-regulated in green leaf vs. purple leaf comparison group. 1,223 differentially expressed genes were up-regulated and 2,014 DEGs were down-regulated in lilac vs purple leaves. 570 DEGs were up-regulated and 505 DEGs were down-regulated in the comparison group of green leaf vs lilac leaf (Figure 5a). Functional annotation and enrichment analysis of differentially expressed transcripts obtained the following results. Among the results, 6,242 transcripts were annotated in the comparison group between green leaf and purple leaf. The difference between green leaf and lilac leaf was the least annotated transcript which was 1,039. The KEGG pathway enrichment analysis of DEGs in green leaves and purple leaves indicated that a total of 1,762 genes were enriched in the Carbon metabolism pathway, as well as in multiple metabolic pathways such as Phenylalanine metabolic pathway, flavonoid metabolic pathway, and branching-antenna proteins (Figure 5- b, c, d).

Leaf color-related metabolic pathways in *Pennisetum setaceum* 'Rubrum'.

The *Pennisetum setaceum* 'Rubrum' leaves are green when they are young, lilac when they are in the middle of development and purple when they are mature. Therefore, we used leaf transcriptome data at different developmental stages to explore changes in anthocyanin biosynthesis. We examined structural genes for anthocyanin biosynthesis pathways in other species (Figure 6).

The 89,450 single gene clusters were annotated in the NR database, and the enzymes related to flavonoid biosynthesis were further searched. A total of 39 flavonoid related

information was obtained, among which CHS, F3H, F3'H, ANR, ANS were found to have a large number of gene families, while the DFR had two, one each for CHI and FLS. Further the 39 flavonoids single gene cluster in *P. setaceum* 'Rubrum' green leaves, lilac leaves and purple leaf (the three repeat), a total of nine samples of the relative expression quantity heat analysis (Figure 6). The expression levels of flavonoid synthetase genes involved in *P. setaceum* 'Rubrum' were more similar in lilac and purple leaves, but significantly different from those in green leaves. In green leaves, the expression levels of Mix_transcript_111598, Mix_transcript_28073, Mix_transcript_49688, Mix_transcript_21896 were significantly lower than those in lilac and purple leaves. The expression of Mix_transcript_8959 in purple leaves was 61.45 times higher than that in green leaves. The expression levels of Mix_transcript_31964, Mix_transcript_34079 and other genes in green leaves were significantly higher than those in lilac and purple leaves. There was no significant difference in gene expression between the lilac leaf and the purple leaf. In the Mix_transcript_32044, Mix_transcript_21020, and Mix_transcript_21896 gene clusters, the lilac leaf had higher gene expression than the purple leaf. At the same time, the expression levels of some repeat samples of single gene clusters were significantly different. Therefore, when collecting relevant data for reference, 2 of the 3 repeat samples in the same part should be selected as much as possible for reference, so as to be closer to the real value.

Also, the analysis of KEGG metabolic pathway omics showed that cyanoidin-3-O-rutin and paeoniflorin-3, 5-O-diglucoside were the main anthocyanins in *P. setaceum* 'Rubrum'. Cyanidin-3-O-rutinoside is a product of glycation modification, while peonidin-3, 5-O-diglucoside is a product of methylation.

Photosynthetic Data Analysis

According to the determination of anthocyanin content in the leaves of different colors (Table 2), the purple leaves had the highest content, followed by lilac leaves, and the green leaves had the lowest content. The results showed that the darker the leaf color, the deeper the anthocyanin content, and the less the anthocyanin content of green leaves (Figure 7). The results of chlorophyll content were consistent with that of anthocyanin content, but the total content of chlorophyll was higher than that of anthocyanin content. At the same time, we measured the net photosynthetic rate, in which purple leaves had the highest net photosynthetic rate, followed by lilac leaves and green leaves.

Flavonoid Content Expression

Through the determination of flavonoids, we detected the contents of 18 substances in total (Table S3). The substances tested include Delphinidin, Delphindin 3-O-B-D-glucoside, Delphinidin-3.5-diglucoside, Cyanin, Cyanidin, Pelargonidin, Pelargonin, Peonidin, Peonidin 3-O-glcucoside, Malvidin, Malvin, Malvidin 3-g alactoside, Kuromanin, Callistephin, Petunidin, Petunidin 3-O-B-D-chloride， Delphindin 3-O-rutinoside, Peonidin-3.5-di-O-glucoside. Among them, Delphinidin-3.5-diglucoside, Pelargonidin, Pelargonin, Malvin were not detected in all the samples. Chloride Delphinidin, Callistephin, Petunidin, Petunidin 3-O-B-D-chloride were detected in purple and lilac leaves but not in green leaves. Malvidin was highly expressed in all leaves. K-means clustering (Figure 8) was performed on all the indicators, and it could be seen that the four repeated samples of purple leaves were clustered into one category, and the repeated samples of lilac leaves and green leaves were clustered into one category.

448 **qRT-PCR**

449 In order to verify the reliability of the transcriptome data of *P. setaceum* 'Rubrum', 5
450 up-regulated genes and 5 down-regulated genes were randomly selected for qRT-PCR
451 analysis from the anthocyanin biosynthesis pathway (Table S4), and their expression
452 levels were consistent between qRT-PCR and RNA-seq data, verifying the reliability
453 of the transcriptome data (Figure 9). The results showed that the expression trends of
454 the 10 genes involved in flavonoid biosynthesis in green leaves, lilac leaves and
455 purple leaves were consistent with the results of FPKM value of the transcriptome.

456

457 **Dissusions**

458 *P. setaceum* 'Rubrum' is a cultivated grass with an elegant shape and light and
459 delicate panicles. It gives people a unique sense of beauty with its natural and simple
460 appearance, and attracts people's attention to landscape design with its beautiful leaf
461 color. However, the current studies on the transcriptome of *P. setaceum* species are
462 mainly based on the analysis of gene expression by NGS technology, and lack of
463 understanding and research on the full-length DNA, so it is difficult to explain the
464 anthocyanins accumulation mechanism. Few studies on the molecular mechanism of
465 anthocyanins in *P. setaceum* 'Rubrum' , and only the second-generation data were
466 used (Zhu et al., 2020). However, in this study, the single-molecule long-read
467 sequencing were combined with the Illumina sequencing data to increase the data
468 volume. PacBio SMRT provides complete transcript sequence information while
469 avoiding assembly errors. The readings are longer and more accurate (Workman et al.,
470 2018;Chao et al., 2019). For species with no reference or bad reference genome, full-
471 length transcriptome can obtain full-length sequence of transcripts and improve the
472 information of genome annotation. It has been demonstrated in many plants, such as
473 *Trifolium pratense* (Chao et al., 2018), *Allium sativum* (Chen et al., 2018) and

Cynodon dactylon (Bing et al., 2018). Both of them have revealed the mechanism of unique plant traits at the transcriptome level. The primary purpose of this study was to investigate the mechanism of anthocyanin accumulation and biosynthesis related genes in *P. setaceum*. In this study, we analyzed the differences of the full-length transcriptome and the combined transcriptome of the second generation in the leaves of *P. setaceum* at different growth stages.

Full-length transcriptome sequencing indicated that there were 316,274 CCS sequences, and 280,413 FLNC sequences. The full-length non-chimeric sequences were clustered to obtain 143,505 consistent sequences, and the consistent sequences were polished to obtain 140,633 high-quality consistent sequences. The combined high-quality consistent sequences were de-redundancy analyzed to obtain 97,450 transcript sequences. In a variety of biological reactions, AS acts as an effector mechanism to increase the complexity and flexibility of the entire transcriptome and proteome. Due to the advantages of long-read sequences, SMRT sequences can accurately identify the complexity of AS at the genome-wide level. The discovery of alternative splicing and translation is part of the abscisic acid reaction in *Arabidopsis seedlings* (Zhu et al., 2017). In *Pennisetum giganteum* , AS were found in both samples treated at low temperature and samples at room temperature, but the proportion of two or more subtypes of Unitransmodels was higher than that of room temperature samples, suggesting that alternative splicing events induced by low temperature stress may exist in *Pennisetum giganteum* (Li et al., 2020b). In this study, 5,352 transcripts were identified as natural splicing altered, accounting for 5.49% of the redundant transcripts. It suggested that the AS event of color change may exist in *P. setaceum* 'Rubrum'. However, its frequency was lower than that of *Carex* (Teng K et al., 2019) and *Arabidopsis* (Zhu et al., 2017). lncRNAs, a recently identified class

of non-coding RNAs function as essential regulators in a wide range of biological processes. A total of 2,910 lncRNAs were predicted for the de-redundant transcripts, and 89,758 sequences were annotated for functional annotation of the de-redundant transcripts.

Flavonoids are important compounds in plant secondary metabolites. The flavonoid constituents and their anabolic enzyme genes have always been the focus of research. In terms of the research on flavonoid related components, it has been reported that there are 4,000 kinds of flavonoids in plants (Yonekura-Sakakibara et al., 2019). In this study, the content of anthocyanins in the leaves of different colors of *P. setaceum* was determined. The results showed that the content of anthocyanins in purple leaves was the highest, followed by lilac leaves and green leaves. At the same time, the content of flavonoid compounds and the expression of enzymes involved in flavonoid synthesis were analyzed by K-means clustering. The relative expression of enzymes involved in flavonoid synthesis in *P. setaceum* was closer to the total content of lilac leaves and purple leaves than that in green leaves. This may be related to the intensity of light received. Studies have shown that anthocyanin pigments in plant tissues can play a role in reducing oxidative stress response under light exposure (Beckwith et al., 2004). Beckwith et al. have determined the content of two anthocyanins in *C. violacea*. The content of cyanidin 3-rutinoside is the highest under low light, while the content of cyanidin 3-glucoside is higher under strong light (Beckwith et al., 2004). Cyanidin-3-O-rutin and paeoniflorin-3, 5-O-diglycoside were the main anthocyanins in the pericarp of 'Taili' jujube under open air (Shi et al., 2020). In this study, Malvidin was highly expressed in all leaves. The detection content of Cyanidin related substances increases gradually with the deepening of leaf color. At this point, anthocyanins provide protection to the photosynthetic organs by absorbing excess

radiation and providing a “barrier” (Steyn et al., 2002). The protective effect is most effective when anthocyanins are located in the epidermis (Gould et al., 2000). This could explain the increase in anthocyanin content in purple leaves. The variation trend of chlorophyll content was consistent with that of anthocyanin. The highest content was found in purple leaves and the lowest in green leaves. Typically, the process of anthocyanin elevation is accompanied by the degradation of chlorophyll. Tian et al. founded that the yellow of carotenoids and the red of anthocyanins overlay each other, combined with the reduction of chlorophyll, making the leaves turn golden (Tian et al., 2021). We think this phenomenon is related to the position of the leaf. Most of the green leaves were in the tender development stage, and the chloroplasts were few and the structure was not perfect. Purple leaves, on the other hand, were in their mature stages of leaf development and grow on the periphery of the plant, where they are exposed to more sunlight. Photosynthesis was enhanced, and the contents of chlorophyll and anthocyanin were increased accordingly.

During normal growth and maturation, the leaf color of ‘Rubrum’ changed from green to purple, which was associated with significantly upregulated anthocyanin biosynthesis genes. KEGG database was used to analyze the metabolic pathway of gene products and the enzyme genes involved in flavonoid metabolism pathway, and elucidate the complex mechanism of gene action. The transcriptional levels of DFR, ANS and UFGT in fruits of various plants are closely related to the content of anthocyanins, such as *Ficus carica* (Li et al., 2020a), *Brassica rapa* (Zhuang et al., 2019) and *Solanum tuberosum* (Hu et al., 2011). KEGG enrichment analysis in this study showed that CHS, F3H, F3’H, ANR and ANS were involved in the flavonoid metabolism pathway, and calorimetric analysis of the expression of related genes showed that both up-regulated and down-regulated genes existed in this process.

Physiological experiments showed that compared with green leaves, the content of chlorophyll and anthocyanin in purple leaves increased, and the photosynthetic efficiency was enhanced. The results confirmed that the up-regulated DEGs were mainly enriched in the pathways related to anthocyanin synthesis, including Flavonoid biosynthesis, Phenylpropanoid biosynthesis, Glutathione metabolism, indicating an increased anthocyanin content, which was consistent with the physiological experimental results of flavonoid content and anthocyanin content reported previously (Zhu et al., 2019). In addition, KEGG pathways related to Photosynthesis, including Carbon metabolism and Carbon fixation in photosynthetic organisms, also contain up-regulated DEGs. Similar conclusions were reported in studies on flower and pericarp color. For example, Sun et al. found that the expression pattern of *FhCHS1* in *Freesia hybrida* flowers was significantly related to the accumulation patterns of anthocyanins during flower development. Furthermore, the function of FhCHS1 was verified by ectopic expression in *Petunia hybrida*, and it was found that CHS1 was responsible for the flower color transition from the original white to pink (Wei et al., 2015). Christopher et al. reported higher anthocyanin content in orange than in yellow *Clivia miniata*, and observed the same trend for the expression of CHS and DFR (Viljoen et al., 2013). Therefore, we believe that these genes play an important role in the purple leaf phenotype of *P. setaceum* 'Rubrum'.

Conclusion

This research first provided a full-length transcriptome of *P. setaceum* 'Rubrum' using the SMRT sequencing method, meanwhile, obtained 140,633 HQ transcripts and identified 5,352 AS events. In addition, 2,910 lncRNA sequences were screened out. The key genes in anthocyanin accumulation were obtained by NGS. These results

will expand studies on gene level and provided basis for breeding of new varieties of *P. setaceum* 'Rubrum'.

Declarations

Ethics approval and consent to participate

The materials used in this study were collected by ourselves. And we complied with all relevant institutional, national and international guidelines and specify the appropriate permissions obtained. We also have acquired a permission to collect all of the plant materials.

Consent for publication

Not applicable.

Availability of sequence data

The Sequencing data used in this study has been submitted to the BioProject database of National Center for Biotechnology Information (PRJNA741859, PRJNA744323).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KT and JW conceived the study and designed the experiments. LL performed the experiment. LL analyzed the data with suggestions by JW, HZ, CH. LL wrote the manuscript. All authors read and approved the final version of the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (31901397), Beijing Natural Science Foundation (No.6204039), and Scientific Funds of Beijing Academy of Agriculture and Forestry Sciences (KJCX20210431). Each of the funding bodies granted the funds based on a research proposal. They had no

602 influence over the experimental design, data analysis or interpretation, or writing the
603 manuscript.

604 **Acknowledgements**

605 We acknowledge the Biomarker Corporation (Beijing, China) for the facilities and
606 expertise of Illumina platform for libraries construction and sequencing.

607

608 **References**

609 Altschul, S.F., Madden, T.L., Schffer, A.A., Zhang, J., Zhang, Z., Webb, M., and Lipman,
610 D.J.J.N.a.R. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein
611 database search programs. *Nucleic Acids Research* 25(17)**,** 3389-3402.

612 Beckwith, A.G., Zhang, Y., Seeram, N.P., Cameron, C., A., and Nair, M.G. (2004).
613 Relationship of light quantity and anthocyanin production in Pennisetum setaceum
614 Cvs. rubrum and red riding hood. *J Agric Food Chem* 52**,** 456-461.

615 Bella, S., and D'urso, V. (2012). First record in the Mediterranean basin of the alien
616 leafhopper Balclutha brevis living on invasive Pennisetum setaceum. *Bulletin of*
617 *Insectology* 65**,** 195-198.

618 Bing, Z., Liu, J., Wang, X., Wei, Z.J.P.P., and Biochemistry (2018). Full-length RNA
619 sequencing reveals unique transcriptome composition in bermudagrass. *Plant Physiol*
620 *Biochem* 132**,** 95-103.

621 Chao, Y., Yuan, J., Guo, T., Xu, L., Mu, Z., and Han, L. (2019). Analysis of transcripts and
622 splice isoforms in Medicago sativa L. by single-molecule long-read sequencing. *Plant*
623 *Mol Biol* 99**,** 219-235.

624 Chao, Y., Yuan, J., Li, S., Jia, S., Han, L., and Xu, L. (2018). Analysis of transcripts and splice
625 isoforms in red clover (Trifolium pratense L.) by single-molecule long-read
626 sequencing. *BMC Plant Biol* 18**,** 300.

627 Chen, X., Xia, L., Zhu, S., Tang, S., Mei, S., Chen, J., Shan, L., Liu, M., Gu, Y., and Dai,
628 Q.J.D.R. (2018). Transcriptome-referenced association study of clove shape traits in
629 garlic. *DNA Res* 25**,** 587-596.

630 Dewey, C.N., and Bo, L. (2011). RSEM: accurate transcript quantification from RNA-Seq
631 data with or without a reference genome. *BMC Bioinformatics* 12**,** 323-323.

632 Fleming, M.B., Patterson, E.L., Reeves, P.A., Richards, C.M., and Walters, C.J.J.O.E.B.
633 (2018). Exploring the fate of mRNA in aging seeds: protection, destruction, or slow
634 decay? *Journal of Experimental Botany* 69**,** 4309-4321.

635 Fossen, T., and Andersen, Y.M. (1998). Cyanidin 3-O-(6″-Succinyl-β-glucopyranoside) and
636 other anthocyanins from phragmites australis. *J. Agric. Food Chem* 49**,** 1065-1068.

637 Fossen, T., Slimestad, R., ?Vstedal, D.O., and Andersen, Y.M. (2002). Anthocyanins of
638 grasses. *Biochemical Systematics & Ecology* 30**,** 855-864.

639 Fossen, T., Slimestad, R., and Andersen, O.M. (2001). Anthocyanins from maize (Zea mays)

and reed canarygrass (Phalaris arundinacea). *Journal of Agricultural & Food Chemistry* 49**,** 2318-2321.

Godzik, L.a.J.B. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22**,** 1658-1659.

Goergen, E., and Daehler, C.C. (2002). Factors affecting seedling recruitment in an invasive grass (Pennisetum setaceum) and a nativegrass (Heteropogon contortus) in the Hawaiian Islands. *Plant Ecology* 161**,** 147-156.

Gould, K.S., Markham, K.R., Smith, R.H., and Goris, J.J.J.J.O.E.B. (2000). Functional role of anthocyanins in the leaves of Quintinia serrata A. Cunn. *J Exp Bot* 51**,** 1107-1115.

Harborne, J.B., and Self, R. (1987). Malonated cyanidin 3-glucosides in Zea mays and other grasses. *Phytochemistry* 26**,** 2417-2418.

He, J., and Giusti, M.M. (2010). Anthocyanins: natural colorants with health-promoting properties. *Annu Rev Food Sci Technol* 1**,** 163-187.

Hoch, W.A., Zeldin, E.L., and Mccown, B.H. (2001). Physiological significance of anthocyanins during autumnal leaf senescence. *Tree Physiology* 21**,** 1-8.

Hu, C., Gong, Y., Jin, S., and Zhu, Q. (2011). Molecular analysis of a UDP-glucose: flavonoid 3-O-glucosyltransferase (UFGT) gene from purple potato (Solanum tuberosum). *Molecular Biology Reports* 38**,** 561-567.

Jaakola, L. (2013). New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends in Plant Science* 18**,** 477-483.

Jia, D., Wang, Y., Liu, Y., Hu, J., Guo, Y., Gao, L., and Ma, R.J.S.R. (2018). SMRT sequencing of full-length transcriptome of flea beetle Agasicles hygrophila (Selman and Vogt). *Scientific reports* 8**,** 2197.

Jia, N., Wang, J., Wang, Y., Ye, W., Liu, J., Jiang, J., Sun, J., Yan, P., Wang, P., Wang, F., and Fan, B. (2021a). The Light-Induced WD40-Repeat Transcription Factor DcTTG1 Regulates Anthocyanin Biosynthesis in Dendrobium candidum. *Front Plant Sci* 12**,** 633333.

Jia, N., Wang, J.J., Liu, J., Jiang, J., Sun, J., Yan, P., Sun, Y., Wan, P., Ye, W., and Fan, B. (2021b). DcTT8, a bHLH transcription factor, regulates anthocyanin biosynthesis in Dendrobium candidum. *Plant Physiol Biochem* 162**,** 603-612.

Jian, W., Cao, H., Yuan, S., Liu, Y., Lu, J., Lu, W., Li, N., Wang, J., Zou, J., Tang, N., Xu, C., Cheng, Y., Gao, Y., Xi, W., Bouzayen, M., and Li, Z. (2019). SlMYB75, an MYB-type transcription factor, promotes anthocyanin accumulation and enhances volatile

673 aroma production in tomato fruits. *Hortic Res* 6**,** 22.
674 Kyoungwon, C., Kwang-Soo, C., Hwang-Bae, S., Jin, H.I., Su-Young, H., Hyerim, L., Young-
675 Mi, K., and Hee, N.M. (2016). Network analysis of the metabolome and
676 transcriptome reveals novel regulation of potato pigmentation. *Journal of*
677 *Experimental Botany* 67**,** 1519-1533.
678 Lai, B., Du, L.N., Rui, L., Bing, H., Su, W.B., Qin, Y.H., Zhao, J.T., Wang, H.C., and Hu, G.B.
679 (2016). Two LcbHLH Transcription Factors Interacting with LcMYB1 in Regulating
680 Late Structural Genes of Anthocyanin Biosynthesis in Nicotiana and Litchi chinensis
681 During Anthocyanin Accumulation. *Front Plant Sci* 7**,** 166.
682 Lepiniec, Loic, Xu, Wenjia, Dubos, and Christian (2015). Transcriptional control of flavonoid
683 biosynthesis by MYB-bHLH-WDR complexes. *Trends in Plant Science* 20**,** 176-185.
684 Li, J., An, Y., and Wang, L.J.I.J.O.M.S. (2020a). Transcriptomic Analysis of Ficus carica
685 Peels with a Focus on the Key Genes for Anthocyanin Biosynthesis. *Int J Mol Sci* 21**,**
686 1245.
687 Li, Q., Xiang, C., Xu, L., Cui, J., Fu, S., Chen, B., Yang, S., Wang, P., Xie, Y., and Wei,
688 M.J.B.G. (2020b). SMRT sequencing of a full-length transcriptome reveals transcript
689 variants involved in C18 unsaturated fatty acid biosynthesis and metabolism
690 pathways at chilling temperature in Pennisetum giganteum. *BMC Genomics* 21**,** 52.
691 Lichtenthaler, Hartmut, K., Wellburn, Alan, and R. (1983). Determinations of total
692 carotenoids and chlorophylls a and b of leaf extracts in different solvents.
693 *Biochm.soc.trans*.
694 Ma Gonzalez-Rodriguez, A., Baruch, Z., Palomo, D., Cruz-Trujillo, G., Soledad Jimenez, M.,
695 and Morales, D. (2010). Ecophysiology of the invader Pennisetum setaceum and
696 three native grasses in the Canary Islands. *Acta Oecologica-International Journal of*
697 *Ecology* 36**,** 248-254.
698 Minoru, K., Susumu, G., Shuichi, K., Yasushi, O., and Masahiro, H. (2004). The KEGG
699 resource for deciphering the genome. *Nucleic Acids Research* 32**,** D277.
700 Nakatsuka, T., Haruta, K.S., Pitaksutheepong, C., Abe, Y., and Nishihara, M. (2008).
701 Identification and Characterization of R2R3-MYB and bHLH Transcription Factors
702 Regulating Anthocyanin Biosynthesis in Gentian Flowers. *Plant and Cell Physiology*
703 49**,** 1818-1829.
704 Robinson, M.D., Mccarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for
705 differential expression analysis of digital gene expression data. *Biogeosciences* 26**,**

139-140.

Shi, Q., Du, J., Zhu, D., Li, X., and Li, X. (2020). Metabolomic and Transcriptomic Analyses of Anthocyanin Biosynthesis Mechanisms in the Color Mutant Ziziphus jujuba cv. Tailihong. *Journal of Agricultural and Food Chemistry* 68**,** 15186-15198.

Steyn, W.J., Wand, S., Holcroft, D.M., and Jacobs, G.J.N.P. (2002). Anthocyanins in vegetative tissues: a proposed unified function in photoprotection. *New Phytol* 155**,** 349-361.

Sun, M., Lin, C., Zhang, A., Wang, X., Yan, H., Khan, I., Wu, B., Feng, G., Nie, G., Zhang, X., and Huang, L. (2021). Transcriptome sequencing revealed the molecular mechanism of response of pearl millet root to heat stress. *Journal of Agronomy and Crop Science* 207**,** 768-773.

Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. *Plant Journal* 54**,** 733-749.

Teng K, Teng Wj, Wen Hf, Yue Ys, Guo We, Wu Jy, and Xf, F. (2019). PacBio single-molecule long-read sequencing shed new light on the complexity of the Carex breviculmis transcriptome. *BMC Genomics* 20(suppl_2):158-164.

Teng, K., Tan, P., Xiao, G., Han, L., Chang, Z., and Chao, Y. (2016). Heterologous expression of a novel Zoysia japonica salt-induced glycine-rich RNA-binding protein gene, ZjGRP, caused salt sensitivity in Arabidopsis. *Plant Cell Reports* 36**,** 1-13.

Tian, Y., Rao, S., Li, Q., Xu, M., Wang, A., Zhang, H., and Chen, J. (2021). The coloring mechanism of a novel golden variety in Populus deltoides based on the RGB color mode. *Forestry Research* 1**,** 1-13.

Viljoen, C.D., Snyman, M.C., and Spies, J.J. (2013). Identification and expression analysis of chalcone synthase and dihydroflavonol 4-reductase in Clivia miniata. *South African Journal of Botany* 87**,** 18-21.

Wang, M., Wang, P., Liang, F., Ye, Z., Li, J., Shen, C., Pei, L., Wang, F., Hu, J., and Tu, L.J.N.P. (2017a). A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytol* 217**,** 163-178.

Wang, N., Xu, H., Jiang, S., Zhang, Z., Lu, N., Qiu, H., Qu, C., Wang, Y., Wu, S., and Chen, X. (2017b). MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (Malus sieversii f. niedzwetzkyana). *Plant J* 90**,** 276-292.

Wang, T., Wang, H., Cai, D., Gao, Y., and Gu, L.J.T.P.J. (2017c). Comprehensive profiling of

739 rhizome-associated alternative splicing and alternative polyadenylation in moso
740 bamboo (Phyllostachys edulis). *Plant J* 91**,** 684-699.
741 Wei, S., Meng, X., Liang, L., Jiang, W., Huang, Y., He, J., Hu, H., Almqvist, J., Xiang, G., and
742 Wang, L. (2015). Molecular and Biochemical Analysis of Chalcone Synthase from
743 Freesia hybrid in Flavonoid Biosynthetic Pathway. *Plos One* 10**,** e0119054.
744 Workman, R.E., Myrka, A.M., William, W.G., Elizabeth, T., Welch, K.C., and Winston, T.J.G.
745 (2018). Single molecule, full-length transcript sequencing provides insight into the
746 extreme metabolism of ruby-throated hummingbird Archilochus colubris.
747 *Gigascience* 7**,** 1-12.
748 Wu, B., Sun, M., Zhang, H., Yang, D., Lin, C., Khan, I., Wang, X., Zhang, X., Nie, G., Feng,
749 G., Yan, Y., Li, Z., Peng, Y., and Huang, L. (2021). Transcriptome analysis revealed
750 the regulation of gibberellin and the establishment of photosynthetic system promote
751 rapid seed germination and early growth of seedling in pearl millet. *Biotechnol*
752 *Biofuels* 14**,** 94.
753 Xu, H., Zou, Q., Yang, G., Jiang, S., Fang, H., Wang, Y., Zhang, J., Zhang, Z., Wang, N., and
754 Chen, X. (2020). MdMYB6 regulates anthocyanin formation in apple both through
755 direct inhibition of the biosynthesis pathway and through substrate removal.
756 *Horticulture Research* 7.
757 Yokozawa, T., Chen, C.P., Dong, E., Tanaka, T., and Nishioka, I. (1998). Study on the
758 Inhibitory Effect of Tannins and Flavonoids against the 1,1-Diphenyl-2-
759 picrylhydrazyl Radical. *Biochemical Pharmacology* 56.
760 Yonekura-Sakakibara, K., Higashi, Y., and Nakabayashi, R. (2019). The Origin and Evolution
761 of Plant Flavonoid Metabolism. *Frontiers in Plant Science* 10**,** 943.
762 Zhang, W., Guo, Y., Li, J., Huang, L., Kazitsa, E.G., and Wu, H. (2016). Transcriptome
763 analysis reveals the genetic basis underlying the seasonal development of keratinized
764 nuptial spines in Leptobrachium boringii. *Bmc Genomics* 17**,** 978.
765 Zhu, F.Y., Chen, M.X., Ye, N.H., Shi, L., Ma, K.L., Yang, J.F., Cao, Y.Y., Zhang, Y., Yoshida,
766 T., and Fernie, A.R.J.P.J. (2017). Proteogenomic analysis reveals alternative splicing
767 and translation as part of the abscisic acid response in Arabidopsis seedlings. *Plant*
768 *Journal* 91**,** 518-533.
769 Zhu, T., Wang, X., Xu, Z., Xu, J., Li, R., Liu, N., Ding, G., and Sui, S. (2020). Screening of
770 key genes responsible for Pennisetum setaceum 'Rubrum' leaf color using
771 transcriptome sequencing. *PLoS One* 15**,** e0242618.

772 Zhu, T., Zhang, L.Y., Zhi-Min, X.U., Hao, L., Chen, L., Liu, Y.Z., and Ding, G.C. (2019).
773 Effects of Shading and Regaining Natural Light Treatment on Leaf Color and
774 Physiological Changes of Pennisetum setaceum 'Rubrum'. *Acta Agrestia Sinica*.
775 Zhuang, H., Lou, Q., Liu, H., Han, H., Wang, Q., Tang, Z., Ma, Y., and Wang, H.J.I.J.O.M.S.
776 (2019). Differential Regulation of Anthocyanins in Green and Purple Turnips
777 Revealed by Combined De Novo Transcriptome and Metabolome Analysis. *Int J Mol*
778 *Sci* 20**,** 4387.
779
780

781 **Table 1 SMRT sequencing statistics**

Samples	Mix
cDNA size	1-6K
CCS Number	316,274
Read Bases of CCS	728,366,915
Mean Read Length of CCS	2,302
Mean Number of Passes	40
Number of full-length non-chimeric reads	280,413
Number of consensus isoforms	143,505
Number of polished high-quality isoforms	140,633
Number of polished low-quality isoforms	2,683

782

783 **Table 2 Annotated transcripts numbers based on eight databases**

Annotated Database	Isoform Number	300<=length<1000	length>=1000	Percentage (%)
COG_Annotation	37,402	2,144	35,257	41.67
GO_Annotation	73,264	5,860	67,368	81.62
KEGG_Annotation	40,056	3,334	36,686	44.63
KOG_Annotation	55,251	3,591	51,640	61.56
Pfam_Annotation	71,148	4,812	66,332	79.27
Swissprot_Annotation	63,343	4,552	58,760	70.57
eggNOG_Annotation	86,788	6,908	79,819	96.69
nr_Annotation	89,450	7,249	82,118	99.66
All_Annotated	89,758	7,306	82,363	100.00

784

785

Figure Legends

Figure 1

(a) Growth status of the mature *Pennisetum setaceum* 'Rubrum'. **(b)** Dynamic change of leaf color in different periods. The leaves of the same color are repeated three times.

Figure 2

(a) Circular consensus (CCS) read length distribution for 1-6K size bin. **(b)**The distribution of full passes of each cDNA database. **(c)** FLNC sequences read length distribution for mix size bin. **(d)** Consensus isoform sequence length distribution.

Figure 3

The gene annotation results based on Gene Ontology (GO) databases.

Figure 4

The results of NR alignment between *Pennisetum setaceum* 'Rubrum' and other species.

Figure 5

(a) Different expressed gene numbers distribution in three colors leaves of *Pennisetum setaceum* 'Rubrum'. **(b)** Rich distribution plots of differentially expressed transcripts between green leaves and lilac leaves in the KEGG pathway. **(c)** Rich distribution plots of differentially expressed transcripts between lilac leaves and purple leaves in the KEGG pathway. **(d)** Rich distribution plots of differentially expressed transcripts between green leaves and purple leaves in the KEGG pathway.

Figure 6

The differential expression of KEGG anthocyanin biosynthesis pathway annotated diagram. The red box indicates upregulated genes, the blue box indicates down regulated genes.

Figure 7

Anthocyanin content, chlorophyll content and net photosynthetic rate of the statistical graph results between green, lilac and purple leaves.

Figure 8

K-means clustering using 18 substances flavonoids in green, lilac and purple leaves.

Figure 9

qRT-PCR results of the expression levels of 10 randomly selected different genes in leaves of green, lilac and purple colors. R^2 represents the correlation between qRT-PCR and RNA sequencing results.

Supplementary Material

Table S1

Summary of the candidate alternative splice (AS) events

Table S2

Prediction results of the target mRNAs for lncRNAs

Table S3

Pigment contents of different flavonoids in green leaves, lilac leaves and purple leaves

Table S4

Sequences of the primers used for qRT-PCR verification

Figure S1

The length distribution of the predicted CDS-encoded protein.

Figure S2

Statistical graph of lncRNAs predicted by CPC analysis, CNCI analysis, PFAM protein domain analysis, and CPAT analysis.

Figure S3

Distribution of transcription factor types.

Figure S4

(a) Comparison diagram of FPKM density distribution of each sample. The abscissa represents the logarithm of the corresponding sample FPKM, and the ordinate of the point represents the probability density. **(b)** The logarithmic distribution of sample expression level FPKM. **(c)** The correlation of expression dependent heat map of the sample.

Figures

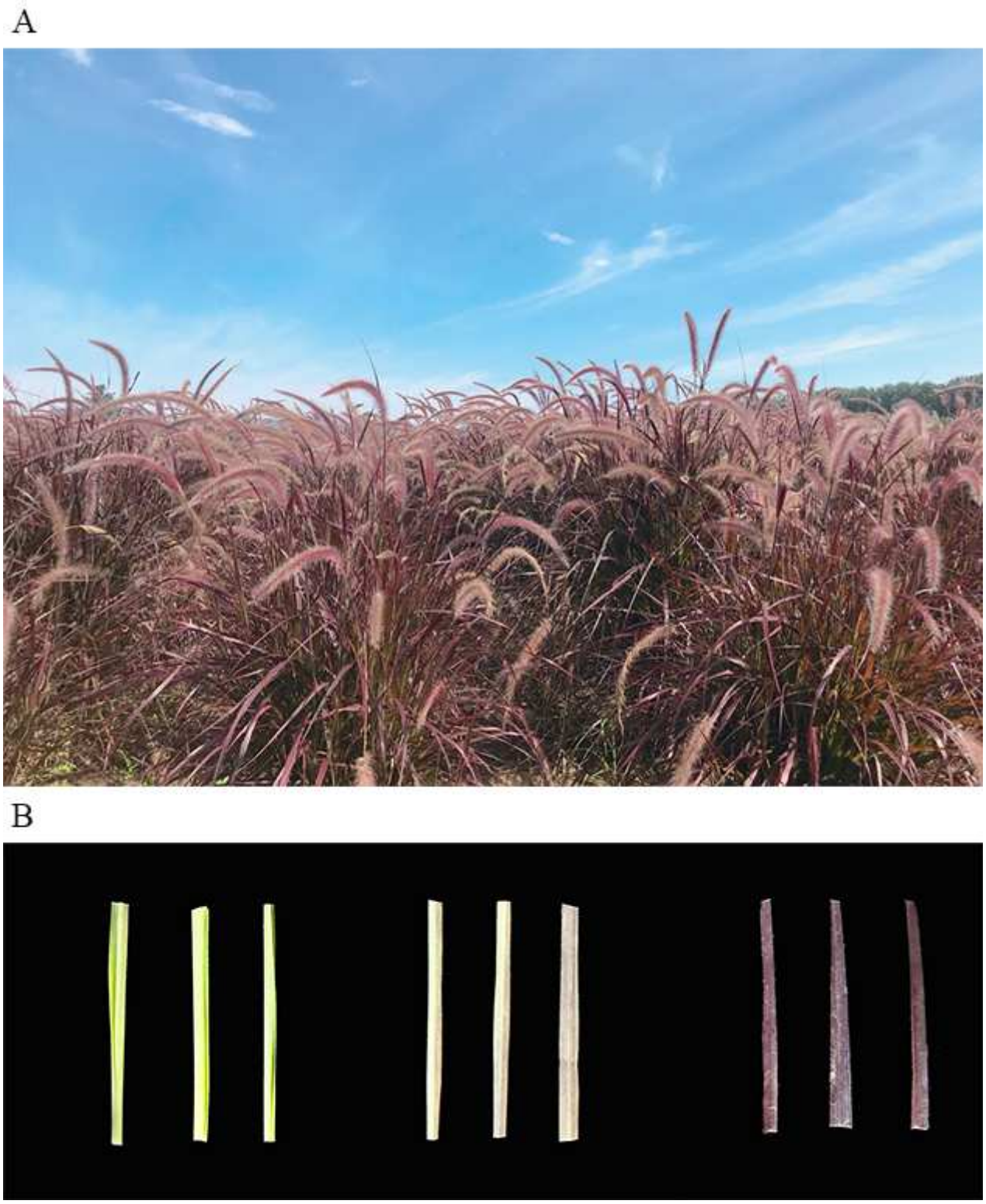

Figure 1

(a) Growth status of the mature Pennisetum setaceum 'Rubrum'. (b) Dynamic change of leaf color in different periods. The leaves of the same color are repeated three times.

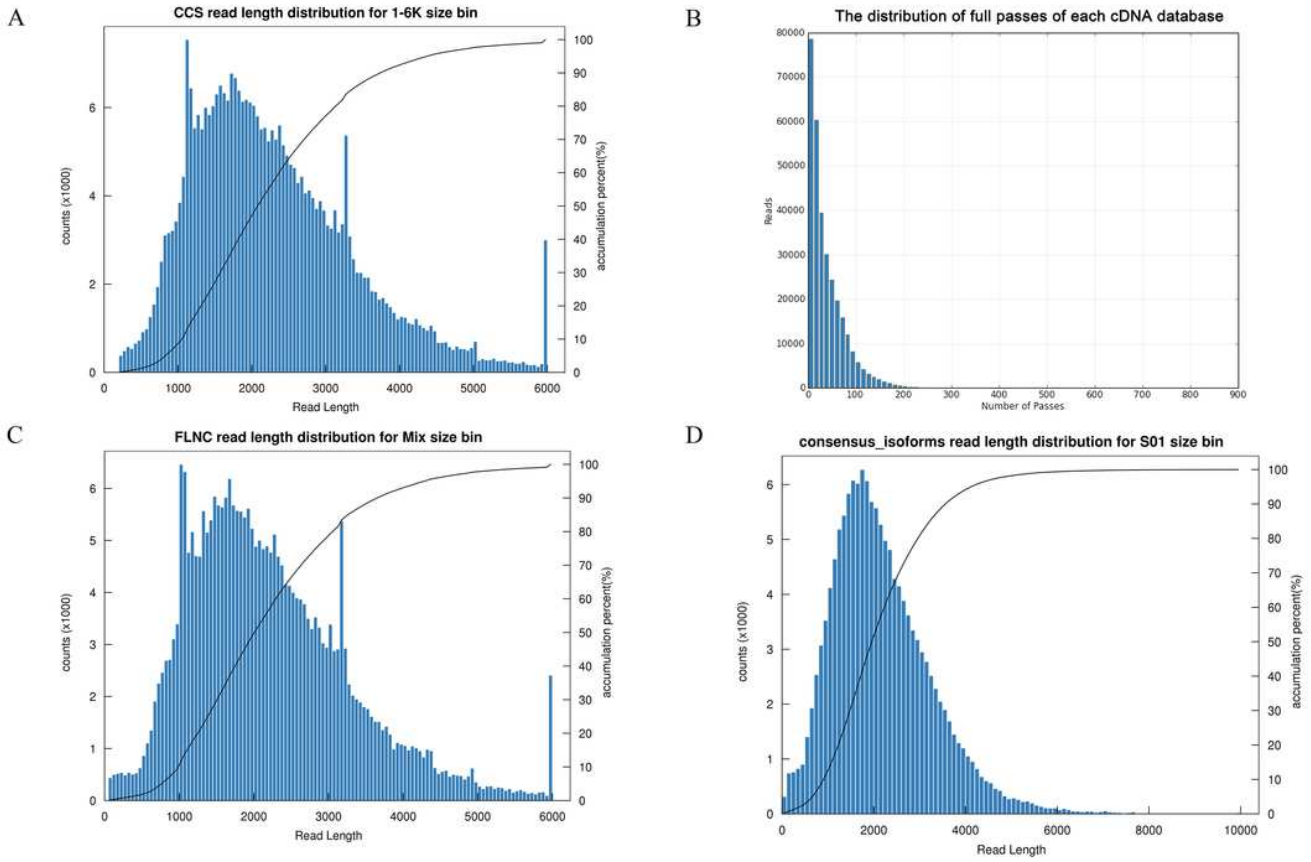

Figure 2

(a) Circular consensus (CCS) read length distribution for 1-6K size bin. (b)The distribution of full passes of each cDNA database. (c) FLNC sequences read length distribution for mix size bin. (d) Consensus isoform sequence length distribution.

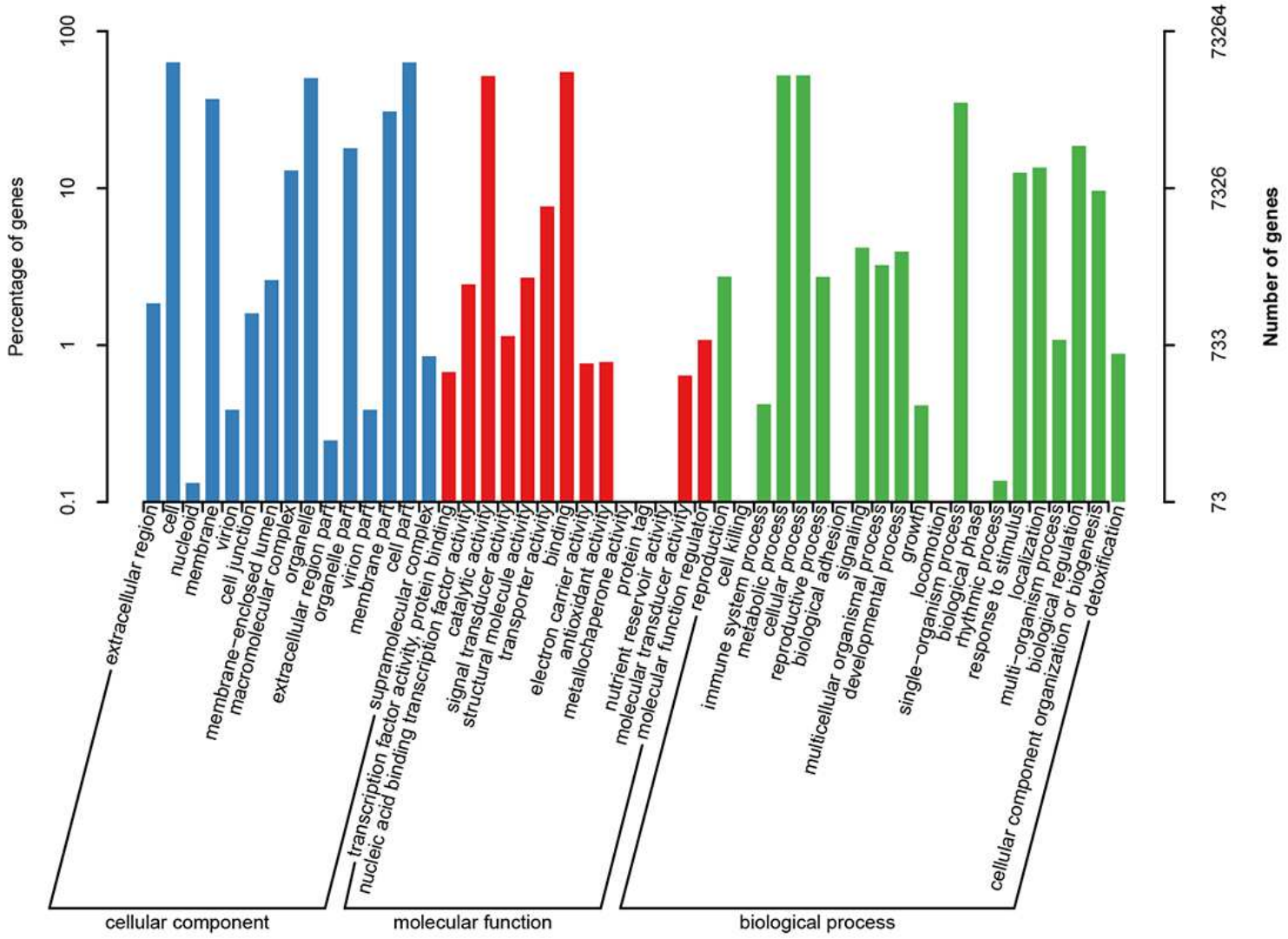

Figure 3

The gene annotation results based on Gene Ontology (GO) databases.

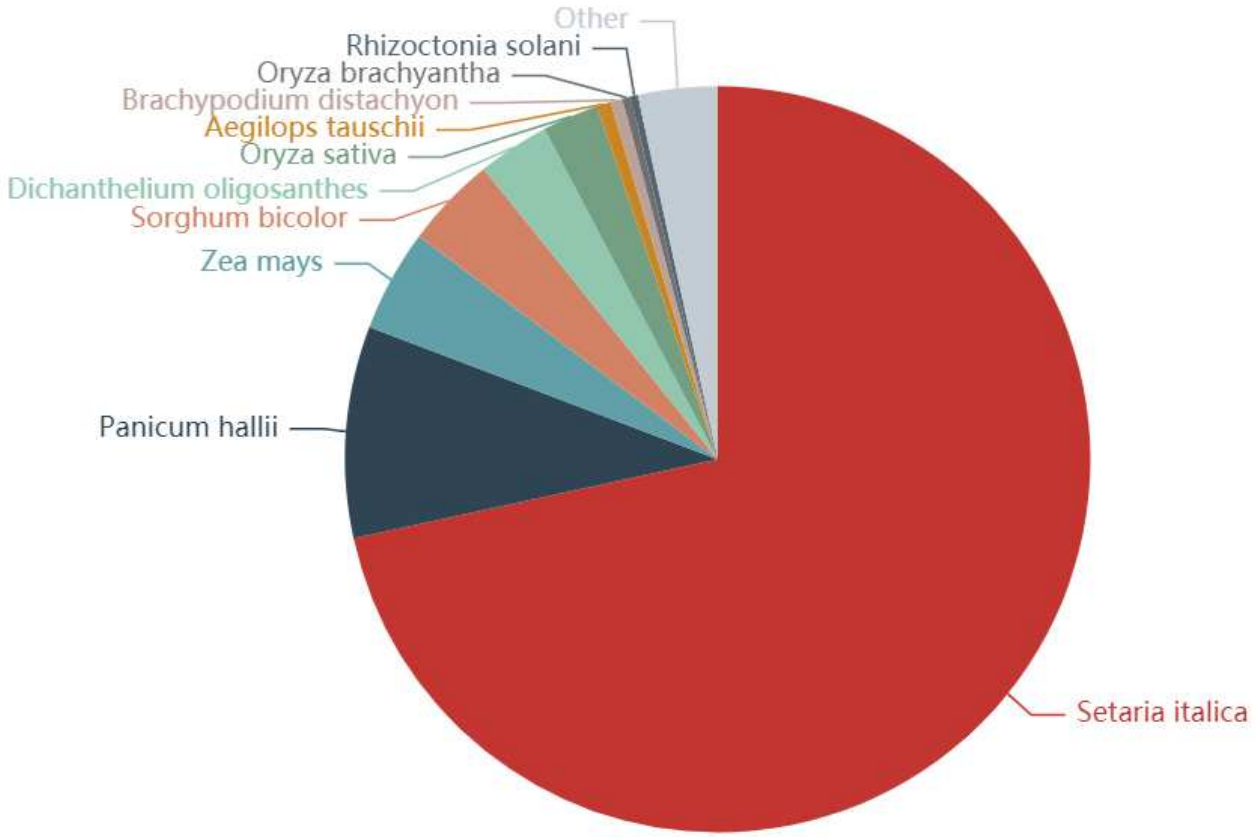

Figure 4

The results of NR alignment between Pennisetum setaceum ‘Rubrum’ and other species.

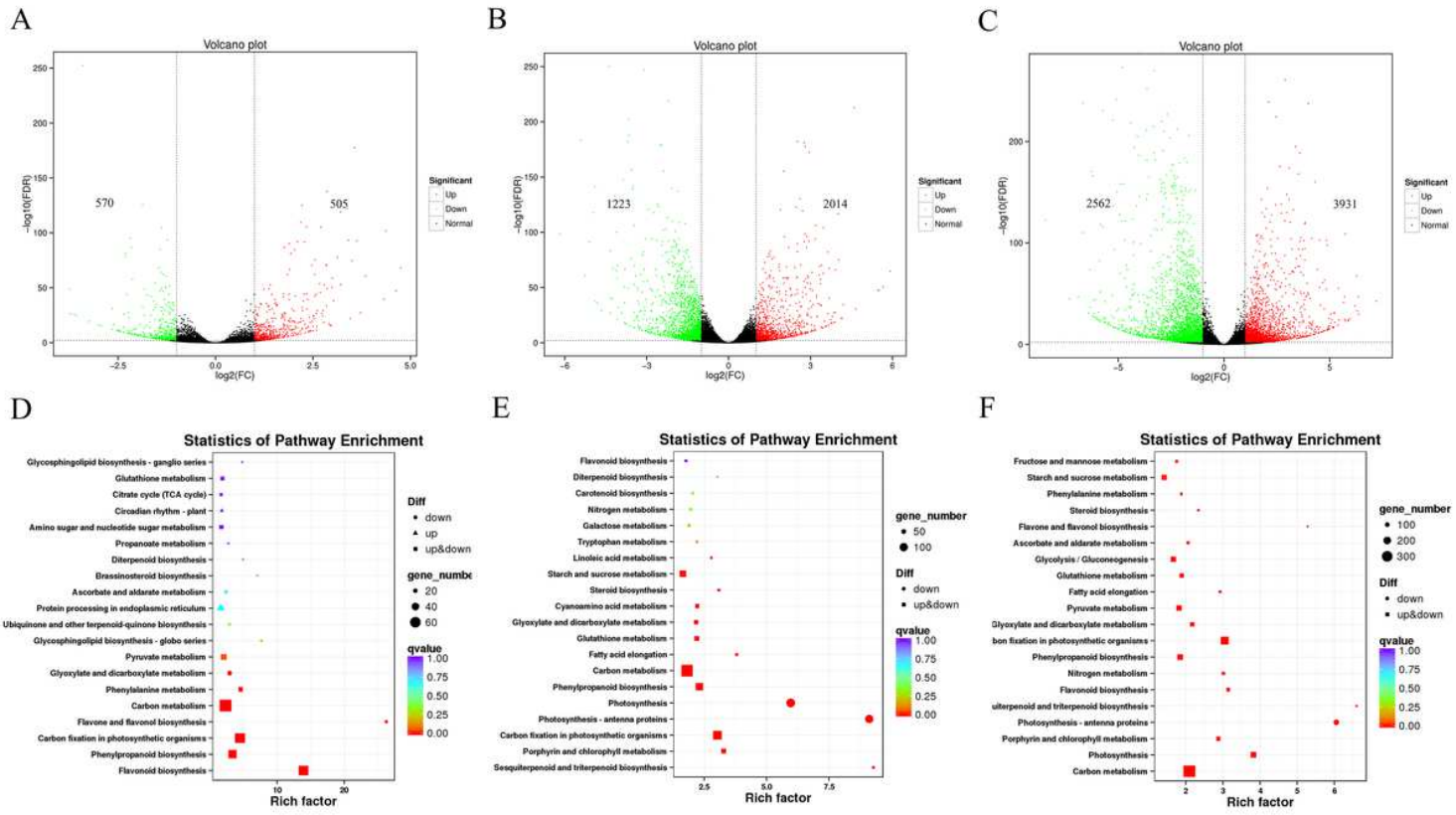

Figure 5

(a) Different expressed gene numbers distribution in three colors leaves of Pennisetum setaceum ‘Rubrum’. (b) Rich distribution plots of differentially expressed transcripts between green leaves and lilac leaves in the KEGG pathway. (c) Rich distribution plots of differentially expressed transcripts between lilac leaves and purple leaves in the KEGG pathway. (d) Rich distribution plots of differentially expressed transcripts between green leaves and purple leaves in the KEGG pathway.

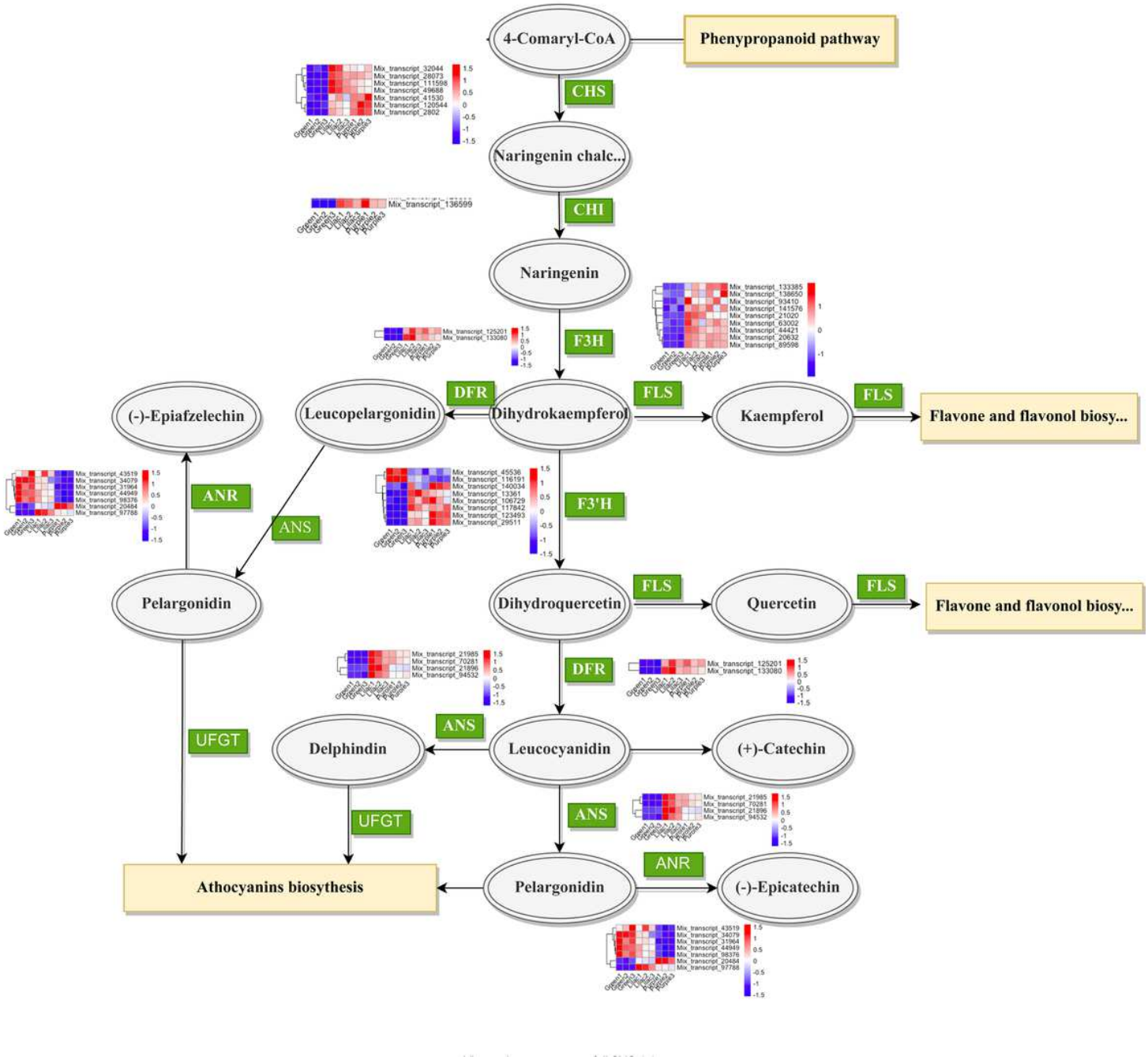

Figure 6

The differential expression of KEGG anthocyanin biosynthesis pathway annotated diagram. The red box indicates upregulated genes, the blue box indicates down regulated genes.

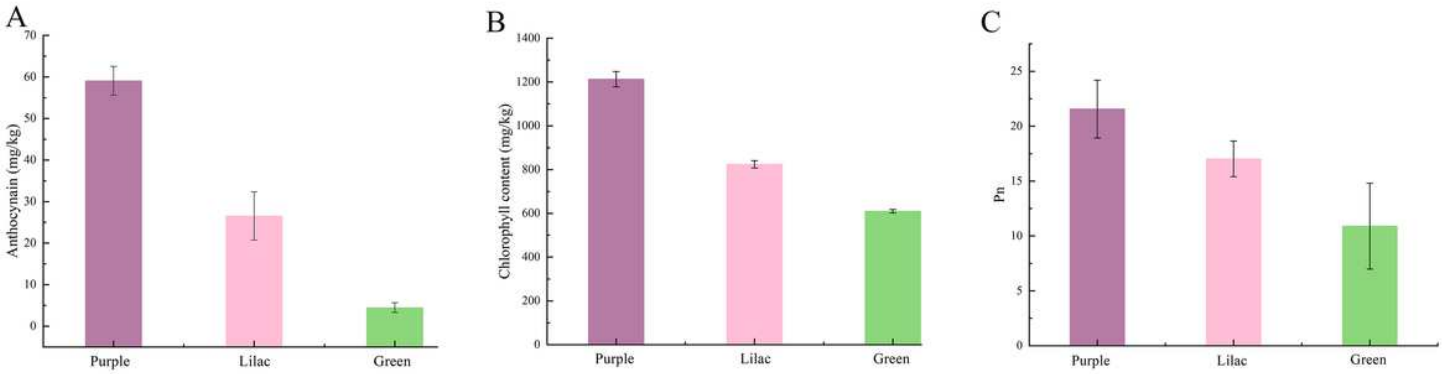

Figure 7

Anthocyanin content, chlorophyll content and net photosynthetic rate of the statistical graph results between green, lilac and purple leaves.

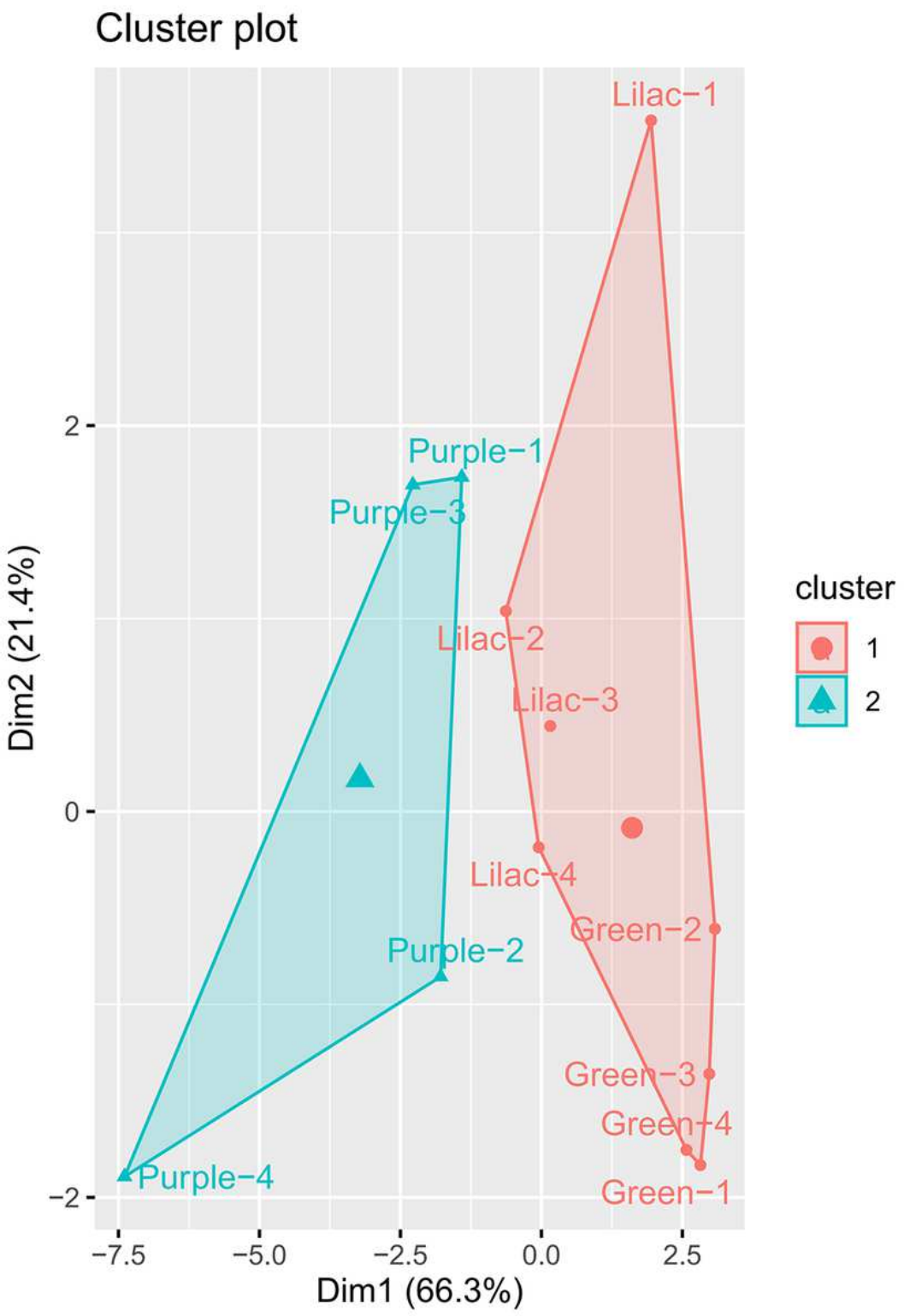

Figure 8

K-means clustering using 18 substances flavonoids in green, lilac and purple leaves.

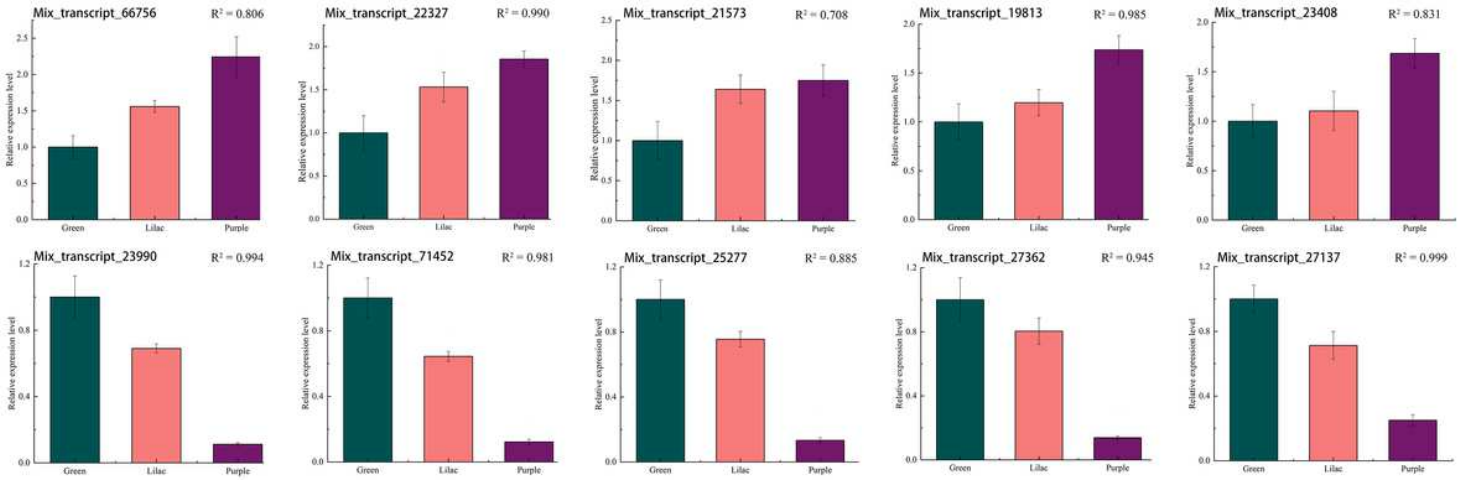

Figure 9

qRT-PCR results of the expression levels of 10 randomly selected different genes in leaves of green, lilac and purple colors. R2 represents the correlation between qRT819 PCR and RNA sequencing results.

Supplementary Files

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

- FigureS1.png
- FigureS2.png
- FigureS3.png
- FigureS4.tif
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx