# Integrated sRNAome and RNA-Seq Analysis Reveals miRNA Effects on Betalain Biosynthesis in Pitaya

**Canbin Chen**
South China Agricultural University College of Horticulture
**Fangfang Xie**
South China Agricultural University College of Horticulture
**Qingzhu Hua**
South China Agricultural University College of Horticulture
**Noemi Tel Zur**
Jacob Blaustein Institutes for Desert Research
**Lulu Zhang**
South China Agricultural University College of Horticulture
**Zhike Zhang**
South China Agricultural University College of Horticulture
**Rong Zhang**
South China Agricultural University College of Horticulture
**Jietang Zhao**
South China Agricultural University College of Horticulture
**Guibing Hu**
South China Agricultural University College of Horticulture
**Yonghua Qin** ( qinyh@scau.edu.cn )
https://orcid.org/0000-0001-6505-0138

**Research article**

**Keywords:** Betalain biosynthesis, Gene expression, Hylocereus, miRNA, sRNAome and RNA-Seq, 5'RACE

**Posted Date:** February 20th, 2020

**DOI:** https://doi.org/10.21203/rs.2.24037/v1

**Version of Record:** A version of this preprint was published on September 22nd, 2020. See the published version at https://doi.org/10.1186/s12870-020-02622-x.

# Abstract

**Background:** MicroRNAs (miRNAs) and their regulatory functions in anthocyanin, carotenoid, and chlorophyll accumulation have been extensively characterized in many plant species. However, the miRNA regulatory mechanism in betalain biosynthesis remains mostly unknown.

**Results:** In this study, 126 conserved miRNAs and 41 novel miRNAs were first isolated from *Hylocereus monacanthus*, among which 95 conserved miRNAs belonged to 53 miRNA families. 34 candidate miRNAs related to betalain biosynthesis were found to be differentially expressed. The expression patterns of those differential expressed miRNAs were analyzed in various tissues of the pitaya by RT-qPCR. A significantly negative correlation was detected between the expression levels of half those miRNAs and corresponding target genes. Target genes of miRNAs i.e. aly-miR157d-5p_L+1_1ss4AC-comp25631_c0, aau-miR160_L-4R+ 1-comp36993_c0_seq3, nta-miR6020b-comp234190_c0, PC-5p-192_7269-comp29967_c0, PC-5p-23845_39-comp28219_c0, mdm-miR828a_1ss22AT-comp24967_c0, mdm-miR858- comp15143_c0, mdm-miR858-comp24362_c0 and mdm-miR858-comp403340_c0 were verified by 5'RACE and transient expression system in tobacco.

**Conclusions:** aly-miR157d-5p_L+1_1ss4AC, aau-miR160_L-4R+1, nta-miR6020b PC-5p-192_7269, PC-5p-23845_39, mdm-miR828a_1ss22AT and mdm-miR858 may play important roles in pitaya fruit coloration and betalain accumulation. Our findings provide insights into the roles of miRNAs and their target genes of regulatory functions involved in betalain biosynthesis of pitaya.


# Background

Mature micro RNAs (miRNAs) are a type of endogenous non-coding small RNAs with 20–24 nucleotide (nt) length. miRNAs regulate their target genes by mRNA at the post-transcriptional level via the RNA-induced silencing complex (RISC) by binding with the ARGONAUTE (AGO) protein to cleavage target mRNA or repress translation of target mRNA [1]. miRNAs play vital roles in plant growth and development, (a)biotic stress response, post-transcriptional regulation, and pigment regulation [2–4].

miRNAs are involved in chlorophyll, carotenoid, and anthocyanin biosynthesis. miRNAs can regulate coloration and chlorophyll accumulation [3, 5–7]. Overexpression of osa-miR171b by an artificial miRNA could enhance chlorophyll accumulation in rice leaves [7]. Based on the reduction of chlorophyll concentration under severe drought stress, a miRNA regulatory network consisting of 2 up-regulated miRNAs and 19 down-regulated miRNAs was constructed in Camellia sinensis [3]. miRNAs are involved in regulating carotenoid pathways [8–12]. miR1857 affected carotenogenesis in a sweet orange red-flesh mutant and its wild type [8]. In different rose cultivars, miRNAs may negatively regulate target genes to prevent carotenoid accumulation resulting in white flowers according to expression analyses of five miRNAs [9]. miRNAs can also regulate anthocyanin accumulation through their target genes [13–25]. Anthocyanin accumulation is promoted with increasing of miR156 abundance in Arabidopsis [13] and litchi [20]. However, the reduction of miR156 abundance could lead to the accumulation of flavonols [13].

miR156 positively regulates anthocyanin accumulation by targeting the SPL transcription factors (TFs) [23]. miR828 negatively regulates anthocyanin accumulation by inhibiting the expression of MYB75, MYB90, and MYB113 in Arabidopsis [14]. Moreover, miR858 and its targets involved in anthocyanin accumulation have been identified from apple, cotton, and tomato [15, 17, 18].

Pitaya, also known as pitahaya or dragon fruit, is a perennial climbing fruit crop belonging to the genus Hylocereus (Cactaceae). As a member of the Cactaceae, these species exhibit a range of specific adaptations to arid lands, i.e., spines instead of leaves, succulent shoots, and the crassulacean acid metabolism (CAM) pathway for $CO_2$ fixation, mechanisms that result in high water use efficiency [26]. Pitaya is an excellent plant material for basic and applied biological research since it is a spiny succulent plant that can adapt to various ecological environments, such as heat, drought, and poor soil [27, 28]. The potential economic impact of pitaya lies in their diverse uses not only as agricultural produce and processed foods but also in industrial and medicinal products. Pitaya is also a fast-return fruit crop with production in the second year after planting and full production in 3–4 years. Pitaya fruit is ripening 28–50 days (28–35 days in summer and 35–50 days in autumn) after flowering and has 7–9 separate fruiting cycles per year due to climatic or nutritional limitations. Therefore, pitaya has become a favorite of many farmers and home gardeners in Southeast Asia, China, the United States, Israel, Australia, Cyprus and the Canary Islands.

The color of pitaya is attributed to the presence of betalain [29–32]. Betalains are red and yellow alkaloid pigments that are found in some families, including Cactaceae. Betalains in pitaya fruit are not only good for human health but also can help consumers distinguish cultivars [29, 33, 34]. Betalains also play vital roles in the protection against drought, UV radiation, high the protection against drought, UV radiation, high saline soils, and diseases [34–38]. Pitaya is the only at large-scale commercially grown fruit containing abundant betalains for the consumer. Previous studies are mainly focused on characterizations of key genes and TFs involved in betalain biosynthesis. Key genes such as tyrosinase (TYR), cytochrome P450 (Cyt P450), 4,5-dihydroxy-phenylalanine (DOPA)-dioxygenase (DOD) and glucosyltransferases (GTs) [39] and TFs such as WRKY and MYB involved in betalain biosynthesis have been investigated in detail [40–42]. However, the roles of miRNAs in betalain biosynthesis has not been reported yet. In this study, candidate miRNAs and their target genes related to betalain biosynthesis were identified based on small RNA and transcriptome databases at different developmental stages of pitaya pulp. The present study aims to explore the roles of miRNA in pitaya betalain biosynthesis, which may contribute to a better understanding of betalain biosynthesis in Hylocereus.

# Results

### Sequencing of sRNAs and the transcriptome

Six sRNA libraries were generated to identify miRNAs involved in the fruit development of the pitaya. The Illumina sequencing data of sRNAs from the $19^{th}$, $25^{th}$ and $29^{th}$ day after flowering (DAF) showed that 24 nt sRNAs are the most abundant, followed by 21 nt sRNAs (Fig. 1). A total of 82,318,241 reads were

obtained from the datasets. After removal of the adaptor, insert, polyA and short RNAs of <18 nt in length, 62,729,725 (76.20%) valid reads were obtained, including the rRNA, tRNA, snRNA, snoRNA and some other Rfam RNA (Table 1). 11,455,464, 10,976,049, 12,337,213, 8,608,731, 9,222,106 and 10,130,162 unique matches for the sRNAs were identified as valid reads in Hp19d_1, Hp19d_2, Hp25d_1, Hp25d_2, Hp29d_1 and Hp29d_2, respectively.

Three transcriptome libraries of pitaya pulps were constructed to identify the target genes of miRNAs. 7.46 Gb, 9.10 Gb and 10.23 Gb nucleotide data (Q20 < 20) were obtained from Hp29d, Hp25d and Hp29d libraries, respectively. All screened reads were de novo assembled into 68,505 transcripts with an N50 of 1,700 bp. And then, these transcripts were assembled into 39,737 unique sequences with an average length of 879 bp. The length distributions of those transcripts and unigenes were shown in Fig. 2. The distribution of assembled transcripts and genes with different GC contents in the transcriptome datasets were summarized in Fig. 3. The majority of transcripts and genes were in the range of 35%–50% in GC contents.

## Identification of conserved miRNAs

A total of 126 conserved miRNAs were identified based on BLAST searches and sequence analyses in comparison of sRNA sequences with known mature plant miRNAs in miRBase (Table S1). Among them, 95 known miRNAs belonging to 53 families were obtained. The number of miRNA members for each family varied from 1 (for MIR535) to 7 (for MIR482) (Fig. 4). MIR160, MIR156, MIR159, MIR166, MIR164 and MIR396 families consisted of 2, 3, 4, 4, 5 and 6 members, respectively. The higher expression levels of the miRNA, the more copies of the miRNAs were sequenced. Among the 126 conserved miRNAs, four miRNAs named as sly-miR164a-3p_1ss12TC, ppe-miR535a, ptc-miR164a and mtr-miR396b-5p had the most reads, reaching up to 10,000 (Table S1). Additionally, nine miRNAs (ama-miR396-5p_1ss4CA, gma-miR396a-5p_L+1_1ss22GT, mtr-miR167b-5p, aly-miR168a-5p, peu-MIR2916-p5_1ss5AG, peu-MIR2916-p3_1ss8TC, osa-miR529b_R-1_1ss8AG, gma-miR6300, and gma-miR393a_R+1) were sequenced in thousands. While another six miRNAs (nta-miR6147, stu-miR7122-5p_1ss18TG, mdm-miR403a_R+1_2ss20CT21GT, sly-miR403-5p_L-2_1ss15TC, osa-miR319a-3p.2- 3p_1ss9AG, and gma-miR172b-5p_L-2R+2) had only one read sequenced at one stage. Changes in expression levels of miRNAs during fruit development of pitaya possibly reflected their potential functional differences.

## Identification of novel miRNAs

Novel miRNAs were predicted using the sRNA valid reads based on structure and expression criteria [43]. 41 novel candidate miRNAs with a clear precursor including stem-loop secondary structure were identified (Table S2). The length of novel miRNAs was between 20-25 nt. The most abundant sequences (49%) were 21 nt-length, and followed by 24 nt-length (29%), which was consistent with typical length distribution of mature miRNAs. The abundance of known miRNAs was higher than that of putative novel miRNAs, except PC-5p-192_7269, PC-5p-2548_357, PC-5p-1835_491 and PC-3p-3086_289, which possessed more than 100 normalized reads (Table S1 and S2).

## Analyses of Differentially Expressed miRNAs

High-throughput sequencing (HTS) was performed to explore the expression level changes of miRNAs involved in betalain biosynthesis of 'Guanhuahong' pitaya. Only miRNAs with expression values at $p<0.05$ were considered to be significantly regulated. In Hp29d/Hp25d/Hp19d, 33 known miRNAs and five novel miRNAs were found to be differentially expressed (Table 2).

In different pulp coloration stages of 'Guanhuahong' pitaya, the expression levels of mtr-miR396b-3p_L+2R-2, ppe-miR535a, PC-5p-2975_303 and PC-3p-53338_18 at the red pulp stages ($25^{th}$ and $29^{th}$ DAP) were significantly higher than that of the white pulp stage ($19^{th}$ DAP) (Table 2). Twelve differentially expressed miRNAs i.e. aly-miR390a-3p, nta-miR6149a, zma-miR159c-3p_L+2R-1, lus-miR530a_R+1_1ss20TG, gma-miR6300, aly-miR390a-5p, nta-miR6020b, gma-miR394a-5p_1ss2TG, osa-miR5072_L-4_1ss12CT, PC-5p-23845_39, gma-miR171c-3p_1ss19AC and stu-miR482a-5p_2ss3AT12AG showed a lower expression level at the red pulp stages ($25^{th}$ and $29^{th}$ DAP) compared to a relatively higher expression level at white pulp stage ($19^{th}$ DAP). Expression levels of ssp-miR156_L+1R-1_1ss4AC, aly-miR157d-5p_L+1_1ss4AC, ptc-miR164a, sly-miR164a-3p_1ss12TC and ppe-miR399a increased gradually from $19^{th}$ DAF to $29^{th}$ DAF in 'Guanhuahong' pitaya. However, the expression levels of ssl-miR171b_1ss21TC, mdm-miR408a_1ss1AT, cme-miR393a_R+1, gma-miR398c, ath-miR8175_L+4, aly-miR398a-3p, aly-miR168a-3p_L+3_1ss19CT, osa-miR529b_R-1_1ss8AG, zma-miR398a-3p_R-2, gma-miR172a_1ss1AC, nta-miR159 and lus-miR397a_L-1R+2 decreased gradually from $19^{th}$ DAF to $29^{th}$ DAF during fruit maturation of 'Guanhuahong' pitaya.

## Bioinformatics of transcriptome analyses

All 39,737 unique sequences were annotated according to BLAST (cut-off E-value $\leq 10^{-5}$) searches of Nr, Swiss-Prot Protein, Pfam, Gene Ontology (GO), euKaryotic Ortholog Groups (KOG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Table 3). Of these, 21,783 unique sequences (54.8%) could be annotated in the Nr database, while 7,966 unigenes (20.1%) were annotated using the KEGG databases.

The GO, KEGG and KOG databases were used to classify the functions of the predicted pitaya pulp unigenes. 12,559 unigenes were classified into three main categories: 'biological process', 'cellular component', and 'molecular function'. More than 150 unigenes involved in the annotation components were integrated (Fig. 5). As for the 'molecular function' category, the largest number of unigenes was annotated as 'ATP binding', while the major groups for the 'cellular component' category were 'integral to membrane' (2,878 unigenes, 23%), 'nucleus' (2,211 unigenes, 18%) and 'plasma membrane' (1,646 unigenes, 13%). In the category of 'biological process', 'regulation of transcription, DNA-dependent' (729 unigenes, 5.8%) and 'transcription, DNA-dependent' (706 unigenes, 5.6%) were the two most abundant subcategories.

7,822 unigenes were mapped onto KEGG pathways. The highest number of unigenes was carbohydrate metabolism (796 unigenes), followed by energy metabolism (596 unigenes), amino acid metabolism (595 unigenes) and translation (510 unigenes) (Fig. 6). Differential expression genes (DEGs) in the Hp25d-VS-Hp19d, Hp29d-VS-Hp19d, and Hp29d-VS-Hp25d were analyzed. In the Hp25d-VS-Hp19d, 24, 20, 16 and 13 unique sequences that significantly differentially expressed from phenylpropanoid, stilbenoid, diarylheptanoid and gingerol biosynthesis, DNA replication and flavonoid biosynthesis pathway were detected ($p<0.0001$), respectively (Fig. 7A). In Hp29d-VS-Hp19d, 140 genes were found to be significantly differentially expressed ($p<0.0001$), including biosynthesis of phenylpropanoid, flavonoid, stilbenoid, diarylheptanoid and gingerol, metabolism of phenylalanine, alpha-Linolenic acid, starch and sucrose, xenobiotics by Cyt P450, as well as drug metabolism-Cyt P450 (Fig. 7B). The expression levels of DEGs among different groups in one pathway showed different scales of change. 13 DEGs were detected from the flavonoid biosynthesis pathway in the Hp25d-VS-Hp19d and Hp29d-VS-Hp19d, respectively. However, no DEGs related to the flavonoid biosynthesis pathway were found in Hp29d-VS-Hp25d (Fig. 7C). The KOG database is used to study the classification and evolutionary rates of orthologous proteins. As shown in Fig. 8, group R (general function prediction only), group O (posttranslational modification, protein turnover, chaperones) and group T (signal transduction mechanisms) are the three most abundant groups in pitaya dataset, suggesting that pitaya fruit development involved in a large number of transcriptional and posttranslational regulation of gene expression and function.

DEGs from three development stages (Hp25d-VS-Hp19d, Hp29d-VS-Hp19d, and Hp29d-VS-Hp25d) were evaluated by pairwise comparisons using the expression fold ($|\log_2 \text{fold change}| \geq 1$) and values $p<0.05$ as the thresholds (Figure 9). In the pairwise comparisons between any two stages, 3,988 genes were found to be significantly differentially expressed. The highest amount of DEGs was obtained between the Hp19d and Hp29d libraries, including 1,486 down-regulated and 733 up-regulated (Fig. 9A). The lowest number of DEGs (1835) was detected between the Hp25d and Hp29d libraries (987 down-regulated and 848 up-regulated). followed by the Hp19d and Hp25d libraries (1,961 DEGs, 1,274 down-regulated and 687 up-regulated). Among those DEGs, 124 genes were significantly differentially expressed in all three fruit development stages of pitaya (Fig. 9B).

## Verification of the accuracy of the sRNAome and RNA-Seq data

To verify the reliability of the sRNAome and RNA-Seq results, the expression of fifteen miRNAs and fifteen genes related to betalain biosynthesis were analyzed by RT-qPCR. The IDs, reads per kilobase of exon model per million mapped read (RPKM) value, and primers for the 30 transcripts were shown in Table S1, S2, S4, S5, and S7, respectively. The overall correlation coefficient of 0.717** and 0.719** were obtained by linear regression [(Q-PCR value) = a (sRNAome or RNA-Seq value) + b] analysis (Fig. 10), respectively, indicating that the results of sRNAome and transcriptome analyses were consistent with those of RT-qPCR. Those results suggested that sRNAome and RNA-Seq data can be used for subsequent experiments.

## Prediction of targets for differentially expressed miRNAs

To investigate the functions of miRNAs in betalain biosynthesis of pitaya, it is crucial to predict their target genes. Target genes of differentially-expressed miRNAs were predicted by Target Finder software. Transcripts of transcriptome were identified as possible target genes for the majority of miRNAs. A total of 124 target mRNAs, 22 target mRNAs for conserved miRNAs and novel miRNAs were obtained, respectively (Table S8). More than one target gene was predicted for most miRNAs. miRNAs have the potential to regulate targets belonging to certain gene families with different biological functions. For example, the predicted target genes of osa-miR529b_R-1_1ss8AG were found to be involved in SPL6 (squamosa promoter-binding-like protein 6), histone-lysine N-methyltransferase and ubiquitin-protein ligase. Six targets may be involved in betalain biosynthesis. Among them, *comp29967_c0* (*Cyt P450 83B1*), *comp234190_c0* (*Cyt P450 71A8*), *comp28219_c0* (*protein-tyrosine sulfotransferase*), *comp27657_c0* (*WD and tetratricopeptide repeats protein 1*) and *comp25650_c0* (*SPL*) were respectively targeted by PC-5p-192_7269, nta-miR6020b, PC-5p-23845_39, ptc-miR164a and ssp-miR156_L+1R-1_1ss4AC while *comp25631_c0* (*SPL6*) was co-targeted by ssp-miR156_L+1R-1_1ss4AC, aly-miR157d-5p_L+1_1ss4AC and osa-miR529b_R-1_1ss8AG.

## Tissue-specific analyses of differentially expressed miRNAs

miRNA preferentially expressed in specific tissues might provide clues to its physiological functions. Thirty differentially expressed miRNAs were analyzed their functions by RT-qPCR using ten different tissues from ‘Guanhuabai’ and ‘Guanhuahong’ pitayas. As shown in Figure 11, the 30 differentially expression miRNAs showed different expression levels in those tissues. Gma-miR172a_1ss1AC, gma-miR394a-5p_1ss2TG, lus-miR530a_R+1_1ss20TG, nta-miR6020b, lus-miR397a_L-1R+2, ttu-miR160_1ss21AG and zma-miR398a-3p_R-2 displayed similar expression patterns. However, ssp-miR156_L+1R-1_1ss4AC strongly expressed in pitaya roots and fruits. Aly-miR398a-3p strongly expressed in pitaya stamens compared to weak expression in the other tissues. PC-5p-2975_303 and gma-miR171c-3p_1ss19AC strongly expressed in petals but moderately or weakly expressed in the other tissues. Ptc-miR164a and sly-miR164a-3p_1ss12TC displayed expression in all tissues and the highest expression level was detected in pitaya fruits. Mtr-miR396b-3p_L+2R-2 and osa-miR5072_L-4_1ss12CT had higher expression in petals of ‘Guanhuabai’ pitaya, but ssl-miR171b_1ss21TC and PC-5p-192_7269 showed higher expression in petals of ‘Guanhuahong’ pitaya. PC-3p-53338_18, PC-5p-1835_491, PC-5p-23845_39, cme-miR393a_R+1, aly-miR390a-5p, nta-miR159 and mdm-miR408a_1ss1AT had preferential expression in petals and receptacles of ‘Guanhuahong’ pitaya. Aly-miR157d-5p_L+1_1ss4AC and gma-miR6300 were highly expressed in ‘Guanhuahong’ and ‘Guanhuabai’ pitayas, respectively. Results from RT-qPCR indicated that expression patterns of the 30 pitaya miRNAs had tissue- and/or cultivar- characteristics in terms of different expression levels.

## Validation of miRNAs and target genes related to belatain biosynthesis

Fourteen target genes predicted from 11 miRNAs were assayed by RT-qPCR. The target genes from the ‘Guanhuahong’ and ‘Guanhuabai’ pitayas had the same sequences (Fig. S2). Sixteen targets showed decreased or increased expression trends along with the increased or decreased expression of the miRNA,

suggesting that they might be actively cleaved by miRNAs (Fig. 12). For example, *comp36993_c0_seq3* (Fig. 12A9 and B9) and *comp35191_c0* (Fig. 12A10 and B10) targeted by aau-miR160_L-4R+1, *comp15143_c0* (Fig. 12A5), *comp24362_c0* (Fig. 12A6) and *comp403340_c0* (Fig. 12A7) targeted by mdm-miR858, and *comp234190_c0* (Fig. 12A2) targeted by nta-miR6020b showed increasing at first and decreasing thereafter while aau-miR160_L-4R+1, mdm-miR858 and nta-miR6020b showed decreasing at first and increasing thereafter at all pulp coloration stages of pitaya. *comp25631_c0* (Fig. 12A15) targeted by ssp-miR156_L+1R-1_1ss4AC showed decreasing while ssp-miR156_L+1R-1_1ss4AC showed increasing at all pulp coloration stages of pitaya. However, some targeted genes and their miRNAs such as *comp28219_c0* (Fig. 12B11) targeted by PC-5p-23845_39 in 'Guanhuahong' pitaya and *comp25631_c0* (Fig. 12A13) targeted by osa-miR529b_R-1_1ss8AG in 'Guanhuabai' pitaya had similar expression patterns at all pulp stages of pitaya.

5'RACE analyses were further verified the fourteen candidate targets. The *comp25631_c0* (SPL6), *comp24967_c0* (TF TT2), *comp15143_c0* (TF MYB12), *comp24362_c0* (anthocyanin regulatory C1 protein), *comp36993_c0_seq3* (*Hpcyt P450-like3*), *comp29967_c0* (*cyt P450 83B1*) and *comp28219_c0* (protein-tyrosine sulfotransferase) were confirmed to be cleaved by their corresponding miRNAs (Fig. 13). The cleavage sites of PC-5p-192_7269 on *comp29967_c0* and PC-5p-23845_39 on *comp28219_c0* were both occurred at the $10^{th}$ nucleotide from the 5'-end of miRNAs in the binding region. The cleavage frequency of PC-5p-192_7269 on *comp29967_c0* and PC-5p-23845_39 on *comp28219_c0* was up to 10/10 in 'Guanhuabai' and in 'Guanhuahong' pitayas, respectively. This finding confirmed that PC-5p-192_7269 and PC-5p-23845_39 can guide the cleavage of the mRNA of *comp29967_c0* and *comp28219_c0*, respectively. Cleavage occurred mostly at the $10^{th}$ nucleotide from the 5'-end in the binding sites. However, *comp24362_c0* occurred at the $9^{th}$ nucleotide in 'Guanhuahong' pitaya, which may be due to the wide range of miRNA cutting mRNA caused by siRNA interference. Besides, the same gene from 'Guanhuahong' pitaya is different from 'Guanhuabai' pitaya (Fig. 13), which may be responsible for the different pulp colors of the two pitayas.

### The relationship between miRNAs and their target genes

A total of 5 miRNAs and 7 corresponding target genes i.e. aau-miR160_L-4R+1-*comp35191_c0*, nta-miR6020b-*comp234190_c0*, PC-5p-192_7269- *comp29967_c0*, PC-5p-23845_39-*comp28219_c0*, mdm-miR858-*comp15143_c0*, mdm-miR858-*comp24362_c0* and mdm-miR858-*comp403340_c0* were verified their interaction in tobacco transient expression system (Fig. 14). This interaction could affect miRNA processing resulting in higher precursor accumulation and reduced mature miRNA in pitaya, and further regulate betalain biosynthesis of pitaya.

# Discussion

Great progressed have been made in pitaya betalain in term of physical and chemical properties [44, 45], purification and identification [46–51], antioxidant and radical scavenging capacity [52–54] as well as metabolic and transcriptional analyses [55, 56]. However, no information is availabl about miRNAs

involved in betalain biosynthesis of pitaya. Therefore, identification of pitaya miRNAs associated with their target genes will help our understanding of molecular regulatory mechanisms of betalain biosynthesis of pitaya. In this study, transcriptome and sRNAome were performed to explore the role of miRNAs in coloration mechanisms of pitaya at different developmental stages. The length of sRNAs ranged from 18 to 25 nt (Fig. 1), a narrow range compared to 16–35 nt in previous report [21]. The percentage of 24 nt sRNAs (an average of 58.03%) was much higher than that of 21 nt sRNAs (an average of 12.16%) in pitaya pulp (Fig. 1). The result was consistent with the length distribution of sRNAs in apple and litchi [15, 20, 57] but was inconsistent with the findings in strawberry, orange, Brassica juncea, and apple [21, 58–60]. The distribution of different length sRNAs could be due to different plant species. A total of 95 known miRNAs belonging to 53 miRNA families and 41 new miRNAs were identified from ‘Guanhuahong’ pitaya (H. monacanthus) (Table S1 and S2). miRNAs showed different expression levels during fruit developmental stages of pitaya, indicating miRNAs play essential roles in pitaya fruit development.

Generally, conserved miRNAs had the same or homologous targets as other plant species, and most of them have a similar function. MIR156 plays key role in the biosynthesis of secondary metabolites. MIR156 can positively regulate anthocyanin biosynthesis by SPL TFs, while SPL TFs negatively regulate anthocyanin accumulation in Arabidopsis, apple, litchi, and Pyrus pyrifolia [13, 20, 21, 23]. Bgy-mir156 regulates the target gene phytoene synthase (PSY) and affects the accumulation of carotenoids in carrot (Daucus carota L.) [12]. Betalains are secondary metabolites but cannot co-exist naturally in one plant at the same time [61]. Pitaya is a high-value, functional fruit containing a high level of betalains. In our study, the highest expression level of aly-miR157d-5p_L + 1_1ss4AC was detected on the 23rd DAF (color conversion) in ‘Guanhuahong’ pitaya, suggesting that aly-miR157d-5p_L + 1_1ss4AC may play significant roles in betalain accumulation. Results from 5' RACE showed that comp25631_c0 (SPL6) was targeted by aly-miR157d-5p_L + 1_1ss4AC (Fig. 11). Those results indicated that aly-miR157d-5p_L + 1_1ss4AC might positively regulate betalain biosynthesis by SPL TFs, and SPL TFs negatively regulates betalain accumulation in pitaya.

MiR828 participated in anthocyanin biosynthesis by repressing the expression of MYB TFs in Arabidopsis, apple, and potato [14, 15, 21, 24]. In this study, the expression levels of comp24967_c0 (TF TT2) after 23rd DAF of ‘Guanhuahong’ pitaya was lower than that of the ‘Guanhuabai’ pitaya (Fig. 12A1 and B1). Comp24967_c0 showed a negative correlation with betalain accumulation in pitaya [55]. Mdm-miR828a_1ss22AT was highly active on the 23rd DAF of ‘Guanhuahong’ pitaya (color conversion). Mdm-miR828a_1ss22AT could target comp24967_c0 (Fig. 13B2), a MYB TF, suggesting that mdm-miR828a_1ss22AT positively regulate betalain accumulation in pitaya. miR858 could directly or indirectly control anthocyanin biosynthesis in Arabidopsis, cotton, apple and tomato, and it can negatively regulate anthocyanin accumulation by the MYB TF [14, 15, 17, 18]. In the present study, four target genes i.e., comp15849_c0 (HpMYB15), comp15143_c0 (HpMYB12), comp24362_c0 (HpMYBC) and comp403340_c0 (HpMYB2) had the same negative expression pattern with mdm-miR858 in ‘Guanhuabai’ pitaya (Fig. 12A5-A8, and B5-B8). 5'RACE and transient expression analyses showed that mdm-miR858 targeted HpMYB12, HpMYBC, and HpMYB2 in ‘Guanhuahong’ pitaya (Fig. 13B3-B4 and Fig. 14A12-A25,

B12-B25 and C12-C25). Those results suggested that mdm-miR858 can promote pitaya betalain accumulation by MYB genes.

Comp36993_c0_seq3 (Hpcyt P450-like3) is involved in betalain biosynthesis in H. monacanthus [55]. In this study, comp36993_c0_seq3 was targeted by aau-miR160_L-4R + 1, suggesting that aau-miR160_L-4R + 1 was involved in betalain biosynthesis (Fig. 13B5). Novel miRNAs are involved in accumulations of chlorophyll, carotenoid, and anthocyanin. Ttu-novel-48 could regulate chlorophyll accumulation in leaves of durum wheat [6]. Csi-novel-03 regulates carotenoid pathways by AP2 TFs in Citrus [8]. In strawberry receptacle fruit ripening, ABA-induced Fa_novel23 resulted in the rapid accumulation of fruit anthocyanin [25]. In the present study, two novel miRNAs PC-5p-192_7269 and PC-5p-23845_39 were obtained. PC-5p-192_7269 influenced betalain accumulation via the regulation of comp29967_c0 (Cyt P450 83B1) (Fig. 13A6-B6 and Fig. 14A34-A41, B34-B41, and C34-C41). PC-5p-23845_39 was identified as a regulator of betalain biosynthesis regulating the expression of comp28219_c0 (protein-tyrosine sulfotransferase) (Fig. 13B7 and Fig. 14A42-A49, B42-B49, and C42-C49). Nta-miR6020b-comp234190_c0 was not verified in 5'RACE but proved in the tobacco transient expression system (Fig. 14A4-A11, B4-B11 and C4-C11) suggested that the translation inhibited the regulation modes of miRNAs on their targets rather than the degradation of mRNAs. The rest of the differentially expressed miRNAs and their targets in HTS were not confirmed in 5'RACE and tobacco transient expression system, indicating that these genes might not be targets of these miRNAs. Further work is necessary to elucidate their roles in betalain biosynthesis of pitaya.

# Conclusions

In this study, sRNAome and RNA-Seq were first used to identify differentially expressed miRNAs and their target genes involved in betalain biosynthesis. Comprehensive sRNAome analysis uncovered 126 conserved miRNAs and 41 novel miRNAs were obtained from ‘Guanhuahong’ pitaya (H. monacanthus), among which 95 conserved miRNAs belonged to 53 miRNA families. 26.79 Gb raw RNA-Seq data were generated and de novo assembled into 68,505 transcripts, in which 39,737 were annotated. MiRNAs and their target genes involved in betalain accumulation were compared at different developmental stages of pitaya fruit. Seven target genes were verified by 5'RACE and tobacco transient expression system. Those Hp-miRNAs negatively regulated expression of their target mRNAs through guiding corresponding target mRNA cleavage or inhibiting the translation. Those results suggested that aly-miR157d-5p_L + 1_1ss4AC-comp25631_c0, aau-miR160_L-4R + 1-comp36993_c0_seq3, nta-miR6020b-comp234190_c0, PC-5p-192_7269-comp29967_c0, PC-5p-23845_39- comp28219_c0, mdm-miR828a_1ss22AT- comp24967_c0, mdm-miR858-comp15143_c0, mdm-miR858-comp24362_c0 and mdm-miR858-comp403340_c0 may be functionally involved in betalain biosynthesis in pitaya. The present study provides new information that miRNAs are actively involved in betalain accumulation of pitaya fruit by regulating the upstream TFs, which may contribute to further study on the roles of miRNAs in betalain biosynthesis of pitaya.

# Materials And Methods

## Plant materials

Two pitaya cultivars, i.e., 'Guanhuahong' (red skin with red flesh, *H. monacanthus*) and 'Guanhuabai' (red skin with white flesh, *H. undatus*) and *Nicotiana benthamiana* were used as plant materials. 'Guanhuahong' and 'Guanhuabai' pitayas, authenticated by Professor Guibing Hu and Yonghua Qin (College of Horticulture, South China Agricultural University), were selected through seedling selection from 860 seedlings of 'Hongshuijing' (*H. monacanthus*) [62-64]. *Nicotiana benthamiana* was grown in a greenhouse with a condition of 16 h/8 h day/night at 25 ℃ and was used for interactions between miRNAs and their target gene assays *in vivo*. The South China Agricultural University provided all plant materials used in this study, and no specific permissions were required for the collection of those samples for research purposes following institutional, national, and international guidelines. Fruits of 'Guanhuahong' and 'Guanhuabai' pitayas from the same orchard of Dalingshan Forest Park were separated into peels and pulps on the $13^{th}$, $16^{th}$, $19^{th}$, $23^{rd}$, $25^{th}$, $27^{th}$ and $29^{th}$ DAF (Figure S1) for expression analyses of crucial miRNAs and their targets. Pulps from 'Guanhuahong' pitaya on the $19^{th}$ DAF (white pulp stage, Hp19d, Hp19d_1, and Hp19d_2), $25^{th}$ DAF (pulp coloration stages, Hp25d, Hp25d_1, and Hp25d_2) and $29^{th}$ DAF (mature stage, Hp29d, Hp29d_1, and Hp29d_2) were used for RNA-Seq and sRNAome. All samples were frozen immediately in liquid nitrogen and stored at −80 ℃ until use.

## SRNAome and RNA-seq

Total RNA was extracted using the TruSeq Small RNA Sample Prep Kits (Illumina, San Diego, USA) according to the manufacturer's instructions. RNA-Seq and sRNAome libraries were performed according to the procedures of Han et al. (2016) and Liu et al. (2016), respectively [20, 65]. Six small RNA libraries (Hp19d_1, Hp19d_2, Hp25d_1, Hp25d_2, Hp29d_1 and Hp29d_2) from two biological replicates and three RNA-Seq libraries (Hp19d, Hp25d and Hp29d) were constructed (https://dataview.ncbi.nlm.nih.gov/object/PRJNA588519? reviewer=ko91rr55muepqo1d25plnp0kv4). All HTS was performed by LC-BIO (Hangzhou, China).

## Bioinformatic analysis

Clean sRNA sequences were obtained from sRNAome raw data (raw reads) by removing adapters, low-quality tags, and contaminants. Clean sRNA sequences were compared with the Rfam database (http://rfam.sanger.ac.uk/) after removing rRNA, tRNA, snRNA, and snoRNA. To screen known miRNAs, the clean data were mapped to the reference sequence in miRBase21.0 by Bowtie [66]. miRNA precursor was submitted to RNAfold software (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) to identify novel miRNA [43].

Transcriptome raw data was similar to pre-analyzed cropping of sRNAome raw data. Clean transcriptome data were assembled into nonredundant unigenes using Trinity (http://trinityrnaseq.github.io/). Unigenes were tentatively identified based on the best hits against known sequences in the database.

## Prediction and functional annotation of target gene

Target genes of differentially-expressed miRNAs were predicted by Target Finder software. GO, KOG and KEGG were used to explore the functions of target genes further. *E*-value ≤ $10^{-5}$ was considered as significant enrichment.

### Sequence alignments and phylogenetic analyses

Multiple sequences were aligned using DNAMan software (version 8). Phylogenetic trees were constructed using the neighbor-joining method of the MEGA 5.1 program. The miRNA target sites were identified using sequence alignments and manual analyses.

### Analysis of miRNAs and target genes by RT-qPCR

Stem-loop RT-qPCR was used to confirm expression of miRNAs since it is a highly sensitive method for detection of miRNAs [67]. cDNAs were produced from 1.0 μg of total RNA samples using the MMLV-reverse transcriptase (Invitrogen) with miRNA specific stem-loop and oligo(dT) primers, respectively. The specific primers and PCR reactions were performed according to our previous method [20]. RT-qPCR was conducted in ABI 7500 real-time PCR System (Applied Biosystems, CA, USA) using the SYBR qPCR Mix (Vazyme). Twenty microliters reaction mixture contained 2 μL of diluted cDNAs, 10 μL 2×SYBR qPCR Mix (Vazyme), 0.5 μL of each primer (10 μM) and 7 μL dd$H_2O$. All experiments were performed in triplicate. *U6* and *actin* gene were used as reference genes. The sequences of miRNAs and target genes were shown in Table S1, S2 and S9, respectively. All primers used for RT-qPCR analyses were listed in Table S3, S4 and S5. The expression levels of miRNAs and target genes were calculated by $2^{-\triangle\triangle C}{}_T$ method [68].

### 5'RACE analyses

To verify the miRNA-mediated cleavage events, RNA ligase-mediated 5'RACE (RLM-RACE) was performed using the SMARTer® RACE 5'/3'Kit User Manual (012615) (TaKaRa, Dalian, China) according to the manufacturer's manual. 1 μg total RNA from the pulps of the 29$^{th}$ DAF 'Guanhuahong' and 'Guanhuabai' pitayas was ligated with 5' RNA adapters, respectively. The ligated mRNA was reversely transcribed by oligo (dT) primer. 5' end products were obtained using 5' adaptor primers and 3' gene-specific primers. PCR products were inserted into the pMD18-T vector (TaKaRa). Specific primers used for nested PCR were shown in Table S6.

### Transient expression analysis

A transient expression system was used to confirm the interaction between miRNAs and their target genes *in vivo*. Construction of expression vectors were constructed following the procedure of Liu et al. [69]. Overexpression vectors of miRNAs related to the betalain biosynthesis and a control miRNA were constructed respectively. Co-expression of miRNAs and their targets in *N. benthamiana* leaves using *Agrobacterium tumefaciens* GV3101 infiltration. Transient expression in *N. benthamiana* was performed as described by Sparkes et al. (2006) [70]. Three days after infiltration, leaves were observed with a

fluorescence microscope (Zeiss Axio Observer D1). The transient expression assays were repeated three to five times.

# Abbreviations

5'RACE:5' rapid amplification of cDNA ends; AGO:ARGONAUTE; cDNA:complementary DNA; Cyt P450:cytochrome P450; DAF:day after flowering; DEG:differential expression gene; DOD:dioxygenase; DOPA:4,5-dihydroxy-phenylalanine; eGFP:enhanced green fluorescent protein; GO:gene ontology; GTs:glucosyltransferases; HTS:high-throughput sequencing; KEGG:kyoto encyclopedia of genes and genomes; KOG:eukaryotic ortholog groups; miRNA:microRNA; nt:nucleotide; PSY:phytoene synthase; RISC:RNA-induced silencing complex; RNA-Seq:RNA sequencing; RPKM:reads per kilobase of exon model per million mapped read; RT-qPCR:reverse transcription quantitative real-time polymerase chain reaction; TFs:transcription factors; TYR:tyrosinase.

# Declarations


### Acknowledgments

The authors would also like to thank fellow members of the orchard of Dalingshan Forest Park for management of experimental materials.


### Authors' contributions

YHQ and CBC conceived the study, and designed the experiments. CBC, FFX and QZH conducted the experiment. LLZ, ZKZ, RZ, JTZ and GBH analyzed data. CBC, NTZ and YHQ wrote the manuscript. All Authors read and approved the manuscript.


### Funding

This work was supported by the Key Realm R&D Program of Guangdong Province (2018B020202011), National Natural Science Foundation of China (31972367), Key Science and Technology Planning Project of Guangzhou (201904020015), Science and Technology Program of Guangzhou (201704020003 and 2014Y2-00164), and YangFan Innovative and Entepreneurial Research Team Project (2014YT02H013). The funding organizations paid the experimental costs and publication fees for this research but did not play any role in the design of the study nor in the collection, analysis and interpretation of data, nor in the writing of the manuscript.


### Availability of data and materials

The transcriptome clean raw reads data that support the findings of this study have been submitted to NCBI Sequence Read Archive (SRA) under Accession (SAMN13253272- SAMN13253280), Bioproject: PRJNA588519. All data generated or analyzed during this study are included in this published article and its supplementary information files. The authors are pleased to share with the data upon request.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no conflicts of interest.

# References


1. Rogers K, Chen XM. Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell. 2013; 25:2383-2399.
2. Liu YL, Ke LL, Wu GZ, Xu YT, Wu XM, Xia R, *et al*. miR3954 is a trigger of phasiRNAs that affects flowering time in *Citrus*. Plant J. 2017; 92:263-275.
3. Guo Y, Zhao S, Zhu C, Chang X, Yue C, Wang Z. Identification of drought-responsive miRNAs and physiological characterization of tea plant (*Camellia sinensis*) under drought stress. BMC Plant Biol. 2017; 17:211.
4. Feng H, Xu M, Zheng X, Zhu TY, Gao XN, Huang LL. MicroRNAs and their targets in apple (*Malus domestica* "Fuji") involved in response to infection of pathogen *Valsa mali*. Front. Plant Sci. 2017:8.
5. Canto-Pastor A, Molla-Morales A, Ernst E, Dahl W, Zhai J, Yan Y, *et al*. Efficient transformation and artificial miRNA gene silencing in *Lemna minor*. Plant Biol. 2015; 171:59-65.
6. De Paola D, Zuluaga DL, Sonnante G. The miRNAome of durum wheat: isolation and characterisation of conserved and novel microRNAs and their target genes. BMC Genom. 2016; 17:505.
7. Tong AZ, Yuan Q, Wang S, Peng JJ, Lu YW, Zheng HY, *et al*. Altered accumulation of osa-miR171b contributes to rice stripe virus infection by regulating disease symptoms. J. Exp. Bot. 2017; 68:4357-4367.
8. Xu Q, Liu YL, Zhu AD, Wu XM, Ye JL, Yu KQ, *et al*. Discovery and comparative profiling of microRNAs in a sweet orange red-flesh mutant and its wild type. BMC Genom. 2010; 11:246.
9. Kim J, Park JH, Lim CJ, Lim JY, Ryu JY, Lee BW, *et al*. Small RNA and transcriptome deep sequencing proffers insight into floral gene regulation in *Rosa* BMC Genom. 2012; 13:657.
10. Zeng SH, Liu YL, Pan LZ, Hayward A, Wang Y. Identification and characterization of miRNAs in ripening fruit of *Lycium barbarum* using high-throughput sequencing. Front. Plant Sci. 2015; 6:778.
11. Koul A, Yogindran S, Sharma D, Kaul S, Rajam MV, Dhar MK. Carotenoid profiling, in silico analysis and transcript profiling of miRNAs targeting carotenoid biosynthetic pathway genes in different developmental tissues of tomato. Plant Physiol. Biochem. 2016; 108:412-421.

12. Bhan B, Koul A, Sharma D, Manzoor MM, Kaul S, Gupta S, *et al*. Identification and expression profiling of miRNAs in two color variants of carrot (*Daucus carota*) using deep sequencing. PLoS ONE. 2019; 14:e212746.
13. Gou JY, Felippes FF, Liu CJ, Weigel D, Wang JW. Negative regulation of anthocyanin biosynthesis in arabidopsis by a miR156-targeted SPL transcription factor. Plant Cell. 2011; 23:1512-1522.
14. Luo QJ, Mittal A, Jia F, Rock CD. An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in *Arabidopsis*. Plant Mol. Biol. 2012; 80:117-129.
15. Xia R, Zhu H, An YQ, Beers EP, Liu ZR. Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol. 2012; 13:R47.
16. Yan J, Gu YY, Jia XY, Kang WJ, Pan SJ, Tang XQ, *et al*. Effective small RNA destruction by the expression of a short tandem target mimic in *Arabidopsis*. Plant Cell. 2012; 24:415-427.
17. Guan XY, Pang MX, Nah G, Shi XL, Ye WX, Stelly DM, *et al*. miR828 and miR858 regulate homoeologous *MYB2* gene functions in *Arabidopsis* trichome and cotton fibre development. Nat. Commun. 2014; 5:3050.
18. Jia XY, Shen J, Liu H, Li F, Ding N, Gao CY, *et al*. Small tandem target mimic-mediated blockage of microRNA858 induces anthocyanin accumulation in tomato. Planta. 2015; 242:283-293.
19. Wang L, Zengj HQ, Song J, Feng SJ, Yang ZM. miRNA778 and SUVH6 are involved in phosphate homeostasis in *Arabidopsis*. Plant Sci. 2015; 238:273-285.
20. Liu R, Lai B, Hu B, Qin YH, Hu GB, Zhao JT. Identification of microRNAs and their target genes related to the accumulation of anthocyanins in *Litchi chinensis* by high-throughput sequencing and degradome analysis. Front. Plant Sci. 2016; 7:e179.
21. Qu D, Yan F, Meng R, Jiang XB, Yang HJ, Gao ZY, *et al*. Identification of microRNAs and their targets associated with fruit-bagging and subsequent sunlight re-exposure in the "Granny Smith" apple exocarp using high-throughput sequencing. Front. Plant Sci. 2016; 7:27.
22. Roy S, Tripathi AM, Yadav A, Mishra P, Nautiyal CS. Identification and expression analyses of miRNAs from two contrasting flower color cultivars of *Canna* by deep sequencing. PLoS ONE. 2016; 11:e0147499.
23. Qian MJ, Ni JB, Niu QF, Bai SL, Bao L, Li JZ, *et al*. Response of miR156-*SPL* module during the red peel coloration of bagging-treated Chinese Sand Pear (*Pyrus pyrifolia* Nakai). Front. Physiol. 2017; 8:550.
24. Bonar N, Liney M, Zhang RX, Austin C, Dessoly J, Davidson D, *et al*. Potato miR828 is associated with purple tube skin and flesh color. Front. Plant Sci. 2018; 9:1742.
25. Li DD, Mou WS, Xia R, Li L, Zawora C, Ying TJ, *et al*. Integrated analysis of high-throughput sequencing data shows abscisic acid-responsive genes and miRNAs in strawberry receptacle fruit ripening. Res. 2019; 6:26.
26. Nobel PS, De La Barrera E. $CO_2$ uptake by the cultivated hemiepiphytic cactus, *Hylocereus undatus*. Ann. Appl. Biol. 2004; 144:1-8.

27. Mizrahi Y, Nerd A, Nobel PS. Cacti as crops. Hortic. Rev. 1997; 18:291-319.
28. Ortiz-Hernández YD, Carrillo-Salazar Pitaya (*Hylocereus* spp.): a short review. Comunicata Sci. 2012; 3:220-237.
29. Esquivel P, Stintzing FC, Carle R. Comparison of morphological and chemical fruit traits from different pitaya genotypes (*Hylocereus*) grown in Costa Rica. J. Appl. Bot. Food Qual. 2007; 81:7-14.
30. Stintzing FC, Carle R. Betalains-emerging prospects for food scientists. Trends Food Sci. Technol. 2007; 18:514-525.
31. Lee EJ, An D, Nguyen C, Patil BS, Kim J, Yoo KS. Betalain and betaine composition of greenhouse- or field-produced beetroot (*Beta vulgaris*) and inhibition of HepG2 cell proliferation. J. Agric. Food Chem. 2014; 62:1324-1331.
32. Suh DH, Lee S, Heo DY, Kim YS, Cho SK, Lee S, *et al*. Metabolite profiling of red and white pitayas (*Hylocereus polyrhizus* and *Hylocereus undatus*) for comparing betalain biosynthesis and antioxidant activity. J. Agric. Food Chem. 2014; 62:8764-8771.
33. Adnan L, Osman A, Hamid AA. Antioxidant activity of different extracts of red pitaya (*Hylocereus polyrhizus*) seed. J. Food Prop. 2011; 14:1171-1181.
34. Song H, Zheng Z, Wu J, Lai J, Chu Q, Zheng X. White pitaya (*Hylocereus undatus*) juice attenuates insulin resistance and hepatic steatosis in diet-induced obese mice. PLoS ONE. 2016; 11:e01496702.
35. Solovchenko AE, Merzlyak MN. Screening of visible and UV radiation as a photoprotective mechanism in plants. Russian J. Plant Physiol. 2008; 55:719-737.
36. Jamaludin NA, Ding P, Hamid AA. Physico-chemical and structural changes of red-fleshed dragon fruit (*Hylocereus polyrhizus*) during fruit development. J. Sci. Food Agric. 2011; 91:278-285.
37. Jain G, Gould KS. Functional significance of betalain biosynthesis in leaves of *Disphyma australe* under salinity stress. Environ. Exp. Bot. 2015; 109:131-140.
38. Jain G, Schwinn KE, Gould KS. Betalain induction by l-DOPA application confers photoprotection to saline-exposed leaves of *Disphyma australe*. New Phytol. 2015; 207:1075-1083.
39. Polturak G, Aharoni A. "la vie en rose": Biosynthesis, sources, and applications of betalain pigments. Mol. Plant. 2018; 11:7-22.
40. Stracke R, Holtgrawe D, Schneider J, Pucker B, Sorensen TR, Weisshaar B. Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet (*Beta vulgaris*). BMC Plant Biol. 2014; 14:249.
41. Hatlestad GJ, Akhavan NA, Sunnadeniya RM, Elam L, Cargile S, Hembd A, *et al*. The beet Y locus encodes an anthocyanin *MYB-like* protein that activates the betalain red pigment pathway. Nat. Genet. 2014; 47:92.
42. Cheng MN, Huang ZJ, Hua QZ, Shan W, Kuang JF, Lu WJ, *et al*. The WRKY transcription factor *HpWRKY44* regulates *CytP450-like1* expression in red pitaya fruit (*Hylocereus polyrhizus*). Hortic. Res. 2017; 4:17039.

43. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, *et al*. Criteria for annotation of plant MicroRNAs. Plant Cell. 2008; 20:3186–3190.
44. Esquivel P, Stintzing FC, Carle R. Pigment pattern and expression of colour in fruits from different *Hylocereus* genotypes. Innov. Food Sci. Emerg. 2007; 8:451–457.
45. Woo KK, Wong FNF, Chua HSC, Tang PY. Stability of the spray-dried pigment of red dragon fruit [*Hylocereus polyrhizus* (Weber) Britton and Rose] as a function of organic acid additives and storage conditions. Agric. Sci. 2011; 94:264–269.
46. Stintzing FC, Schieber A, Carle R. Identification of betalains from yellow beet (*Beta vulgaris*) and cactus pear [*Opuntia ficus-indica* (L.) Mill.] by high-performance liquid chromatography-electrospray ionization mass spectrometry. J. Agric. Food Chem. 2002; 50:2302-2307.
47. Wybraniec S, Stalica P, Jerz G, Klose B, Gebers N, Winterhalter P, *et al*. Separation of polar betalain pigments from cacti fruits of *Hylocereus polyrhizus* by ion-pair high-speed countercurrent chromatography. Chromatogr. A. 2009; 1216:6890–6899.
48. Naderi N, Stintzing FC, Ghazali HM, Manap YA, Jazayeri SD. Betalain extraction from *Hylocereus polyrhizus* for natural food coloring purposes. Prof. Assoc. Cactus. 2010; 12:143–154.
49. Rebecca OPS, Boyce AN, Chandran S. Pigment identification and antioxidant properties of red dragon fruit (*Hylocereus polyrhizus*). J. Biotechnol. 2010; 9:1450–1454.
50. Rebecca OPS, Harivaindaran KV, Boyce AN, Chandran S. Potential natural dye with antioxidant properties from red dragon fruit (*Hylocereus polyrhizus*). Acta Hort. 2010; 875:477–485.
51. Lim SD, Yusof YA, Chin NL, Talib RA, Endan J, Aziz MG. Effect of extraction parameters on the yield of betacyanins from pitaya fruit (*Hylocereus polyrhizus*) pulps. Food Agric. Environ. 2011; 9:158–162.
52. Wu LC, Hsu HW, Chen YC, Chiu CC, Lin YI, Ho JAA. Antioxidant and antiproliferative activities of red pitaya. Food Chem. 2006; 95:319–327.
53. Tenore GC, Novellino E, Basile A. Nutraceutical potential and antioxidant benefits of red pitaya (*Hylocereus polyrhizus*) extracts. Funct. Foods. 2012; 4:129–136.
54. Garcia-Cruz L, Valle-Guadarrama S, Salinas-Moreno Y, Joaquin-Cruz E. Physical, chemical, and antioxidant activity characterization of pitaya (*Stenocereus pruinosus*) fruits. Plant Food Hum. Nutr. 2013; 68:403–410.
55. Hua Q, Chen CJ, Chen Z, Chen PK, Ma YW, Wu JY, *et al*. Transcriptomic analysis reveals key genes related to betalain biosynthesis in pulp coloration of *Hylocereus polyrhizus*. Plant Sci. 2015; 6:1179.
56. Hua Q, Zhou QJ, Gan SS, Wu JY, Chen CB, Li JQ, *et al*. Proteomic analysis of *Hylocereus polyrhizus* reveals metabolic pathway changes. J. Mol. Sci. 2016; 17:1606.
57. Yao FR, Zhu H, Yi C, Qu HX, Jiang YM. MicroRNAs and targets in senescent litchi fruit during ambient storage and post-cold storage shelf life. BMC Plant Biol. 2015; 15:181.
58. Li H, Mao WJ, Liu W, Dai HY, Liu YX, Ma Y, *et al*. Deep sequencing discovery of novel and conserved microRNAs in wild type and a white-flesh mutant strawberry. 2013; 238:695–713.

59. Yang J, Liu X, Xu B, Zhao N, Yang X, Zhang M. Identification of miRNAs and their targets using high-throughput sequencing and degradome analysis in cytoplasmic male-sterile and its maintainer fertile lines of *Brassica juncea*. BMC Genom. 2013; 14:9.
60. Zhang XN, Li X, Liu JH. Identification of conserved and novel cold-responsive microRNAs in trifoliate orange [*Poncirus trifoliata* (L.) Raf.] using high-throughput sequencing. Plant Mol. Biol. Rep. 2014; 32:328-341.
61. Stafford HA. Anthocyanins and betalains: Evolution of the mutually exclusive pathways. Plant Sci. 1994; 101:91-98.
62. Ye YX, Hu GB, Li JQ, Qin YH, Gu WQ, Zhai X, *et al*. A new red-fleshed pitaya cultivar 'Guanhuahong'. Acta Hortic. Sinica. 2018; 45 (11): 2267-2268.
63. Ye YX, Hu GB, Li JQ, Qin YH, Zhou YW, Zhai X, *et al*. A new white-fleshed pitaya cultivar 'Guanhuabai'. Acta Hortic. Sinica. 2018; 45 (S2): 2739–2740.
64. Chen CB, Wu PY, Xie FF, Sun LY, Xing YM, Hua QZ, *et al.* Breeding of 'Hongguan No. 1' and 'Shuangse No. 1' Pitayas with superior quality. HortScience. 2018; 53(3): 404-409.
65. Han X, Yin H, Song X, Zhang Y, Liu M, Sang J, *et al*. Integration of small RNAs, degradome and transcriptome sequencing in hyperaccumulator *Sedum alfredii* uncovers a complex regulatory network and provides insights into cadmium phytoremediation. Plant Biotechnol. J. 2016; 14:1470-1483.
66. Langmead B. Aligning short sequencing reads with Bowtie. Curr. Prot. Bioinform. 2010; 11:Unit 11.7.
67. Chen CF, Ridzon DA, Broomer AJ, Zhou ZH, Lee DH, Nguyen JT, *et al*. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 2005; 33:e179.
68. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\triangle\triangle C}{}_T$ Methods. 2001; 25:402-408.
69. Liu YL, Wang L, Chen DJ, Wu XM, Huang D, Chen LL, *et al*. Genome-wide comparison of microRNAs and their targeted transcripts among leaf, flower and fruit of sweet orange. BMC Genom. 2014; 15:695.
70. Sparkes IA, Runions J, Kearns A, Hawes C. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat. Protoc. 2006; 1:2019–2025.

# Supplementary Files Legend

**Additional file 1: Figure S1.** Different fruit developmental stages of 'Guanhuahong' (A) and 'Guanhuabai' (B) pitayas. A1 and B1, 13 d; A2 and B2, 16 d; A3 and B3, 19 d; A4 and B4, 23 d; A5 and B5, 25 d; A6 and B6, 27 d; A7 and B7, 29 d. Bar=2.0 cm.

**Additional file 2: Figure S2.** The sequences of 17 target genes from the two pitaya cultivars. '-W', 'Guanhuabai' pitaya. '-R', 'Guanhuahong' pitaya. Red lines indicate sequence of target genes primers for real-time PCR. Green lines indicate outer specific primers for nested PCR. Blue lines indicate inner specific primers for nested PCR.

**Additional file 3: Table S1.** The information of conserved miRNAs in pitaya.

**Additional file 4: Table S2.** The information of putative novel miRNAs in pitaya.

**Additional file 5: Table S3.** Sequence of miRNA-specific primers for reverse transcription.

**Additional file 6: Table S4.** Sequence of miRNA primers for real-time PCR.

**Additional file 7: Table S5.** Sequence of target genes primers for real-time PCR.

**Additional file 8: Table S6.** Sequence of specific primers for 5'RACE.

**Additional file 9: Table S7.** The information of part RNA-Seq data in pitaya.

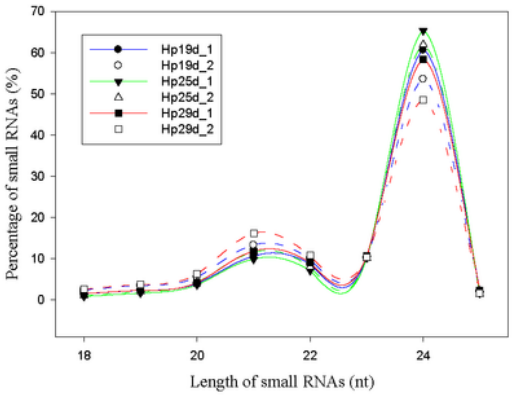
**Additional file 10: Table S8.** Prediction of targets for differentially expression miRNAs in pitaya.

**Additional file 11: Table S9.** cDNA sequences of 17 target genes.

# Tables

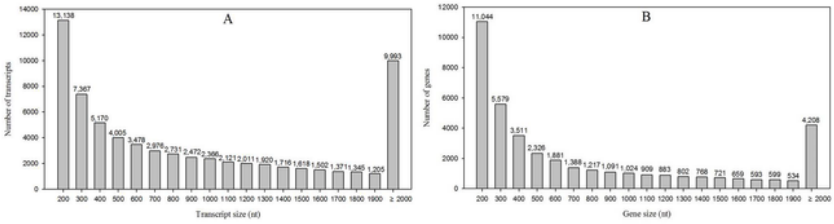
Due to technical limitations, Tables 1 - 3 are only available for download from the Supplementary Files section.

# Figures



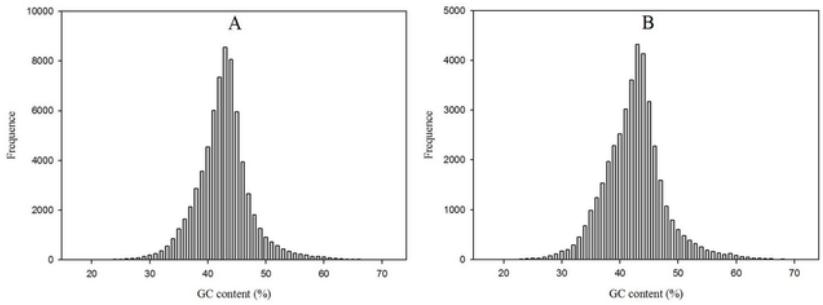

## Figure 1

The length distribution of miRNA fragments in pitaya (unique).



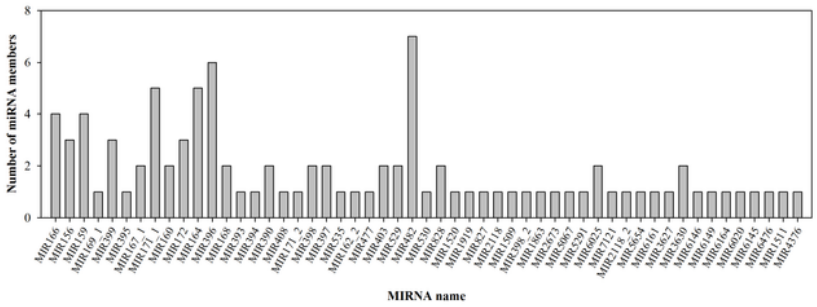

## Figure 2

The length distributions of assembled transcripts and genes. (A) The length distributions of transcripts. (B) The length distributions of genes.



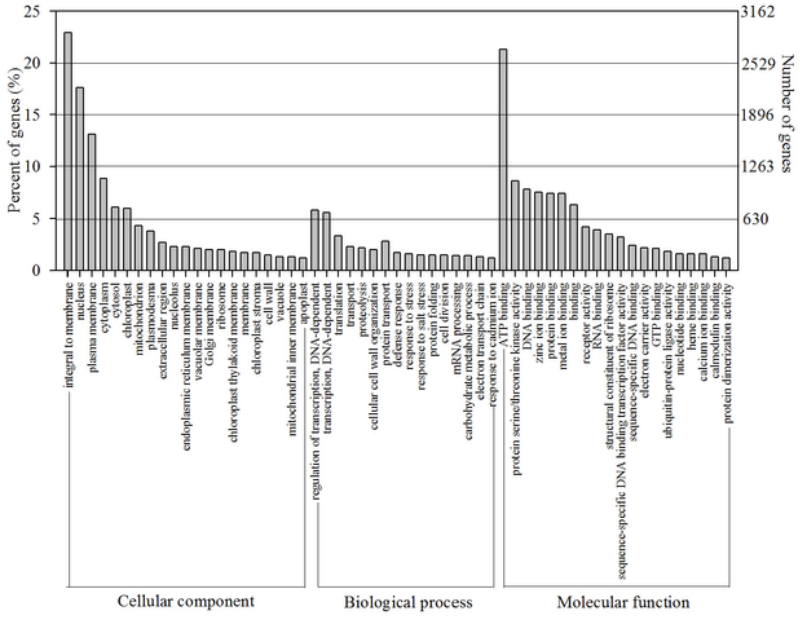

## Figure 3

GC content distributions of assembled transcripts and genes. (A) GC content distributions of transcripts. (B) GC content distributions of genes.



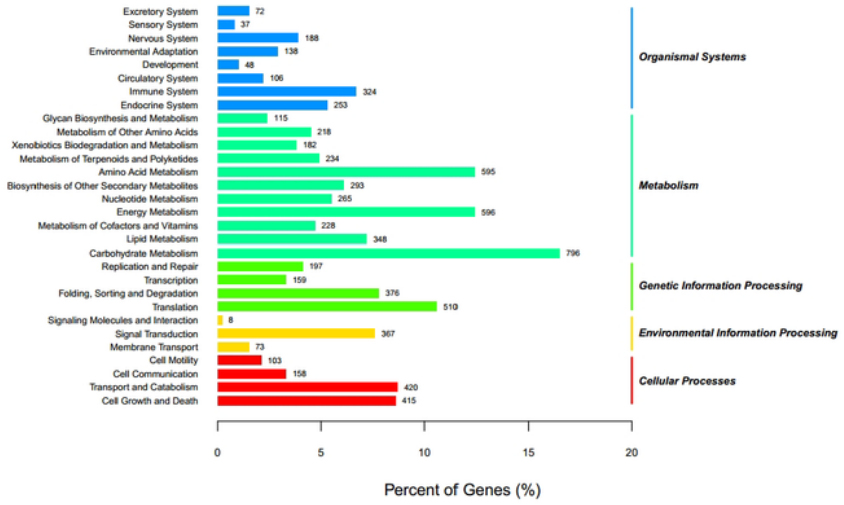

## Figure 4

Number of miRNA family members in pitaya pulp.



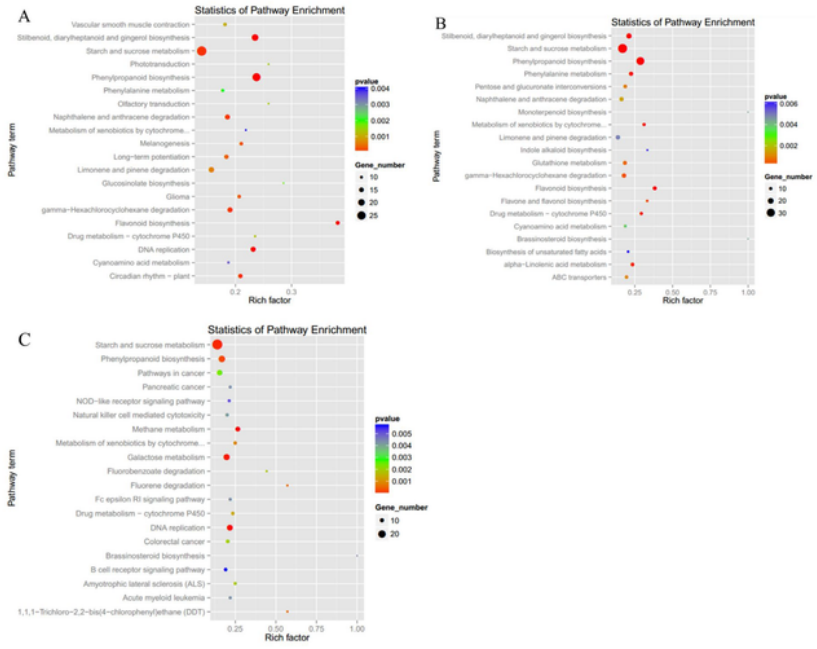

## Figure 5

Histogram of GO classifications for transcripts of pitaya pulp.



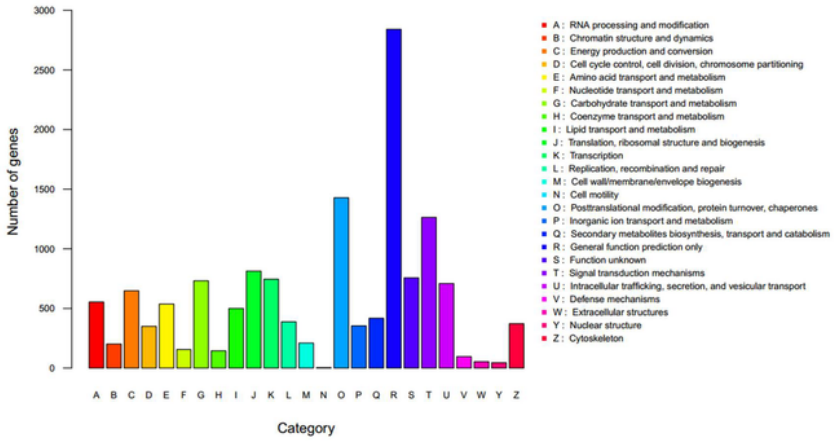

## Figure 6

KEGG pathway classifications for transcripts of pitaya pulp.



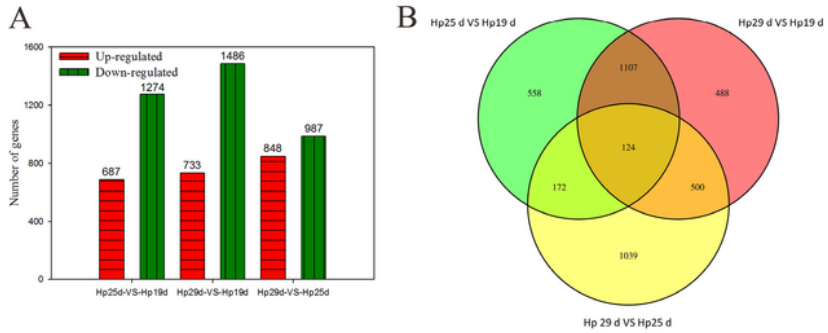

## Figure 7

KEGG pathway enrichment for differential gene expression of the three coloration stages. (A) Hp25 d-VS-Hp19 d. (B) Hp29 d-VS-Hp19 d. (C) Hp29 d-VS-Hp25 d.



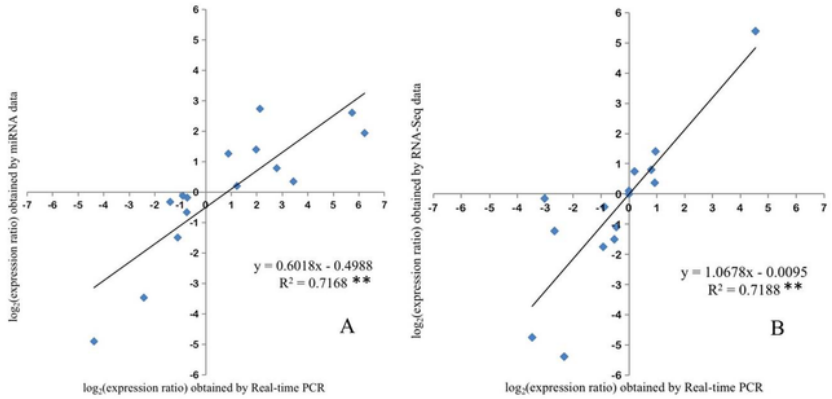

## Figure 8

Histogram of KOG functional categories for transcripts of pitaya pulp.



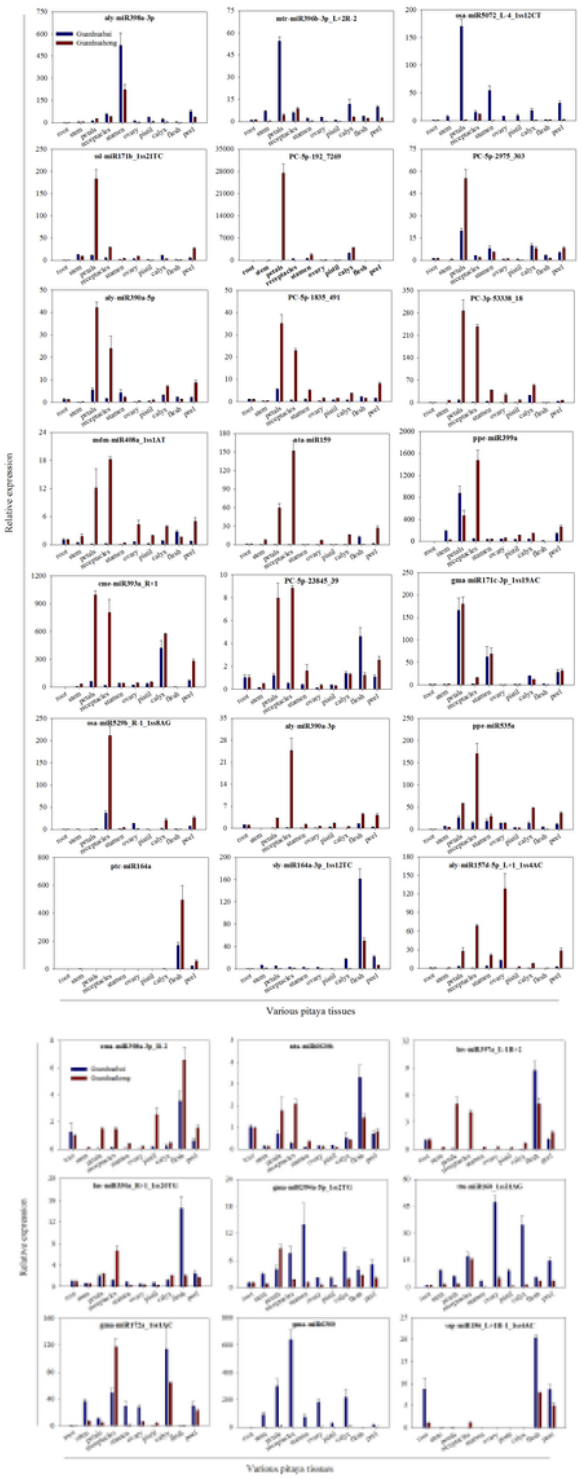

## Figure 9

Differential gene expression profiles based on the libraries of the three coloration stages. (A) The numbers of up- and down-regulated genes in comparisons of the Hp25 d-VS-Hp19 d, Hp29 d-VS-Hp19 d and Hp29 d-VS-Hp25 d. (B) Venn diagram showing the comparison of DEGs between any two stages of the pitaya pulp.



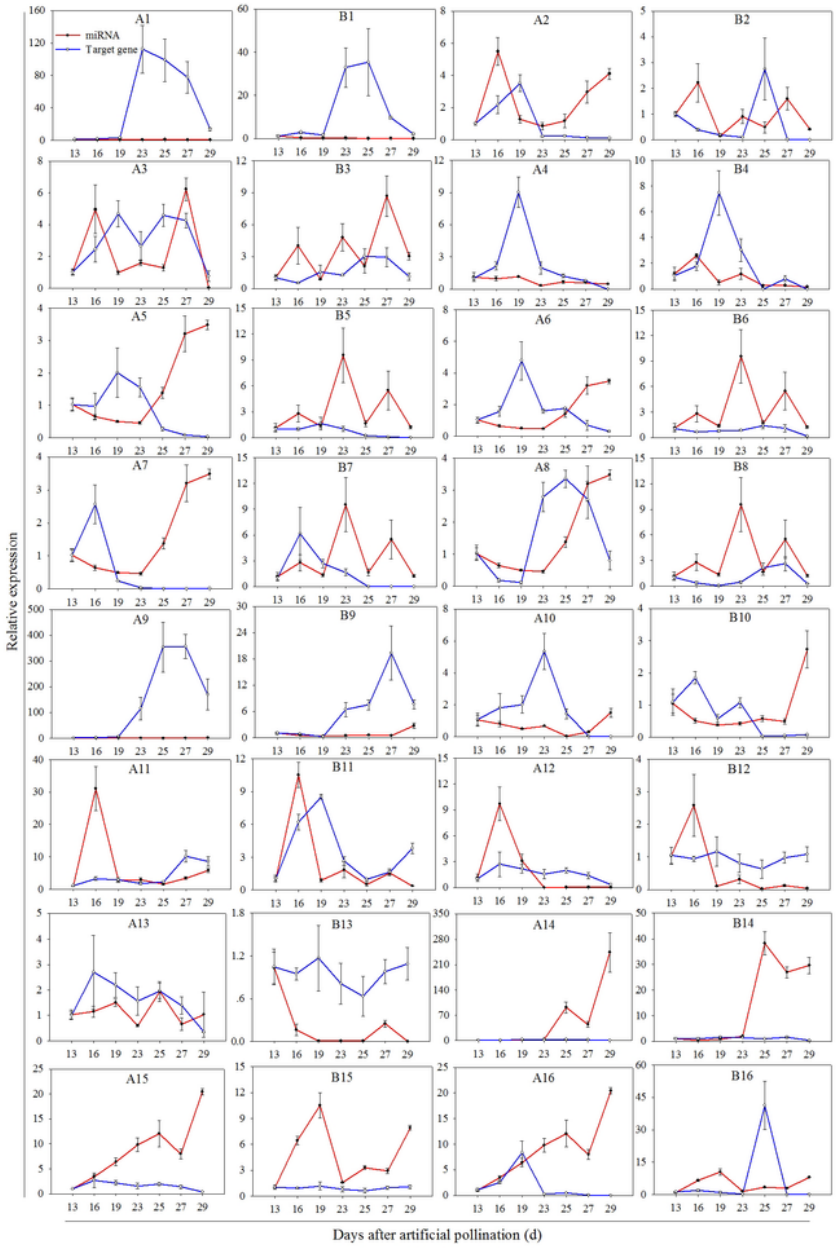

## Figure 10

Verification of the accuracy of the sRNAome and RNA-Seq data by coefficient analyses. (A) Coefficient analyses of miRNA expression levels from sRNAome and RT-qPCR data. (B) Coefficient analyses of gene expression levels from RNA-Seq and RT-qPCR data. The RT-qPCR log2 values (x-axis) were plotted against high-throughput data (y-axis). **indicates a significant difference at p□0.01 (n=15).




## Figure 11

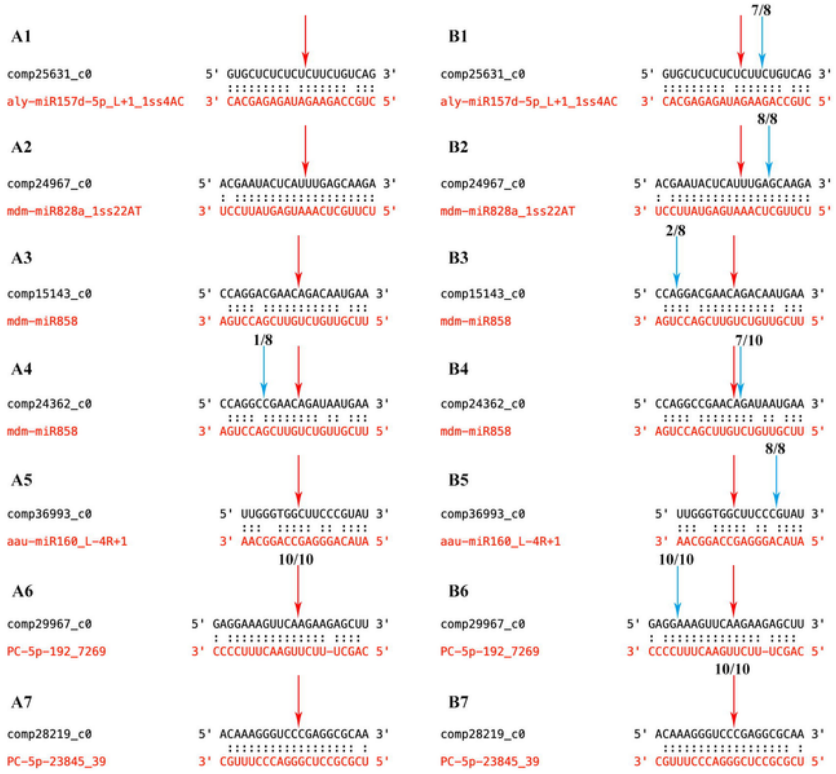
RT-qPCR analyses of miRNAs in various pitaya tissues.




## Figure 12

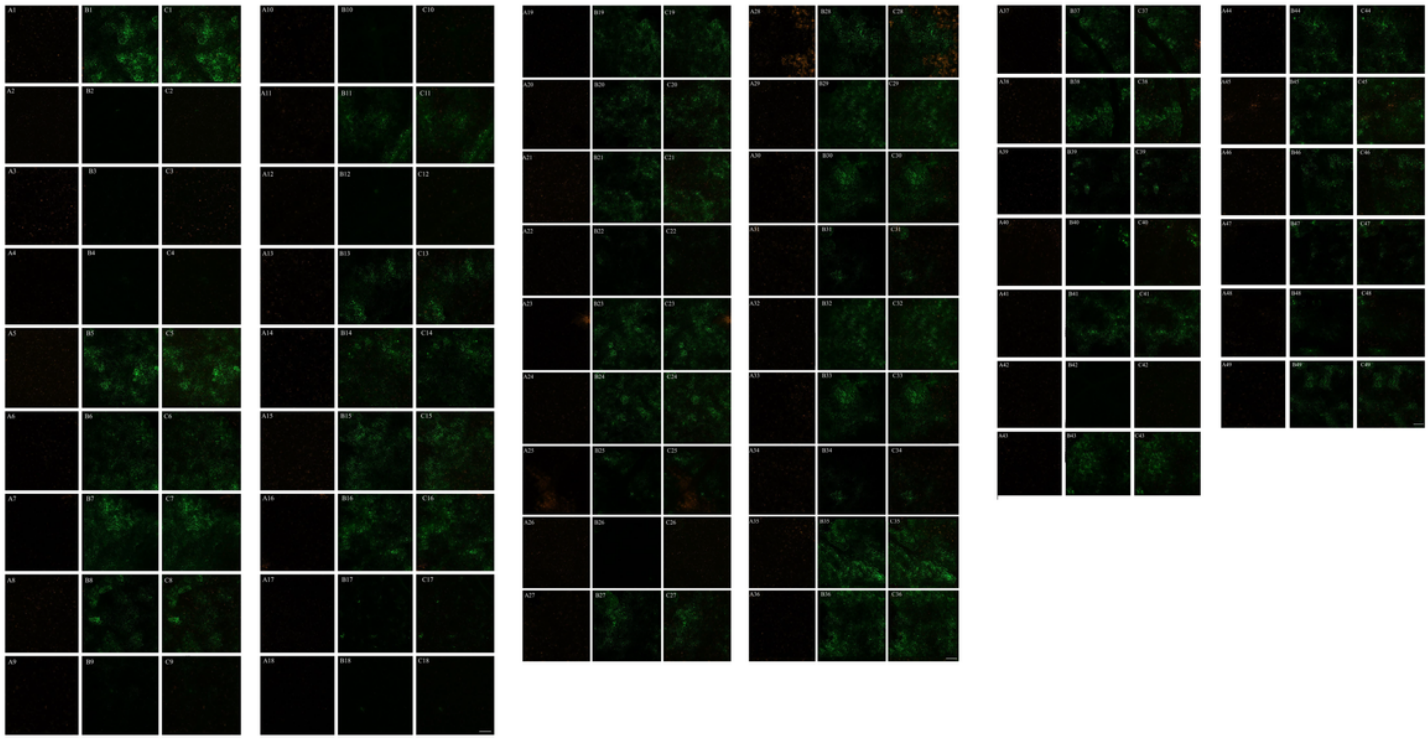
Expression analyses of miRNAs and their target genes by RT-qPCR. (A) ‘Guanhuabai’ pitaya. (B) ‘Guanhuahong’ pitaya. (A1-B1) mdm-miR828a_1ss22AT and comp24967_c0. (A2-B2) nta-miR6020b and comp234190_c0. (A3-B3) PC-5p-192_7269 and comp29967_c0. (A4-B4) zma-miR159c-5p_L-1R-2_1ss13AG and comp24676_c0. (A5-B5) mdm-miR858 and comp15143_c0. (A6-B6) mdm-miR858 and comp24362_c0. (A7-B7) mdm-miR858 and comp403340_c0. (A8-B8) mdm-miR858 and comp15849_c0.

(A9-B9) aau-miR160_L-4R+1 and comp36993_c0_seq3. (A10-B10) aau-miR160_L-4R+1 and comp35191_c0. (A11-B11) PC-5p-23845_39 and comp28219_c0. (A12-B12) aly-miR157d-5p_L+1_1ss4AC and comp25631_c0. (A13-B13) osa-miR529b_R-1_1ss8AG and comp25631_c0. (A14-B14) ptc-miR164a and comp27657_c0. (A15-B15) ssp-miR156_L+1R-1_1ss4AC and comp25631_c0. (A16-B16) ssp-miR156_L+1R-1_1ss4AC and comp25650_c0.




## Figure 13

RLM-5'-RACE validation of miRNA-mediated cleavage of target genes. (A1-A7) ‘Guanhuabai’ pitaya. (B1-B7) ‘Guanhuahong’ pitaya. Top strand (black) depicts a miRNA complementary site, and bottom strand (red) depicts a miRNA complementary site the miRNA. Watson–Crick pairing (:) are indicated. Blue arrows indicated the cleavage sites of targets and the numbers showed the frequency of the clones sequenced. Red arrows indicated that the cleavage site of miRNA on target gene occurs at the 10th nucleotide from the 5'-end of miRNA in the binding region and the numbers showed the frequency of the clones sequenced. The cleavage sites outside of the displayed sequence are not shown.



## Figure 14

Transient expressions of miRNAs and their target genes. (A1-A49) bright field. (B1-B49) eGFP. (C1-C49) merge. (A1-C1) 35s::eGFP. (A2-C2) WT. (A3-C3) 35s::Pre-miR164. (A4-C4) 35s::Pre-miR6020b. (A5-C5) 35s::comp234190_c0::eGFP. (A6-C6) 35s::positive control comp234190_c0::eGFP. (A7-C7) 35s::negative control comp234190_c0::eGFP. (A8-C8) 35s::comp234190_c0::eGFP + 35s::Pre-miR164. (A9-C9) 35s::comp234190_c0::eGFP+35s::Pre-miR6020b. (A10-C10) 35s::positive control comp234190_c0::eGFP+35s::Pre-miR6020b. (A11-C11) 35s::negative control comp234190_c0::eGFP+35s::Pre-miR6020b. (A12-C12) 35s::Pre-miR858. (A13-C13) 35s::comp15143_c0::eGFP. (A14-C14) 35s::positive control comp15143_c0::eGFP. (A15-C15)

35s::negative control comp15143_c0::eGFP. (A16-C16) 35s::comp15143_c0::eGFP+35s::Pre-miR164. (A17-C17) 35s::comp15143_c0::eGFP +35s::Pre-miR858. (A18-C18) 35s::positive control comp15143_c0::eGFP+35s::Pre- miR858. (A19-C19) 35s::negative control comp15143_c0::eGFP+35s::Pre-miR858. (A20-C20) 35s::comp24362_c0::eGFP. (A21-C21) 35s::comp24362_c0::eGFP+35s:: Pre-miR164. (A22-C22) 35s::comp24362_c0::eGFP+35s::Pre-miR858. (A23-C23) 35s::comp403340_c0::eGFP. (A24-C24) 35s::comp403340_c0::eGFP+35s::Pre- miR164. (A25-C25) 35s::comp403340_c0::eGFP+35s::Pre-miR858. (A26-C26) 35s::Pre-miR160. (A27-C27) 35s::comp35191_c0::eGFP. (A28-C28) 35s::positive control comp35191_c0::eGFP. (A29-C29) 35s::negative control comp35191_c0::eGFP. (A30-C30) 35s::comp35191_c0::eGFP+35s::Pre-miR164. (A31-C31) 35s::comp35191_c0::eGFP+35s::Pre-miR160. (A32-C32) 35s::positive control comp35191_c0::eGFP+35s::Pre-miR160. (A33-C33) 35s::negative control comp35191_c0::eGFP+35s::Pre-miR160. (A34-C34) 35s::Pre-PC-5p-192_7269. (A35-C35) 35s::comp29967_c0::eGFP. (A36-C36) 35s::positive control comp29967_c0::eGFP. (A37-C37) 35s::negative control comp29967_c0::eGFP. (A38-C38) 35s::comp29967_c0::eGFP+35s::Pre-miR164. (A39-C39) 35s:: comp29967_c0::eGFP+35s::Pre-PC-5p-192_7269. (A40-C40) 35s::positive control comp29967_c0::eGFP+35s::Pre-PC-5p-192_7269. (A41-C41) 35s::negative control comp29967_c0::eGFP+35s::Pre-PC-5p-192_7269. (A42-C42) 35s::Pre- PC-5p-23845_39. (A43-C43) 35s::comp28219_c0::eGFP. (A44-C44) 35s::positive control comp28219_c0::eGFP. (A45-C45) 35s::negative control comp28219_c0::eGFP. (A46-C46) 35s::comp28219_c0::eGFP+35s::Pre-miR164. (A47-C47) 35s:: comp28219_c0::eGFP+35s::Pre-PC-5p-23845_39. (A48-C48) 35s::positive control comp28219_c0::eGFP+35s::Pre-PC-5p-23845_39. (A49-C49) 35s::negative control comp28219_c0::eGFP+35s::Pre-PC-5p-23845_39. Bar=200 μm.

## Supplementary Files

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

- Tables13.docx
- TableS4.docx
- TableS9.docx
- TableS5.docx
- TableS7.docx
- TableS6.docx
- TableS2.docx
- TableS3.docx
- FigureS2.docx
- FigureS1.docx
- TableS1.docx
- TableS8.docx