Contribution of anthocyanin pathways to fruit flesh coloration in pitayas

Ruiyi Fan, Qingming Sun, Jiwu Zeng and Xinxin Zhang*

Abstract

Background: Color formation in *Hylocereus* spp. (pitayas) has been ascribed to the accumulation of betalains. However, several studies have reported the presence of anthocyanins in pitaya fruit and their potential role in color formation has not yet been explored. In this study, we profiled metabolome and transcriptome in fruit of three cultivars with contrasting flesh colors (red, pink and white) to investigate their nutritional quality and the mechanism of color formation involving anthocyanins.

Results: Results revealed that pitaya fruit is enriched in amino acid, lipid, carbohydrate, polyphenols, vitamin and other bioactive components with significant variation among the three cultivars. Anthocyanins were detected in the fruit flesh and accumulation levels of Cyanidin 3-glucoside, Cyanidin 3-rutinoside, Delphinidin 3-O-(6-O-malonyl)-beta-glucoside-3-O-beta-glucoside and Delphinidin 3-O-beta-D-glucoside 5-O-(6-coumaroyl-beta-D-glucoside) positively correlated with the reddish coloration. Transcriptome data showed that the white cultivar tends to repress the anthocyanin biosynthetic pathway and divert substrates to other competing pathways. This perfectly contrasted with observations in the red cultivar. The pink cultivar however seems to keep a balance between the anthocyanin biosynthetic pathway and the competing pathways. We identified several active transcription factors of the MYB and bHLH families which can be further investigated as potential regulators of the anthocyanin biosynthetic genes.

Conclusions: Collectively, our results suggest that anthocyanins partly contribute to color formation in pitaya fruit. Future studies aiming at manipulating the biosynthetic pathways of anthocyanins and betalains will better clarify the exact contribution of each pathway in color formation in pitayas. This will facilitate efforts to improve pitaya fruit quality and appeal.

Keywords: Anthocyanins, Betalains, Plant coloration, Fruit quality, RNA-seq, Widely-targeted metabolomics

Background

As an important part of human diet, fruits are reservoirs of nutrients and phytochemicals with a wide range of health benefits [1]. *Hylocereus* spp. (pitaya) is a new fruit crop with exotic aesthetic characteristics and is getting very popular among consumers [2]. It belongs to the Cactaceae family and Caryophyllales order [3]. The species originates in Central America [4] but with the increasing demand, the growing areas have now expanded throughout tropical and subtropical regions, particularly in countries, such as Malaysia, Vietnam, Thailand and China [5].

The major determinant in the expansion of pitaya cultivation is its high adaptation to dry environments and poor soils [3]. In addition, examination of the nutritional quality of pitaya fruit revealed that it is rich in various nutrients such as vitamin C, sugars, organic acids, phytalbumin, amino acids and minerals [6–8]. Besides, the high betalain, polyphenol and flavonoid content in pitaya fruit [9] have been shown to protect against some oxidative stress-related...
disorders [10, 11], lower cholesterol concentration, treat anemia among pregnant women, prevent colon cancer, inhibit anxiety and strengthen kidney function [8, 12–14].

Betalains have been extensively characterized in pitayas [15–19]. Betalains are only found in some species of Caryophyllales and include betacyanin (red-purple) and betaxanthins (yellow) [20]. The occurrence of betalains and anthocyanins (the dominant plant natural pigment) is thought to be mutually exclusive as both pigments have never been detected in the same plant species or tissues [21–24]. Several mechanisms have been invoked to explain this uncommon phenomenon in betalain-producing plants such as transcriptional down-regulation of anthocyanin genes [25–27], non-functional anthocyanin biosynthetic genes [28], loss of MYB-bHLH-WD40 transcriptional complex essential for the anthocyanin regulatory pathway [29] and fundamental imbalance between tyrosine pathway (leading to betalain biosynthesis) and phenylalanine-derived pathway (leading to anthocyanin biosynthesis) [30]. It is documented that pitaya fruit peel or flesh pigmentation is due to the presence of high level of betacarotenes and this was shown through metabolome and transcriptome investigations [17, 31–33]. Curiously, several studies also detected the presence of anthocyanins in pitayas, although the specific anthocyanin compounds have not been characterized [34–37]. This suggests that both betalains and anthocyanins may be present in *Hylocereus spp.*, therefore, anthocyanins may also partly contribute to the fruit pigmentation. Unfortunately, no study has explored the changes in expression levels of genes participating and regulating the anthocyanin biosynthetic pathway in pitaya during fruit development or between fruits with contrasting colors.

Color of fruits is a major quality criteria governing consumer’s preference and determining market value [38]. There is a large diversity of fruit colors in *Hylocereus spp.* but the red-colored cultivars (*Hylocereus polyrhizus*) are the most preferred by consumers. Unfortunately, detailed metabolic profiling to unravel important bioactive components in *Hylocereus polyrhizus* is very limited [17, 19, 39]. Similarly, the metabolic potentials of other colored cultivars such as *Hylocereus undatus* have been neglected. In this study, we employed the widely targeted metabolomics approach to comprehensively detect and compare hundreds of metabolites between three *Hylocereus spp.* cultivars with contrasting fruit flesh colors. By integrating transcriptome data, we further examined the possible contribution of anthocyanin metabolites and related biosynthetic genes to color formation in pitaya fruit.

**Results**

**Metabolome profiling in three *Hylocereus spp.* cultivars**

In the present work, three *Hylocereus spp.* cultivars with different phenotypes (Table 1) were used for widely targeted metabolomics based on six biological replicates. The cultivars Da Hong (DH), Fen Rou (FR), and Bai Rou (BR) were selected mainly because they display different fruit flesh colors (Fig. 1a). In total, 443 metabolites belonging to various classes of metabolites were successfully detected in the samples. Metabolites belonging to the classes of amino acid, lipid, carbohydrate, cofactors and vitamins were the most enriched in *Hylocereus spp.* fruits (Fig. 1b). Metabolite peak identification, filtration, alignment were performed using the XCMS package of R (v3.3.2). The intensity of each metabolite was presented in Table S1. To assess the quality of the metabolic profiling data, we performed a principal component analysis (PCA) of all replicates together with the quality control (QC) samples. All QC samples clustered together in the PCA with very little variability, indicating that the data is reliable. Furthermore, the three cultivars could be clearly distinguished by the first two PCs showing that large differences exist in their fruit metabolome.

**Variations in fruit metabolome among the three cultivars**

In order to explore the fruit nutritional quality of the three cultivars, we compared their concentrations in major classes of metabolites. Globally, the red samples (DH) contained less amino acid and lipid metabolites than the pink and white samples (Fig. 2). In contrast, DH samples were more endowed with carbohydrates, energy, nucleotides and cofactors and vitamins related metabolites as compared to FR and BR samples; FR displayed a strong content of phenolic metabolites. These results suggest a high variation in the nutritional properties of these three cultivars.

**Analysis of metabolites related to the flavonoid-anthocyanin pathway**

The major phenotypic difference between the three cultivars is the flesh coloration. Anthocyanins are the

---

**Table 1: Characteristics of the three *Hylocereus* spp. cultivars used in this study**

| Species name | Fruit name | Origin          | Fruit size/weight (g) | Peel/flesh color | Fertilization mode  | Soluble solid content (°Brix) | Cultivation in mainland China |
|--------------|------------|-----------------|-----------------------|------------------|--------------------|------------------------------|-----------------------------|
| *Hylocereus polyrhizus* | Da Hong (DH) | Taiwan          | Large/850              | Red/ Red         | Self-fertilization | 12.8                         | Major                        |
| Hybrid (*Hylocereus polyrhizus* x *Hylocereus undatus*) | Fen Rou (FR) | Mainland China  | Medium/325            | Pink/Pink        | Self-incompatible     | 12.5                         | Minor                        |
| *Hylocereus undatus* | Bai Rou (BR) | Vietnam         | Large/700              | Pink/White       | Self-fertilization | 11.8                         | Minor                        |
main pigments in plants [15]. To understand the contribution of anthocyanins to pitaya flesh coloration, we investigated changes in the concentrations of metabolites related to the flavonoid-anthocyanin pathway between the three cultivars. We identified five anthocyanin compounds, including Cyanidin 3-glucoside, Cyanidin 3-rutinoside, Delphinidin 3-rutinoside, Delphinidin 3-O-(6-O-malonyl)-beta-glucoside-3-O-beta-glucoside and Delphinidin 3-O-beta-D-glucoside 5-O-(6-coumaroyl-beta-D-glucoside). In addition, two upstream metabolites were identified: Naringenin and Quercetin. As shown in Fig. 3, except Delphinidin 3-rutinoside which was more enriched in the white flesh cultivar (BR), the other four anthocyanins were more concentrated in the red sample (DH) and pink sample (FR) with a more pronounced accumulation in DH. Concerning the two upstream metabolites, we observed that they were more accumulated in BR as compared to FR and DH. We further determined the quantities of three selected anthocyanins namely, Cyanidin 3-glucoside, Cyanidin 3-rutinoside and Delphinidin 3-rutinoside using the electrospray ionization/high-performance liquid chromatography/tandem mass spectrometry (ESI-HPLC-MS/MS) method. As presented in Figure S1, the patterns of accumulation of these anthocyanins based on the ESI-HPLC-MS/MS analysis perfectly matched the report of the widely targeted metabolomics (Fig. 3).

**Transcriptome sequencing and assembly**

In order to get insight into the expression patterns of anthocyanin biosynthetic genes between the different pitaya cultivars, we profiled gene expression in the flesh samples (three biological replicates). A total of 85.70 Gb raw data was generated. After removing low quality reads, 99% of the raw data was kept as clean data for
downstream analyses (Table 2). Overall the quality of the sequencing was high as evidenced by the high Q30 score (> 91%) and the quasi- absence of unknown nucleotides (N). Using the Trinity tool, we assembled a total of 49,212,589 bp sequence containing 53,850 unigenes with an N5 length of 1647 bp (Table 3; Figure S2). Transcriptome assembly validation was done using Benchmarking Universal Single-Copy Orthologs (BUSCO) v.3 [40]. 70% of complete BUSCOs were present in the de novo transcriptome, indicating a good quality assembly (Figure S3).

Fig. 2 Comparison of the total ion intensity of various classes of metabolites between three Hylocereus spp. fruits. DH, FR, and BR represent the red, pink and white colored flesh samples, respectively. Data represent average values from six biological replicates. The error bar represents the SD of six biological replicates. Different letters above bars mean significant difference at \( P < 0.05 \).

Fig. 3 Comparison of the total ion intensity of metabolites related to the flavonoid-anthocyanin pathway between three Hylocereus spp. fruits. DH, FR, and BR represent the red, pink and white colored flesh samples, respectively. Data represent average values from six biological replicates. The error bar represents the SD of six biological replicates. Different letters above bars mean significant difference at \( P < 0.05 \).
Table 2 Overview of the transcriptome sequencing dataset and quality check

| Samples | Raw data (bp) | Clean data (bp) | Q30 (%) | N (%) | GC (%) |
|---------|---------------|----------------|---------|-------|--------|
| DH-1    | 7,313,628,900 | 7,275,654,049  | 91.93%  | 0.00% | 51.18% |
| DH-2    | 8,471,004,900 | 8,434,243,349  | 91.53%  | 0.00% | 51.14% |
| DH-3    | 7,895,180,700 | 7,860,911,481  | 91.77%  | 0.00% | 51.29% |
| FR-1    | 8,714,710,800 | 8,673,510,516  | 91.81%  | 0.00% | 51.60% |
| FR-2    | 11,753,924,700| 11,695,962,338 | 93.75%  | 0.00% | 51.88% |
| FR-3    | 11,898,625,500| 11,848,132,870 | 92.70%  | 0.00% | 51.92% |
| BR-1    | 12,352,945,500| 12,285,577,982 | 92.55%  | 0.00% | 51.18% |
| BR-2    | 8,355,256,800 | 8,300,684,267  | 93.37%  | 0.00% | 51.60% |
| BR-3    | 8,942,012,100 | 8,797,612,590  | 92.68%  | 0.00% | 50.88% |

Functional annotation of 25,830 unigenes was obtained using KEGG, SwissProt, COG and Nr databases with 12, 153 unigenes commonly annotated in all these databases (Fig. 4a). Blast search of the transcripts against published genome sequences revealed that Hylocereus spp. shared a significant number of genes with Beta vulgaris (Figure S4).

Gene expression was estimated based on the Reads Per kb per Million reads (RPKM) method. Overall, DH and FR exhibited similar gene expression profiles in contrast to BR (Fig. 4b). Based on the gene expression profiles, we performed a PCA of the nine samples to assess the quality of the biological replicates and clustering patterns of samples from the three cultivars. Figure 4c shows that most biological replicates are closely related and a clear separation of the three cultivars could be observed, indicating that large discrepancies in the transcriptional activity exist within these three cultivars. Pairs of cultivars were compared in order to detect differentially expressed genes (DEGs). The lowest number of DEGs was obtained between the red (DH) and pink (FR) samples while the highest number of DEGs was detected between colored samples (DH and FR) and the white flesh cultivar (BR) (Fig. 4d). In particular, we observed that BR tends to repress the expression levels of hundreds of genes as compared to DH and FR, a mechanism which may be associated with the differential flesh coloration.

Analysis of genes involved in the flavonoid-anthocyanin pathway

Various structural genes are known to catalyze the biosynthesis of anthocyanins in plants. In total, 33 genes involved in the flavonoid-anthocyanin pathway were annotated in the Hylocereus spp. transcriptomes. Of these, 12 genes were differentially expressed between the three cultivars. The flavonoid-anthocyanin pathway was then reconstructed based on these DEGs (Fig. 5). Results showed that structural genes involved in both early and late committed steps of anthocyanin biosynthesis were differentially modulated. All the genes directly involved in the anthocyanin biosynthesis, such as trans-cinnamate 4-monooxygenase (C4H: Unigene0008383), chalcone synthase (CHS: Unigene004474), naringenin 3-dioxygenase (F3H: Unigene0001653), flavonoid 3’-monooxygenase (F3′H: Unigene0044459), flavonoid 3’,5’-hydroxylase (F3′5′H: Unigene0020548), dihydroflavonol 4-reductase (DFR: Unigene005595), anthocyanidin synthase (ANS; Unigene0029010) were higher expressed in the red (DH) and pink (FR) samples as compared to the white samples (BR). Distinctly, C4H, DFR, F3′5′H and ANS were not expressed in BR, a mechanism which may limit the biosynthesis of anthocyanins. The anthocyanin biosynthesis pathway is competed by other related pathways such as shikimic and quinic acids pathway, flavonol biosynthesis, quercetin biosynthesis and proanthocyanidin biosynthesis [41]. Here, we identified several DEGs (shikimate O-hydroxycinnamoyltransferase (HCT: Unigene00021892), flavonol synthase (FLS: Unigene0031287, Unigene0051509), leucoanthocyanidin reductase (LAR: Unigene0037637) and anthocyanidin reductase (ANR: Unigene0044348), catalyzing these competing pathways and these genes were higher expressed in FR and BR samples than DH samples.

Transcription factors differentially expressed between the three cultivars

Transcription factors (TF) are special genes that modulate the expression levels of other genes and represent the main players for the determination of spatiotemporal transcriptional activity [42]. Among the 12,153 annotated unigenes, 921 genes encoded for TFs, the majority being members of bHLH, bZIP, C2H2, ERF, MYB and WRKY families (Figure S5). All DEGs (223 genes) encoding TFs were retrieved and the results revealed that bHLH, MYB and ERF were the more active TFs which may differentially regulate structural genes participating in the color formation of Hylocereus spp. fruit flesh (Table S2). Remarkably, most of these TFs were found down-regulated in white and pink cultivars (BR and FR) when compared to the red cultivar (DH), suggesting that a high transcriptional activity is required to form the red color.

Table 3 Statistics of the unigene assembly results

| Genes Number | GC percentage | N50 number | N50 length (bp) | Max length (bp) | Min length (bp) | Average length (bp) | Total assembled bases (bp) |
|--------------|---------------|------------|-----------------|----------------|-----------------|---------------------|---------------------------|
| 53,850       | 43.279        | 9089       | 1647            | 14,515         | 201             | 913                 | 49,212,589                |
qRT-PCR validation of selected genes
Nine DEGs including six DEGs involved in the flavonoid-anthocyanin pathway (Fig. 6) and three TF encoding genes (Unigene0001134 (MYB), Unigene0007067 (MYB), Unigene0053482 (bHLH)) were selected and their transcript levels were estimated by qRT-PCR. All the tested genes were significantly and differentially expressed between the cultivars and the trends of expression fold change matched well the RNA-seq report (Fig. 6). This result showed that the RNA-seq report presented in this study and subsequent interpretations are reliable.

Discussion
The goal of this study was to characterize the fruit metabolome of three *Hylocereus* spp. and explore the role of anthocyanins in color formation. Fruits at 30 days after flowering were harvested and used in this study because color change mainly occurs at this stage [19]. Although previous studies investigated the metabolome of *Hylocereus* spp. fruits, they did not provide deep insights into the diversity of primary and secondary bioactive components available in this exotic species [17, 19, 34]. By using the widely targeted metabolomics approach, we identified and quantified extensive metabolites in *Hylocereus* spp. and our results suggest high concentrations of amino acid, lipid, carbohydrate, polyphenols, vitamin, and other bioactive molecules. Interestingly, the concentrations of the major classes of metabolites detected in *Hylocereus* spp. fruits were higher than some commonly consumed fruits such as Goji [43] and jujube [44], making it an excellent nutritious fruit. Red *Hylocereus polyrhizus* is particularly appreciated by consumers and the cultivar *Da Hong* used in this study is widely grown and highly consumed in China. However, our results showed that it has low levels of several amino acids and lipids.
diversity in the nutritional components observed among the three cultivars is interesting as it could be harnessed in breeding programmes aiming at developing cultivars with enhanced nutritional quality [33, 45]. Recently, metabolite-based genome wide association study (mGWAS) has emerged as an efficient approach to pinpoint functional genes associated with variation of metabolite concentrations in plants [46]. Therefore, as perspective of our study, we will design a comprehensive mGWAS involving a large and diverse panel of Hylocereus spp. cultivars to decipher the genetic basis of the fruit nutritional quality [47].

Color of fruits is a major determinant of consumer’s preference and market value [38]. Anthocyanins are the main secondary metabolites responsible for a variety of colors observed in plant organs such as fruits and flowers [48]. However, in some Caryophyllales plant species, it has been suggested that betalains are the main pigments, replacing anthocyanins [49]. Since the two pigments have not been detected in the same plant or tissue, the hypothesis of their mutual exclusion has been postulated [21–24]. However, no clear evidences have been provided to date to explain this curious phenomenon in the plant kingdom. Previous studies stated that the coloration of Hylocereus spp. fruit peel and pulp is ascribed to the presence of betalains, particularly betacyanins [17, 31–33]. Meanwhile, other studies have reported the presence of anthocyanins in Hylocereus spp. [34–37]. This suggests the presence of both pigments in Hylocereus spp., which refutes the paradigm of mutual exclusion of anthocyanins and betalains within the...
same species/tissue. In this study, we identified five different anthocyanins in fruit peel samples of three different Hylocereus spp. cultivars (Fig. 3), proving that except the well characterized betalains, Hylocereus spp. fruit also contains anthocyanins. Therefore, in depth biochemical and functional genomics studies will be required to explore and elucidate the mechanism of anthocyanin and betalain double production in Hylocereus spp. fruit. The accumulation levels of these anthocyanins in particular, Cyanidin 3-glucoside, Cyanidin 3-rutinoside, Delphinidin 3-O-(6-O-malonyl)-beta-glucoside-3-O-beta-glucoside and Delphinidin 3-O-beta-D-glucoside 5-O-(6-coumaroyl-beta-D-glucoside) positively correlated with the red flesh coloration, implying that these compounds partly contribute to the red color. To evidence this assertion, we further sequenced and de novo assembled transcriptomes in the three cultivars. In a recent transcriptome assembly from Hylocereus polyrhizus stem, Xu et al. [50] strangely reported Vitis vinifera as the species sharing the greatest number of transcripts, although both species are relatively distant. Beta vulgaris shared the highest transcriptome similarity with our assembled transcriptome, a result which is consistent by the fact that both species belong to the order Caryophyllales. Notably, all genes encoding enzymes participating in the early and late flavonoid-anthocyanin biosynthesis pathway were identified in Hylocereus spp. transcriptomes, suggesting the presence of a functional pathway.

Variation of anthocyanin content in plant tissues is ascribed to the differential expression levels of key genes participating in the biosynthetic pathway [51]. Comparison of cultivars with contrasting fruit flesh colors allowed us to identify differentially expressed anthocyanin biosynthetic genes that may affect the color formation as previously documented in turnip [52], Prunus mira [53] and Lagerstromemia indica [54]. Interestingly, 12 key genes were highlighted in this study, seven of which directly involved in the anthocyanin biosynthetic pathway, were found significantly higher expressed in the colored cultivars (red and pink) as compared to the white cultivar. Conversely, we observed that the other five genes participating in the anthocyanin biosynthetic competing pathways (shikimic and quinic acids pathway, flavonol biosynthesis, quercetin biosynthesis and proantocyanidin biosynthesis pathways)

**Fig. 6** Quantitative real time PCR validation of nine candidate genes predicted to differentially affect the anthocyanin profiles in fruit flesh of three Hylocereus spp. cultivars. DH, FR, and BR represent the red, pink, and white colored flesh samples, respectively. Data represent average values from three biological replicates. The error bar represents the SD of three biological replicates. The Actin gene was used as the internal reference gene for normalization. Different letters above bars mean significant difference at $P < 0.05$
were all repressed in the red cultivar but significantly increased in the pink and white cultivars. Globally, it was clear that the white cultivar tends to repress the anthocyanin biosynthetic pathway and divert substrates to other competing pathways. The high content of naringenin, which is the key upstream substrate for anthocyanin pathway [55]; the high level of quercetin, which biosynthesis competes with leucoanthocyanidin biosynthesis [56] and the low level of anthocyanins fully support our conclusion. In contrast, the red cultivar prioritizes the anthocyanin biosynthetic pathway over the competing pathways, with high level of anthocyanin and low levels of naringenin and quercetin production (Fig. 3). Finally, the pink material seems to keep a balance between the anthocyanin biosynthetic pathway and the competing pathways. These mechanisms involving competition between anthocyanin biosynthesis and other pathways were also reported in plant species such as turnip [52], Mimulus lewisii [57], Petunia [58, 59] and Lisianthus [60].

The molecular regulation of anthocyanin biosynthetic genes has been extensively studied in plants as an effective approach for engineering materials with tailored coloration [61, 62]. Transcription factors (TF) are particular genes that modulate the expression level of other genes [42]. It has been reported that MYB and bHLH are the main TFs which control the expression levels of anthocyanin biosynthetic genes [63–65]. In our study, we also observed several MYB and bHLH genes differentially expressed between the three cultivars, most of them were up-regulated in the red cultivar. This indicates their positive control of anthocyanin biosynthetic genes. Predicting the specific target genes of these MYB and bHLH TFs would facilitate the selection of candidate genes to target for controlling flesh color in Hylocereus spp. As proposed by Qiao et al. [55], the construction of gene co-expressed network and connecting the candidate regulators and their target could be an effective approach.

Conclusions
Metabolic profiling in three Hylocereus spp. cultivars with contrasting flesh colors revealed a large set of metabolites and a significant variation in fruit nutritional quality. We detected several anthocyanin molecules with varying levels in the three cultivars and integrative analysis with transcriptome indicates their probable contribution to flesh coloration. This involves a competition between anthocyanin biosynthesis and other pathways. Overall, this study provides an important theoretical basis for further in-depth dissection of the importance of anthocyanins versus betalains for color formation in Hylocereus spp. The findings from this study will benefit molecular breeders in their efforts to improve fruit appeal and quality in Hylocereus spp.

Methods
Plant materials
Three Hylocereus spp., including Hylocereus polyrhizus cv. Da Hong (DH), Hylocereus undatus cv. Bai Rou (BR) and a hybrid Hylocereus polyrhizus × Hylocereus undatus cv. Fen Rou (FR) with various flesh colors were used in this study and materials were provided by the Institute of Fruit Tree Research, Guangdong Academy of Agricultural Sciences, Guangzhou, China. The formal identification of the cultivars materials was undertaken by the corresponding author of this article (Professor Xinxin Zhang). No voucher specimen of this material has been deposited in a publicly available herbarium. All cultivars were grown at the farm of the Institute of Fruit Tree Research, Guangdong Academy of Agricultural Sciences, Guangzhou, China. Fruits were harvested 30 days after flowering on September 5th 2019. Flesh samples were cut from six fruits (biological replicates) of each cultivar and immediately frozen in liquid nitrogen for later use.

Metabolic profiling
Extract preparation, metabolite extraction, identification and quantification were performed following standard procedures of Suzhou BioNovoGene Metabolomics Platform, Suzhou, China.

Metabolite extraction
In total, 18 samples (six biological replicates) were used for metabolite extraction. Approximately 200 mg of each sample was accurately weighed and inserted in 2 mL EP tube. Then, 0.6 mL 2-chlorophenylalanine (4 ppm) methanol (−20 °C) was added and vortexed for 30 s; 100 mg glass beads were added and the samples were put into a Tissuelysis II tissue grinding machine. Samples were grind at 25 Hz for 60 s. The tubes were placed in an ultrasound bath at room temperature for 15 min; Then, centrifuged at 25 °C for 10 min at 1750 g, and the supernatant was filtered through 0.22 μm membrane to obtain the prepared samples for liquid chromatography-mass spectrometry (LC-MS); 20 μL from each sample was mixed and used as quality control (QC) samples (These QC samples were used to monitor deviations of the analytical results from these pool mixtures and compare them to the errors caused by the analytical instrument itself). The remaining extracts were used for LC-MS detection.

Chromatographic conditions
Chromatographic separation was performed in a Thermo Ultimate 3000 system equipped with an ACQUITY UPLC® HSS T3 (150 × 2.1 mm, 1.8 μm, Waters) column maintained at 40 °C. The temperature of the autosampler was 8 °C. Gradient elution of analytes was carried out with 0.1% formic acid in water (D) and 0.1% formic acid in
acetonitrile (C) or 5 mM ammonium formate in water (B) and acetonitrile (A) at a flow rate of 0.25 mL/min. Injection of 2 μL of each sample was done after equilibration. An increasing linear gradient of solvent A (v/v) was used as follows: 0 ~ 1 min, 2% A/C; 1 ~ 9 min, 2% ~ 50% A/C; 9 ~ 12 min, 50% ~ 98% A/C; 12 ~ 13.5 min, 98% A/C; 13.5 ~ 14 min, 98% ~ 2% A/C; 14 ~ 20 min, 2% C-positive model (14 ~ 17 min, 2% A-negative model).

**Metabolite data analysis**

Sample data processing was performed as described by Smith et al. [66]. The original data was converted into the mzXML format (xcms input file format) through the Proteowizard software (v3.0.8789). Based on metabolites information in public metabolomic databases and the self-built MetWare database (http://www.metware.cn/), the metabolites were qualitatively analyzed with the secondary spectrum information. Peak identification, filtration, alignment were performed using the XCMS package of R (v3.3.2) as described by Liu et al. [67]. Principal component analysis (PCA) was performed with R package (http://www.r-project.org/).

**Determination of selected anthocyanins using ESI-HPLC-MS/MS**

Standards of three anthocyanins including, Cyanidin 3-glucoside (CAS Number: 7084-24-4), Cyanidin 3-rutinoside (CAS Number: 18719-76-1) and Delphinidin 3-rutinoside (CAS Number 15674-58-5) were purchased from Merck (Germany). Sample preparation, metabolite extraction and quantification were performed as described above using the electrospray ionization/high-performance liquid chromatography/tandem mass spectrometry (ESI-HPLC-MS/MS) method on the 1260 Infinity II HPLC (Agilent Systems, USA).

**Transcriptome sequencing and analysis**

**RNA extraction, library construction and sequencing**

Trizol reagent (Invitrogen, Carlsbad, CA, USA) was used for RNA extraction. RNA quality check, enrichment and library construction were performed following standard procedures of Gene Denovo Biotechnology Co. (Guangzhou, China) and previously described by Hu et al. [52]. The sequencing was conducted on Illumina HiSeqTM 4000 platform.

**Filtering of raw reads and de novo assembly**

Raw reads were filtered by fastp [68] (version 0.18.0) and the obtained clean reads were de novo assembled using Trinity [69] software. Benchmarking Universal Single-Copy Orthologs (BUSCO) v.3 [40] was used for transcriptome assembly validation.

Next, the transcripts were realigned to construct unigenes and annotated in different databases, including Nr, Swiss-Prot, KEGG and COG/KOG. Protein coding sequences of unigenes submitted to Plant TFdb (http://planttfdb.cbi.pku.edu.cn/) to predict transcription factors. The unigene expression was calculated and normalized to Reads Per kb per Million reads [70].

**Differential expression**

Differential expression analysis was performed by edgeR Bioconductor package [71] between two samples based on false discovery rate (FDR) < 0.05 and absolute fold change ≥2.

**Gene expression using quantitative real time-PCR (qRT-PCR)**

The qRT-PCR was performed on RNA extracted from flesh samples (in triplicate) of the three cultivars and specific primer pairs. Table S3 following descriptions of Dossa et al. [72]. We used the Actin gene as the endogenous control for gene expression normalization. The gene relative expression level was estimated based on the 2^-ΔΔCt method [73].

**Statistical analysis**

Analysis of variance followed by the mean comparison test of Tukey HSD was performed to compare the three cultivars. Data analysis was performed using R version 3.2.2.

**Supplementary information**

Supplementary information accompanies this paper at https://doi.org/10.1186/s12870-020-02566-2.
dragon fruit peels (Hylocereus polyrhizus) extract. Drug Innovation Today. 2019;12:670–8.

38. Moser R, Raffaeli R, Thilmann-McFadden D. Consumer preferences for fruit and vegetables with credence-based attributes: A review. Int Food Agribusiness Manage Rev. 2011;14:121–42.

39. Hua QZ, Chen CB, Zur TN, Wanga HE, Wu JY, Chen JY, Zhang ZK, Zhao JT, Hu GB, Qin QH. Metabolic characterization of pitaya fruit from three red-skinned cultivars with different pulp colors. Plant Physiol Biochem. 2018;126:117–25.

40. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2018;35:943–8.

41. Mouradov A, Spangenberg G. Flavonoids: a metabolic network mediating plants adaptation to their real estate. Front Plant Sci. 2014;5:620.

42. Banerjee N, Zhang MQ. Identifying cooperativity among transcription factors controlling the cell cycle in yeast. Nucleic Acids Res. 2003;31:7024–31.

43. Wang Y, Liang X, Li Y, Fan Y, Li Y, Cao Y, An W, Shi Z, Zhao J, Guo S. Changes in Metabolome and nutritional quality of Lycium barbarum fruits from three typical growing areas of China as revealed by widely targeted metabolomics. Metabolomics. 2020;10:46.

44. Zhang Q, Wang L, Liu Z, Zhao Z, Zhao J, Wang Z, Zhou G, Liu P, Liu M. Transcriptome and metabolome profiling unveil the mechanisms of Zyphus jujuba Mill coloration. Food Chem. 2020;321:125903.

45. Tel-Zur N, Dudai M, Ravish E, Mizrahi Y. Selection of interspecific vine cacti in controlling the cell cycle in yeast. Nucleic Acids Res. 2003;31:7024–31.

46. Loo J. Metabolite-based genome-wide association studies in plants. Curr Opin Plant Biol. 2015;24:31–8.

47. Fang C, Luo J, Wang S. The diversity of nutritional metabolites: origin, dissection, and application in crop breeding. Front Plant Sci. 2019;10:1028.

48. Khoo HE, Azlan A, Tang ST, Lim SM. Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. Food Nutr Res. 2017;61:161779.

49. Mabry T. The betacyanins, a new class of red violet pigments, and their phylogenetic significance. New York: Roland Press; 1964.

50. Xu M, Liu C-L, Luo J, Qi Z, Yan Z, Fu Y, Wei S-S, Tang H. Transcriptome and metabolic profiling of pitaya (Hylocereus polyrhizus) cancer disease caused by Neoscytalidium dimidiatum. BMC Genomics. 2019;20:2019.

51. Zheng T, Tan W, Yang H, Zhang L, Li T, Liu B, Zhang R, Yin H. Regulation of anthocyanin accumulation via MYB75/HAT1/TPL-mediated transcriptional repression. PLoS Genet. 2019;15:e1007993.

52. Zhang H, Lou Q, Liu H, Han H, Wang Q, Tang H, Ma Y, Wang H. Differential regulation of anthocyanin in green and purple peaches revealed by combined de novo transcriptome and metabolome analysis. Int J Mol Sci. 2019;20:4387.

53. Ying H, Shi J, Zhang S, Pengcuo G, Wang Y, Bao Wei Q, Yizi Z, Zeng X. Transcriptomic and metabolic profiling provide novel insights into fruit development and flesh coloration in Prunus triloba Koehne, a special wild peach species. BMC Plant Bio. 2020;20:340.

54. Qiao Z, Liu S, Zeno KH, Li Y, Wu Q, Chen Y, Wang X, Cai N. Exploring the molecular mechanisms underlying the stable purple-red leaf phenotype in Lagerstroemia indica. Front Genet. 2019;10:772.

55. Dixon RA, Chua NZ, Zur TN, Wanga HE, Wu JY, Chen JY, Zhang ZK, Zhao JT, Hu GB, Qin QH. Metabolic characterization of pitaya fruit from three red-skinned cultivars with different pulp colors. Plant Physiol Biochem. 2018;126:117–25.

56. Crozier A, Jaganath IB, Clifford MN. Dietary phenolics: chemistry, bioavailability and effects on health. Nat Prod Rep. 2009;26:1001–43.

57. Fan YW, Ren YZ, Ab, Sagawa JM, Stanley LE, Bradshaw HD Jr. Transcriptome between anthocyanin and flavanol biosynthesis produces data about pattern variation of flavonol pigments between Mimulus species. Proc Natl Acad Sci U S A. 2016;113(9):2448–53. https://doi.org/10.1073/pnas.1520941113.

58. Davies KM, Schwinn KE, Deroles SC, Manson DG, Lewis DH, Bloor SJ, Bradley JM. Enhancing anthocyanin production by altering competition for substrate between flavonol synthase and dihydroflavonol 4-reductase. Euphytica. 2003;131:259–68.

59. Saito R, Kuchitsu K, Ozeki Y, Nakayama M. Spatiotemporal metabolic regulation of anthocyanin and related compounds during the development of marginal picotee petals in Petunia hybrid (Solanaceae). J Plant Res. 2007;120(4):563–8.