Molecular Plant Breeding 2021, Vol.12, No.6, 1-14
http://genbreedpublisher.com/index.php/mpb

Research Article **Open Access**

Differential Expression Analysis of Genes Related to Flesh Color in *Hylocereu polyrhizus* and *Hylocereu undatus*

Guo Panyang, Zeng Canbin, Liu Chengli, Wei Shuangshuang, Huang Jiaquan, Tang Hua
Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, College of Tropical Crops, Hainan University, Haikou, 570228, China
Corresponding author Email: thtiger@163.com
Molecular Plant Breeding, 2021, Vol.12, No.6 doi: 10.5376/mpb.2021.12.0006
Received: 01 Dec., 2020
Accepted: 31 Jan., 2021
Published: 19 Feb., 2021
Copyright © 2021 Guo et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Preferred citation for this article:
Guo P.Y., Zeng C.B., Liu C.L., Wei S.S., Huang J.Q., and Tang H., 2021, Differential expression analysis of genes related to flesh color in *Hylocereu polyrhizus* and *Hylocereu undatus*, Molecular Plant Breeding, 12(6): 1-14 (doi: 10.5376/mpb.2021.12.0006)

Abstract Pitaya is a burgeoningly tropical fruit, there are two common varieties, red peel with red pulp and red peel with white pulp. To investigate the reason of different colors in two kinds of pitaya, explore differencially expressed genes (DEGs), in this study, we profiled transcriptome in flesh of two varieties (*Hylocereu polyrhizus* and *Hylocereu undatus*) in green stage and mature stage respectively. 53 240 reads with high quality were obtained. Analysing the DEGs, we found there were bigger differences in different stage rather than different variety. Gene ontology (GO) functional analysis and KEGG pathway analysis were adopted in R1 VS R3 and W3 VS W3. Through GO analysis, we found many genes enriched in protein binding transcription factor activity in molecule founction. Through KEGG pathway analysis, we found most genes were enriched in biosynthesis of secondary metabolism, amino acid and nucleic acid metabolism, and tyrosine metabolism. In addintion, tyrosine was the precursor of betalaines. In conclusion, the development of *Hylocereu polyrhizus* accompanying by the synthesis of a large amount of betalaines, tyrosine played a key role as a precursor, and the synthesis of this pigment required a mass of amino acids, enzymes, and transcription factors. Betalaines synthesis pathway has not cleared yet, this study explored key genes related to betalaines synthesis and provided useful information for optimizing the betalaines synthesis pathway, which was the basis of later experiments.
Keywords Flesh color development of pitaya; RNA-Seq; Differentially expressed genes; Betalaines

Pitaya is a plant belonging to *Cactaceae Hylocereus*. It originated from Central America, such as Costa Rica, Mexico and Colombia (Shen et al., 2015). There is large-scale cultivation in Hainan, Guangxi, Guangdong, Fujian, Yunnan, Guizhou and Taiwan in China. The fruit is rich in betalanins, vegetable albumin, Vc and dietary fiber, which integrates edible, ornamental and medicinal values, and is very popular all over the world.

At present, there are 3 varieties are mainly planted, red flesh with red skin, white flesh with red skin and white flesh with yellow skin, of which the second is the most widespread, and the outstanding varieties are “Dahong” and “Jindu No.1”. “Dahong” was bred by Chen Yongchi and Cai Dongxun from Taiwan, China. It tastes sweet and delicate, and is a variety of self-pollination (Liu et al., 2014); Jindu No.1 is bred by Wang Jindu of Taiwan, with large fruit, high sweetness and long fruit period, and is also a self-pollinating variety (Liu et al., 2017). In 2019, the total planting area of dragon fruit in China has exceeded 66 666.67hm2, with an output value of more than 10 billion, which has high economic value. In the process of cultivation, pitaya will produce abundant variations through artificial selection and natural mutation, such as pink flesh with red skin, red flesh with green skin, pink-white flesh with red skin and so on. Different variations of character are discovered due to the differences of pigment content. Therefore, pigment content is an indispensable and important evaluation method in the breeding of new varieties.

Betalanins is a kind of water-soluble natural plant pigment, which can show different colors such as red, purple and yellow in plants under different pH. It is an important component of dragon fruit color (Lee et al., 2014; Clifford et al., 2015; Martinez et al., 2015; Hua et al., 2016). Some studies have shown that in the rhizome of sugar beet, the synthesis of betalanins requires four enzymes: tyrosinase, 4-ray-5-ray DOPA dioxygenase, cytochrome P450 and glucosyltransferase (Hua et al., 2016). However, the synthesis mechanism of beet pigment in pitaya is not clear. At

present, gas chromatography-mass spectrometry (GC-MS) technology has been used to study the metabolic components of the peel and pulp of red flesh pitaya. It is found that *Hylocereu polyrhizus* will accumulate a large amount of betalanins (Wu et al., 2019) during mature period. Betalanins is a kind of water-soluble pigment, which has two forms: betaine (purplish red) and betaxanthin (orange) (Duan et al., 2017). Both of them are synthesized from tyrosine, which is a precursor. Betalanins has good effects in health, such as antioxidation, anticancer, scavenging free radicals and reducing blood lipids, but it will easily degrade under the conditions of light, heat, oxygen and metal ions (Fe^{3+}, Cu^{2+}, Al^{3+}). When pH is 4-5, it is the most stable, and exists in many plants such as amaranth, Chenopodiaceae, Jasminaceae and Phytolaccaceae (Chen et al., 2013). Betalanins is composed of beet aldehyde acid and imine of cyclodopa, in which the conjugated double bond system forms its chromophoric group. The conjugated double bond of cyclic dopa in aromatic group makes the absorbance of betaine 50~70nm higher than that of beet flavin. The difference of pigment content may be an important factor leading to the difference among varieties (Wu et al., 2019). The biosynthesis of this natural plant pigment is more complex. At present, the studies on its physiological activity at home and abroad have been more thorough, but the metabolic pathway and molecular mechanism are still not clear. For many years, academia has speculated that glucosyltransferase (GT) and tyrosinase (TYR) are involved in the metabolic pathway of betaine (Wyler et al., 1984; Gandía and García, 2013; Sakuta, 2014), but it has not been strongly verified. In the squama of dragon fruit, vigs (virus-mediated gene silencing) technique was used to make the breaking stop and change color to yellow by silencing *HmB5GT* or *HmHCGT2*. After detecting the content of betaine and betaxanthin respectively, it was found that they all decreased, among which the content of betaine decreased more significantly (Xie et al., 2020).

Transcriptome sequencing technology can determine the amount of all mRNA transcribed by a certain tissue or cell of a species in a specific period, which can be used as a basis to determine the expression of some genes in this period, which can provide an important reference for gene transcription in a certain state. At present, it has been widely used in the field of molecular biology, and it is a convenient and fast technical method (Wilhelm and Landry, 2009). In this experiment, *Hylocereu polyrhizus* (R) and *Hylocereu undatus* (W) were used as research materials. Through transcriptome sequencing and follow-up analysis, the DEGs in the pulp of two different types of pitaya fruit was compared, and the related gene function and metabolic pathway enrichment analysis were carried out. The DEGs were screened out, and the metabolic differences of pigments in red and white pitaya fruits were preliminarily analyzed. Through the functional verification experiment of the screened genes, we can provide reliable information for the synthesis and metabolism of sugar betalanins.

1 Results and Analysis

1.1 Results of assembly of transcriptome data

The transcriptome sequencing data of *Hylocereu polyrhizus* (R) and *Hylocereu undatus* (W) in green stage and mature stage were processed respectively. After the Raw Reads filtered and the low-quality Reads removed, W1, W3, R1, R3 got 57 575 196, 55 265 234, 59 436 460, 43 087 310 Clean Reads, Q30 was used to assemble reference sequences. The assembly ratios of each sample were 91.13%, 91.52%, 91.81% and 93.42%, respectively. The final high-quality sequences assembled in the four samples were 5 914 708, 13 946 140, 14 364 654 and 5 296 884, respectively, and the assembly ratios of each sample were 91.13%, 91.52%, 91.81% and 93.42%, the final high-quality sequences were 5 914 708, 13 946 140, 14 364 654 and 5 296 884, respectively. The transcriptome data were spliced by Trinity software (Table 1). The N50 was 1 429 and the average length was 84.092 bp. Among them, 35 618 (66.90%) were assembled in W1, 45 142 (84.79%) in W3, 41 265 (77.51%) in R1 and 38 588 (72.48%) in R3 (Table 2). Through the statistical analysis of the length of the assembled sequence (Figure 1), excluding the head and tail, the length distribution of the sequence is uniform, which shows that the assembly result is reliable.

Table 1 Analysis of general assembly

Total Unigenes	All samples genes num	GC percentage (%)	N50	Max length	Min length	Average length
53 240	48 253	43.42	1 429	15 381	228	840.92

Table 2 Overview of the sequencing and assembly

Item	W1	W3	R1	R3
Total raw reads	59 276 174	56 792 504	60 929 858	44 011 092
Adaptor percentage (%)	0.01	0.03	0.02	0.02
Low Quality percentage (%)	2.85	2.66	2.43	2.08
Total Clean Reads	57 575 196	55 265 234	59 436 460	43 087 310
Total raw length (bp)	7 409 521 750	7 099 063 000	7 616 232 250	5 501 386 500
Total clean length (bp)	7 196 899 500	6 908 154 250	7 429 557 500	5 385 913 750
Q30 percentage (%)	91.13	91.52	91.81	93.42
N percentage (%)	0.00	0.00	0.00	0.00
GC percentage (%)	51.16	50.16	50.60	50.65
Unmapped Reads	7 454 416	19 091 220	19 578 556	9 564 356
Unique Mapped Reads	5 447 531	13 131 590	12 761 953	4 811 709
Multiple Mapped Reads	467 177	814 550	1 602 701	485 175
Total Mapped Reads	5 914 708	1 3946 140	14 364 654	5 296 884
Total unigenes	35 618	45 142	41 265	38 588
Percentage of total Unigenes (%)	66.90	84.79	77.51	72.48

Note: W1: White meat and fruit stage; W3: White meat maturity stage; R1: Red meat and fruit stage; R3: Red meat maturation stage

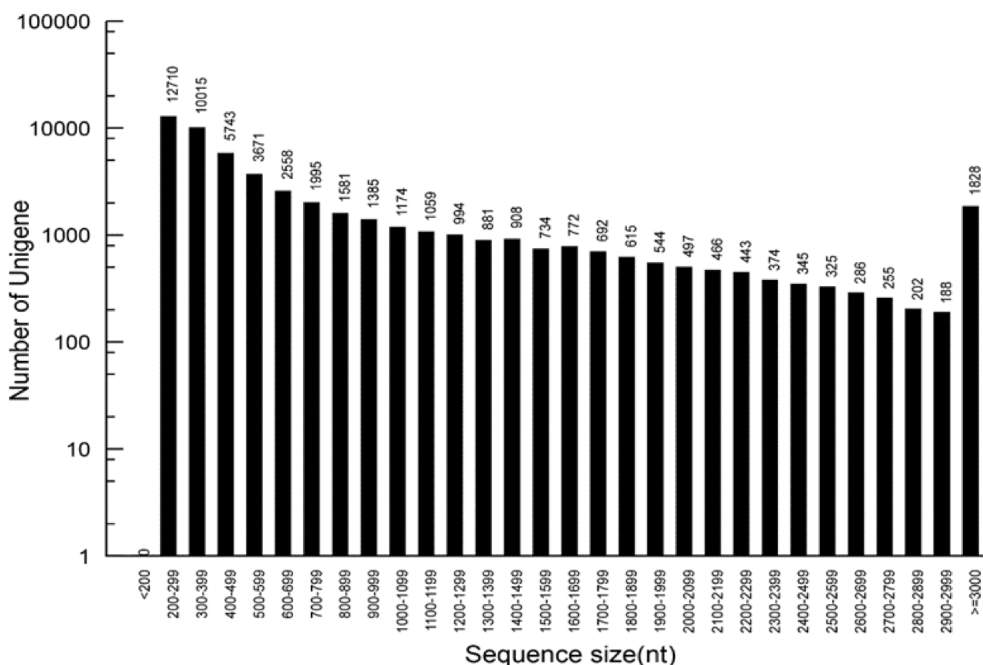

Figure 1 Length distribution of four samples' all Unigenes

1.2 Results of gene function annotation

53 240 gene sequences were BLAST in Nr, Swiss-Prot, KEGG and KOG databases respectively (Figure 2). A total of 27 907 sequences were annotated, including 27 725 in Nr database, 19 279 in Swissprot database, 16 942 in KOG database and 7 394 in KEGG database. There are 6 065 sequences annotated in four major databases at the same time. Species comparison analysis was carried out in the Nr database (Figure 3). A total of 389 species were compared, with the top ten being *Beta vulgaris subsp* (13 410), *Theobroma cacao* (1476), *Vitis vinifera* (1 193), *Gossypium arboreum* (872), *Brassica napus* (853), *Medicago truncatula* (774), *Arabidopsis thaliana* (440), *Jatropha curcas* (438), *Nelumbo nucifera* (421), *Nicotiana tomentosiformis* (383). It can be concluded that pitaya and *Beta vulgaris subsp* have high homology in gene sequence. Based on the comparative analysis of gene homologues in the

KOG database (Figure 4), 6 498 genes were found to be associated with general functional prediction and 3 224 genes associated with post-translational modification, protein transport and molecular chaperone. The top 10 categories showed that these proteins were mainly related to gene transcription, translation, protein processing and intracellular transport, and 705 genes were related to secondary metabolites.

GO (Geneontology) functional annotation analysis (Figure 5) enriched genes in biological processes, cellular components and molecular functions in all samples, in which more genes were enriched in cellular components and less in molecular functions. In the biological process, 7 892 genes are enriched in the metabolic process, 7 637 genes in the intercellular process, 5 774 genes in the single organism process, and 2 165 genes in biological regulation and response to stimulus. These processes are closely related to the metabolism of pigment. Among the cellular component, it can be seen that cells or part of cells, organelles or part of organelles, and membranes or partial membranes are enriched into more genes; in molecular function, 6 496 genes are enriched in catalytic activity, 6 065 genes are enriched in binding function, and 6 065 genes are enriched in transport activity. In addition, some genes are enriched in the binding activities of proteins and transcription factors.

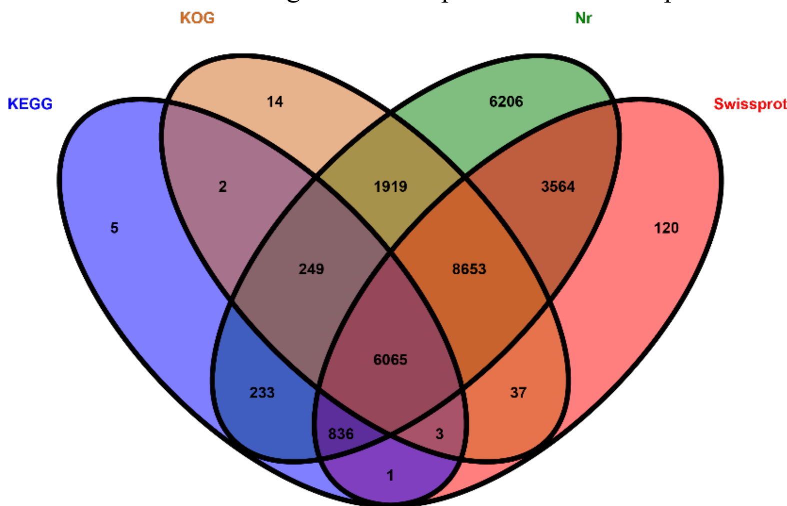

Figure 2 Venn diagram of Unigenes annotated in four databases

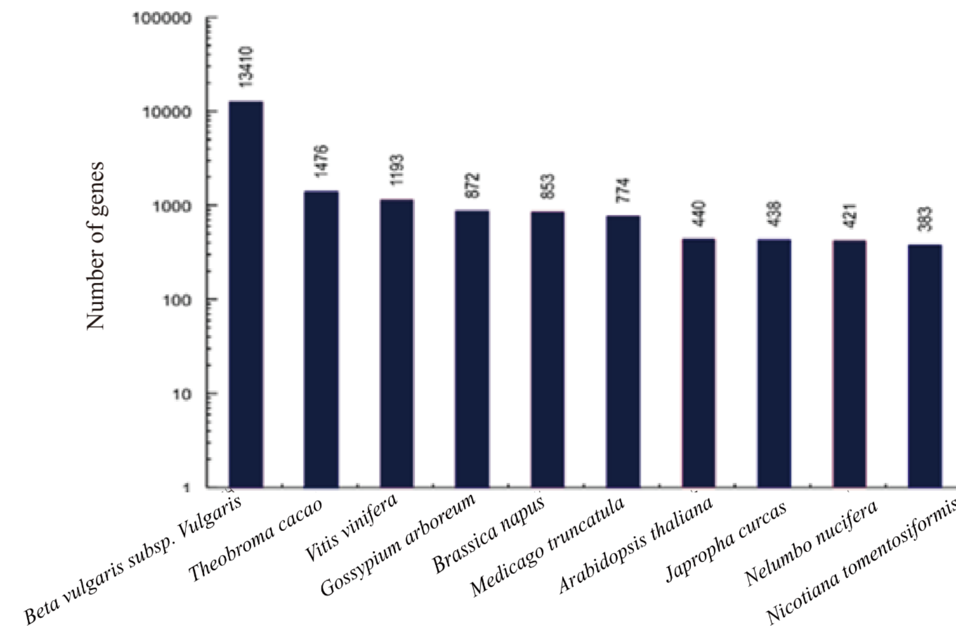

Figure 3 Species comparison in Nr database of Unigenes

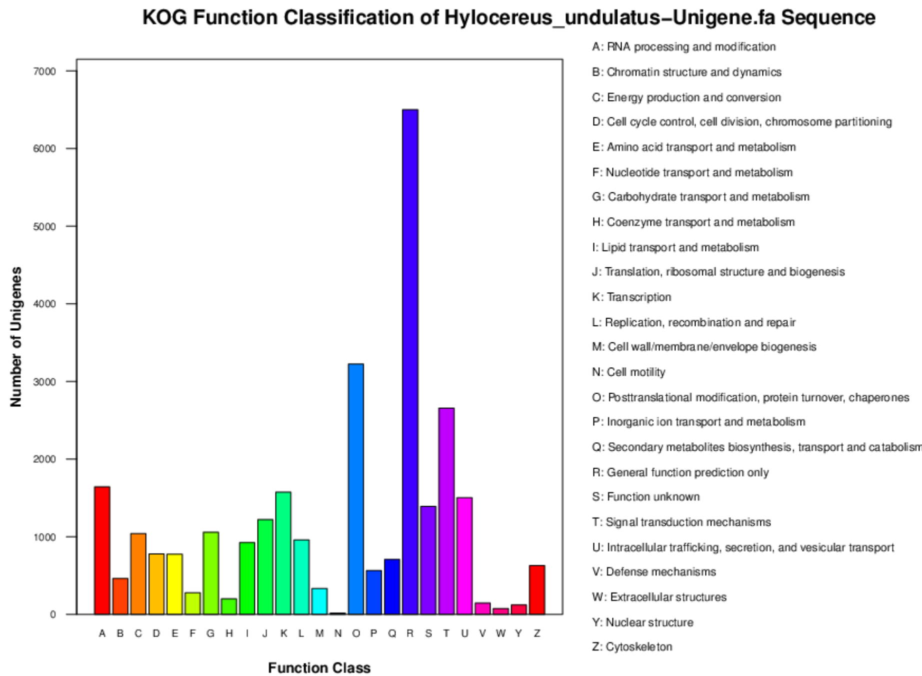

Figure 4 KOG function classification of Unigenes

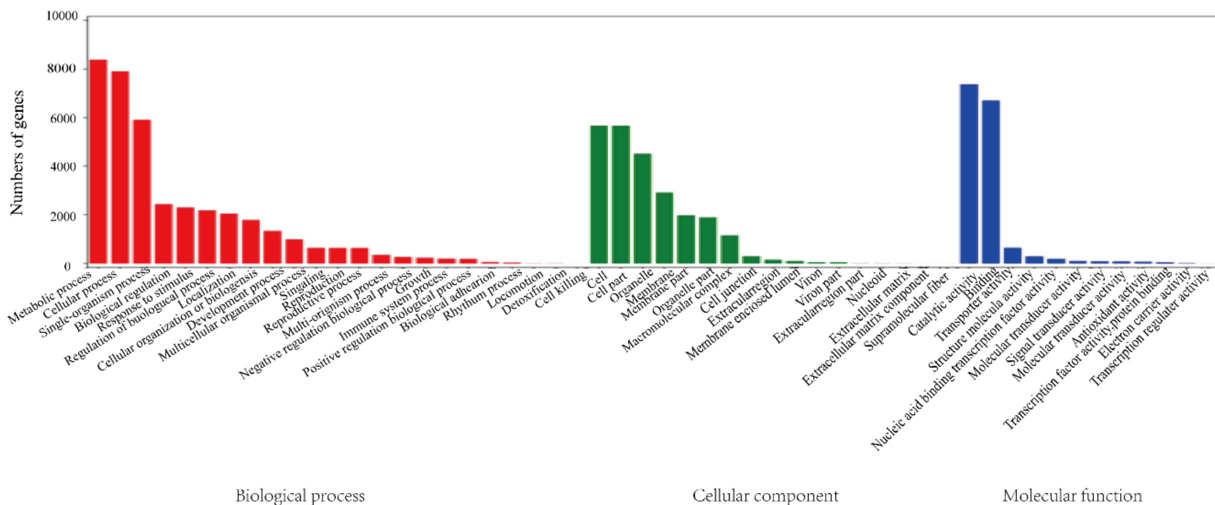

Figure 5 GO analysis of Unigenes

1.3 Results of differentially expressed genes (DEGs) among four samples

35 618, 45 142, 41 265 and 38 588 genes were assembled from W1, W3, R1 and R3 respectively (Table 2). According to the criteria of FDR < 0.05 and | log2FC | > 1, pairwise comparison was used to filter among the four samples (Table 3). It can be seen that the DEGs of the same variety in different periods are more than those of different varieties in the same period.

Table 3 Statistics of differentially expressed genes between four groups

Item	W1 VS W3	R1 VS R3	W1 VS R1	W3 VS R3
FDR<0.05, $\|log_2FC\|$>1	11 131	10 077	5 090	7 067
Up	8 653	6 651	2 425	2 714
Down	2 478	3 426	2 665	4 353

In R1 VS R3, a total of 10 077 genes were significantly expressed, of which 6 651 were up-regulated and 3 426 down-regulated (Table 3). GO (GeneOntology) analysis of DEGs (Figure 6) showed that four parts were obviously enriched to more genes in molecular function, catalytic activity, connection, transport activity, nucleic acid binding activity with transcription factor, which had an important relationship with pigment metabolism. In the KEGG metabolic pathway analysis (Figure 7), more genes were enriched in the secondary metabolite synthesis pathway, amino acid and nucleic acid metabolic pathway and tyrosine metabolic pathway.

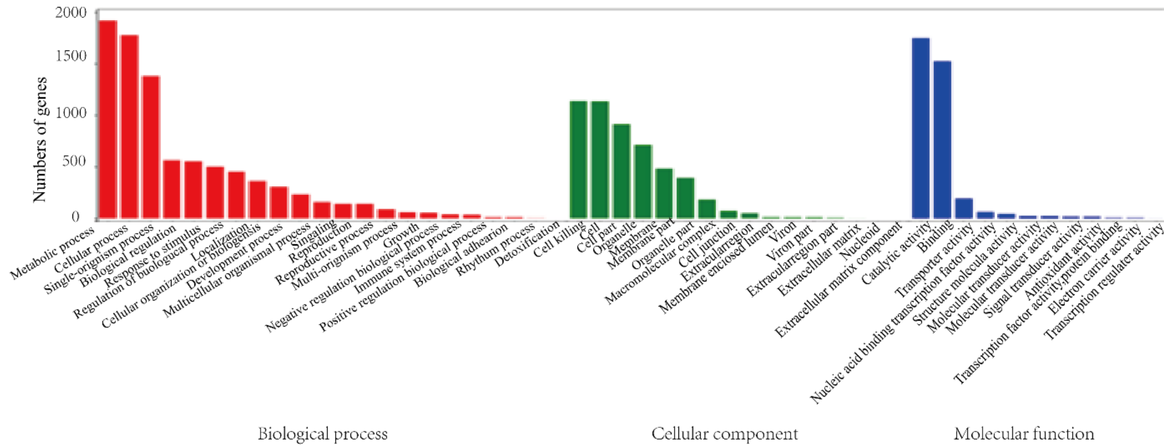

Figure 6 GO analysis of DEGs in R1 VS R3

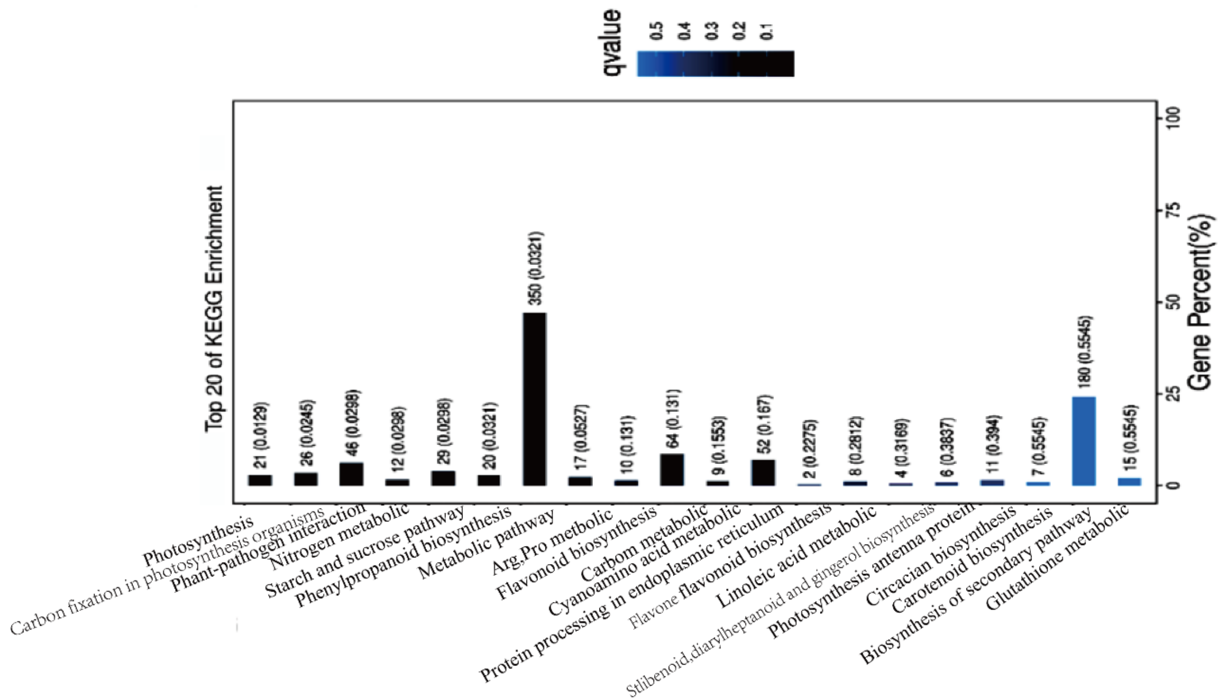

Figure 7 KEGG pathway analysis of DEGs in R1 VS R3

In the comparative analysis of W3 VS R3, we got 7 067 DEGs, of which 2 714 were up-regulated and 4 353 down-regulated (Table 3). Based on the progressive GO enrichment analysis of all DEGs (Figure 8), we found that the up and down regulated genes were basically enriched in the same functional region, similar to R1 VS R3 in category, and enriched in catalysis, transport, binding, transcription factor activity, and protein binding. Three up-regulated genes were significantly enriched in antioxidant function (Unigene 0 022 428; Unigene 0 023 466; Unigene 0 037 666), and no down-regulated genes were enriched in this molecular function. Further enrichment analysis of KEGG metabolic pathways was carried out on 7 067 DEGs (Figure 9). It was found that in addition to

metabolic pathways and secondary metabolic pathways, some genes were enriched in phenylalanine metabolic pathways, flavonoid and flavonol metabolic pathways.

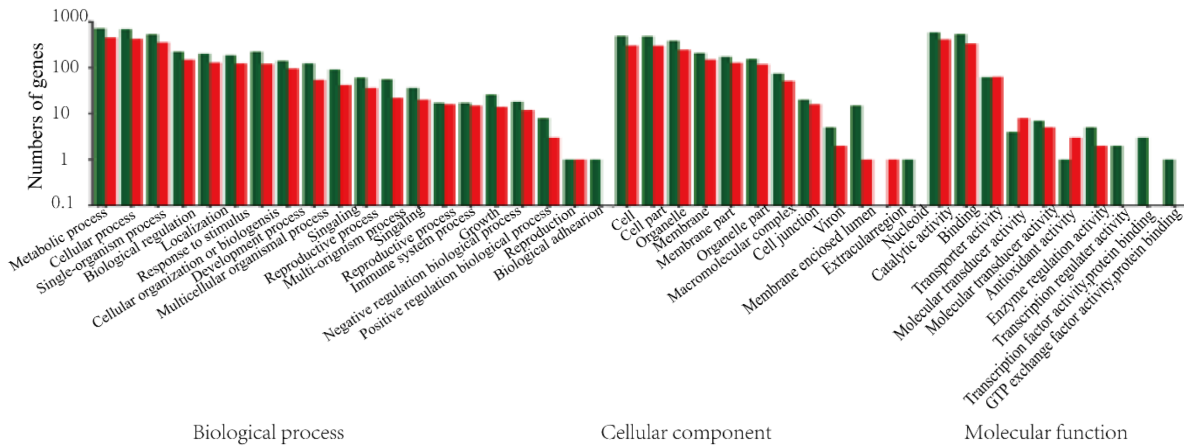

Figure 8 GO analysis of Up and Down DEGs in W3 VS R3

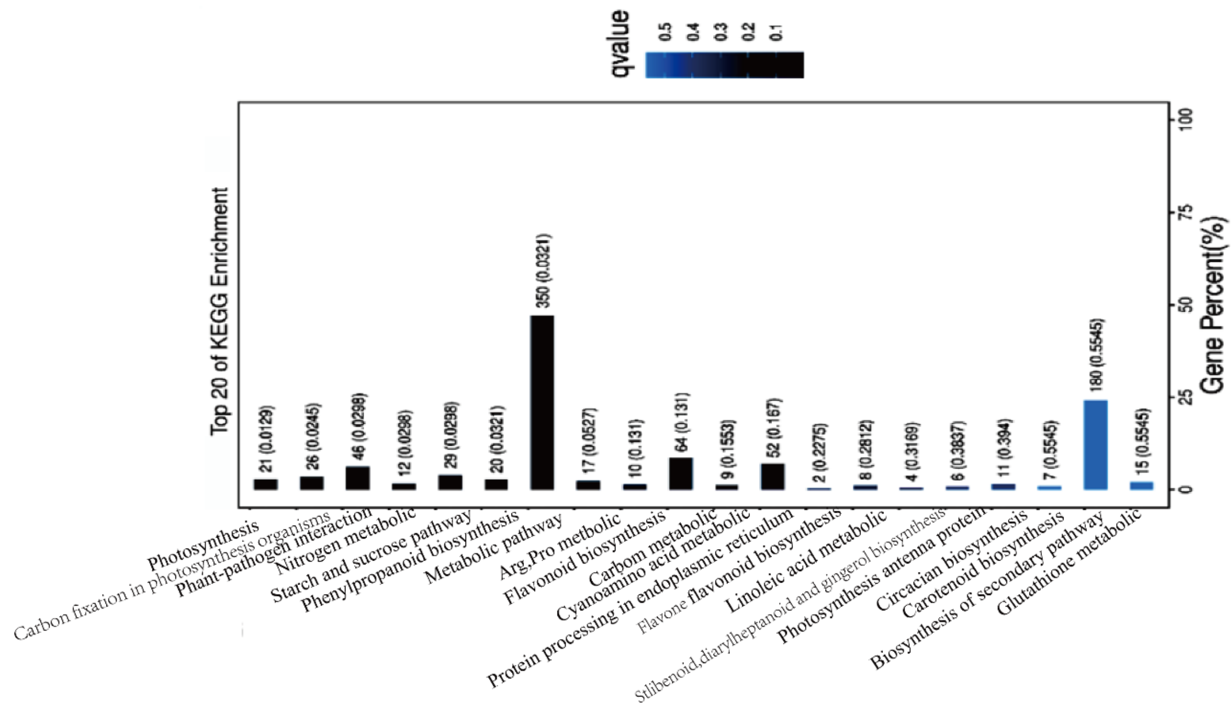

Figure 9 KEGG analysis of Up and Down DEGs in W3 VS R3

1.4 Verification of transfer group data by qRT-PCR

Based on the analysis of KEGG metabolic pathway, this study selected the key enzyme genes involved in betalanins metabolic pathway or upstream pathway. Genes' ID were used to obtain the sequence data of related annotated genes. Through the KEGG database, the key enzyme genes and other key genes in the related metabolic pathway were screened out. With the RPKM as the reference, 18 transcripts were randomly selected for qRT-PCR verification (Table 4), with UBQ as the internal reference gene (Table 5). The results showed that the expression patterns of the 18 transcripts were consistent with the RNA-Seq data (Figure 10). These results show that the transcriptome data of this study are reliable and can be studied further.

Table 4 Genes selected for qRT-PCR verification

Unigene	W3	R3	Log2 fold change (R3/W3)	Up/down regulation	Annotation
Unigene0017646	2.57	4.23	0.73	Up	Tryptophan/tyrosine permease isoform 1
Unigene0051695	0.001	82.34	16.33	Up	Full=4,5-DOPA dioxygenase extradio
Unigene0003029	2.22	21.50	3.28	Up	Glucosyltransferase
Unigene0004253	60.45	216.99	1.84	Up	Cytochrome P450 CYP72A219-like
Unigene0006446	0.61	555.88	9.83	Up	Cytochrome P450 76AD1-like protein, partial
Unigene0023505	28.53	63.80	1.16	Up	Prephenate aminotransferase
Unigene0028648	9.86	63.01	2.68	Up	Phenylalanine-4-hydroxylase
Unigene0039772	13.06	112.66	3.11	Up	Arogenate dehydrogenase 2, chloroplastic-like
Unigene0025394	4.57	8.84	0.96	Up	Probable chalcone--flavonone isomerase 3
Unigene0003132	16.67	15.60	-0.09	Down	Cyclo-DOPA 5-O-glucosyltransferase
Unigene0012439	6.99	5.14	-0.43	Down	Cyclo-DOPA 5-O-glucosyltransferase
Unigene0047695	5.39	5.27	-0.03	Down	Aromatic-L-amino-acid decarboxylase-like
Unigene0037073	11.75	6.03	-0.97	Down	Glucosyltransferase
Unigene0019344	18.20	17.87	-0.03	Down	Flavanone-3-hydroxylase
Unigene0031027	16.25	3.40	-2.26	Down	Glucosyltransferase
Unigene0015977	15.17	7.42	-1.04	Down	Anthocyanidin 3-O-glucosyltransferase 7-like
Unigene0042389	3.835 4	0.697 9	-2.46	Down	Leucoanthocyanidin reductase-like
Unigene0048533	6.963 1	0.001	-12.77	Down	Ethylene-responsive transcription factor ERF061

2 Disscussion

Pulp's conversion of color is a complex process, which requires the regulation of various genes, enzymes and transcription factors (Hatlestad et al., 2015), as well as light, temperature and other natural factors. As the key substance, betalanins is synthesized from tyrosine (Duan et al., 2017). Tyrosine forms levodopa (L-Dopa) under the catalysis of tyrosinase (TYR) and other enzymes. On the one hand, L-dopa forms betaine (Betalamic acid), under the catalysis of 4, 5-DOPA dioxygenase (DOD) and other enzymes, in which the conjugated double bond forms the chromogenic group (Polturak et al., 2018). On the other hand, cyclodopa (Cyclo-Dopa) is formed under the catalysis of cytochrome P450 (CytP450). Cyclodopa spontaneously forms betaine ligands (betanidin), and betaine ligands to form betaine (Betanin) (Polturak et al., 2018 under the catalysis of GT et al. Wu et al., 2019), stored in vacuoles (Chen et al., 2017) (Figure 11).

According to the quantitative results, this study found that the transcriptional Unigene 0017646 of tyrosine permease isomer 1 related to tyrosine transport was up regulated in R3, but not significantly. This expression pattern is similar to the transcriptional Unigene 0051695. The L-Dopa catalyzed by Unigene 0051695 encoding DOD, TYR is the substrate (Hua et al., 2016) of DOD. Unigene0004253 and Unigene0006446 can encode and synthesize CytP450. These two transcripts are up-regulated in R3, and the Cyclo-Dopa catalyzed by them is necessary for the synthesis of beet pigment in sugar beet (Hatlestad et al., 2012). GT can transfer carbohydrate groups to form betaine. We randomly selected four GT-related transcripts for verification, but only one showed an up-regulation trend in R3, which was not consistent with the prediction of this study, indicating that GT has a more complex expression pattern and needs further research. With the development of *Hylocereu polyrhizus*, the gene expression level of some key enzymes related to the synthesis of betalanins increased, betalanins gradually accumulated in the flesh, and the flesh turned red, but there was no betalanins in *Hylocereu undatus* for the expression of these key genes was too low, its color did not change (Hua et al., 2016).

Table 5 Primers information of genes for qRT-PCR

Number of unigenes	Primer	Name	Product length (bp)	primer length (bp)	Annealing temperature (°C)	GC percentage (%)	Sequence of primers (5’-3’)
Unigene0017646	Forward primer	U17646P1	164	22	60.9	50	5'GAAGAAGGGTGGGTTCATTGTG 3'
	Reverse primer	U17646P2		21	60.7	52.4	5'TTTCGCTGTACTCCTCGGTCA 3'
Unigene0051695	Forward primer	U51695P1	299	19	56.8	57	5'CCAAAGCCCTCCCTATGTC 3'
	Reverse primer	U51695P2		21	62	57	5'GCCTTCTGTATCTGCCGGTGA 3'
Unigene0003029	Forward primer	U3029P1	158	22	60.9	54.5	5'CCTACCCTAGACATCGCAAAGC 3'
	Reverse primer	U3029P2		23	61.4	47.8	5'GTCGGGAACCTAAACACCTCAAT 3'
Unigene0004253	Forward primer	U4253P1	131	22	60.9	45.5	5'GCTTATTGGGTTGTGCTTGGAT 3'
	Reverse primer	U4253P2		23	60.9	52.2	5'TGGGACTGTCACTTGGCTACTTC 3'
Unigene0006446	Forward primer	U6446P1	138	24	64.7	54.2	5'AGGCTATCCCTGTCATCTGGCTCT 3'
	Reverse primer	U6446P2		20	62.6	55	5'TGGTGTCTGTTCCGGCATCA 3'
Unigene0023505	Forward primer	U23505P1	152	18	58.4	66.7	5'CCACCGCCCTAAGTCCAG 3'
	Reverse primer	U23505P2		18	58.1	61.1	5'ATCCTCGTCACCGTCGCT 3'
Unigene0028648	Forward primer	U28648P1	125	21	60	48	5'CCCCACTGAAACAACCACAAA 3'
	Reverse primer	U28648P2		21	60	52	5'ACTGGGCTCAGGAATCAAACC 3'
Unigene0039772	Forward primer	U39772P1	165	19	57.8	47	5'GCAAGAAACTGCCCGAAAT 3'
	Reverse primer	U39772P2		19	61.8	57	5'TCCGCTGCTCCGTAAATGG 3'
Unigene0025394	Forward primer	U25394P1	181	20	54.6	50	5'CTGTTCTTTCTCCTCCATCG 3'
	Reverse primer	U25394P2		21	62.8	57	5'GGGTGCTGTTAGGGACCGTTT 3'
Unigene0003132	Forward primer	U3132P1	261	23	63	48	5'TGATCCCATTTCTAGCACAAGGC 3'
	Reverse primer	U3132P2		23	60	54	5'AGTGGAAGACGGAGATGACGAG 3'
Unigene0012439	Forward primer	U12439P1	149	23	62.9	52.2	5'CATCAGATGCTGTTCGACCAGTG 3'
	Reverse primer	U12439P2		21	61.4	52.4	5'TCTCCTCCCGCAGTTCCTTTA 3'
Unigene0047695	Forward primer	U47695P1	142	21	64	57	5'TCGGGTTCTTCCCAGCCACTA 3'
	Reverse primer	U47695P2		21	66	57	5'GACGCAAGCCAATCCATGACG 3'
Unigene0037073	Forward primer	U37073P1	238	22	60	50	5'TTCCTCACCCTCCATCAACTCT 3'
	Reverse primer	U37073P2		19	55	53	5'GATTCTAAGCCATGCGTCC 3'

Continued Table 5

Number of unigenes	Primer	Name	Product length (bp)	primer length (bp)	Annealing temperature (℃)	GC percentage (%)	Sequence of primers (5'-3')
Unigene0019344	Forward primer	U19344P1	233	23	66.5	56.5	5'TCGTTAGACCTGCCTCGACCACA 3'
	Reverse primer	U19344P2		22	62.5	50	5'CCCTTCCAAAGCCGTTAGTGTT 3'
Unigene0031027	Forward primer	U31027P1	239	22	61	54	5'GCACATAGGCCCTGTTTCACTC 3'
	Reverse primer	U31027P2		21	62.6	57	5'CATTGCTGTTCCTCACGACCC 3'
Unigene0015977	Forward primer	U15977P1	191	23	55.3	43.5	5'GAAGAACTACACCCAACCATAAC 3'
	Reverse primer	U15977P2		23	61.9	47.8	5'GCAACTGACCCAAAGCTCACATA3'
Unigene0042389	Forward primer	U42389P1	128	21	60	52	5'CGGTTGTCATCCACCACTTTC3'
	Reverse primer	U42389P2		21	61	52	5'ACTCACCCTGCCGATGTTCTT 3'
Unigene0048533	Forward primer	U48533P1	118	24	62.6	46	5'TGATTTGAAGGTTGGCAGTAGGAG 3'
	Reverse primer	U48533P2		24	61.8	46	5'CACCATTCACAGAAGCCATAGACA 3'
UBQ	Forward primer	Ubq P1	192	18			5'TGAATCATCCGACACCAT 3'
	Reverse primer	Ubq P2		19			5'TCCTCTTCTTAGCACCACC 3'

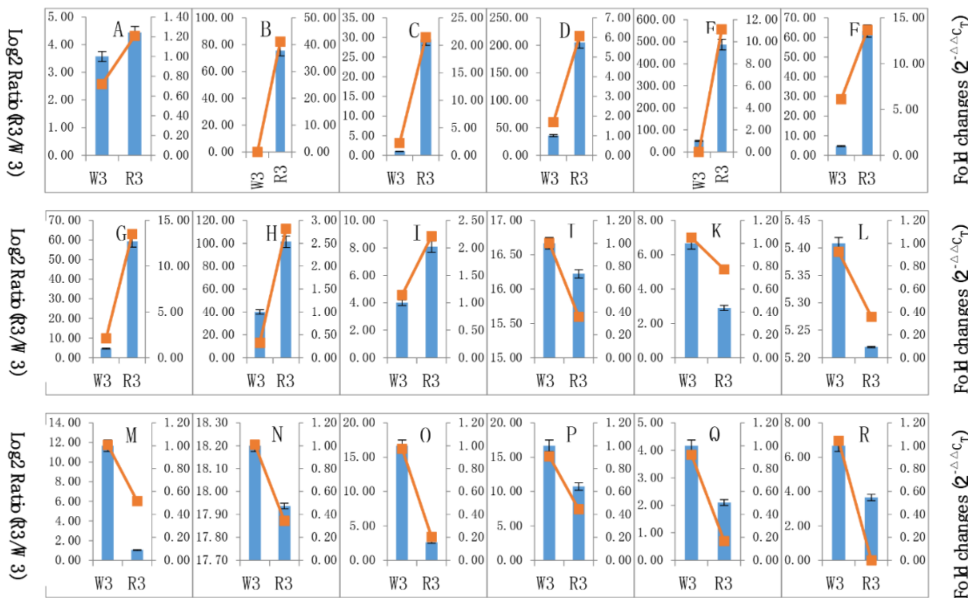

Figure 10 Comparison of DEGs' expression results from RNA-seq and qRT-PCR analyses in W3 VS R3.

Note: W3: White pulp mature stage; R3: Red pulp mature stage; Blue histograms represent the fold changes of genes (R3/W3) by qRT-PCR; Whereas orange line charts represent gene expression according to the log2 ratio (Reads Per Kilobase per Million mapped reads (RPKM) of R3/RPKM of W3) in RNA-Seq; A: Unigene0017646; B: Unigene0051695; C: Unigene0003029; D: Unigene0004253; E: Unigene0006446; F: Unigene0023505; G: Unigene0028648; H: Unigene0039772; I: Unigene0025394; J: Unigene0003132; K: Unigene0012439; L: Unigene0047695; M: nigene0037073; N: Unigene0019344;O: Unigene0031027; P: Unigene0015977; Q: Unigene0042389; R: Unigene0048533

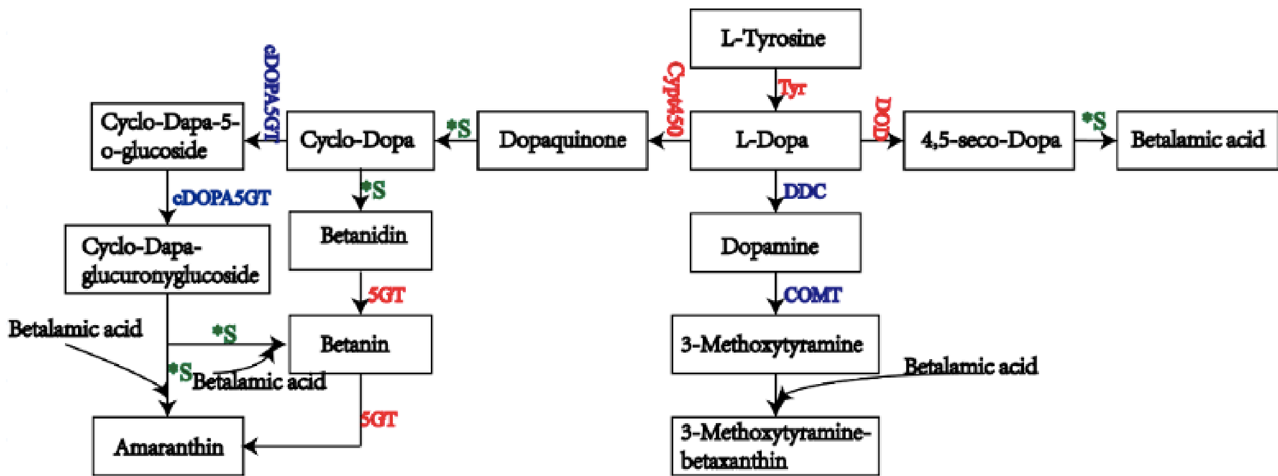

Figure 11 Betalaine metabolic pathway

Note: The red represent up regulated; The blue represent down regulated

The genes of all samples were analyzed by GO, and it was found that more genes would be enriched in catalytic activity, ligation, transport activity, nucleic acid and transcription factor binding activity, indicating that pigment metabolism needs a variety of enzymes and acting factors to regulate together. The pairwise comparison of four samples (W1 VS W3, R1 VS R3, W1 VS R1, W3 VS R3) showed that there were more DEGs in the same variety at different stages (R1 VS R3, 10 077), which indicated that the difference of flesh color was more closely related to the fruit development stage. With the growth and development of *Hylocereu polyrhizus*, some key genes related to color may play a role in promoting the synthesis of betalanins. Therefore, in this study, GO and KEGG analysis of

R1 VS R3 showed that most of the enriched genes were closely related to enzyme activity, and more genes were enriched in secondary metabolic pathway and tyrosine metabolic pathway. Betalanins is a secondary metabolite synthesized from tyrosine (Duan et al., 2017). There may be key genes related to fruit color development in R1 VS R3. We also analyzed different mature samples (W3 VS R3) and found that some genes were enriched in phenylalanine metabolic pathway, flavonoid and flavonol metabolic pathway. Phenylalanine is the precursor of tyrosine synthesis (Zhao et al., 2017), while flavonoids and flavonols are closely related to the synthesis of flavonoid pigments. The color composition of pitaya is more complex, not only betalanins, there may also be flavonoid pigments such as anthocyanins (Zeng, 2017). Recent studies have shown that there may be competition between betalanins and anthocyanin metabolic pathway in pitaya. Anthocyanin contributes to the development of pitaya's color partly (Fan et al., 2020).

To sum up, among the four genes related to tyrosinase or tyrosine synthesis (Unigene0017646 may be related to tyrosinase synthesis; Unigene0023505 may be related to prephenylalanine aminotransferase synthesis; Unigene0028648 is related to phenylalanine hydroxylase synthesis; Unigene0039772 is related to preaniline dehydrogenase synthesis. All three are key enzymes in upstream metabolism of tyrosine synthesis). Tyrosine permeability enzyme isomer 1 encoded by Unigene0017646 contributes to the transport of tyrosine and provides raw materials for the synthesis of beet pigment; two (Unigene0004253 and Unigene0006446) are related to the formation of cytochrome P450 (CytP450), and their expression patterns are consistent with previous studies. (Hua et al., 2016); two (Unigene0025394 and Unigene0019344) are related to the synthesis of flavonoids. Seven (Unigene0003029, Unigene0003132, Unigene0012439, Unigene0047695, Unigene0037073, Unigene0031027 and Unigene0015977) were related to the synthesis of glucosyltransferase (GT) and one (Unigene0051695) was related to the synthesis of DOPA dioxygenase (DOD). In some *Caryophylla* plants, some genes in the anthocyanin biosynthesis pathway are deleted, which triggers the betalanins biosynthesis pathway. DOD is the key enzyme for the formation of chromogenic groups in the sugarbeet pigment metabolism pathway, and it is also the key enzymes related to the expression of key genes synthetising betalanins (Xu et al., 2013; Wu et al., 2019). One (Unigene0048533) belongs to the ERF (ethylene response factor, ERF) family gene of ethylene response factor, ERF is closely related to plant fruit ripening (Ma et al. 2017), and one (Unigene0042389) is related to the synthesis of colorless anthocyanin reductase. We can select the genes with higher multiple of difference and more significantly expressed in R3 samples for more in-depth study.

3 Materials and Methods

3.1 Collection and processing of test materials

The experimental materials were collected from the dragon fruit plantation in Liguo Town, Dongfang City, Hainan Province (19°09'N, 108°64'E), and the tree age was 5 years old. Two varieties' Vietnam White flesh and "Jindu No.1" (Red flesh) were selected, and the fruit samples of green stage (15 days after pollination) and mature stage (30 days after pollination) were collected and marked as W1 (white flesh in green stage), W3 (white flesh in mature stage), R1 (red flesh in green stage) and R3 (red flesh in flesh stage). They were cutted into pieces and packed in 50 mL centrifuge tube in laboratory and stored in -80°C refrigerator after liquid nitrogen quick freezing.

3.2 Experimental steps for RNA-Seq

The pulp tissue samples of green and mature fruit were sent to Shenzhen Chideo Biological Company for transcriptome sequencing based on IlluminaHiSeqTM 2 000. After the total RNA was extracted from the sent samples, all mRNA, was enriched by magnetic beads method, and then the collected transcripts were decomposed into short fragments of 200~700nt with decomposition buffer. Using them as templates, random primers and reverse transcriptase were used to synthesize cDNA. After PCR amplification with suitable fragments, the complete piyata library was obtained, and the library was sequenced by IlluminaHiSeqTM 2 000.

3.3 Processing of transcriptome data

After sequencing, all the original datas (Raw Data) are analyzed, and after filtering, the Clean Data is obtained, and then the invalid low-quality sequences are removed. After reassembly, high-quality Unigenes are obtained. According to Swissprot, Nr, KOG, GO and KEGG databases, all Unigenes were annotated, including expression comments and function comments. Finally, enrichment analysis was carried out.

3.3 Differential expression analysis

The expression of Unigene was calculated by RPKM (Reads Per kb per Million reads) method. The screening criteria for differentially expressed genes were FDR < 0.05, | log2FC | > 1, and the corrected P value was less than 0.05. Including the analysis of differential expression, GO and KEGG Pathway enrichment analysis.

3.4 qRT-PCR verification

Pitaya UBQ gene was selected as internal reference gene, through transcriptome data analysis, combined with KEGG database, we selected the key enzymes of betalaine pathway, 21 Unigenes were used for qRT-PCR verification. Primer 5.0 was used to design primers, TaKaRa company's PrimeScriptTMRT REAGENT Kit was used to synthetise cDNA, gDNA Eraser (Perfect Real Time) (RR047A) kits and 7 500 Fast real-time fluorescence quantitative PCR system were used for qPCR test, in order to prevent technical errors, each reaction was repeated three times, and the results were analyzed by relative quantitative method ($2^{-\triangle\triangle Ct}$).

Authors' contributions

Zeng Canbin is the experimental designer and executor of this study; Guo Panyang completes the data analysis and writes the first draft of the paper; Guo Panyang, Zeng Canbin and Liu Jiaquan participate in the experimental design and analysis of the experimental results; Wei Shuangshuang and Huang Jiaquan participate in the data analysis; Tang Hua is the architect and person in charge of the project, guiding experimental design, data analysis and paper writing and revision. All the authors read and approved the final manuscript.

Acknowledgements

This study is supported by the National key Project of improved varieties of Pitaya (HD-KYH-2019042), the innovative Research Project of postgraduates of Hainan Province in 2019 (Hys2019-138) and the project of "Protection and Innovation of Germplasm Resources of Pitaya" of the Ministry of Agriculture and Village (18200032-1).

Reference

Chen C.Q., Wang M.W., Liu X.H., Ding H.R., Zhu X.M., Liu C., Xin J.C., Zhao B.Q., Dong J., and Hong L.Z., 2013, Research progress in plant secondary metabolism betacyanin, Anhui Nongye Kexue (Journal of Anhui Agricultural Sciences), 41(34): 13115-13116, 13131

Chen N., Yu Z.H., and Xiao X.G., 2017, Cytosolic and nuclear co-localization of betalain biosynthetic enzymes in tobacco suggests that betalains are synthesized in the cytoplasm and/or nucleus of betalainic plant cells, Front Plant Sci., 8: 831
https://doi.org/10.3389/fpls.2017.00831
PMid:28572813 PMCid:PMC5435750

Clifford T., Howatson G., West D.J., and Stevenson E.J., 2015, The potential benefits of red beetroot supplementation in health and disease, Nutrients, 7(4): 2801-2282
https://doi.org/10.3390/nu7042801
PMid:25875121 PMCid:PMC4425174

Duan X.Y., Tian Y., and Deng F.M, 2017, Advances on bioactivity and separating methods of pitaya fruit pigment, Shipin yu Jixie (Food & Machinery), 33(10): 214-219

Fan R.Y., Sun Q.M., Zeng J.W., and Zhang X.X., 2020, Contribution of anthocyanin pathways to fruit flesh coloration in pitayas, BMC Plant Biol., 20(1): 361
https://doi.org/10.1186/s12870-020-02566-2
PMid:32736527 PMCid:PMC7394676

Gandía-Herrero F., and García-Carmona F., 2013, Biosynthesis of betalains: yellow and violet plant pigments, Trends Plant Sci., 18(6): 334-343
https://doi.org/10.1016/j.tplants.2013.01.003
PMid:23395307

Hatlestad G.J., Akhavan N.A., Sunnadeniya R.M., Elam L., Cargile S., Hembd A., Gonzalez A., McGrath J.M., and Lloyd A.M., 2015, The beet *Y* locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway, Nat. Genet., 47(1): 92-96
https://doi.org/10.1038/ng.3163
PMid:25436858

Hatlestad G.J., Sunnadeniya R.M., Akhavan N.A., Gonzalez A., Goldman I.L., McGrath J.M., and Lloyd A.M., 2012, The beet R locus encodes a new cytochrome P450 required for red betalain production, Nat. Genet., 44(7): 816-820
https://doi.org/10.1038/ng.2297
PMid:22660548

Hua Q.Z., Chen C.J., Chen Z., Chen P.K., Ma Y.W., Wu J.Y., Zheng J., Hu G.B., Zhao J.T., and Qin Y.H., 2016, Transcriptomic analysis reveals key genes related to betalain biosynthesis in pulp coloration of hylocereus polyrhizus, Front Plant Sci., 6: 1179
https://doi.org/10.3389/fpls.2015.01179
PMid:26779215 PMCid:PMC4700300

Lee E.J., An D., Nguyen C.T., Patil B.S., Kim J., and Yoo K.S., 2014, Betalain and betaine composition of greenhouse- or field-produced beetroot (*Beta vulgaris* L.) and inhibition of HepG2 cell proliferation, J. Agric. Food Chem., 62(6): 1324-1331
https://doi.org/10.1021/jf404648u
PMid:24467616

Liu Y.J., Lin S.M., Huang X.F., and Xiong Y.M., 2014, A preliminary study on 'DAHONG' pitaya introduceton, Dongnan Yuanyi (Southeast Horticulture), 2(6): 86-88

Liu Y.J., Xiong Y.M., Huang X.F., Yang L., and Sheng C.G., 2017, A preliminary study on the introduction of pitaya cv.'Jing du No.1', Dongnan Yuanyi (Southeast Horticulture), 5(3): 11-12

Ma Y., Zhang H.X., Tu Y., and Cheng X.L., 2017, Advances in research on transcription factors related to fruit development and ripening, Jiyinzuxue yu Yingyomg Shengwuxue (Genomics and Applied Biology), 36(11): 4836-4846

Martinez R.M., Longhi-Balbinot D.T., Zarpelon A.C., Staurengo-Ferrari L., Baracat M.M., Georgetti S.R., Sassonia R.C., Verri W.A. Jr, Casagrande R., 2015, Anti-inflammatory activity of betalain-rich dye of Beta vulgaris: effect on edema, leukocyte recruitment, superoxide anion and cytokine production, Arch. Pharm. Res., 38(4): 494-504
https://doi.org/10.1007/s12272-014-0473-7
PMid:25173360

Polturak G., Heinig U., Grossman N., Battat M., Leshkowitz D., Malitsky S., Rogachev I., and Aharoni A., 2018, Transcriptome and metabolic profiling provides insights into betalain biosynthesis and evolution in *Mirabilis jalapa*, Mol. Plant, 11(1): 189-204
https://doi.org/10.1016/j.molp.2017.12.002
PMid:29247705

Sakuta M., 2014, Diversity in plant red pigments: anthocyanins and betacyanins, Plant Biotechnol. Rep., 8(1): 37-48
https://doi.org/10.1007/s11816-013-0294-z

Wilhelm B.T., and Landry J.R., 2009, RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing, Methods, 48(3): 249-257
https://doi.org/10.1016/j.ymeth.2009.03.016
PMid:19336255

Wu Y.W., Xu J., He Y.Z., Shi M.Y., Han X.M., Li W.Y., Zhang X.W., and Wen X.P., 2019, Metabolic profiling of pitaya (*Hylocereus polyrhizus*) during fruit development and maturation, Molecules, 24(6): 1114
https://doi.org/10.3390/molecules24061114
PMid:30897852 PMCid:PMC6470951

Wyler H., Meuer U., Bauer J., and Stravs-Mombelli L., 1984, Cyclodopa glucoside (=(2S)-5-(β-D-glocopyranosyloxy)-6-hydroxyindoline-2-carboxylic acid) and its occurrence in red beet (*beta vulgaris* var. *rubra* L.), Helv. Chim. Acta, 67(5): 1348-1355
https://doi.org/10.1002/hlca.19840670520

Xie F.F., Hua Q.Z., Chen C.B., Zhang L.L., Zhang Z.K., Chen J.Y., Zhang R., Zhao J.S., Hu G.B., Zhao J.T., and Qin Y.H., 2020, Transcriptomics-based identification and characterization of *glucosyltransferases* involved in betalain biosynthesis in *Hylocereus megalanthus*, Plant Physiol. Biochem., 152: 112-124
https://doi.org/10.1016/j.plaphy.2020.04.023
PMid:32413806

Xu S.X., Lin C.S., Huang Q.Y., Zeng B.Y., and Lin F.C., 2013, Separation and expression analysis of the betalain synthase *4, 5-DOPA-dioxygenase* gene in *Bougainvilea*, Redai Zuowu Xuebao (Chinese Journal of Tropical Crops), 34(11): 2214-2221

Zeng C.B., 2017, Study on pigment comparison and differentially expressed genes between red and white pitaya fruit, Thesis for M.S., College of Tropical Crops, Hainan University, Supervisor: Tang H., pp.39-62

Zhao S.J., Zhao X.Z., Xiang H., Qin X.M., and Tian J.S., Regulation of seabuckthorn seed oil on metabolic pathway disturbances on rats exposed to chronic unpredictable mild stress, Zhongcaoyao (Chinese Traditional and Herbal Drugs), 48(13): 2682-2690