

RESEARCH ARTICLE

Comparative transcriptome analysis of purple-fleshed sweet potato provides insights into the molecular mechanism of anthocyanin biosynthesis

Hongyuan ZHAO, Shanshan ZHANG, Feibing WANG, Ning ZHAO, Shaozhen HE, Qingchang LIU,
Hong ZHAI (✉)

Key Laboratory of Sweetpotato Biology and Biotechnology of Ministry of Agriculture/Beijing Key Laboratory of Crop Genetic Improvement/Laboratory of Crop Heterosis and Utilization of Ministry of Education, China Agricultural University, Beijing 100193, China

Abstract Sweet potato, *Ipomoea batatas*, is a globally important food crop. The purple-fleshed sweet potato, rich in anthocyanins, has great potential for both nutritional and pharmaceutical uses. In this study, we characterized the root transcriptomes of the purple-fleshed sweet potato cv. Jingshu 6 and its mutant JS6-5 with high anthocyanin content by high-throughput RNA sequencing. A total of 22873364 and 27955097 high quality reads were obtained from Jingshu 6 and JS6-5, respectively, and assembled into 35592 unigenes. In all, we obtained 1566 differentially expressed genes (DEGs). Among them, 994 were upregulated and 572 were downregulated in JS6-5 compared to the expression in Jingshu 6. A total of 1436 DEGs were annotated, in which 847 DEGs had gene ontology (GO) terms and 329 DEGs were assigned to 84 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Most importantly, 23 differentially expressed genes and 24 transcription factors were identified as candidate genes involved in anthocyanin biosynthesis. In addition, 2349 SSRs were detected. This study not only provides the candidate genes but also provides insights into the molecular mechanism of anthocyanin biosynthesis in sweet potato.

Keywords anthocyanin, gene expression, mutant, purple-fleshed sweet potato, transcriptome

1 Introduction

Sweet potato, *Ipomoea batatas*, a globally important food crop, has high yielding potential and wide adaptability. It also contains abundant nutritional materials including

complex carbohydrates, dietary fibers, minerals, vitamins and various antioxidants such as carotenoids and anthocyanins^[1]. The storage roots of purple-fleshed sweet potato are rich in anthocyanins. Now, high anthocyanin content has become one of the most important breeding goals for sweet potato.

Anthocyanins belong to the flavonoid family. They are important natural colorants widely distributed among flowers, fruits, seeds, leaves and storage roots, and have been implicated in tolerance to biotic and abiotic stresses in plants^[2]. Anthocyanins also have great potential to be utilized in food colorants, nutritional foods and drug development, such as hepatoprotective, antimutagenic, antineoplastic, antihyperglycemic, antiinflammatory and antioxidant agents^[3–6].

The structural genes and transcription factors involved in the synthesis and metabolism of anthocyanins have been identified in some plant species such as *Arabidopsis*, maize, petunia and snapdragon^[7]. Transcription factors of the MYB, basic helix–loop–helix (bHLH) and WD40 classes can form a MBW complex which binds to promoters and induces transcription of the phenylpropanoid^[8] and anthocyanin^[9] biosynthetic pathway genes, including those encoding phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate CoA ligase (4CL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavonoid 3'-hydroxylase (F3'H), flavonoid 3'5'-hydroxylase (F3'5'H), dihydroflavonol reductase (DFR), leucoanthocyanidin dioxygenase/anthocyanidin synthase (LDOX/ANS) and UDPglucose-flavonoid 3-O-glucosyltransferase (UFGT). Among these, PAL, C4H and 4CL catalyze the primary steps in the biosynthesis of phenylpropanoids, which convert phenylalanine to a wide variety of phenolic compounds including flavonoids. CHS, CHI, F3H, F3'H, F3'5'H, DFR, ANS/LDOX and UFGT are key enzymes controlling the metabolism of flavonoids, of which DFR,

Received January 6, 2018; accepted March 15, 2018

Correspondence: zhaihong@cau.edu.cn

ANS/LDOX and UFGT are more anthocyanin-specific enzymes^[10]. So far, several anthocyanin biosynthesis-associated genes including *IbCHI*, *IbF3'H*, *IbDFR*, *IbANS*, *IbMYB1*, *IbWD40* and *IbMADS10* have been isolated from purple-fleshed sweet potato^[11–17]. Their overexpression or downregulation was found to significantly affect anthocyanin levels. Although the anthocyanin biosynthetic pathway is well characterized, the molecular mechanisms regulating flux through the pathway are still unclear.

Sweet potato is an autohexaploid ($2n = 6x = 90$) with an estimated genome of 2.4 Gb^[18]. Due to the complexity of the genetics and the lack of genome resources, the breeding of sweet potato has been constrained. Therefore, transcriptome sequencing has become an efficient way to discover important genes and transcription factors in sweet potato. Transcriptome sequencing of sweet potato has offered important transcriptional data resources to study flower development, storage root formation, carotenoid biosynthesis, and cSSR and SNP marker development of this crop^[19–23]. Also, there have been several reports related to transcriptome sequencing of purple-fleshed sweet potato. Xie et al.^[24] performed RNA sequencing on the tuberous roots of the purple sweet potato and the results were compared with the RNA database of the yellow sweet potato. Ma et al.^[25] characterized the root transcriptomes of a mutant of purple sweet potato and its wild type by high-throughput RNA sequencing.

In this study, we conducted a transcriptome sequencing of the purple-fleshed sweet potato cv. Jingshu 6 and its mutant JS6-5 with high anthocyanin content on the Illumina HiSeq™ 2500 platform. We analyzed the differentially expressed genes (DEGs) involved in anthocyanin biosynthesis and provided insights into the molecular mechanism of anthocyanin biosynthesis.

2 Materials and methods

2.1 Plant materials

The purple-fleshed sweet potato cv. Jingshu 6 and its high anthocyanin mutant JS6-5 were used in this study (Fig. S1). Jingshu 6 is a commercial cultivar with anthocyanin content of 27.02 mg per 100 g FW (fresh weight), whereas in JS6-5, the content of anthocyanin is up to 79.80 mg per 100 g. Three growing tuberous roots (diameter 30–35 mm) of Jingshu 6 and JS6-5 were harvested 110 d after planting. The samples were immediately frozen in liquid nitrogen and stored at -80°C .

2.2 RNA extraction

We used the RNeasy Pure Plant Kit (Qiagen Biotech, Beijing, China) to extract total RNA from the storage roots of Jingshu 6 and JS6-5. After DNase treatment, the purified RNA concentrations were quantified using a Nanodrop

spectrophotometer (Thermo Nanodrop Technologies, Wilmington, DE, USA). The quality of the total RNA was subsequently examined using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

2.3 cDNA library construction and Illumina sequencing

Poly (A) mRNA was enriched from the total RNA using magnetic beads with Oligo (dT). Poly (A) mRNA was subsequently fragmented by an RNA fragmentation kit (Ambion, Austin, TX, USA). The fragmented RNA was transcribed into first-strand cDNA using reverse transcriptase and random hexamer primers. The second-strand cDNA was synthesized using DNA polymerase I and RNase H (Invitrogen, Carlsbad, CA, USA). After end repair and the addition of a poly (A) tail, suitable length fragments were isolated and connected to the sequencing adaptors. The fragments were sequenced on Illumina HiSeq™ 2500 platform.

2.4 De novo assembly of sequences

To acquire high quality reads, the raw reads were processed by removing adaptor sequences and low quality reads with unidentified bases. The clean reads were *de novo* assembled using the Trinity software which combined three components: Inchworm, Chrysalis and Butterfly^[26]. Firstly, clean reads produced from the two libraries were assembled together into contigs using the Inchworm program^[26]. Secondly, based on the paired-end sequences information, the contigs were linked into sets of connected parts using the Chrysalis program followed by transcript construction using the Butterfly program^[26]. Finally, the transcripts were clustered and processed with multiple sequence alignment tool BLAT according to their nucleotide identity^[26]. We took the longest transcript of each cluster units as the unigene. The unigenes formed into the non-redundant database used for annotation.

2.5 Functional annotation and classification of non-redundant unigenes

To confirm the functional annotation of the unigenes, they were aligned to various database including the NCBI non-redundant (Nr) database, the protein family (Pfam) database, the Swiss-Prot protein database, the Cluster of Orthologous Groups (COG) database, the euKaryotic Clusters of Orthologous Groups (KOG) database and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database by using BLASTX ($E < 1 \times 10^{-5}$). Using the Blast2GO program, we obtained gene ontology (GO) annotation of the unigenes according to the BLASTX results against the Nr databases. Each unigene was named based on the highest BLAST score. The WEGO software was used for executing function classification of all GO-annotated unigenes^[27]. The GO classification provides an

ontology of defined terms related to gene product properties, such as cellular component, molecular function, and biological process. It enhances our understanding of gene classification in a species-independent manner and provides an overview of gene functions of the species^[28].

2.6 Analysis of differentially expressed genes

Gene expression levels in Jingshu 6 and JS6-5 were normalized by calculating FPKM (fragments per kilobase of transcript per million fragments mapped)^[26]. The FPKM method minimizes the influence of sequencing depth and gene length when estimating gene expression levels. The EBSeg software was used to identify differentially expressed genes between the two samples. With the standard of Benjamini–Hochberg false discovery rate, the results of all the statistical tests were corrected by multiple testing. The unigenes were deemed significantly differentially expressed at adjusted $P < 0.01$ with at least a twofold change in FPKM.

2.7 Expression analysis of genes involved in anthocyanin biosynthesis by qRT-PCR

Total RNA of the storage roots of Jingshu 6 and JS6-5 was reverse-transcribed using the Quantscript Reverse Transcriptase Kit (Tiangen Biotech, Beijing, China). The cDNA solution was used as template to validate gene expression. The specific primers for *PAL*, *C4H*, *4CL*, *CHI*, *F3H*, *LDOX*, *UFGT*, *MYBI* and 12 randomly picked genes, and the sweet potato β -actin gene used as internal control are listed in Table S1. We performed the PCR amplifications by ABI PRISM 7500 (software for 7500 and 7500 Fast Real-Time PCR Systems, V2.0.1, Foster City, CA, USA) using SYBR[®] Premix Ex Taq[™] II (TaKaRa Code No. RR820A, Takara Biomedical Technology (Beijing) Co., Ltd.). The comparative CT method ($2^{-\Delta\Delta C_T}$ method) was used to quantify gene expression^[29].

2.8 SSR detection

Using the MISA software ((MicroSATellite identification

tool), six types of SSRs including mono-, di-, tri-, tetra-, penta- and hexa-nucleotide repeats were detected among the unigenes with length > 1000 bp.

3 Results

3.1 Transcriptome sequencing and *de novo* assembly

In this study, 23693146 raw reads with a total of 3659735905 nt and an average GC content of 48.7% were produced from tuberous roots of Jingshu 6, and 28751513 raw reads totalling of 4272814403 nt and an average GC content of 48.6% from tuberous roots of JS6-5. After removing adaptor sequences and discarding low quality reads, 22873364 (96.5%) and 27955097 (97.2%) high quality reads were obtained from Jingshu 6 and JS6-5, respectively. Further assembly of the high quality reads yielded 1910521 contigs with mean length of 59.61 bp. These contigs were assembled into 35592 unigenes with a mean length of 697.22 bp (Table 1). The length distributions of contigs and unigenes are shown in Fig. 1.

3.2 Function annotation and classification of non-redundant unigenes

For functional annotation, the sequences of 35592 assembled unigenes were searched in the Nr, Swiss-Prot, Pfam, KOG, GO, COG and KEGG databases. According to the BLASTX results, a total of 28183 (79.2%) unigenes were annotated with putative functions based on hits from at least one database. Among them, 28072 (78.9%) of the putative proteins showed similarity to sequences in the Nr database. Also, 18669 (52.5%), 17247 (48.5%), 16247 (45.7%), 14669 (41.2%), 7924 (22.3%) and 6113 (17.2%) of the unigenes had functional annotation in the Swiss-Prot, Pfam, KOG, GO, COG and KEGG databases, respectively (Table 2). In particular, the predicted result of the Nr database revealed that 5731 (16.1%) and 5431 (15.3%) unigenes showed significant homology with sequences of *Nicotiana sylvestris* and *N. tomentosiformis*, respectively (Fig. 2).

Table 1 Length distribution of assembled contigs and unigenes from Jingshu 6 and JS6-5

Item	Contig	Unigene
0–300 bp	1881281 (98.5%)	10383 (29.2%)
300–500 bp	11829 (0.62%)	9143 (25.7%)
500–1000 bp	10082 (0.53%)	8519 (23.9%)
1000–2000 bp	5865 (0.31%)	5899 (16.6%)
> 2000 bp	1464 (0.08%)	1648 (4.63%)
Total number	1910521	35592
Total length/bp	113887400	24815314
N50 length/bp	55	1032
Mean length/bp	59.61	697.22

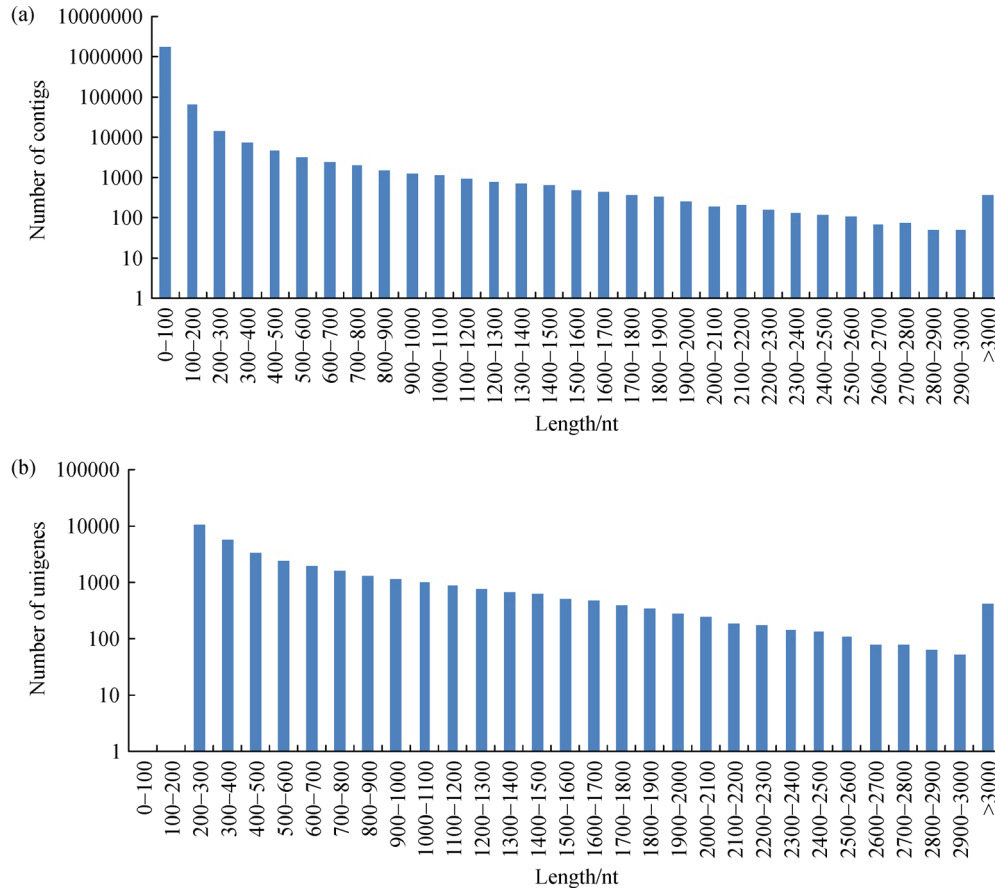


Fig. 1 Overview of the sweet potato transcriptome assembly. Length distribution of contigs (a) and unigenes (b).

Table 2 Summary of the annotation of the unigenes in the transcriptome of Jingshu 6 and JS6-5

Public protein database	No. of annotated unigenes	Percentage of annotated unigenes/%
COG annotation	7924	22.3
GO annotation	14669	41.2
KEGG annotation	6113	17.2
KOG annotation	16247	45.7
Pfam annotation	17247	48.5
Swiss-Prot annotation	18669	52.5
Nr annotation	28072	78.9
Total	28183	79.2

Note: COG, Cluster of Orthologous Groups; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; KOG, the euKaryotic Clusters of Orthologous Groups; Pfam, Protein family; Nr, Non-redundant.

3.3 Analysis and functional classification of DEGs

In all, 35592 unigenes had detectable levels of expression in Jingshu 6 and JS6-5. Using the IDEG6 software, 1566 genes were differentially expressed between Jingshu 6 and JS6-5 (Fig. 3). Among them, 994 were upregulated and 572 were downregulated in JS6-5 compared with Jingshu 6.

In total, 1436 (91.7%) of the 1566 DEGs were annotated. The numbers of DEGs annotated in the Nr, Swiss-Prot, Pfam, GO, KOG, COG and KEGG databases were 1433, 1130, 1073, 847, 762, 491 and 329, respectively.

According to the GO function and significant enrichment analyses (Fig. 4), a total of 847 DEGs were classified into three categories including 53 functional groups. The most common assignments in the cellular component

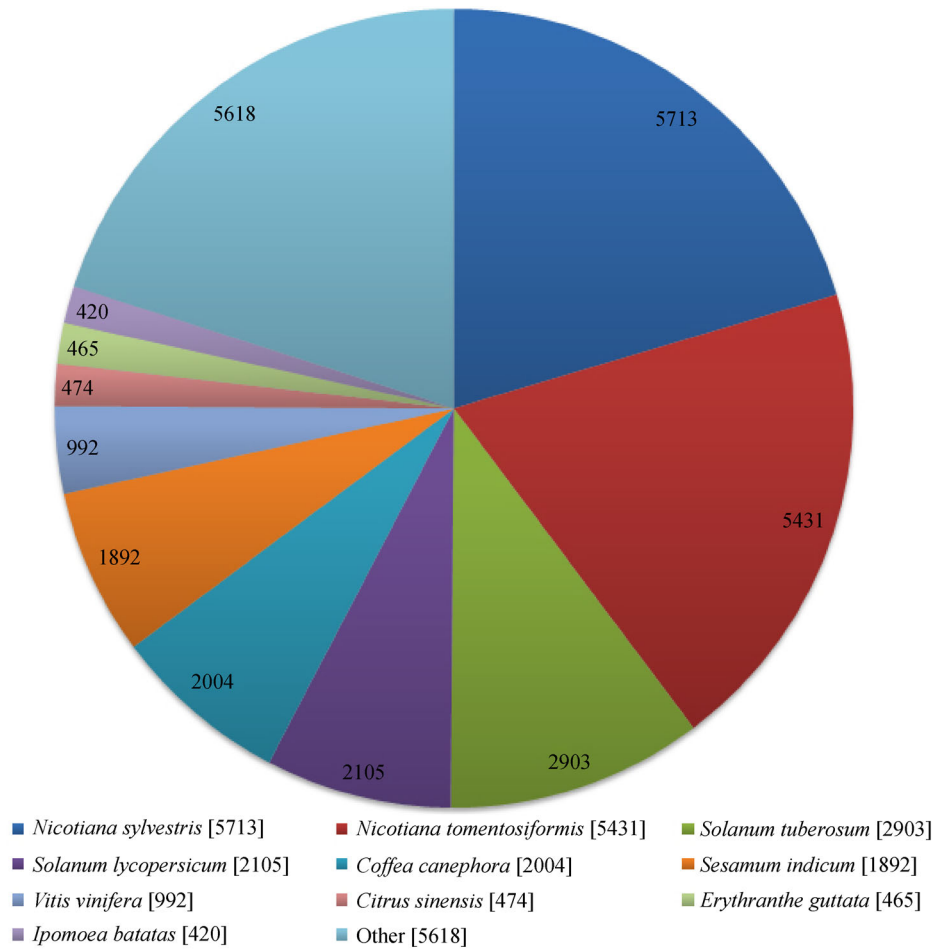


Fig. 2 Species distribution of BLAST hits for each unigene in the Nr database

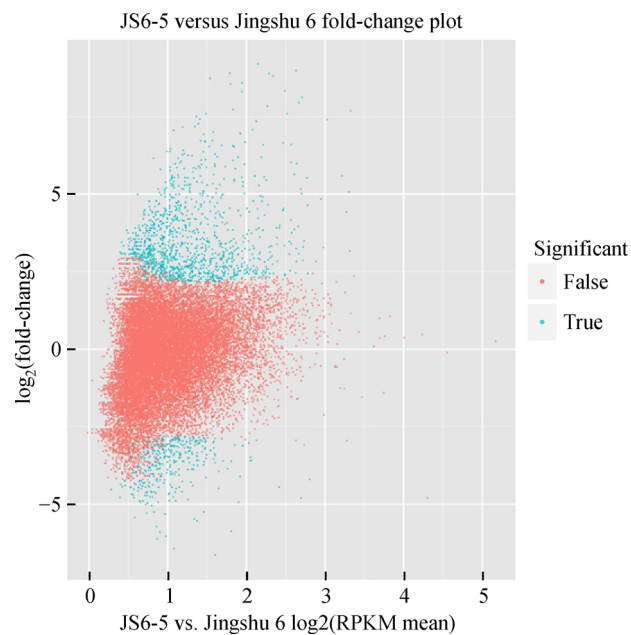


Fig. 3 Comparison of gene expression levels between Jingshu 6 and JS6-5. Blue dots represent genes that had significant differences and red dots represent genes where no significant differences observed.

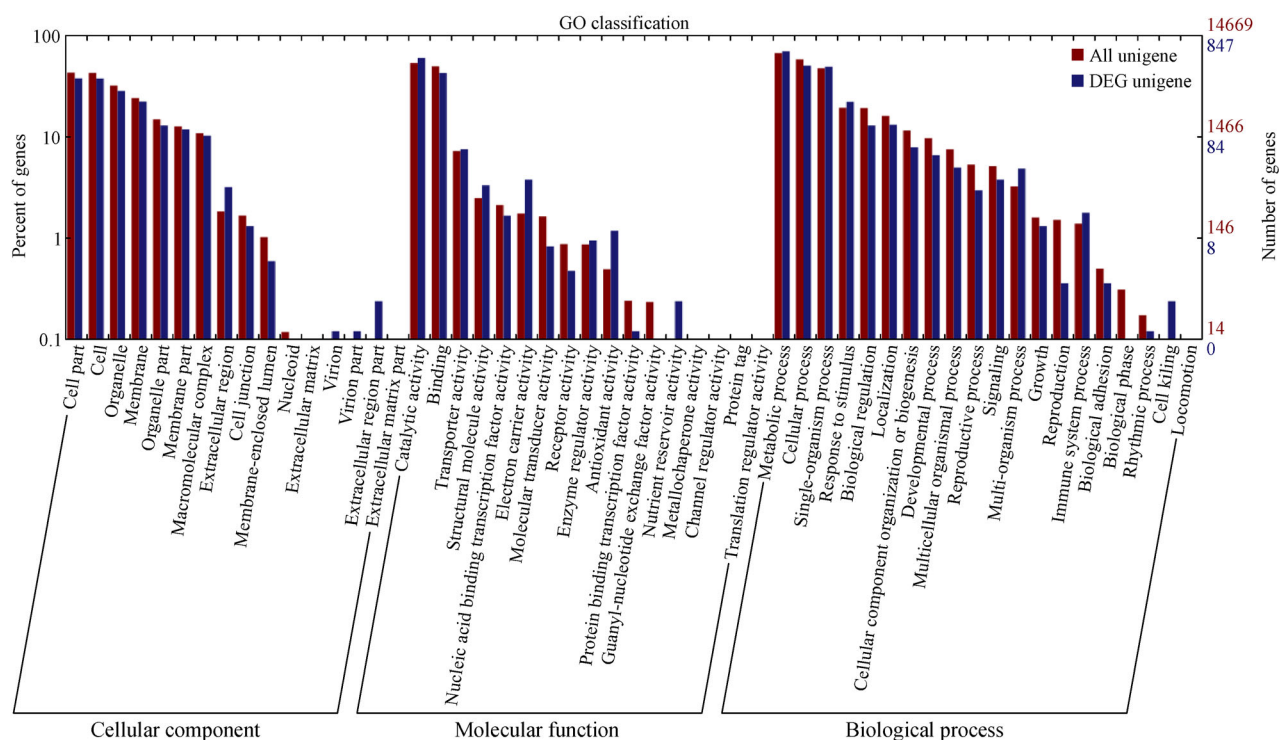


Fig. 4 Gene ontology (GO) classification of unigenes in the transcriptomes of Jingshu 6 and JS6-5. The red bars represent all the unigenes and the blue bars represent the differentially expressed genes (DEGs).

category were cell part (320 DEGs) and cell (319 DEGs) subcategories. In the molecular function category, the majority of DEGs were grouped into the catalytic activity (513 DEGs) and binding (364 DEGs) subcategories. Among all related molecular function, there were five terms related to anthocyanin biosynthesis including phenylalanine ammonia lyase activity (GO:0045548), 4-coumarate CoA ligase activity (GO:0016207), chalcone isomerase activity (GO:0045430), flavonoid 3',5'-hydroxylase activity (GO:0033772) and flavonoid 3'-monooxygenase activity (GO:0016711). DEGs in the biological process category were primarily sorted into the metabolic process (596 DEGs) and cellular process (430 DEGs). Among all related biological processes, there were six terms related to anthocyanin biosynthesis including L-phenylalanine catabolic process (GO:0006559), cinnamic acid biosynthetic process (GO:0009800), anthocyanin-containing compound biosynthetic process (GO:0009718), anthocyanin accumulation in tissues in response to UV light (GO:0043481), positive regulation of flavonoid biosynthetic process (GO:0009963) and flavonoid biosynthetic process (GO:0009813).

The 491 DEGs were annotated with the COG database and subdivided into 25 COG classifications (Fig. 5). The largest group is the cluster for general functional prediction (137 DEGs), followed by amino acid transport and metabolism (66 DEGs), carbohydrate transport and

metabolism (58 DEGs) and secondary metabolite biosynthesis (57 DEGs).

A total of 117 relevant metabolic pathways were identified by comparing the 35592 unigenes with the KEGG database. Among them, 329 DEGs were assigned to the 84 pathways. The most noticeable groups were phenylpropanoid biosynthesis (ko00940) and phenylalanine metabolism (ko00360) which are related to anthocyanin biosynthesis (Fig. 6).

3.4 Detection of genes related to anthocyanin biosynthesis

According to the results of the GO, COG, KEGG pathway enrichment, 14 DEGs were directly involved in anthocyanin biosynthesis (Table 3). Among them, we identified five genes related to PAL activity, three genes related to C4H, two genes related to 4CL and four genes related to CHI, F3H, LDOX and UFGT. Nine DEGs encoding the cytochrome P450 family, N-hydroxycinnamoyl/benzoyl-transferase (HCBT), ATP-citrate lyase (ACL) and transmembrane ascorbate ferrireductase were also found to be related to anthocyanin biosynthesis. All the genes above were significantly upregulated in JS6-5 compared with Jingshu 6 (Table 3). It is noteworthy that 24 important transcription factors, including MYB, bHLH, MADs, NAC, AP2/ERF and bZIP were significantly upregulated in JS6-5 compared with Jingshu 6 (Table 3).

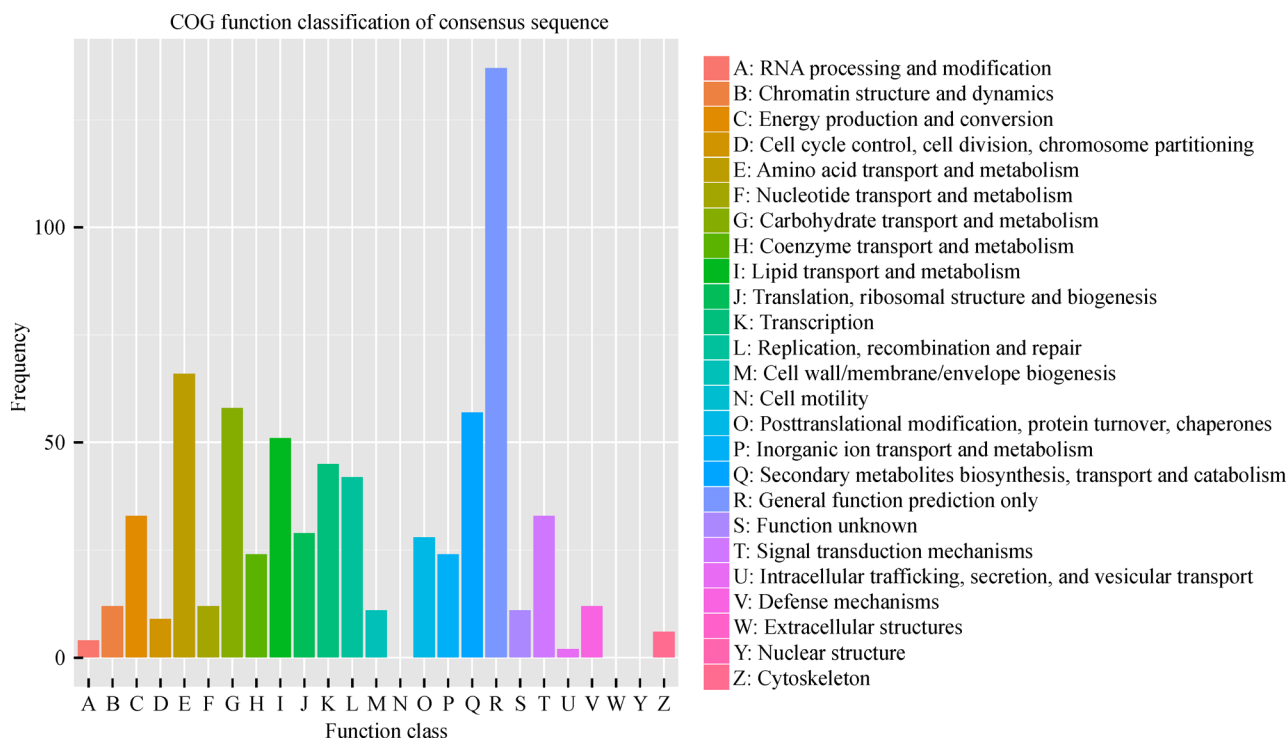


Fig. 5 COG-based functional classification of DEGs between Jingshu 6 and JS6-5

3.5 Expression analysis of genes involved in anthocyanin biosynthesis by qRT-PCR

To verify the transcriptome data, we used qRT-PCR to analyze the expression levels of 12 randomly selected genes, such as *c33937.graph_c0*, *c41003.graph_c0*, *c31758.graph_c0*, *c37779.graph_c0*, *c36415.graph_c0*, *c38523.graph_c0*, *c57881.graph_c0*, *c56053.graph_c0*, *c54151.graph_c0*, *c5336.graph_c02*, *c40801.graph_c0* and *c53825.graph_c0* in the two materials (Table S2). All these genes were significantly upregulated in JS6-5 and the results were highly concordant with the RNA-seq results (Fig. S2).

In addition, qRT-PCR results showed that the expression level of anthocyanin biosynthetic pathway genes including *PAL* (*c17852.graph_c0*), *C4H* (*c36515.graph_c0*), *4CL* (*c40730.graph_c0*), *CHI* (*c32008.graph_c0*), *F3H* (*c39514.graph_c0*), *LDOX* (*c38710.graph_c0*), *UFGT* (*c40432.graph_c0*) and *MYBI* (*c30141.graph_c0*) were also significantly upregulated in JS6-5 and the results were highly concordant with the RNA-seq results (Fig. 7).

3.6 SSR markers

Using the MISA software to search for potential SSRs from 7547 unigenes, 1810 of them contained SSR markers and 415 had more than one SSR. In total, 2349 SSRs were identified including 894 mono-, 686 di-, 714 tri-, 41 tetra-,

nine penta- and five hexa-nucleotide repeats.

4 Discussion

High-throughput transcriptome sequencing technology can generate large amounts of genome wide transcription data. Transcriptome analysis is an effective method to identify the candidate genes involved in complex biosynthetic pathways, especially in non-model plant for which reference genome sequences are unavailable^[26]. To date, several sweet potato transcriptomes have been sequenced^[19–23]. Only two reports were related to transcriptome sequencing of purple-fleshed sweet potato^[24,25]. In this study, we characterized the root transcriptomes of the purple-fleshed sweet potato cv. Jingshu 6 and its mutant JS6-5 with high anthocyanin content on the Illumina HiSeq™ 2500 sequencing platform. A total of 22873364 and 27955097 high quality reads were produced from Jingshu 6 and JS6-5, respectively. These reads were assembled into 35592 unigenes. There were 1566 DEGs between JS6-5 and Jingshu 6. Among them, 994 were upregulated and 572 were downregulated in JS6-5 compared with Jingshu 6. This study provides the genomic resources for discovering the candidate genes and enabling further research on the molecular mechanism of anthocyanin biosynthesis in sweet potato.

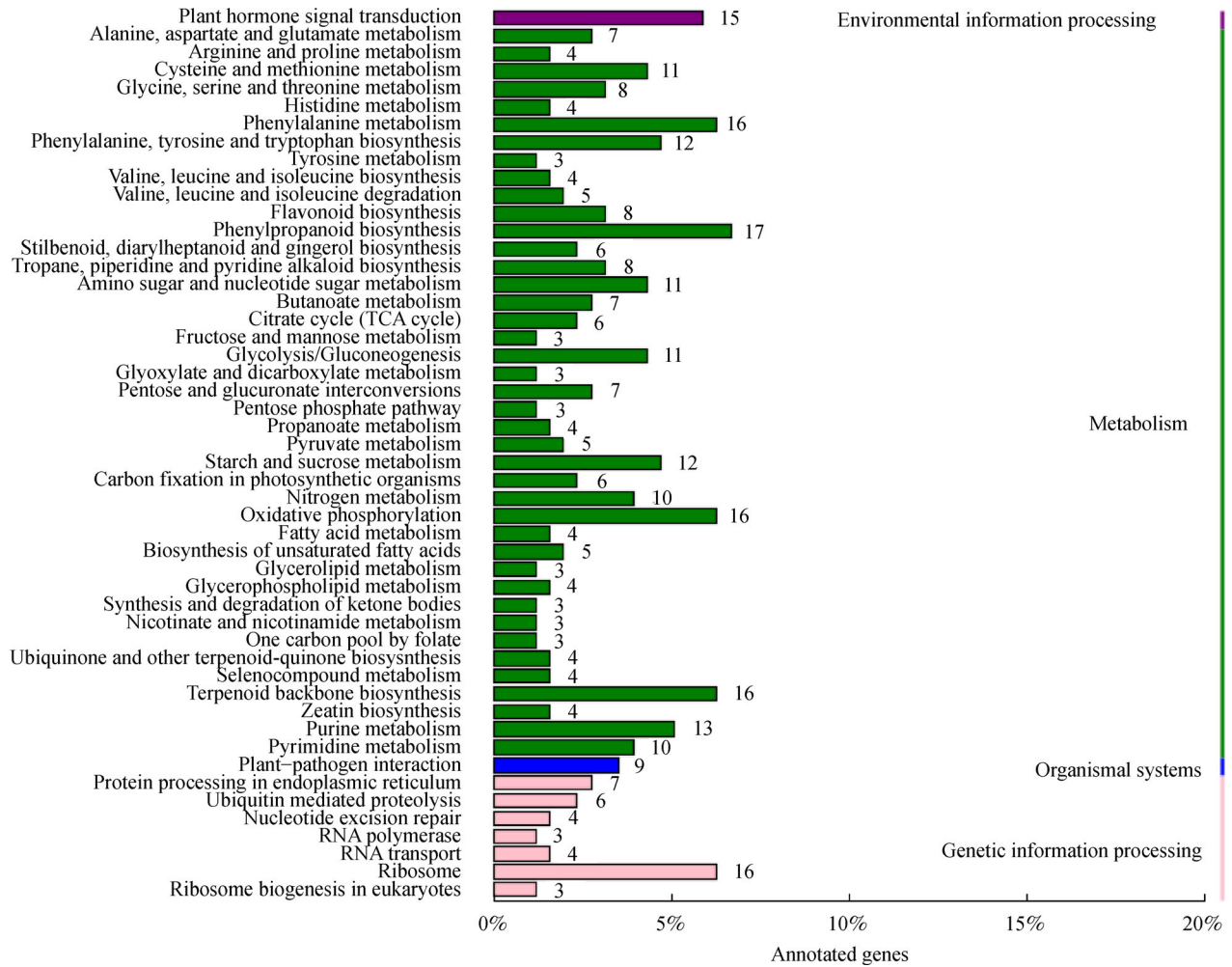


Fig. 6 KEGG-based functional classification of DEGs between Jingshu 6 and JS6-5. Numbers beside each bar indicate the actual number of DEGs classified in that descriptive term.

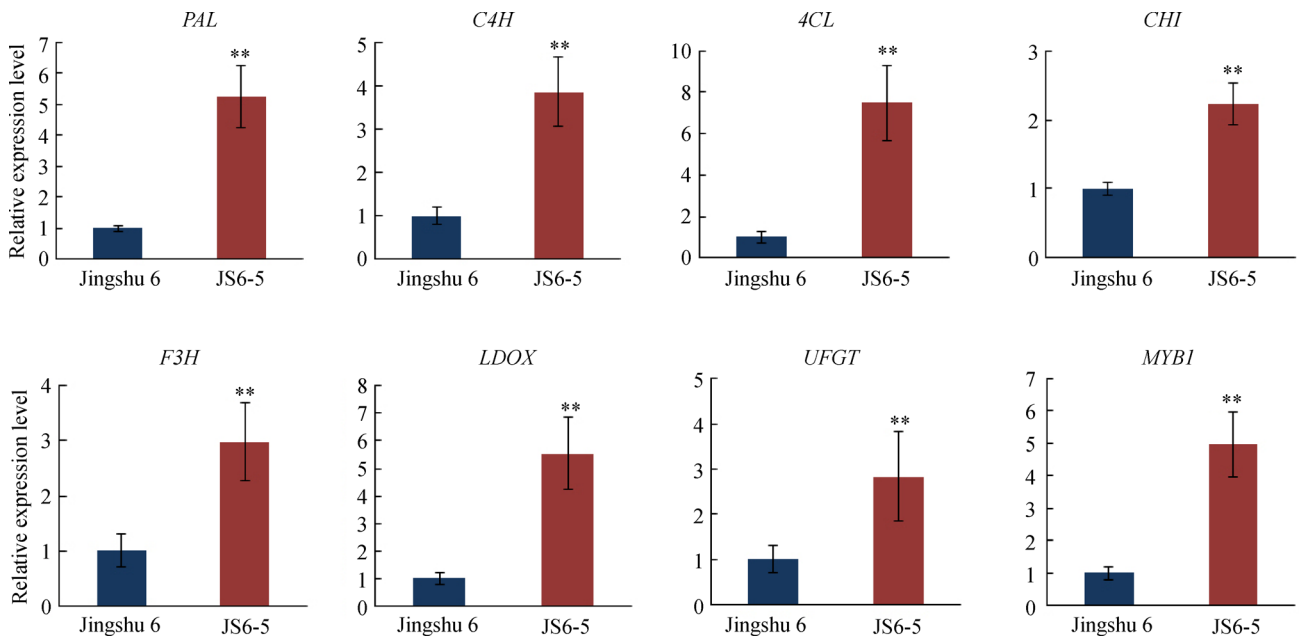


Fig. 7 Expression level of genes in anthocyanin biosynthetic pathway. **, significantly different at $P < 0.01$.

The results showed that 14 DEGs including *PAL* (5 unigenes), *C4H* (3 unigenes), *4CL* (2 unigenes), *CHI* (1 unigene), *F3H* (1 unigene), *LDOX* (1 unigene) and *UFGT* (1 unigene) encoded the enzymes that directly participated in anthocyanin biosynthesis (Fig. 8). In addition, 9 DEGs encoding the cytochrome P450 family, HCBT, ACL and transmembrane ascorbate ferrereductase were also found to be related to anthocyanin biosynthesis (Table 3). Among them, the four cytochrome P450 enzymes were annotated for flavonoid biosynthesis. HCBT can produce anthramides working on acylated anthocyanins^[30]. ACL catalyzes citrate to acetyl-CoA, and it plays important roles in the biosynthesis of flavonoids^[31]. Transmembrane ascorbate ferrereductase could be the link between iron levels and tissue ascorbate^[32]. Ascorbate influences anthocyanin accumulation during high light acclimation^[33]. The 23 genes above were all significantly upregulated in JS6-5 compared with Jingshu 6. This study provides the candidate genes involved in anthocyanin biosynthesis, most of which have not previously been reported in sweet potato.

Anthocyanin biosynthesis in plants is governed by a regulatory network that consists of the MBW complex and other transcription factors. In *Arabidopsis*, the MBW complex associated with anthocyanin biosynthesis is formed by an R2R3 MYB transcription factor (PAP1, PAP2, MYB113 or MYB114), a bHLH transcription factor (TT8, GL3 or EGL3) and the WD40-repeat protein TTG1^[8]. In addition, some members of the MADs, NAC, AP2/ERF and bZIP families have been reported to

have important roles in regulating anthocyanin biosynthesis. For example, MADs box genes have been reported to positively regulate anthocyanin accumulation in sweet potato and bilberry fruits^[17,34]. In peach, *PpNAC1* activates the expression of *PpMYB10.1* resulting in anthocyanin accumulation^[35]. *Arabidopsis* NAC transcription factor (ANAC078) acts as a positive regulator to enhance anthocyanin accumulation under high light acclimation^[36]. The ERF/AP2 transcription factor PyERF3 is found to interact with PyMYB114 and PybHLH3 to coregulate anthocyanin biosynthesis^[37]. In *Arabidopsis*, HY5 (a bZIP protein) positively regulates anthocyanin biosynthesis by transcriptional activation of the PAP1 transcription factor^[38]. In this study, 24 important transcription factors, including MYB, bHLH, MADs, NAC, AP2/ERF and bZIP were significantly upregulated in JS6-5 compared with Jingshu 6 (Table 3). Among them, a MYB transcription factor (c30141.graph_c0) was highly homologous to *IbMYB1* which was identified as a key positive regulator of anthocyanin biosynthesis^[15]. There are few reports of others likely to regulate anthocyanin biosynthesis in sweet potato and the most likely ones are shown in Fig. 8. Thus, this study provides important information for discovering and isolating the transcription factors and further illuminating the molecular mechanisms of anthocyanin biosynthesis in sweet potato.

The genome of the sweet potato is still unavailable now and molecular genetic markers are important for sweet potato breeding and genetic linkage maps. Due to their

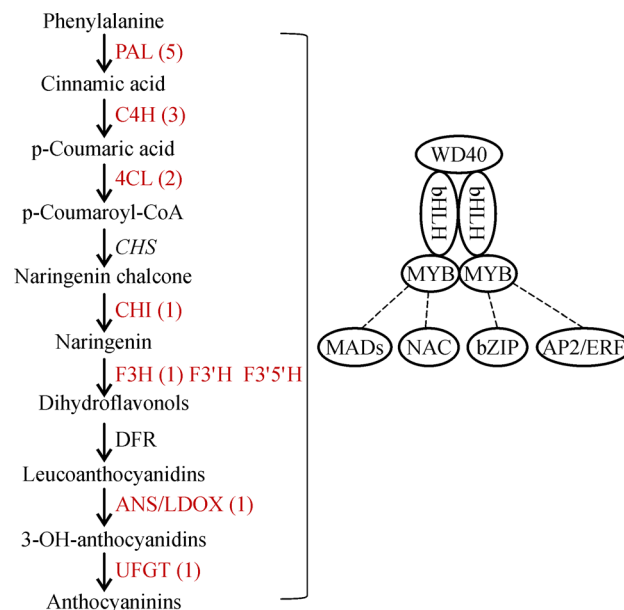


Fig. 8 Model of anthocyanin biosynthesis and the regulatory networks. The DEGs related to the enzymes (red font) are all upregulated in JS6-5 compared with Jingshu 6, and the red numbers in parentheses indicate the number of DEGs for each enzyme in the transcriptome library.

Table 3 Differentially expressed genes and transcription factors related to anthocyanin biosynthesis between Jingshu 6 and JS6-5

Gene ID	Jingshu 6 FPKM	JS6-5 FPKM	log ₂ (JS6-5/ Jingshu 6 FPKM)	Gene expres- sion level	Predicted function
Gene					
c17852.graph_c0	2.014	90.012	5.482	Up	Phenylalanine ammonia lyase
c33937.graph_c0	5.044	175.193	5.118	Up	Phenylalanine ammonia lyase
c41003.graph_c0	28.621	1152.011	5.331	Up	Phenylalanine ammonia lyase
c28448.graph_c0	2.384	101.592	5.413	Up	Phenylalanine ammonia lyase
c28750.graph_c0	5.062	98.558	4.283	Up	Phenylalanine ammonia lyase
c36515.graph_c0	34.481	970.542	4.815	Up	Cinnamate 4-hydroxylase
c31758.graph_c0	0.919	20.438	4.475	Up	Cinnamate 4-hydroxylase
c37779.graph_c0	33.881	169.078	2.319	Up	Cinnamate 4-hydroxylase
c40730.graph_c0	2.311	56.231	4.604	Up	4-Coumarate CoA ligase
c36415.graph_c0	6.598	53.089	3.008	Up	4-Coumarate CoA ligase
c32008.graph_c0	2.911	21.592	2.891	Up	Chalcone isomerase
c39514.graph_c0	3.478	137.511	5.305	Up	Flavanone 3-hydroxylase-like
c38710.graph_c0	52.111	353.301	2.761	Up	Leucoanthocyanidin dioxygenase isoform X2
c40432.graph_c0	5.626	78.999	3.812	Up	UDPglucose-flavonoid 3-O-glucosyltransferase 3
c41049.graph_c0	12.976	90.484	2.802	Up	Cytochrome P450 71A1-like
c38523.graph_c0	58.866	693.698	3.559	Up	Cytochrome P450 98A2-like
c36598.graph_c0	2.262	20.238	3.161	Up	Cytochrome P450 CYP736A12-like
c31759.graph_c0	0.318	11.874	5.224	Up	Cytochrome P450 CYP73A100
c37234.graph_c0	53.121	372.168	2.809	Up	N-hydroxycinnamoyl/benzoyltransferase
c27267.graph_c0	0.389	12.137	4.963	Up	ATP-citrate lyase
c30672.graph_c0	0.691	11.929	4.109	Up	ATP-citrate lyase
c35307.graph_c0	59.242	755.203	3.672	Up	ATP-citrate lyase
c37771.graph_c0	8.786	47.906	2.447	Up	Transmembrane ascorbate ferrireductase 1
Transcription factor					
c30141.graph_c0	1.624	39.055	4.588	Up	MYB transcription factor 1
c33320.graph_c0	4.627	20.454	2.144	Up	MYB transcription factor 4-like
c33451.graph_c0	7.535	30.975	2.039	Up	MYB transcription factor 330-like
c37043.graph_c0	5.154	30.518	2.566	Up	MYB transcription factor 48
c49061.graph_c0	1.692	30.437	4.169	Up	MYB transcription factor 21
c35393.graph_c0	2.126	46.986	4.466	Up	bHLH transcription factor 3-like
c34197.graph_c0	10.011	43.136	2.107	Up	bHLH transcription factor 130-like
c34229.graph_c0	1.132	11.398	3.331	Up	bHLH transcription factor14-like
c49373.graph_c0	1.452	17.871	3.622	Up	Mads-box transcription factor 27-like
c54533.graph_c0	4.048	24.894	2.620	Up	Mads-box transcription factor soc1-like
c38071.graph_c0	18.392	215.323	3.551	Up	NAC transcription factor 1
c28958.graph_c0	0.897	28.898	5.009	Up	NAC transcription factor 29-like
c30930.graph_c0	1.536	10.038	2.708	Up	NAC transcription factor 11
c31917.graph_c0	1.888	15.299	3.019	Up	NAC transcription factor 21
c28862.graph_c0	0.353	17.243	5.611	Up	NAC transcription factor 90
c32530.graph_c0	1.145	20.928	4.192	Up	Ethylene-responsive transcription factor 114-like
c29651.graph_c0	24.663	115.511	2.228	Up	Ethylene-responsive transcription factor 2
c34644.graph_c0	59.854	295.081	2.302	Up	Ethylene-responsive transcription factor ABR1-like

(Continued)

Gene ID	Jingshu 6 FPKM	JS6-5 FPKM	log ₂ (JS6-5/Jingshu 6 FPKM)	Gene expression level	Predicted function
c35202.graph_c0	4.867	28.852	2.567	Up	Ethylene-responsive transcription factor
c31579.graph_c0	6.798	76.979	3.501	Up	Ethylene-responsive transcription factor 4
c32097.graph_c0	0.713	22.035	4.950	Up	bZIP transcription factor 43-like
c50431.graph_c0	0.128	7.749	5.920	Up	bZIP transcription factor
c11820.graph_c0	1.825	11.102	2.605	Up	bZIP transcription factor
c32899.graph_c0	5.774	42.997	2.897	Up	bZIP transcription factor 114

functionality, abundance, high polymorphism and excellent reproducibility, SSRs are useful resources for genome analysis^[39] and in this study, a total of 2349 potential SSRs were identified from 7547 unigenes.

5 Conclusions

A total of 35592 unigenes were obtained from the purple-fleshed sweet potato using high-throughput sequencing technology. There were 1566 DEGs between Jingshu 6 and JS6-5. Among them, 23 differentially expressed genes and 24 important transcription factors might be involved in anthocyanin biosynthesis of sweet potato. In addition, 2349 SSRs were identified. This study not only provides the candidate genes but also provides insights into the molecular mechanism of anthocyanin biosynthesis in sweet potato.

Supplementary materials The online version of this article at <https://doi.org/10.15302/J-FASE-2018219> contains supplementary materials (Tables S1–S2; Figs. S1–S2).

Acknowledgements This work was supported by National Natural Science Foundation of China (31271777; 31461143017), China Agriculture Research System (CARS-11), and Science and Technology Planning Project of Guangdong Province, China (2015B020202008).

Compliance with ethics guidelines Hongyuan Zhao, Shanshan Zhang, Feibing Wang, Ning Zhao, Shaozhen He, Qingchang Liu, and Hong Zhai declare that they have no conflicts of interest or financial conflicts to disclose.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Loebenstein G, Thottappilly G. The Sweetpotato. 11nd ed. Berlin: Springer Netherlands, 2009, 189–234
- Holton T A, Cornish E C. Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell*, 1995, **7**(7): 1071–1083
- Wang L, Zhao Y, Zhou Q, Luo C L, Deng A P, Zhang Z C, Zhang J L. Characterization and hepatoprotective activity of anthocyanins from purple sweet potato (*Ipomoea batatas* L. cultivar Eshu No. 8). *Journal of Food and Drug Analysis*, 2017, **25**(3): 607–618
- Yoshimoto M, Okuno S, Yoshinaga M, Yamakawa O, Yamaguchi M, Yamada J. Antimutagenicity of sweetpotato (*Ipomoea batatas*) roots. *Nippon Nogeikagaku Kaishi*, 1999, **63**(3): 537–541
- Matsui T, Ebuchi S, Kobayashi M, Fukui K, Sugita K, Terahara N, Matsumoto K. Anti-hyperglycemic effect of diacylated anthocyanin derived from *Ipomoea batatas* cultivar Ayamurasaki can be achieved through the alpha-glucosidase inhibitory action. *Journal of Agricultural and Food Chemistry*, 2002, **50**(25): 7244–7248
- Johnson M, Pace R D. Sweet potato leaves: properties and synergistic interactions that promote health and prevent disease. *Nutrition Reviews*, 2010, **68**(10): 604–615
- Lattanzio V, Cardinali A, Linsalata V. Recent Advances in Polyphenol Research, Volume 3. 6nd ed. New Jersey: Wiley-Blackwell, 2012, 143–160
- Borevitz J O, Xia Y, Blount J, Dixon R A, Lamb C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell*, 2000, **12**(12): 2383–2394
- Gonzalez A, Zhao M, Leavitt J M, Lloyd A M. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant Journal*, 2008, **53**(5): 814–827
- Winkel-Shirley B. Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. *Physiologia Plantarum*, 2002, **107**(1): 142–149
- Guo J, Zhou W, Lu Z, Li H, Li H, Gao F. Isolation and functional analysis of chalcone isomerase gene from purple-fleshed sweet potato. *Plant Molecular Biology Reporter*, 2015, **33**(5): 1451–1463
- Zhou W, Gong Y, Lu X, Huang C, Gao F. Molecular cloning and characterization of a flavonoid 3'-hydroxylase gene from purple-fleshed sweet potato (*Ipomoea batatas*). *Molecular Biology Reports*, 2012, **39**(1): 295–302
- Wang H, Fan W, Li H, Yang J, Huang J, Zhang P. Functional characterization of dihydroflavonol-4-reductase in anthocyanin biosynthesis of purple sweet potato underlies the direct evidence of anthocyanins function against abiotic stresses. *PLoS One*, 2013, **8** (11): e78484
- Zhou W, Huang C, Gong Y, Feng Q, Gao F. Molecular cloning and expression analysis of an *ANS* gene encoding anthocyanidin synthase from purple-fleshed sweet potato (*Ipomoea batatas* (L.) Lam). *Molecular Biology Reports*, 2009, **37**(5): 2525–2531
- Tanaka M, Takahata Y, Kurata R, Nakayama H, Yoshinaga M.

- Structural and functional characterization of *IbMYB1* genes in recent Japanese purple-fleshed sweetpotato cultivars. *Molecular Breeding*, 2012, **29**(3): 565–574
16. Dong W, Niu L, Gu J, Gao F. Isolation of a WD40-repeat gene regulating anthocyanin biosynthesis in storage roots of purple-fleshed sweet potato. *Acta Physiologiae Plantarum*, 2014, **36**(5): 1123–1132
 17. Lalusin A G, Nishita K, Kim S H, Ohta M, Fujimura T. A new MADS-box gene (*IbMADS10*) from sweet potato (*Ipomoea batatas* (L.) Lam) is involved in the accumulation of anthocyanin. *Molecular Genetics and Genomics*, 2006, **275**(1): 44–54
 18. Tao X, Gu Y H, Wang H Y, Zheng W, Li X, Zhao C W, Zhang Y Z. Digital gene expression analysis based on integrated *de novo* transcriptome assembly of sweet potato [*Ipomoea batatas* (L.) Lam]. *PLoS One*, 2012, **7**(4): e36234
 19. Tao X, Gu Y H, Jiang Y S, Zhang Y Z, Wang H Y. Transcriptome analysis to identify putative floral-specific genes and flowering regulatory-related genes of sweet potato. *Bioscience, Biotechnology, and Biochemistry*, 2013, **77**(11): 2169–2174
 20. Wang Z, Fang B, Chen X, Liao M, Chen J, Zhang X, Huang L, Luo Z, Yao Z, Li Y. Temporal patterns of gene expression associated with tuberous root formation and development in sweetpotato (*Ipomoea batatas*). *BMC Plant Biology*, 2015, **15**(1): 180–192
 21. Qin Z, Li A, Hou F, Wang Q, Dong S, Zhang L. Gene identification using RNA-seq in two sweetpotato genotypes and the use of mining to analyze carotenoid biosynthesis. *South African Journal of Botany*, 2017, **109**: 189–195
 22. Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, Chen X, Li Y. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics*, 2010, **11**(1): 726–739
 23. Kou M, Xu J L, Li Q, Liu Y J, Wang X, Tang W, Yan H, Zhang Y G, Ma D F. Development of SNP markers using RNA-seq technology and tetra-primer ARMS-PCR in sweetpotato. *Journal of Integrative Agriculture*, 2017, **16**(2): 464–470
 24. Xie F, Burklew C E, Yang Y, Liu M, Xiao P, Zhang B, Qiu D. De novo sequencing and a comprehensive analysis of purple sweet potato (*Ipomoea batatas* L.) transcriptome. *Planta*, 2012, **236**(1): 101–113
 25. Ma P, Bian X, Jia Z, Guo X, Xie Y. De novo sequencing and comprehensive analysis of the mutant transcriptome from purple sweet potato (*Ipomoea batatas* L.). *Gene*, 2016, **575**(2): 641–649
 26. Grabherr M G, Haas B J, Yassour M, Levin J Z, Thompson D A, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceci E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren B W, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 2011, **29**(7): 644–652
 27. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research*, 2006, **34**(Web Server issue): W293–W297
 28. Conesa A, Götz S, García-Gómez J M, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 2005, **21**(18): 3674–3676
 29. Schmittgen T D, Livak K J. Analyzing real-time PCR data by the comparative C_T method. *Nature Protocols*, 2008, **3**(6): 1101–1108
 30. Yang Q, Reinhard K, Schiltz E, Matern U. Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA: anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L. *Plant Molecular Biology*, 1997, **35**(6): 777–789
 31. Hu X M, Shi C Y, Liu X, Jin L F, Liu Y Z, Peng S A. Genome-wide identification of citrus ATP-citrate lyase genes and their transcript analysis in fruits reveals their possible role in citrate utilization. *Molecular Genetics and Genomics*, 2015, **290**(1): 29–38
 32. Su D, Asard H. Three mammalian cytochromes *b561* are ascorbate-dependent ferrireductases. *FEBS Journal*, 2006, **273**(16): 3722–3734
 33. Page M, Sultana N, Paszkiewicz K, Florance H, Smirnoff N. The influence of ascorbate on anthocyanin accumulation during high light acclimation in *Arabidopsis thaliana*: further evidence for redox control of anthocyanin synthesis. *Plant, Cell & Environment*, 2012, **35**(2): 388–404
 34. Jaakola L, Poole M, Jones M O, Kämäräinen-Karppinen T, Koskimäki J J, Hohtola A, Häggman H, Fraser P D, Manning K, King G J, Thomson H, Seymour G B. A SQUAMOSA MADS box gene involved in the regulation of anthocyanin accumulation in bilberry fruits. *Plant Physiology*, 2010, **153**(4): 1619–1629
 35. Zhou H, Lin-Wang K, Wang H, Gu C, Dare A P, Espley R V, He H, Allan A C, Han Y. Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *Plant Journal*, 2015, **82**(1): 105–121
 36. Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, Shigeoka S. Arabidopsis NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant & Cell Physiology*, 2009, **50**(12): 2210–2222
 37. Yao G, Ming M, Allan A C, Gu C, Li L, Wu X, Wang R, Chang Y, Qi K, Zhang S, Wu J. Map-based cloning of the pear gene *MYB114* identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. *Plant Journal*, 2017, **92**(3): 437–451
 38. Shin D H, Choi M, Kim K, Bang G, Cho M, Choi S B, Choi G, Park Y I. HY5 regulates anthocyanin biosynthesis by inducing the transcriptional activation of the MYB75/PAP1 transcription factor in *Arabidopsis*. *FEBS Letters*, 2013, **587**(10): 1543–1547
 39. Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, 1996, **2**(3): 225–238