Identification of key genes and regulators associated with carotenoid metabolism in apricot (*Prunus armeniaca*) fruit using weighted gene co-expression network analysis

Lina Zhang

Southwest University

Qiuyun Zhang

Zhejiang University

Wenhui Li

Xinjiang Academy of Agricultural Sciences

Shikui Zhang

Xinjiang Academy of Agricultural Sciences

Wanpeng Xi (✉ xwp1999@zju.edu.cn)

Southwest University

Research article

Keywords: carotenoids, apricot, color, WGCNA, *Prunus armeniaca* L.

Posted Date: May 24th, 2019

DOI: <https://doi.org/10.21203/rs.2.9772/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Carotenoids are the main class of terpenoid pigments that contribute to the color and nutritional value of many fruits. However, while their biosynthetic pathway have been well established in a number of model plant species, many details of the regulatory mechanism controlling carotenoid metabolism remain to be elucidated. Apricot (*Prunus armeniaca*) fruit are rich in carotenoids and substantial amounts accumulate during ripening, making them a valuable system to investigate carotenoid metabolism. **Results:** In this study, we compared the apricot transcriptome profiles of bulked samples of ripening fruit. A total of 44,754 unigenes and 6,916 differentially expressed genes were detected during ripening, and 33,498 unigenes were annotated using public protein databases. Weighted gene co-expression network analysis (WGCNA) showed that two of thirteen gene expression modules were highly correlated with carotenoid metabolism, and a total of 33 structural genes involved in the carotenoid biosynthetic pathway were identified. Further, network analysis revealed that the transcription factors ERF4/5, PIF3/4, HY5, MADS14, NAC25, GLK1/2, MYC2 and MYB21/44 have potential regulatory roles in carotenoid metabolism during ripening. **Conclusions:** Based on these results, a regulatory network model is proposed. Our findings provide a platform for future functional genomic analysis and provide new insights into carotenoid metabolism.

Background

Carotenoids are widely distributed secondary metabolites that play important roles in both plant physiology and in human health as dietary components. They confer flowers and fruits with yellow, orange and red colors, thereby helping attract insects or animals to disperse seed and pollen, and are involved in photosynthesis and photoprotection. Carotenoids also scavenge free radicals in plants and help overcome biotic and abiotic stress. In humans, dietary carotenoids provide vitamin A precursors, scavenge free radicals, strengthen immunity and prevent various diseases, such as certain cancers and cardiovascular diseases[1]. Carotenoid biosynthesis has been extensively studied in a number of plant species, and a number of structural genes in the biosynthetic pathway have been characterized[2]. In addition, various regulators of carotenoid accumulation have been identified[2], but it is clear that many details of the regulatory control networks remain to be uncovered.

Apricot (*Prunus armeniaca* L.) is an important fruit crop that is grown in temperate climates, and can be consumed fresh or as purées, providing a valuable important dietary component. Fresh apricots are excellent sources of several nutrients, including fibre and carotenoids, polyphenols, ascorbic acid and different microelements[3]. Its processed by-products have also been suggested to have value in human nutrition and for the treatment of different diseases[3]. The 'Luntaixiaobaixing' apricot cultivar from Xinjiang, China is commercially popular and valued by consumers due to its attractive flavor, shape and color. During ripening, its color changes from green to orange and then to light yellow, which involves a range of carotenoid metabolic reactions, including synthesis, accumulation and cleavage. Thus, this apricot cultivar represents an excellent model for studying the regulatory mechanism of carotenoid metabolism in fruit trees.

RNA sequencing (RNA-seq) is a powerful technology for studying gene expression, identify genes and perform functional analysis[4], due to its high-throughput, low cost, high accuracy and high sensitivity[5]. In recent years, RNA-seq has been applied to the study of many fruit tree species, to elucidate molecular processes underlying fruit ripening and secondary metabolism[6-9]. It represents a particularly powerful approach for species for which there is not yet whole genome sequence data, such as apricot. At present, transcriptomic data of apricot are still very limited[10, 11].

The apricot fruit carotenoid profile and carotenogenic gene expression has been investigated previously[12-14]; however, no carotenoid metabolism regulators were identified, and the transcriptional regulation remains largely unknown. In this current study, we integrated transcriptomic and metabolomic data by weighted gene co-expression network analysis (WGCNA) to identify candidate genes related to carotenoid metabolism and propose a possible transcriptional regulatory network. Our results provide new insights into the molecular basis of carotenoid metabolism.

Results

Changes in quality parameters and carotenoid content during apricot fruit development and ripening

Apricot fruit color changed from green to yellow, and then yellow to light yellow during ripening (Fig. 1A). The fruit weight increase sharply as the fruit grows, to 19 g at the full ripe stage (FR) (Fig. 1B), while fruit firmness gradually decrease from 42.7 N to 2.8 N at the beginning of the enlargement (E) stage (Fig. 1B). The total soluble solids (TSS) content increased from 7.3 °Brix to 19.9 °Brix from the E to the FR stage, and the titratable acidity (TA) first increased from the fruitlet (F) to the turning (T) stage and then decreased from the T to the FR stage (Fig. 1B). Based on these physiological parameters, our samples were divided into development (from the F to the T stage) and ripening (from the T to the FR stage) sub stages.

The citrus color index (CCI) values of the fruit significantly increased during development and ripening (Fig. 1C), while the levels of lutein, violaxanthin, β -carotenoid and total carotenoids significantly decreased (Fig. 1C), with the total carotenoid content decreasing from 59.6 $\mu\text{g/g}$ to 21.3 $\mu\text{g/g}$. Conversely, the contents of neoxanthin and phytoene increased during the same period (Fig. 1C), suggesting that the color change largely depended on the composition and content of carotenoid compounds.

Transcriptome analysis of fruit ripening

Nine mRNA libraries were generated from triplicate samples at the three apricot ripening stages (T, turning, 57 DPA, days postanthesis; CM, commercial maturation, 65 DPA; FR, full ripe, 74 DPA). An average of 26.3, 26.0 and 27.2 million clean reads for the three stages, respectively, were mapped from approximately 2.37, 2.34 and 2.45 gigabase pairs of nucleotides, with GC percentages of 45.88%, 45.58%

and 46.16%, respectively (Additional file 1). Principal component analysis (PCA) confirmed the reproducibility of the sequence data for each stage (Fig. 2).

Next, a total of 44,754 unigenes, with a mean length of 1,196 bp and an N50 length of 1,860 nt, as well as a total of 71,243 contigs with a mean length of 529 bp and a N50 length of 1,303 nt, were assembled using Trinity software (Additional file 2). A total of 23,784 unigenes were assigned gene ontology (GO) terms, and were classified into 55 sub-categories within three standard categories ('biological processes', 'cellular component' and 'molecular function'). The sub-categories, 'metabolic process' and 'cellular process' were the most highly enriched in the 'biological process' domain, while 'cell' and 'cell part' were the most highly enriched in the 'cellular component' domain, and 'catalytic activity' and 'binding' were the top ranking terms in the 'molecular function' domain (Additional file 3).

Identification of differentially expressed genes (DEGs)

On the basis of the mapped reads, a total of 6,916 DEGs were identified using the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values (Fig. 3A), with 2,436, 4,801 and 5,728 DEGs present in the different ripening periods, respectively. The number of unique DEGs in the different stages was 630, 1,288, and 1,561 at 57-65 DPA, 57-74 DPA and 65-74 DPA, respectively, and a total of 448 genes were expressed at all three stages (Fig. 3B). We identified 452 (T-vs-CM stage), 2,152 (T-vs-FR stage), and 1,334 (CM-vs-FR stage) up-regulated genes, 581 (T-vs-CM stage), 494 (T-vs-FR stage), and 1,146 (CM-vs-FR stage) up-down regulated genes, 764 (T-vs-CM stage), 2,344 (T-vs-FR stage), and 1,049 (CM-vs-FR stage) down-regulated genes and 639 (T-vs-CM stage), 738 (T-vs-FR stage), and 1,272 (CM-vs-FR stage) down-up regulated genes in the three ripening periods, respectively (Additional file 4).

DEG gene ontology (GO) enrichment and kyoto encyclopedia of genes and genomes (KEGG) pathway analysis

To further identify the biological functions of the DEGs, the GO and KEGG databases were used to identify enriched terms/pathways in the three ripening stages from the identified DEGs. 'Cellular process', 'metabolic process' and 'single organism process' were the main categories in the 'biological process' domain, while 'cell', 'cell part' and 'organelle' categories were enriched in the 'cellular component' domain. Finally, molecular functions such as 'binding', 'catalytic activity' and 'transporter activity' were the most highly enriched during the three ripening stages (Additional file 5).

KEGG pathway enrichment analysis was carried out to assess the biological significance of the DEGs during the three ripening stages, and a total of 1,773, 4,719, 3,215 unigenes from the three stages, respectively, were mapped to 20 KEGG pathways. 'Metabolic pathways', 'plant-pathogen interaction' and 'plant hormone signal transduction' were the most significantly enriched metabolic pathways. We identified 19, 33, and 16 DEGs from the carotenoid biosynthetic pathway enriched in the three ripening stages, respectively (Additional file 6).

WGCNA

Thirteen co-expression modules were identified by WGCNA, of which two (blue and turquoise) were significantly associated with carotenoid content during ripening (Fig. 4). Analysis of the module-trait relationships suggested that the 'blue' module, containing 4,981 genes, was highly positively correlated with levels of violaxanthin ($r=0.99$, $P=8e-07$), lutein ($r=0.84$, $P=0.005$) and β -carotenoid ($r=0.94$, $P=2e-04$), and negatively correlated with neoxanthin ($r=-0.47$, $P=0.2$) and phytoene ($r=-0.75$, $P=0.02$) contents during ripening. Moreover, the abundance of phytoene was highly positively correlated with gene expression in the 'turquoise' module, containing 5,258 genes, with a coefficient of 0.95 ($P=7e-05$), but the levels of violaxanthin and lutein were significantly negatively correlated with gene expression, with a coefficient of -0.87 ($P=0.002$) and -0.97 ($P=2e-05$), respectively (Fig. 4B).

Candidate genes involved in carotenoid biosynthesis

We identified 33 DEGs that were annotated as functioning in the carotenoid biosynthesis pathway, and a heat map of their expression was drawn based on their FPKM value during fruit ripening (Fig. 5A, B and Additional file 7). As shown in Figure 5a and b, these included genes from all the major steps of the pathway and were distributed as follows: two isopentenyl diphosphate isomerase (IPI) genes, a geranylgeranyl diphosphate synthase (GGPPS) gene, two phytoene synthase (PSY) genes, four phytoene desaturase (PDS1) genes, five ζ -carotene desaturase (ZDS) genes, carotenoid isomerase (CRTISO), four lycopene β -cyclase (LCYB) genes, two β -carotene hydroxylase (CHYB) genes, five zeaxanthin epoxidase (ZEP) genes, two violaxanthin de-epoxidase 1 (VDE1) genes, neoxanthin synthase (NXS), two 9-cis-epoxycarotenoid dioxygenase 1 (NCED1) genes and the carotenoid cleavage dioxygenase 1 (CCD1) and CCD4.

Visualization of gene networks

A total of 78 genes were present in the blue WGCNA module and 15 genes in the turquoise module, both of which had a module significance between carotenoid contents and gene expression of > 0.4 (Fig. 6, Additional file 8 and 9). In the two networks, 15 ERF, 9 bHLH and 24 WRKY family members were identified as hub genes, and therefore as having a connection with the regulation of carotenoid metabolism. In addition to these genes, we identified other transcription factors from the two modules, including U7919, C4195.1 and U6922, which are homologs of *Arabidopsis thaliana* MYB1R1, MYB21 and MYB44 proteins that are known to be involved in carotenoid metabolism. Homologs of *A. thaliana* transcription factors MYC2, GLK1, GLK2, NAC25, HY5, PIF3, MADS14, WRKY6, WRKY31, WRKY69, ERF003, RAP2-12 (ERF), NAC and bHLH68 were also identified (Fig. 6A, B).

Discussion

An increasing number of studies are highlighting the complexity and diversity of carotenoid metabolism in different plant species[2, 15]. In this study, we used the apricot cultivar 'Luntaixiaobaixing', which has a specific carotenoid accumulation pattern including biosynthesis, sequestration and degradation during fruit development, to identify key candidate genes related to carotenoid metabolism and the associated regulatory network.

Light is a key factor in plant growth and development, since it provides energy for photosynthesis, as well as regulating photomorphogenesis[16], through which it is involved in many biological processes such as seed germination and flowering[15]. The light signal causes a rapid transfer of active phytochrome to the nucleus, where it interacts with phytochrome-interacting factors (PIFs). PIF is a member of the bHLH transcription factor family, and can integrate multiple environmental and intercellular signals to control growth and development, stress responses and secondary metabolism[17]. Light signaling also regulates carotenoid metabolism[18], since PIFs directly bind to the G-box element in the *PSY* promoter to inhibit its expression, resulting in reduced carotenoid accumulation[19]. Two light signaling transcription factors, LONG HYPOCOTYL5 (HY5) and COP1-like, were identified in tomato (*Solanum lycopersicum*) fruit. Hy5 inhibits gene expression and can give rise to decreased carotenoid content and thus negative regulation of fruit coloration. Conversely, down-regulation of COP1-like cause increased carotenoid accumulation and thus positive regulation of fruit coloration[20]. In citrus, blue light enhances carotenoid accumulation by up-regulating *CitPSY* expression[21], while red light has been shown to increase carotenoid synthesis and accumulation in tomato fruit[22]. In the present study, we observed that PIF3, PIF4, HY5 and bHLH transcription factor family members were differentially expressed during ripening, suggesting that they may be involved in modulating carotenoid metabolism during apricot fruit ripening.

Previous studies have shown that the synthesis and accumulation of carotenoids is regulated by various phytohormones. In *A. thaliana*, the RAP2.2 ethylene response factor (ERF) directly binds to the *PSY* promoter to alter carotenoid biosynthesis[23]. Tomato fruit ripening is triggered by an increase in ethylene biosynthesis, and is accompanied by a large increase in β -carotenoid and lycopene content, as well as the expression of *SIPSY1* and *SIPDS*, which is ethylene dependent[24]. ERF genes have been shown to both suppress and induce carotenoid production in tomato[25][26], and we identified members of this family in our study, indicating that the accumulation of carotenoids in apricot is also associated with ethylene signal transduction. Signaling involving the phytohormone abscisic acid (ABA), mediated by the ABA receptor (PYR/PYL/RCAR), can also regulate the composition and content of carotenoids, such as lycopene and β -carotene[27], during tomato fruit ripening. In addition, it was recently shown that jasmonic acid (JA) can regulate fruit carotenoid metabolism in tomato by inducing lycopene production[28]. Finally, the transcription factor BRASSINAZOLE RESISTANT1 (BZR1), involved in brassinosteroid (BR) signaling, has been shown to have a role in carotenoid accumulation in tomato fruit[29]. The identification of a BZR1 homolog in this study suggests that the BR signaling pathway may also be associated with regulation of carotenoids in apricot.

PSY is believed to be a rate-limiting enzyme in the carotenoid biosynthetic pathway[2], and changes in its expression have been reported in response to abiotic stresses, such as an up-regulation during salt and

drought stress in maize (*Zea mays*)[30], and upregulation during salt stress in *A. thaliana*[31]. Lycopene biosynthesis is inhibited and completely blocked, respectively, at temperatures lower than 12°C or higher than 32°C[32], and it has been suggested that high temperatures (35°C) can stimulate the transformation of lycopene into β -carotene, thereby specifically inhibiting lycopene accumulation[33]. Enriched CO₂ concentrations have also been shown to significantly enhance the lycopene and β -carotene content in tomato fruit[34]. Even in the absence of specifically applied stress conditions in this study, we identified two *PSY* genes, which were differentially expressed during apricot ripening.

Past research has revealed that many developmental and ripening related transcription factors are also involved in the synthesis and accumulation of carotenoids. These include the MADS-box genes AGAMOUS-like 1 and FRUITFULL, which modulate carotenoid accumulation during tomato fruit ripening[35], and the MADS-box transcription factors, RIN and FUL1/TDR4, which directly bind to the *SIPSY1* and *SIPDS1* promoters to regulate carotenoid synthesis[36, 37]. Recently, the citrus CsMADS6 transcription factor was found to promote carotenoid accumulation by coordinately modulating the expression of *LCYb1*, *PSY*, *PDS* and *CCD1* [38], and tomato SINAC1 and SINAC4^{48,49} and papaya (*Carica papaya*) CpNAC1^{50,51} are known to have important regulatory effects on carotenoid accumulation during ripening[39, 40]. Finally, the CrMYB68 transcription factor has been found to suppress the expression of *CrBCH2* and *CrNCED5* from the citrus carotenoid pathway[41].

A recent study showed that the MYB7 transcription factor can activate the *AdLCYB* promoter, thereby playing a major role in kiwifruit (*Actinidia deliciosa*) carotenoid biosynthesis[42]. Another MYB transcription factor, GOLDEN2-LIKE (GLK2), was shown to enhance lycopene levels in tomato fruit at the red ripe stage[43, 44],[45]. Here, we identified twenty-four WRKY and nine bHLH transcription factors, which have also been implicated in papaya fruit carotenoid biosynthesis[46], but their regulatory roles are still unknown. In the current study we identified a number of transcription factors, such as MADS14, NAC25, MYB1R1, MYB21, MYB44, MYC2, GLK1 and GLK2, that are likely involved in carotenoid metabolism in apricot fruit.

Conclusions

In summary, WGCNA analysis revealed two modules that were highly correlated with carotenoid metabolism, containing 33 structural genes and 93 transcription factors, including ERF4, NAC25, MADS14, HY5 and PIF3/4. It is known that carotenoid metabolism in apricot fruit is regulated by light, temperature, developmental factors and phytohormones (Fig. 7), and our findings provides and provides new insights into regulation of carotenoid metabolism and control of fruit color. These data sets will also provide a useful platform for further functional studies of candidate genes.

Methods

Plant materials

Fruit from different developmental and ripening stages (fruitlet, F, 21 DPA; enlargement, E, 27 DPA; turning, T, 57 DPA; commercial maturation, CM, 65 DPA and full ripe, FR, 74 DPA) were gathered from the National Fruit Tree Germplasm Repository, Academy of Xinjiang Agricultural Sciences, Luntai, Xinjiang, China (Fig. 1A). All samples were transported to the laboratory on the day of harvest. A set of fruit of uniform size and no obvious mechanical damage were selected for the subsequent experiments. Fifty fruit were regarded as one replicate, and three biological replicates were used for each sample. For each replicate, 20 fruit were used to measure a range of physiological indices. Other fruit were cut into small cubes and then immediately frozen in liquid nitrogen and stored at -80°C for measurement of carotenoid contents, and for RNA-seq and gene expression analyses.

Determination of basic quality parameters

The firmness of flesh was measured by a hardness tester (Model: HL-300, Xianlin Non Detection Device Co., Ltd., Nanjing, China) with 8 mm probe. The three fruits were squeezed juices and mixed in a group, then using handheld digital refractometer (B32T Brix Meter, Guangzhou Ruiqi Trading Company., Guangdong, China) to determine the total soluble solid (TSS). Titratable acid (TA) were determined after juices were diluted 100 times. The color of the flesh at different developmental stages was measured by Hunter's Mini Scanning Colorimeter (Hunter Associates Laboratory, Inc., Reston, VA, USA). And the color index CCI (Citrus Color Index) was calculated by the following formula: $\text{CCI} = 1000 \times a^* / (L^* \times b^*)$. Take three fruits as a repeat and each sample were repeated three times.

Extraction and carotenoid detection

Carotenoids were extracted using a protocol that was modified from a previous methods[47]. Eight grams of fruit peel was placed into screw-top tubes and 50 mL of extraction solvent (hexane/acetone/ethanol, 50:25:25, v/v) was added. After standing for 30 min, the samples were centrifuged for 5 min at 6500 rpm. The colored top layer of hexane was recovered and transferred to a volumetric flask, where it was dried under nitrogen, and dissolved in 2 mL of methyl tert-butyl ether (MTBE), before the solution was transferred to 2 mL of 10% methanol/potassium hydroxide. The mixture was allowed to stand for one hour and separated in a separating funnel, rinsed twice with water and once with 0.1% butylhydroxytoluene (BHT)/MTBE. The rinsed solution was transferred to a brown bottle and dried under nitrogen, before 2 mL methanol/acetone (2:1) was added and the solution filtered through a 0.22 μm filter membrane.

Carotenoids were separated and detected by high pressure liquid chromatography (HPLC, Waters, Milford, MA, USA) with a C30 chromatography column (250 mm $\times$ 4.6 mm, 5 μm , YMC, Wilmington, NC, USA). The mobile phase flow rate was 1 mL/min, and the column temperature was 25°C , with a detection wavelength of 450 nm, and an injection volume of 20 μL . The mobile phase composition was: methanol, MTBE, water. Carotenoids were identified by comparison to standard retention times and UV-visible

spectral peaks. The quantification of carotenoids was performed using a standard curve and expressed as µg/g. Three replicates for each sample were used.

RNA extraction, library construction and data analysis

Total RNA was isolated and extracted from 1 g of fruit peel for each sample using the Tiangen reagent kit ((Tiangen, Beijing, China)) as previously described[48]. RNA purity, concentration and integrity were determined using a Nanodrop 2000 ((NanoDrop 2000, Wilmington, DC, USA)) and denaturing agarose gel electrophoresis. RNA-seq libraries of T, CM and FR staged fruit were constructed using an Illumina TruSeq RNA Library Prep Kit v2 following standard procedures, and three biological replicates were used for each stage. The libraries were sequenced on a Illumina HiSeq™ 2000 system at the Beijing Genomics Institute (BGI), China. Clean reads were obtained by removing the linker, repeating redundant and low-quality sequences from the raw reads. The sequence data from this study have been deposited in the National Center for Biotechnology Information (NCBI) under Sequence Read Archive (SRA) accession number PRJNA530709.

De novo assembly and functional annotation

Transcriptome *de novo* assembly was carried out using the short read assembling program Trinity (<http://trinityrnaseq.sourceforge.net/>). The assembled unigenes were first de-redundant and further spliced using the TGICL package (TIGR tools), and then the sequences were clustered by homologous transcripts to obtain the final unigenes. Subsequently, unigenes were divided into two classes: one with several unigenes with >70% similarity, with the prefix CL and the cluster ID after; and the other consisting of singletons, with the prefix unigene. Finally, blastx alignment (evalue < 0.00001) between unigenes and NR, Swiss-Prot, KEGG and COG protein databases was performed, and the best aligning results were then used to decide sequence direction of the unigenes[49]. According to the NR annotation information, the GO annotation of unigenes was obtained using the Blast2GO software[50]. WEGO software (<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>) was used to performed GO functional classification of all the unigenes [51].

Identification of DEGs

DEGs were identified using FPKM values[52] and pairwise comparisons with a false discovery rate (FDR) threshold of < 0.001 and an absolute value of log2 Ratio ≥ 1. Overall unigene expression patterns, excluding those with no significant changes in expression or with changes > 32 fold as determined by FDR analysis.

GO and KEGG pathway enrichment analysis

First, the DEGs were mapped to each term in the GO database, and the number of genes in each term was determined. Based on gene numbers for each GO term, a hypergeometric test was applied to determine the GO terms significantly enriched for DEGs compared with the genome background. A corrected p-value < 0.05 was used as the threshold and the GO terms meeting this condition were defined as significantly enriched for DEGs. KEGG (<http://www.genome.jp/kegg/>) was used to perform a pathway enrichment analysis using the DEGs. After multiple tests and corrections, we determined that pathways with a Q-value 0.05 were significantly enriched in DEGs[53].

WGCNA and visualization of gene networks

A total of 16,773 unigenes that had FPKM values > 1 were used for WGCNA analysis using the WGCNA package in R[54]. The modules were obtained using the automatic network construction function. Modules were identified using default settings, except that the soft power was 10, min module size was 30, and the merge cut height was 0.25. Finally, Cytoscape 2.8 (<http://www.cytoscape.org/>) was used to construct the carotenoid metabolism regulatory network based on the WGCNA modules.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data sets supporting the results of this article are included within the article and its additional files. All the obtained sequences from apricot were deposited to the NCBI Sequence Read Archive (SRA) repository under accession number PRJNA530709 and were released on April 3, 2019

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (No. 31872046).

Authors' contributions

W.X. designed the project. L.Z. measured the carotenoids and prepared the manuscript. Q.Z. analyzed the data. W.L. participated in collecting the materials. S.Z. participated in assayed the basic physiology indexes. All authors read and approved the final manuscript.

Acknowledgements

We thank Julie Caruana (Department of Cell Biology and Molecular Genetics, University of Maryland) for comments on an earlier version of this manuscript.

References

1. Rao AV, Rao LG: Carotenoids and human health. *Pharmacol Res* 2007, 55(3):207-216.
2. Yuan H, Zhang J, Nageswaran D, Li L: Carotenoid metabolism and regulation in horticultural crops. *Horticulture Research* 2015, 2:15036.
3. Fratianni F, Ombra MN, d'Acierno A, Cipriano L, Nazzaro F: Apricots: biochemistry and functional properties. *Current Opinion in Food Science* 2018, 19:23-29.
4. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J *et al*: A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* 2013, 14(6):671-683.
5. Malone JH, Oliver B: Microarrays, deep sequencing and the true measure of the transcriptome. *Bmc Biology* 2011.
6. Li G, Ma J, Tan M, Mao J, An N, Sha G, Zhang D, Zhao C, Han M: Transcriptome analysis reveals the effects of sugar metabolism and auxin and cytokinin signaling pathways on root growth and development of grafted apple. *Bmc Genomics* 2016, 17(1):150.
7. Garcia-Seco D, Zhang Y, Gutierrez-Mañero FJ, Martin C, Ramos-Solano B: RNA-Seq analysis and transcriptome assembly for blackberry (*Rubus* sp. Var. Lochness) fruit. *BMC Genomics* 2015, 16(1):5.
8. Yu K, Xu Q, Da X, Guo F, Ding Y, Deng X: Transcriptome changes during fruit development and ripening of sweet orange (*Citrus sinensis*). *BMC Genomics* 2012, 13(1):10.
9. Sánchez-Sevilla JF, Cruz-Rus E, Valpuesta V, Botella MA, Amaya I: Deciphering gamma-decalactone biosynthesis in strawberry fruit using a combination of genetic mapping, RNA-Seq and eQTL analyses. *BMC Genomics* 2014, 15(1):218.

10. Zhang Q, Feng C, Li W, Qu Z, Zeng M, Xi W: Transcriptional regulatory networks controlling taste and aroma quality of apricot (*Prunus armeniaca* L.) fruit during ripening. *BMC Genomics* 2019, 20(1):45.
11. Manganaris GA, Rasori A, Bassi D, Geuna F, Ramina A, Tonutti P, Bonghi C: Comparative transcript profiling of apricot (*Prunus armeniaca* L.) fruit development and on-tree ripening. *Tree Genetics & Genomes* 2011, 7(3):609-616.
12. Dragovic-Uzelac V, Levaj B, Mrkic V, Bursac D, Boras M: The content of polyphenols and carotenoids in three apricot cultivars depending on stage of maturity and geographical region. *Food Chemistry* 2007, 102(3):966-975.
13. Ruiz D, Egea J, Tomás-Barberán FA, Gil MI: Carotenoids from New Apricot (*Prunus armeniaca* L.) Varieties and Their Relationship with Flesh and Skin Color. *Journal of Agricultural and Food Chemistry* 2005, 53(16):6368-6374.
14. Gouble B, Albagnac G, Sarkissian G, Marty I, Audergon JM, Bureau S: Ethylene regulation of carotenoid accumulation and carotenogenic gene expression in colour-contrasted apricot varieties (*Prunus armeniaca*). *Journal of Experimental Botany* 2005, 56(417):1877-1886.
15. Nisar N, Li L, Lu S, Khin NC, Pogson BJ: Carotenoid metabolism in plants. *Molecular Plant* 2015, 8(1):68-82.
16. Sun T, Yuan H, Cao H, Yazdani M, Tadmor Y, Li L: Carotenoid Metabolism in Plants: The Role of Plastids. *Molecular Plant* 2018, 11(1):58-74.
17. Leivar P, Monte E: PIFs: Systems Integrators in Plant Development. *The Plant Cell* 2014, 26(1):56-78.
18. Llorente B, Martinez-Garcia JF, Stange C, Rodriguez-Concepcion M: Illuminating colors: regulation of carotenoid biosynthesis and accumulation by light. *Current Opinion in Plant Biology* 2017, 37:49-55.
19. Gabriela TO, Enamul H, Manuel RC: Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107(25):11626-11631.
20. Yongsheng L, Sherry R, Zhibiao Y, Cornelius B, Ageeth VT, Julia V, Chris B, Jim G: Manipulation of light signal transduction as a means of modifying fruit nutritional quality in tomato. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101(26):9897-9902.
21. Zhang L, Ma G, Kato M, Yamawaki K, Takagi T, Kiriwa Y, Ikoma Y, Matsumoto H, Yoshioka T, Nesumi H: Regulation of carotenoid accumulation and the expression of carotenoid metabolic genes in citrus juice sacs in vitro. *Journal of Experimental Botany* 2012, 63(2):871.
22. Liu LH, Zabaras D, Bennett LE, Aguas P, Woonton BW: Effects of UV-C, red light and sun light on the carotenoid content and physical qualities of tomatoes during post-harvest storage. *Food Chemistry* 2009,

115(2):495-500.

23. Ralf W, Dirk M, Tanja V, Dean D, Peter B: Transcription factor RAP2.2 and its interacting partner SINAT2: stable elements in the carotenogenesis of Arabidopsis leaves. *Plant Physiology* 2007, 145(3):1073-1085.
24. Marty I, Bureau S, Sarkissian G, Gouble B, Audergon JM, Albagnac G: Ethylene regulation of carotenoid accumulation and carotenogenic gene expression in colour-contrasted apricot varieties (*Prunus armeniaca*). *Journal of Experimental Botany* 2005, 56(417):1877-1886.
25. Chung M-Y, Vrebalov J, Alba R, Lee J, McQuinn R, Chung J-D, Klein P, Giovannoni J: A tomato (*Solanum lycopersicum*) APETALA2/ERF gene, SIAP2a, is a negative regulator of fruit ripening. *The Plant Journal* 2010, 64(6):936-947.
26. Lee JM, Joung JG, Mcquinn R, Chung MY, Fei Z, Tieman D, Klee H, Giovannoni J: Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor SIERF6 plays an important role in ripening and carotenoid accumulation. *Plant Journal* 2012, 70(2):191-204.
27. Sun L, Yuan B, Zhang M, Wang L, Cui M, Wang Q, Leng P: Fruit-specific RNAi-mediated suppression of SINCED1 increases both lycopene and β -carotene contents in tomato fruit. *Journal of Experimental Botany* 2012, 63(8):3097-3108.
28. Liu L, Wei J, Zhang M, Zhang L, Li C, Wang Q: Ethylene independent induction of lycopene biosynthesis in tomato fruits by jasmonates. *Journal of Experimental Botany* 2012, 63(16):5751.
29. Maria S, Bradbury LMT, Monaco RR, Wurtzel ET: Plastid localization of the key carotenoid enzyme phytoene synthase is altered by isozyme, allelic variation, and activity. *Plant Cell* 2012, 24(9):3725-3741.
30. Faqiang L, Ratnakar V, Jane Y, Torbert R, Wurtzel ET: The maize phytoene synthase gene family: overlapping roles for carotenogenesis in endosperm, photomorphogenesis, and thermal stress tolerance. *Plant Physiology* 2009, 147(3):1334-1346.
31. M Águila RS, Vicent A, Aurelio GC, Manuel RC, Antía RV: A root specific induction of carotenoid biosynthesis contributes to ABA production upon salt stress in arabidopsis. *Plos One* 2014, 9(3):805-813.
32. Dumas Y, Dadomo M, Lucca GD, Grolier P: Effects of environmental factors and agricultural techniques on antioxidant content of tomatoes. *Journal of the Science of Food & Agriculture* 2003, 83(5):369-382.
33. Hamauzu Y, Chachin K, Ueda Y: Effect of postharvest storage temperature on the conversion of 14C-mevalonic acid to carotenenes in tomato fruit. *Journal of the Japanese Society for Horticultural Science* 1998, 67(4):549-555.

34. Zhang Z, Liu L, Zhang M, Zhang Y, Wang Q: Effect of carbon dioxide enrichment on health-promoting compounds and organoleptic properties of tomato fruits grown in greenhouse. *Food Chemistry* 2014, 153(12):157-163.
35. Julia V, Pan IL, Arroyo AJM, Ryan MQ, Miyoung C, Mervin P, Jocelyn R, Graham S, Silvana G, James G: Fleshy fruit expansion and ripening are regulated by the Tomato SHATTERPROOF gene TAGL1. *Plant Cell* 2009, 21(10):3041-3062.
36. Masaki F, Toshitsugu N, Yoko S, Yasuhiro I: A large-scale identification of direct targets of the tomato MADS box transcription factor RIPENING INHIBITOR reveals the regulation of fruit ripening. *Plant Cell* 2013, 25(2):371-386.
37. Masaki F, Yoko S, Hiroyuki N, Mamiko K, Junji K, Toshitsugu N, Takafumi K, Yasuhiro I: Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and associated MADS box proteins. *Plant Cell* 2014, 26(1):89-101.
38. Lu S, Zhang Y, Zhu K, Yang W, Ye J, Chai L, Xu Q, Deng X: The Citrus Transcription Factor CsMADS6 Modulates Carotenoid Metabolism by Directly Regulating Carotenogenic Genes. *Plant Physiology* 2018, 176(4):2657-2676.
39. Mingku Z, Guoping C, Shuang Z, Yun T, Yi W, Tingting D, Zongli H: A new tomato NAC (NAM/ATAF1/2/CUC2) transcription factor, SINAC4, functions as a positive regulator of fruit ripening and carotenoid accumulation. *Plant & Cell Physiology* 2014, 55(1):119-135.
40. Ma N, Feng H, Meng X, Li D, Yang D, Wu C, Meng Q: Overexpression of tomato SINAC1 transcription factor alters fruit pigmentation and softening. *BMC Plant Biology* 2014, 14(1):351.
41. Zhu F, Luo T, Liu C, Wang Y, Yang H, Yang W, Zheng L, Xiao X, Zhang M, Xu R: An R2R3-MYB transcription factor represses the transformation of α - and β -branch carotenoids by negatively regulating expression of CrBCH2 and CrNCED5 in flavedo of Citrus reticulata. *New Phytologist* 2017, 216(1).
42. Ampomah-Dwamena C, Thrimawithana AH, Dejnopratt S, Lewis D, Espley RV, Allan AC: A kiwifruit (*Actinidia deliciosa*) R2R3-MYB transcription factor modulates chlorophyll and carotenoid accumulation. *New Phytologist* 2019, 221(1):309-325.
43. Waters MT, Moylan EC, Langdale JA: GLK transcription factors regulate chloroplast development in a cell-autonomous manner. *Plant Journal* 2010, 56(3):432-444.
44. Waters MT, Peng W, Muris K, Capper RG, Saunders NJ, Langdale JA: GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. *Plant Cell* 2009, 21(4):1109-1128.
45. Powell ALT, Bennett AB: Uniform ripening Encodes a Golden 2-like Transcription Factor Regulating Tomato Fruit Chloroplast Development. *Science* 2012, 336(6089):1711-1715.

46. Shen YH, Yang FY, Lu BG, Zhao WW, Jiang T, Feng L, Chen XJ, Ming R: Exploring the differential mechanisms of carotenoid biosynthesis in the yellow peel and red flesh of papaya. *BMC Genomics* 2019, 20(1):49.
47. Lee HS, Castle WS: Seasonal changes of carotenoid pigments and color in Hamlin, Earlygold, and budd blood orange juices. *Journal of Agricultural & Food Chemistry* 2001, 49(2):877.
48. Shan L, Li X, Wang P, Cai C, Zhang B, De Sun C, Zhang W, Xu C, I, Chen K: Characterization of cDNAs associated with lignification and their expression profiles in loquat fruit with different lignin accumulation. *Planta* 2008, 227(6):1243-1254.
49. Grabherr MG, Haas BJ, Moran Y, Levin JZ, Thompson DA, Ido A, Xian A, Lin F, Raktima R, Qiandong Z: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 2011, 29(7):644.
50. Ana C, Stefan GT, Juan Miguel GG, Javier T, Manuel T, Montserrat R: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005, 21(18):3674-3676.
51. Jia Y, Lin F, Hongkun Z, Yong Z, Jie C, Zengjin Z, Jing W, Shengting L, Ruiqiang L, Lars B: WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research* 2006, 34(Web Server issue):W293-W297.
52. Mortazavi A, Williams B, McCue K, Schaeffer L, Wold B: Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008, 5(7):621-628.
53. Minoru K, Michihiro A, Susumu G, Masahiro H, Mika H, Masumi I, Toshiaki K, Shuichi K, Shujiro O, Toshiaki T: KEGG for linking genomes to life and the environment. *Nucleic Acids Research* 2008.
54. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics* 2008.

Figures

Figure 1

Characterization of fruit quality traits during development and ripening. (A) Apricot fruit color changes; (B) Fruit weight, fruit firmness, total soluble solid (TSS) and titratable acid (TA) contents; (C) color index for yellow flesh (CCI) and changes in carotenoid content in apricot fruit. LSD, least significant difference ($P < 0.01$). Error bars are $\pm$ SD of the means of three biological replicates.

Figure 2

Principal component analysis (PCA) of transcriptome data. Three biological replicates per sample were analyzed for each ripening stage. The percentages on the axes indicate the values explained by each PCA.

Figure 3

Expression profiles of differentially expressed genes (DEGs) during apricot fruit ripening. (A) DEG expression dynamics during fruit ripening. (B) Venn diagram representing the number of DEGs in each of the three ripening stages.

Figure 4

Weighted gene co-expression network analysis (WGCNA) of apricot fruit during ripening. (A) Hierarchical clustering showing 13 modules of co-expressed genes. Each leaf in the tree represents one gene. (B) Module-carotenoid correlations and corresponding p-values. The left panel shows the 13 modules and the number of member genes. The right panel shows a color scale for module/trait correlations from -1 to 1.

Figure 5

Carotenoid biosynthesis pathway in apricot fruit. (A) Heatmap of the expression levels of differentially expressed genes (DEGs) involved in carotenoid metabolism, as determined by Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values. Columns and rows represent samples and genes in the heatmap, respectively. Red, blue and white colors indicate high expression, low expression and the absence (or undetectable levels) of detectable transcripts at the corresponding stage, respectively. (B)

Carotenoid biosynthesis pathway in apricot fruit. IPI, isopentenyl diphosphate isomerase; GGPPS, geranylgeranyl diphosphate synthase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ζ -carotene desaturase; CRTISO, carotenoid isomerase; LCYE, lycopene ϵ -cyclase; LCYB, lycopene β -cyclase; CHYB, β -carotene hydroxylase; ZEP, zeaxanthin epoxidase; VDE, violaxanthin de-epoxidase; NXS, neoxanthin synthase; CCD, carotenoid cleavage dioxygenase; NCED, 9-cis-epoxycarotenoid dioxygenase.

Figure 6

Co-expression networks of transcription factors and structural genes involved in carotenoid metabolism. The network includes transcription factors and structural genes from the “blue” (A) and “turquoise” (B) modules. Dot size and color represent the number of connections between genes.

Figure 7

Carotenoid metabolic model in ripening apricot fruit. Red arrows represent positive regulation; blue lines represent negative regulation; blue font in brackets indicates the transcription factors and structural genes identified in this work. PIF, phytochrome interacting factor; HY5, LONG HYPOCOTYL5; COP1-like, cop1-like protein; AP2a; APETALA2-like protein; RAP2-4, Ethylene-responsive transcription factor RAP2-4; ERF, Ethylene response factor ; PYR, Pyrabactin resistance; PYL, PYR1-like; RCAR1, regulatory component of abscisic acid receptors 1; MeJA, methyl-jasmonate; BZR1, BRASSINAZOLE RESISTANT1; MADS, MADS-box transcription factor 6; AGAMOUS-like 1, Agamous-like MADS-box protein; FRUITFULL, FRUITFULL-like MADS-box; NAC, NAC transcription factor ; MYB, MYB transcription factor ; GLK2, Golden 2-like transcription factor; MYC, MYC transcription factor ; bHLH, bHLH transcription factor ; WRKY, WRKY transcription factor; ET, Ethylene; ABA, abscisic acid; JA, jasmonic acid; BR, brassinolide.

Supplementary Files

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

- [supplement1.xlsx](#)
- [supplement2.xlsx](#)

- [supplement3.docx](#)
- [supplement3.xlsx](#)
- [supplement5.xlsx](#)
- [supplement6.xlsx](#)
- [supplement7.xlsx](#)
- [supplement8.docx](#)
- [supplement9.xlsx](#)