Comparative analysis of transcriptional regulation of betalain biosynthesis based on SMRT sequencing of full-length transcriptome in two pitaya cultivars (red pulp and white pulp)

Yawei Wu
Guizhou University

Juan Xu
Huazhong Agriculture University

Xiumei Han
Guizhou Academy of Agricultural Sciences

Guang Qiao
Guizhou University

Kun yang
Guizhou University

Xiaopeng Wen (✉ xpwensc@hotmail.com)
Guizhou University

Research article

Keywords: Pitaya; SMRT; betalains; SSR; IncRNAs

Posted Date: September 24th, 2019

DOI: https://doi.org/10.21203/rs.2.14828/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background: In order to gain more valuable genomic information involved in betalain biosynthesis, the full-length transcriptome of pitaya was analyzed using Single-Molecule Real-Time (SMRT) sequencing corrected by RNA-seq in the present study. Two pitaya cultivars, ‘Zihonglong’ (red pulp) and ‘Jinghonglong’ (white pulp) were selected to analyze betalain transcriptome in four fruit developmental stages.

Results: A total of 65,317 and 91,638 genes coding proteins were identified in ‘Zihonglong’ and ‘Jinghonglong’, respectively. A total of 11,377 and 15,551 genes with more than two isoforms were investigated from ‘Zihonglong’ and ‘Jinghonglong’, respectively. Also, 156,955 genes were acquired after elimination of redundancy, of which, 120,604 genes (79.63%) were annotated, and 30,875 (20.37%) sequences without hits to reference database were probably novel genes in pitaya. Totally, 31,169 and 53,024 SSRs were uncovered from the genes of ‘Zihonglong’ and ‘Jinghonglong’, and 11,650 IncRNAs in ‘Zihonglong’ and 11,113 IncRNAs in ‘Jinghonglong’ were obtained herein. Further, 104 genes involved in betalain metabolism were identified, and HpCYP76AD4 and HpDODA probably responded to betalains biosynthesis.

Conclusions: Conclusively, this is the first study to perform SMRT sequencing of the full-length transcriptome of pitaya, which provides a useful genomic clue for exploring the structure and function of genes in pitaya, particularly for betalain biosynthesis.

Background

Pitaya (Hylocereus), originated from Latin America and West Indies [1], is one of economical and nutritional fruits cultivated in tropical and subtropical regions. H. polyrhizus (with red pulp and peel) and H. undatus (with white pulp and red peel) are the two major species [2], and the red pigment in pitaya is mainly attributed to betanin pigments [3]. Betalains are tyrosine-derived pigments that occur solely in the order of caryophyllales, which largely replaced the anthocyanins in a mutually exclusive manner [4]. Betalains also have high nutritional value and positive effects in health and disease prevention by high antioxidant and anti-inflammatory capabilities [5–6]. Therefore, betalains synthesis has become a research hot spot on scientific interest as well as economic significance [7]. Currently, the metabolic pathway of betalains is more clearly defined [4], the identification of genes is important for betalains biosynthesis. Sequencing platforms is an efficient approach to illustrate putative genes, 9 key transcripts involved in betalains synthesis were identified based on second-generation sequencing (SGS) in pitaya [6]. Whereas, many questions remain open for limited sequence data of pitaya [4].

SGS is a powerful tool for the description of gene expression levels [9–9]. However, it is difficult to identify full-length transcript using the SGS data [10]. High quality transcript sequences is crucial for plant biology research, fortunately, full-length transcriptome is being employed as an effective approach to obtain high quality transcript sequences. SMRT sequencing developed by PacBio, a third-generation
sequencing (TGS), can obtain full-length sequencing gradually without post-sequencing assembly [11–12], which has been used for whole-transcriptome profiling in many plants [13–19], but was not in pitaya so far. However, SMRT sequencing need be corrected with NGS reads to eliminate high error rate [20], hence, a combination of SMRT sequencing and RNA-seq is preferable process.

Currently, a combination of SMRT sequencing and RNA-seq was used for red and white pulp to generate full-length and high quality pitaya transcriptome, and the PacBio transcriptome sequencing of pitaya was reported firstly. Based on the obtained transcriptome data, transcript functional annotation, SSRs analysis and IncRNAs prediction were performed. This study might be a valuable resource for further investigation of pitaya, and might provide a better understanding of betalains biosynthesis in pitaya fruit.

**Result**

**Pulp color change**

Obviously, there is larger difference in pulp colour between the two cultivars. There were not significant differences in $L^*$, $a^*$, $b^*$, $C^*$, $h^*$ value of ‘Zihonglong’ and ‘Jinghonglong’ pulps on 22 DPA. The $L^*$, $a^*$, $b^*$, $C^*$, $h^*$ value of ‘Jinghonglong’ were relatively stable, while those of ‘Zihonglong’ changed significantly during the fruit development stages (Table 1). With the development of ‘Zihonglong’ fruit, $L^*$ value was decreased **gradually**, $a^*$, $C^*$, $h^*$ increased prominently on 25DPA and reached the **highest** values (27.90, 28.53, 343.05 respectively) on 28DPA, and then appeared a slight decline on mature fruit. $b^*$ value decreased from 2.11(yellow pulp) on 22 DPA to −6.73 (blue pulp) on 25DPA.

**Transcriptome analysis using PacBio Sequel**

The full-length transcriptome of pitaya fruit was generated by PacBio Sequel on ‘Zihonglong’ and ‘Jinghonglong’. As shown in Table 2, 9,579,839 subreads from 8.47G bases were obtained from the pulp of ‘Zihonglong’, while 7,245,659 subreads from 7.74G bases from the pulp of ‘Jinghonglong’. After removing adapters and artefacts, 367,001 circular consensus sequence (CCS) (including 314,173 full-length non-chimerics, FLNCs) of ‘Zihonglong’ and 481,602 CCS (including 348,184 FLNCs) of ‘Jinghonglong’ were generated, respectively. The lengths of ‘Zihonglong’ FLNCs ranged from 334 to 14,604 nt with an average length of 950 nt, while ‘Jinghonglong’ FLNCs showed an average length of 1,095 nt and ranged from 374 to 6,988 nt. For ‘Zihonglong’, 184,875 polished consensus sequences transcripts were produced, including 23,669 polished high-quality (HQ) and 161,206 low-quality (LQ) isoform sequences. For ‘Jinghonglong’, 188,215 polished consensus sequences, including 25,299 polished HQ and 162,916 LQ isoform sequences were obtained. After correcting and removing redundant reads, 65,312 and 91,638 genes (non-redundant reads) were obtained from full length transcripts of ‘Zihonglong’ and ‘Jinghonglong’, respectively.
Comparison of SMRT sequencing and next-generation sequencing

The number of SMRT gene obtained from SMRT sequencing were less than that of unigene assembled from NGS reads, whereas, the mean length of SMRT gene is much longer than that of unigene assembled from NGS reads. Additionally, about 65% of the assembled transcripts from NGS reads were <500 bases in ‘Zihonglong’, while about 74% of that in ‘Jinghonglong’. However, only 12% in ‘Zihonglong’ and 6% in ‘Jinghonglong’ of the transcripts from the PacBio Sequel reads were <500 bases. Approximately 80% of the transcripts from the PacBio Sequel reads ranged from 500 bases to 2000 bases (Table 3). Hence, the SMRT sequencing offered significant advantages over NGS in the length of reads.

Clustering analysis

Multiple transcripts corresponded to one gene in the transcriptional group. PacBio long reads clustering analysis demonstrated that 65,317 and 91,638 genes were generated from polished consensus sequences transcripts in ‘Zihonglong’ and ‘Jinghonglong’, respectively. Various isoforms generated by a single gene were widely found among the tested samples. A total of 17.42% genes had more than one isoform in ‘Zihonglong’ pulp, a slightly higher than that (16.97%) of ‘Jinghonglong’. In the former, 11,377 genes showed more than two alternative splice forms (isoforms), of which the majority corresponded to two-to-three isoforms, accounting for 74.78% of the total, and 516 genes contained over 10 isoforms. To the latter, 15,551 genes had more than two isoforms, among which the majority were two-to-three isoforms, accounting for 67.99% of the total, and 767 genes with over 10 isoforms were obtained (Figure 1). Therefore, when alleles and associated homologs were grouped against these results, they typically shared the same alternative splicing patterns [9], indicating that a gene might generated different transcripts via alternative splicing.

Function annotation

Function annotation of pitaya non-redundant FLNC transcripts (genes) was investigated using different databases. As shown in Table 4, a total of 34,601 transcripts were annotated in the Clusters of Orthologous Groups of proteins (COG) database; 54,706 in Gene Ontology (GO); 28,796 in Kyoto Encyclopedia of Genes and Genomes (KEGG); 56,010 in euKaryotic Ortholog Groups (KOG); 88,549 in protein families and domains (Pfam); 72,130 in Swiss-Prot; 95,458 in TrEMBL; 10,5413 in NR; and 63,052 in NT. Moreover, 120,604 transcripts were annotated in all of the nine databases, while 30,875 sequences without hits to reference database were probably novel genes in pitaya.

The homologous species of Hylocereus were predicted by sequence alignment on the basis of the NR database. Of all the genes hits to NR plant proteins from BLASTx, the pitaya genes gave the highest number of hits to the Beta vulgaris (51,879 hits), followed by Spinacia oleraces (21,876 hits) and Vitis
Most hits found in *Beta vulgaris* were probably for the reason that pitaya and *Beta vulgaris* belong to Caryophyllales, and *Beta vulgaris* database is better annotated than those of other species. As shown in Figure 2b, molecular function (MF, 62,439 FLNCs) was more abundant than biological process (BP, 142,635 FLNCs) and cellular component (CC, 104,215 FLNCs). Within these functional groups, the highest number of sequences were annotated with the metabolic process (35,263 sequences, 11.40%), cellular process (28,379 sequences, 9.18%) and catalytic activity (27,507 sequences, 8.89%). A total of 117 pathways with 28,796 genes were annotated by KEGG, associated with 23.88% of the whole annotated dataset (120,640 genes). Among these, 237 genes were identified in phenylalanine, tyrosine and tryptophan biosynthesis pathway, however, KEGG pathway involved in betalain biosynthesis was not founded.

**SSR and IncRNA prediction**

A total of 31,169 SSRs were discovered in 24,889 genes (38.10%) from ‘Zihonglong’, the number of genes containing more than one SSR was 11,885, and the number of SSRs present in compound formation was 4,472. A total of 53,024 SSRs were discovered in 39,793 genes (43.42%) from ‘Jinghonglong’, the number of genes containing more than one SSR was 18,725, and the number of SSRs present in compound formation was 8,868 (Figure 3). In both cases, the most abundant motifs detected was mono-nucleotides, accounting for 41.72% and 40.58% of the total SSRs in ‘Zihonglong’ and ‘Jinghonglong’, respectively, and 4,883 (15.67%) and 6,204 (11.70%) di-nucleotides were detected from ‘Zihonglong’ and ‘Jinghonglong’, respectively.

We obtained 11,650 and 11,113 IncRNAs from 65,317 and 91,638 genes in ‘Zihonglong’ and ‘Jinggonglong’, respectively (Figure 4). Four of these IncRNAs were up to 3,000 nt in ‘Zihonglong’, while 18 up to 3,000 nt were investigated from ‘Jinghonglong’, most of which were single-isoform transcripts presenting in both samples. The functions of these IncRNAs need to be further characterized.

**Genes involved in betalain biosynthesis**

Taken account of the different expression levels of genes between ‘Zihonglong’ and ‘Jinghonglg’, the genes from PacBio sequel were used as the reference dataset. It was shown that 44,109 DEGs were found between ‘Zihonglong’ and ‘Jinghonglong’ during four development stages, among which, the most DEGs were investigated in R2_vs_W2, containing 11,317 up-regulated and 4,788 down-regulated DEGs respectively (Figure 5a). Heat map of all DEGs in both ‘Zihonglong’ and ‘Jinghonglong’ was created, and the four developmental stages of ‘Zihonglong’ and ‘Jinghonglong’ were clustered together, subsequently, both of the two pitaya cultivars showed that the 22 DPA and the 25 DPA got together, accompanying with the 28 DPA and the 30 DPA got together (Figure 5b).

To evaluate the candidate genes involved in betalain biosynthesis, further analysis was performed. According to the known betalain synthesis pathway [3, 21], TYR, MYB1, CYP76AD1, DODA and 5GT are
involved in betalain biosynthesis. Totally, 104 genes including 7 TYR, 38 MYB, 20 CYP76AD, 26 DODA and 13 5GT genes were identified in our SMRT data and these genes were illustrated in a heat map (Figure 5c).

5 CYP76AD and 8 DODA genes with an FPKM > 10 or specially expressed in different colored phenotypes of pulp were selected as candidate genes, among which, \textit{HpCYP76AD4} (i1\_HQ\_R\_c13003/f5p0/1979) and \textit{HpDODA} (i1\_LQ\_R\_c96099/f1p0/1004) expressed highly in the pulp of 'Zihonglong' with the highest FPKM value, but there was almost no expression in the pulp of 'Jinghonglong'. Furthermore, the expression patterns observed here were consistent with their role in betalain production of pitaya pulp. Thereby, we deduced that the low expression of \textit{HpCYP76AD4} (i1\_HQ\_R\_c13003/f5p0/1979) and \textit{HpDODA} (i1\_LQ\_R\_c96099/f1p0/1004) may account for the absence of betalains in pulp of 'Jinghonglong', the length of the two genes were 1979 bp and 1004 bp, respectively. Subsequently, 5 CYP76AD and 8 DOD genes were selected as candidate genes, the BLAST search in NCBI and Neighbor-joining tree of with their homologous gene reported in other plants were performed. The results showed that \textit{HpCYP76AD4} (i1\_HQ\_R\_c13003/f5p0/1979) had higher similarity with \textit{OfCYP76AD8} (KM51679.1), followed by \textit{PoCYP76AD1-likeprotein} (AK133825.1) (Quvery67\%, ident86\%) and \textit{BaCYP76AD14} (AJD87470.1) (Quvery75\%, ident81\%) (Figure 6a). \textit{HpDODA} (i1\_LQ\_R\_c96099/f1p0/1004) has the highest sequence consistency with \textit{PgDODA} (Q7XA48.1) (Quvery79\%, ident88\%), the Quvery and ident between \textit{HpDODA} (i1\_LQ\_R\_c96099/f1p0/1004) and \textit{BvDODA} (AET43288.1) were 78\% and 63\%, respectively (Figure 6b). The length of open reading frame (ORF) of \textit{HpCYP76AD4} (i1\_HQ\_R\_c13003/f5p0/1979) was 1521nt (from 122 nt to 1642 nt), that of \textit{HpDODA} (i1\_LQ\_R\_c96099/f1p0/1004) was 816 nt (from 111 nt to 926 nt).

**Discussion**

Fruit is a major product for plant-derived pigments, and the formation of pigment is closely related to the process of fruit development. The $L^*$ values decreased and $a^*$, $b^*$ values increased with apple fruit development [22]. $L^*$ values declined with increasing betacyanin contents, the $h^*$ shift from 88.2 (yellow stems) to 350.3 (purple stems) [23]. High anthocyanin content may lead to the decrease of fruit brightness [22–23]. In the present work, variation tendency of color data were consistent with reports [22–23]. The color data were in accordance with the visual pigments appearance of 'Zihonglong' pulps. It's worth mentioning that color data varied predominantly on 25 DPA in red pup cultivar while the fruit pulp was in color initiation stage, which indicated that the stage was a crucial period for the accumulation of red pigment.

Transcripts from RNA-seq require assembly and full-length transcripts proportion is very small, and inaccuracy in gene structure characterization resulted from mis-assembly is a common problem, which is exacerbated in the species without a reference genome sequence for the prediction of gene models [24]. Recently, SMRT sequel as a new TGS platform was carried out by PacBio sequencing, non-assembled long-read transcripts with error rate (10\%) can be generated by SMRT sequel, and that of error rate can be overcome by correction of Illumina reads[25]. For example, the mapping rate of long reads in maize can
be increased from 11.6% to 99.1% after correction with Illumina read[14], whereas, the report of reference genome sequence or SMRT sequence on pitaya has not yet been found so far. In the present case, 65,312 and 91,638 genes (non-redundant reads) were generated by SMRT from pooled-stage pulp and corrected by Illumina RNA-seq, and the mean length of SMRT gene were up to 1,175 bp and 1,337 bp in ‘Zihonglong’ and ‘Jinghonglong’, whereas, the mean length of unigene from Illumina were only 681bp and 696bp in ‘Zihonglong’ and ‘Jinghonglong’ (Figure 3), respectively. Also, the pitaya genes had the highest number of hits to the B. vulgaris (50.63%), whereas, the species distribution with the greatest number of H. polyrhizus was Vitis vinifera (50.1%) by RNA-Seq [6]. Both pitaya and B. vulgaris belong to Caryophyllales plants, therefore, the result illustrated that SMRT data are of higher quality than unigenes from RNA-Seq.

Transcriptome reconstruction and annotation has been improved significantly with the development of sequencing techniques [24]. Longer reads are capable of sequencing complete transcripts and qualifying gene features. PacBio corrected reads can provided an efficient reference sequences for plants without reference genome [9]. Different transcription isoforms in pitaya pulp were detected without a reference genome. A total of 17.42% and 16.97% of genes in the red and white pulp were identified, respectively, including more than 10 isoforms in red pulp (516, 0.79%) in comparison with that of the white pulp (767, 0.84%). LncRNAs has been found to be a key functional molecules that can regulate gene expression, which has been a hot topic in biology [26–27], and 11,046 IncRNAs were predicted in Salvia miltiorrhiza [13]. In maize, 867 transcripts with a mean length of 1.1 kb were identified as novel high-confidence IncRNAs [14]. A total of 417 and 531 IncRNAs were identified in sweet potato and l. trifida, respectively [26], 223 and 205 IncRNAs were obtained in leaf and root of Astragalus membranaceus respectively [24], 2,426 transcript sequences including 1,220 non-ORF transcript sequences candidate IncRNAs in sugarcane [17]. Currently, 11,650 and 11,113 IncRNAs were identified with four analytical methods inform the red and white pulp, respectively, which were more than that from other documented species. However, their functions in pitaya required further investigations.

The molecular mechanism of betalains synthesis has been well documented in previous studies. First, tyrosine is hydroxylated to L-DOPA and subsequently converted to cyclo-DOPA by one or more cytochrome P450s [28–32], alternatively, L-DOPA can be converted to betalamic acid by DOPA 4, 5-dioxygenase (DOD) [33]. Next, betalamic acid condensates spontaneously with cyclo-DOPA to form betanidin or with amino acids and other amines to form betaxanthins. Betanidin is further glucosylated by a betanidin glucosyltransferase to form the basic betacyanins betanin or gomphrenin [4]. BvMYB1 is the only currently known betalain-related transcription factor, which has an essential role as a positive regulator of betalain biosynthesis through activation of the CYP76AD1 and BvDODA1 genes [7].

Betalains, an important pigment in most Caryophyllales plants, can be used as a natural colorant in food [34], cosmetics and pharmaceuticals [35]. Intensive attempts have been focused on the betalain biosynthesis and genes function, and much more betalain-related candidate genes, such as TYR [36–37], BvMYB1 [38], CYP76AD1 [28] and BvDODA1 [39] were identified. Even so, research about betalain-related genes, especially on pitaya has been reported little so far. In current study, 104 DEGs involving in betalain
biosynthesis were identified and analyzed by both SMRT and SGS. Betalamic acid is the chromophore molecule of both betacyanins and betaxanthins, and cDOPA as well as its derivatives are essential to produce betacyanin [40]. The formation of betalamic acid and cDOPA are crucial in betacyanin synthesis, especially, the absence of betalamic acid may block the production of betalain. As a white pulp cultivar, neither red betacyanins nor yellow betaxanthins were detected in white pulp. Hence, we hypothesized CYP76AD and DODA were crucial genes to the formation of betalains. In present case, the two novel pitaya genes *HpCYP76AD4* (i1_HQ_R_c13003/f5p0/1979) and *HpDODA* (i1_LQ_R_c96099/f1p0/1004) probably respond to the presence of betalains.

**Conclusions**

In summary, combining SMRT with SGS provide an efficient process to the research of genes. Currently, full-length transcripts were generated by both SMRT Squel and Illumina RNA-seq from pooled-stage pulp, and full-length transcriptome of pitaya pulp provided reference sequence of pitaya. Compared with the reported genes in previous studies on pitaya using cDNA cloning or RNA-Seq, the date from our study have higher quality and much more complete annotation, so as to provide a valuable resource for pitaya research and facilitate the identification of additional betalain-related genes in the near future. Up to now, the mining and utilization of the data was limited in this study, further research will focus on the structure and function of genes involved betalains biosynthesis.

**Methods**

**Pitaya pulps**

The cutting seedling of *H. polyrhizus* cv Zihonghlong (certificate number, Qian guo shen no. 2009005) and *H. undatus* cv Jinghonglong (certificate number, Qian guo shen no. 2009006) cultivated by Guizhou fruit institute were used in this study. All plants had been panted in Langdang fruit professional cooperative (Luodian county, Guizhou province, P. R. China.) in 2009. Three hundred flowers blooming on the same day were marked with tags in 2016, and thirty labelled healthy fruits of each four developmental stages (22nd, 25th, 28th and 30th after anthesis) were collected randomly from different plants (Figure 7). All samples intended for RNA extraction were fresh-frozen in liquid nitrogen immediately after collection and stored at −80 °C until use.

**Color Measurements**

For color analyses, *L*, *a*, *b*, *C* and *h°* of pitaya pulp were measured with CR−10 Chromaportable colorimeter (Konica Minolta Sensing, inc., Osaka, Japan). All determinations were performed in duplicate. *L* value represented the relative lightness of colors ranging from 0 (black) to 100 (white). Values of *a* and *b* ranged from −60 to 60, where *a* was negative for green color and positive for red color, and *b* was negative for blue and positive for yellow [22,41]. Chroma values (*C*) expressing color saturation,
higher $C^*$ value means less color type and brighter color, to the contrary, lower $C^*$ value means more color type and bleak color. Hue angle ($h^\circ$) is expressed on a color wheel, where 0°/360° = red-purple, 90° = yellow, 180° = green, and 270° = blue [23].

**RNA sample preparation**

Total RNA was isolated using the RNeasy Plus Mini Kit (Qiagen, Valencia, CA, USA), respectively. The purity and concentration of RNA were measured using the NanoDrop ND–1000 spectrophotometer (NanoDrop Technologies, Rockland, DE, USA) with an OD260/280 reading. The integrity of the RNA was determined on agarose gel electrophoresis with the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA).

**Library preparation and SMRT sequencing**

The same amount of RNA from each variety sample (pulp of four development stages) were mixed for SMRT sequencing ananlysis. Firstly, mRNA was enriched by Oligo (dT) magnetic beads. Then the enriched mRNA was reverse transcribed into full length 1st strand cDND using Clontech SMARTer PCR cDNA Synthesis Kit. PCR cycle optimization was used to determine the optimal amplification cycle number for the downstream large-scale application. The optimized cycle number was used to generate double-stranded cDNA, followed optional size (≤4kb) selection using the BluePippinTM for combined SMRT bell library. Full length cDNAs were performed DNA damage repaired, end repaired, and ligated to sequencing adapters, and then digested with exonuclease. Qualified libraries were sequenced on the PacBio Sequel (Pacific Bio-science Inc.) platform according to the effective concentration and data output requirements of the library.

**Preprocessing of SMRT reads**

First, the subreads were acquired from raw sequencing reads using the SMRT Link v5.0 (minLength = 200, minReadScore = 0.75) pipeline supported by Pacific Biosciences, and CCS reads were extracted out of subreads BAM file. Through RS_IsoSeq (minPasses = 1, minPredicted Accuracy = 0.8), CCS reads were classified into full-length non-chimeric (FLNC), non-full-length (NFL) based on cDNA primers and polyA tail signal. Subsequently, the FLNC reads were clustered by Iterative Clustering for Error Correction (ICE) software to generate the cluster consensus isoforms. Followed, NFL reads were used to polish the above obtained cluster consensus isoforms by Quiver software to finally obtain the FLNC polished high quality consensus sequences (accuracy≥99%). After corrected by SGS using LoRDEC, non-redundant high-quality full-length transcripts was generated by CD-HIT (c = 0.99) for further analysis.

**Functional annotation of genes**
Non-redundant transcript sequence as genes obtained after CD-HIT deduplication were grouped and mapped to nine protein and nucleic acid database to obtain the annotation information of the gene. These databases included NR, Nt, Swiss-Prot, GO, COG, KOG, Pfam, TrEMBL, and KEGG. GO annotation was analyzed by Blast2GO software with Nr annotation results of genes. Genes ranking the first 20 highest score and no shorter than 33 HSPs (High-scoring Segment Pair) hits were selected to conduct Blast2GO analysis. Then, functional classification of genes was run using WEGO software.

SSR detection

The MicroSAtellite identification tool (MISA; http://pgrc.ipk-gatersleben.de/misa/) was employed for microsatellite mining in the whole transcriptome, mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, hexanucleotide and compound SSR were identified by analyzing transcript sequences.

IncRNAs prediction

The coding potential of transcripts were predicted by predictor of long non-coding RNAs and messenger RNAs based on an improved k-mer scheme(PLEK)[42] and Coding-Non-Coding Index (CNCI)[43], and then, transcriptional sequences predicted from PLEK and CNCI were blasted with the known protein database using Coding Potential Calculator (CPC) software[44]. The transcriptional sequences predicted by PLEK, CNCI and CPC software were performed hmmscan homologous search with Pfam [45] database, finally the LncRNA sequences were obtained.

Second-generation sequence

Total RNA (5μg) was digested by using DNase I (NEB, Frankfurt, Germany). The sample was purified with Agencourt RNAclean XP Beads and fragmented into 130–170 nt. First-strand cDNA was generated by First Strand Master Mix and Super Script II reverse transcription (Invitrogen). Then second-strand cDNA was synthesized using Second Strand Master Mix. After end repairing, adding A and adaptor ligation, several rounds of PCR amplification with PCR Primer Cocktail and PCR MasterMix were performed to enrich the cDNA fragments. The final library is quantitated by using the Agilent 2100 bioanalyzer instrument. The qualified libraries were sequenced pair end on the Illumina HiSeq 4000 System.

Different expressed genes (DEGs) analysis

Data from RNA-seq were mapped to the non-redundant SMRT reference by RSEM software. The expression abundance of unigene was represented as value of FPKM, and differential expression gene (FDR<0.01 and FC≥2) were obtained using EBSeg [46]. The differential expression gene venue was
named by means of “A_vs_B”, among which, the expression level of the up-regulated gene in sample A was higher than that in sample B, and the reverse was the downregulation gene.

**Phylogenetic analysis of DEGs involved in betalain synthesis**

The sequences of 5 CYP76AD and 8 DODA genes were blasted in NCBI, 7 CYP76AD genes and 7 DODA genes from NCBI were selected according to sequence similarity (Query cover ≥ 70% and ident ≥ 80%), and an unrooted phylogenetic tree was constructed with MEGA 6.0 software (MEGA, http://www.megasoftware.net/).

**Abbreviations**

SMRT: Single-Molecule Real-Time; RNA-seq: RNA sequencing; DPA: days post-anthesis; DEGs: different expressed genes; IncRNAs: long non-coding RNAs; DODA: DOPA 4, 5-dioxygenase; SGS: second-generation sequencing; TGS: third-generation sequencing; SSRs: simple sequence repeats; CCS: circular consensus sequence; FLNC: full-length non-chimeric; NFL: non-full-length; ICE: Iterative Clustering for Error Correction; GO: Gene Ontology; COG: Clusters of Orthologous Groups of proteins; KOG: euKaryotic Ortholog Groups; KEGG: Kyoto Encyclopedia of Genes and Genomes; MF: molecular function; BP: biological process; CC: cellular component; HSPs: High-scoring Segment Pair; PLEK: k-mer scheme; CNCI: Coding-Non-Coding Index; CPC: Coding Potential Calculator; Pfam: protein families and domains.

**Declarations**

**Acknowledgements**

For the access to experimental materials we would like to thank Luodian county Langdang fruit professional cooperative in Guizhou Province, China, providing pitaya fruit for the experiment. For the technical assistance, we would like to thank Biobreeding Biotechnology Corporation (Shanxi, China). Likewise, we would like to express great gratitude to the anonymous peer review and critical revise for the manuscript improvement.

**Author Contributions**

YW analyzed the data and wrote the paper. JX provided assistance in conceived and designed the experiments. XH and GQ performed the experiment and analyzed the data. KY provided assistance in data analysis and submission. XW reviewed the manuscript and supervised the whole project. All the authors contributed to discussion and revision of the manuscript.
Funding

This project is supported by grants from the National Natural Science Foundation of China to Xiao-Peng Wen (31760566, 31560549), as well as the 2016 Open Foundation from Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education) to Yawei Wu, Our thanks are also extended to the partial grant from the open funds of the Construction Program of Biology First-class Discipline in Guizhou (GNYL [2017] 009) to Xiao-Peng Wen and Key Laboratory of Horticulture Plant Biology (MOE) to Juan Xu. The funding bodies had no role in the design of the study and no role in the collection, analysis, and interpretation of data or in writing the manuscript.

Availability of data and materials

SMRT sequencing data and ILLUMINA HiSeq4000 data have been submitted to the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under accession numbers PRJNA494058 and PRJNA495654.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

Author details

1 Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Institute of Agro-Bioengineering/College of Life Sciences, Guizhou University, Guiyang (550025), Guizhou, China. 2 Institute of Pomology Science, Guizhou Academy of Agricultural Sciences, Guiyang (550006), Guizhou, China. 3 Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan (430070), Hubei, China.

References
1. Suh DH, Lee S, Heo DY, Kim YS, Cho SM, Lee S, Lee CH. 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(34): 8764–71. https://doi.org/10.1021/jf5020704

2. Bellec FL, Vaillant F, Imbert E. Pitahaya (Hylocereus spp.): a new fruit crop, a market with a future. Fruits. 2006; 61(4): 237–50. https://doi.org/10.1051/ fruits:2006021

3. Stintzing FC, Schieber A, Carle R. Betacyanins in fruits from red-purple pitaya, Hylocereus polyrhizus (Weber) Britton & Rose. Food Chem. 2002: 77(1): 101–6. https://doi.org/10.1016/S0308–8146(01)00374–0

4. Polturak G, Heining U, Grossman N, Battat M, Leshkowitz D, Malisky S, Rogachev I, Aharoni A. Transcriptome and Metabolic Profiling Provides Insights into Betalain Biosynthesis and Evolution in Mirabilis jalapa. Mol. Plant. 2018; 11(1): 189–204. https://doi.org/10.1016/j.molp.2017.12.002

5. Gandiaherrero F, Cabanes J, Escribano J, Garcia-carmona F, Jimenezatienzar M. Encapsulation of the Most Potent Antioxidant Betalains in Edible Matrixes as Powders of Different Colors. J Agric Food Chem. 2013; 61(18): 4294–302. https://doi.org/10.1021/jf400337g

6. Hua QZ, Chen CJ, Chen Z, Chen PK, Ma YW, Wu JY, Zheng J, Hu GB, Qin YH. Transcriptomic Analysis Reveals Key Genes Related to Betalain Biosynthesis in Pulp Coloration of Hylocereus polyrhizus. Front Plant Sci. 2016; 6: 1179. https://doi.org/10.3389/fpls.2015.01179

7. Polturak G, Aharoni A. "La Vie en Rose": biosynthesis, sources and applications of betalain pigments. Mol. Plant. 2017; 11(1): 7–22. https://doi.org/10.1016/j.molp.2017.10.008

8. Cartolano M, Huettel B, Hartwig B, Reinhardt R, Schneeberger K. cDNA Library Enrichment of Full Length Transcripts for SMRT Long Read Sequencing. PLoS one. 2016; 11(6): e0157779. https://doi.org/10.1371/journal.pone.0157779

9. Ning GG, Chen X, Luo P, Liang F, Wang Z, Yu GL, Li X, Wang DP, Bao MZ. Hybrid sequencing and map finding (HySeMaFi): optional strategies for extensively deciphering gene splicing and expression in organisms without reference genome. Sci Rep.2017; 7(1): 43793. https://doi.org/10.1038/srep43793

10. Chen SY, Deng FL, Jia XB, Li C, Lai SJ. A transcriptome atlas of rabbit revealed by PacBio single-molecule long-read sequencing. Sci Rep. 2017; 7(1): 7648. https://doi.org/10.1038/s41598–017–08138-z

11. Rhoads A, Au KF. PacBio sequencing and its applications. Genom Proteom Bioinf. 2015; 13(5): 278–89. https://doi.org/10.1016/j.gpb.2015.08.002

12. Steijger T, Abril JF, Engström PG, Kokocinski F, Consortium TR, Hubbard TJ, Guigó R, Harrow J, Bertone P. Assessment of transcript reconstruction methods for RNA-seq. Nat. Methods. 2013; 10(12): 1177–84. https://doi.org/10.1038/nmeth.2714

13. Xu ZC, Peters RJ, Weirather JL, Luo HM, Liao BS, Zhang X, Zhu, YJ, Ji AJ, Zhang B, Hu SN, Au KF, Chen SL. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis. Plant J. 2015; 82: 951–61. https://doi.org/10.1111/tpi.12865
14. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao YP, Lu ZY, Olson A, Stein JC, Ware D. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nat. Commun. 2016; 7(1): 11708. https://doi.org/10.1038/ncomms11708

15. Abdelghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt NP, Schilkey FD, Benhur A, Reddy AS. A survey of the sorghum transcriptome using single-molecule long reads. Nat. Commun. 2016; 7(1): 11706. https://doi.org/10.1038/ncomms11706

16. Wang TT, Wang HY, Cai DW, Gao YB, Zhang HX, Wang YS, Lin CT, Ma LY, Gu LF. Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (Phyllostachys edulis). Plant J. 2017; 91(4): 684–99. https://doi.org/10.1111/tpj.13597

17. Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJ. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. BMC Genom. 2017; 18(1): 395. https://doi.org/10.1186/s12864–017–3757–8

18. Li YP, Wei W, Feng J, Luo HF, Pi MY, Liu ZC, Kang CY. Global identification of alternative splicing via comparative analysis of SMRT- and Illumina based RNA-seq in strawberry. Plant J. 2017; 90(1): 164–76. https://doi.org/10.1111/tpj.13462

19. Li YP, Dai C, Hu CG, Liu ZC, Kang CY. Genome re-annotation of the wild strawberry Fragaria vesca using extensive Illumina- and SMRT-based RNA-seq datasets. DNA Res. 2018; 25(1): 61–70. https://doi.org/10.1093/dnares/ dsx038

20. Ma JE, Jiang HY, Li LM, Zhang XJ, Li HM, Li GY, Mo DY, Chen JP. SMRT sequencing of the full-length transcriptome of the Sunda pangolin (Manis javanica). Gene. 2019; 692: 208–16. https://doi.org/10.1016/j.gene.2019.01.008

21. Brockington SF, Walker R, Glover BJ, Soltis PS, Soltis DE. Complex pigment evolution in the Caryophyllales. New Phytol. 2011; 190(4): 854–64. https://doi.org/10.1111/j.1469–8137.2011.03687.x

22. Liu YL, Che F, Wang LX, Meng R, Zhang XJ, Zhang ZY. Fruit Coloration and Anthocyanin Biosynthesis after Bag Removal in Non-Red and Red Apples (Malus × domestica Borkh). Molecules. 2013; 18(2): 1549–63. https://doi.org/10.3390/molecules18021549

23. Kugler F, Stintzing F, Carle R. Identification of Betalains from Petioles of Differently Colored Swiss Chard (Beta vulgaris L. ssp. cicla [L.] Alef. Cv. Bright Lights) by High-Performance Liquid Chromatography-Electrospray Ionization Mass Spectrometry, J Agric Food Chem. 2004; 52(10): 2975–81. https://doi.org/10.1021/jf035491w

24. Li J, Harata-Lee Y, Denton MD, Feng QJ, Rathjen JR, Qu ZP, Adelson DL. Long read reference genome-free reconstruction of a full-length transcriptome from Astragalus membranaceus reveals transcript variants involved in bioactive compound biosynthesis. Cell Discov. 2017; 3(1): 17031. https://doi.org/10.1038/celldisc.2017.31

25. An D, Cao HX, Li CS, Humbeck K, Wang WQ. Isoform Sequencing and State-of-Art Applications for Unravelling Complexity of Plant Transcriptomes. Genes. 2018; 9(1): 43.
26. Luo YH, Ding N, Shi X, Wu YX, Wang RY, Pei LQ, Xu RY, Cheng S, Lian YY, Gao JY, Wang AM, Cao QH, Tang J. Generation and comparative analysis of full-length transcriptomes in sweetpotato and its putative wild ancestor I. trifida. BioRxiv. 2017; 30, 112425. https://doi.org/10.1101/112425

27. Jia D, Wang YX, Liu YH, Hu J, Guo YQ, Gao LL, Ma RY. SMRT sequencing of full-length transcriptome of flea beetle Agasicles hygrophila (Selman and Vogt). Sci Reports. 2018; 8(1): 2197. https://doi.org/10.1038/s41598–018–20181-y

28. Hatlestad GJ, Sunnadeniya R, Akhavan N, Gonzalez A, Goldman IL, Mcgrath JM, Lloyd AM. The beet R locus encodes a new cytochrome P450 required for red betalain production. Nat Genet. 2012; 44(7): 816–20. https://doi.org/10.1038/ng.2297

29. Suzuki M, Miyahara T, Tokumoto, H, Hakamatsuka T, Goda Y, Ozeki Y, Sasaki N. Transposon-mediated mutation of CYP76AD3 affects betalain synthesis and produces variegated flowers in four o’clock (Mirabilis jalapa). J. Plant Physiol. 2014; 171(17): 1586–90. https://doi.org/10.1016/j.jplph.2014.07.010

30. Deloache WC, Russ ZN, Narcross L, Gonzales AM, Martin VJ, Dueber JE. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nat Chem Biol. 2015; 11(7): 465–71. https://doi.org/10.1038/nchembio.1816

31. Polturak G, Breite D, Grossman1 N, Sarrión-Perdigones A, Weithorn E, Pliner M, Orzaez D, Granell A, Rogachev I, Aharoni A. Elucidation of the first committed step in betalain biosynthesis enables the heterologous engineering of betalain pigments in plants. New Phytol. 2016; 210(1): 269–83. https://doi.org/10.1111/nph.13796

32. Sunnadeniya R, Bean A, Brown M, Akhavan N, Gregory H, Gonzalez A, Symonds VV, Lloyd AM. Tyrosine hydroxylation in betalain pigment biosynthesis is performed by cytochrome P450 enzymes in beets (Beta vulgaris). PLoS One. 2016; 11(2): e0149417. https://doi.org/10.1371/journal.pone.0149417

33. Christinet L, Burdet FX, Zaiko M, Hinz U, Zryd J. Characterization and Functional Identification of a Novel Plant 4,5-Extradiol Dioxygenase Involved in Betalain Pigment Biosynthesis in Portulaca grandiflora. Plant Physiol. 2004; 134(1): 265–74. https://doi.org/10.1104/pp.103.031914

34. Wybraniec S, Michalowski T. New Pathways of Betanidin and Betanin Enzymatic Oxidation. J Agric Food Chem. 2011; 59(17): 9612–22. https://doi.org/10.1021/jf2020107

35. Esatbeyoglu T, Wagner AE, Schinikerth VB, Rimbach G. Betanin-A food colorant with biological activity. Mol Nutr Food Res. 2015; 59(1): 36–47. https://doi.org/10.1002/mnfr.201400484

36. Steiner U, Schliemann W, BoÈhm H, Strack D. Tyrosinase involved in betalain biosynthesis of higher plants. Planta. 1999; 208(1): 114–124. https://doi.org/10.1007/s004250050541

37. Lopeznieves S, Yang Y, Timoneda A, Wang MM, Feng T, Smith SA, Brockington SF, Maeda H. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. New Phytol. 2018; 217(2): 896–908. https://doi.org/10.1111/nph.14822
38. Hatlestad GJ, Akhavan NA, Sunnadeniya RM, Elam L, Cargile S, Hembd A, Gonzalez A, McGrath JM, Lloyd AM. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. Nat Genet. 2015; 47(1): 92–96. https://doi.org/10.1038/ng.3163

39. Chung HH, Schwinn KE, Ngo HM, Lewis DH, Massey B, Calcott KE, Crowhurst R, Joyce DC, Gould KS, Davies KM, Harrison DK. Characterisation of betalain biosynthesis in Parakeelya flowers identifies the key biosynthetic gene DOD as belonging to an expanded LigB gene family that is conserved in betalain-producing species. Front Plant Sci. 2015; 6: 499. https://doi.org/10.3389/fpls.2015.00499

40. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J. 2008; 54(4): 733–49. https://doi.org/10.1111/j.1365–313X.2008.03447.x

41. McGuire RG. Reporting of objective color measurements. HortScience. 1992; 27(12): 1254–55. https://doi.org/10.21273/hortsci.27.12.1254

42. Li AM, Zhang JY, Zhou ZY. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. BMC Bioinformatics. 2014; 15(1): 311. https://doi.org/10.1186/1471–2105–15–311

43. Sun L, Luo HT, Bu D, Zhao GG, Yu KT, Zhang CH, Liu YN, Chen RS, Zhao Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 2013; 41(17): e166. https://doi.org/10.1093/nar/gkt646

44. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei LP, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007; 36: 345–9. https://doi.org/10.1093/nar/gkm391

45. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016; 44: 279–85. https://doi.org/10.1093/nar/gkv1344

46. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman Al, Smits BMG, Haag JD, Gould MN, Stewart RM, Kendziorski C. EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics. 2013; 29(8): 1035–43. https://doi.org/ 10.1093/bioinformatics/btt087

Tables

Table 1. Color Characteristics of pitaya pulp
Table 2. Comparison of SMRT sequencing between ‘Zihonglong’ and ‘Jinghonglong’

| Data                        | ‘Zihonglong’ | ‘Jinghonglong’ |
|-----------------------------|--------------|----------------|
| Subreads base(G)            | 8.47         | 7.74           |
| Number                      | 9,579,839    | 7,245,659      |
| CCS reads                   |              |                |
| Total                       | 367,001      | 481,602        |
| FL                          | 322,995      | 366,583        |
| FLNC                        | 314,173      | 348,184        |
| NFL                         | 43,599       | 114,621        |
| Polished consensus sequences|              |                |
| Total                       | 184,875      | 188,215        |
| High quality                | 23,669       | 25,299         |
| Low quality                 | 161,206      | 162,916        |
| Corrected consensus         | 184,875      | 188,215        |
| Genes                       | 65,317       | 91,638         |

Table 3. Comparison results between SMRT sequencing transcript and Illumina sequencing unigene.

| Length Distribution (bp) | SMRT gene number ‘Zihonglong’ | Jinhonglong | Illumina sequencing assembled unigene number ‘Zihonglong’ | Jinhonglong |
|--------------------------|-------------------------------|-------------|----------------------------------------------------------|-------------|
| 200-300                  | 503                           | 265         | 43,646                                                   | 38,796      |
| 300-500                  | 7,497                         | 5,081       | 26,714                                                   | 24,097      |
| 500-1000                 | 29,237                        | 33,559      | 18,441                                                   | 17,436      |
| 1000-2000                | 21,462                        | 40,471      | 10,459                                                   | 10,174      |
| 2000-3000                | 4,700                         | 9,579       | 4,714                                                    | 4,424       |
| 3000+                    | 1,918                         | 2,683       | 3,187                                                    | 2,984       |
| Total Number             | 65,317                        | 91,638      | 107,161                                                  | 97,911      |
| Total Length             | 76,752,116                    | 122,496,823 | 72,942,534                                               | 68,104,067  |
| N50 Length               | 975                           | 1,385       | 1,169                                                    | 1,208       |
| Mean length              | 1,175                         | 1,337       | 681                                                      | 696         |

Table 4. Pitaya long-read sequencing transcriptome annotation with different databases
| Annotated-database      | Annotated-number | percentage (%) |
|-------------------------|------------------|----------------|
| COG-Annotation          | 34,601           | 28.69          |
| GO-Annotation           | 54,706           | 45.36          |
| KEGG-Annotation         | 28,796           | 23.88          |
| KOG-Annotation          | 56,010           | 46.44          |
| Pfam-Annotation         | 88,549           | 73.42          |
| Swissprot-Annotation    | 72,130           | 59.81          |
| TrEMBL-Annotation       | 95,458           | 79.15          |
| nr-Annotation           | 105,413          | 87.40          |
| nt-Annotation           | 63,052           | 52.28          |
| All-Annotation          | 120,604          | 100.00         |

**Figures**

**Figure 1**

Isoform number per gene: X-axis: isoform number Y-axis: gene number.

a) ![Gene distribution chart](image1.png)

b) ![Gene distribution chart](image2.png)
Figure 2

The distribution features of annotated genes. (a) Homologous species distribution of pitaya annotated in the NR database. (b) GO functional annotation of pitaya genes. The x-axis represents GO categories; the y-axis (right) represents the number of genes; and the y-axis (left) represents the percentage of genes.

Figure 3

The distribution characteristics of SSRs motifs in ‘Zihonglong’ (a) and ‘Jinghonglong’ (b). The x-axis represents the SSR motif unit; and the y-axis represents the repeat counts; z-axis represents repeat type.

Figure 4
Venn diagram of the number of lncRNAs predicted by CNCI, pfam, Plek and CPC in ‘Zihonglong’ (a) and ‘Jinghonglong’ (b).

Figure 5

The analysis of DEGs expression pattern. (a) Statistical table of the number of DEGs; (b) Heat map of all DEGs in pulp of pitaya by hierarchical clustering; (c) Candidate genes expression in betalains biosynthesis in pulp of pitaya.
Figure 6

Neighbor-joining phylogenetic tree of CYP76AD (a) and DODA (b) using closely related genes from NCBI.
Figure 7

Four fruit development stages from ‘Jinghonglong’ and ‘Zihonglong’. W1 refers to ‘Jinghonglong’ fruit at 22 days post-anthesis (DPA) (stage 1). W2 refers to ‘Jinghonglong’ fruit at 25 DPA (stage 2). W3 refers to ‘Jinghonglong’ fruit at 28 DPA (stage 3). W4 refers to ‘Jinghonglong’ fruit at 30 DPA (stage 4). R1 refers to ‘Zihonglong’ fruit at 22 DPA (stage 1). R2 refers to ‘Zihonglong’ fruit at 25 DPA (stage 2). R3 refers to ‘Zihonglong’ fruit at 28 DPA (stage 3). R4 refers to ‘Zihonglong’ fruit at 30 DPA (stage 4).

Supplementary Files

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

- supplementarymaterial.pdf