Nucleotide Variation In The *Phytoene Synthase* (*ClPsy1*) Gene Contributes To Golden Flesh In Watermelon (*Citrullus Lanatus* L.)

Shi Liu (shiliu@neau.edu.cn)
Northeast Agricultural University https://orcid.org/0000-0003-1364-4043
Zhongqi Gao
Northeast Agricultural University
Feishi Luan
Northeast Agricultural University
Xuezheng Wang
Northeast Agricultural University
Zuyun Dai
Anhui Jianghuai Horticulture Technology Co., Ltd.
Zhongzhou Yang
Anhui Jianghuai Horticulture Technology Co.
Qian Zhang
Horticulture Institute, Anhui Academy of Agricultural Sciences

Research Article

Keywords: Watermelon, Flesh color, Phytoene synthase, β-carotene

Posted Date: August 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-655111/v1

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on October 11th, 2021. See the published version at https://doi.org/10.1007/s00122-021-03958-0.

Abstract

Vitamin A deficiency is a worldwide public nutrition problem, and β-carotene is the precursor for vitamin A synthesis. Watermelon with golden flesh (gf, due to accumulated abundance of β-carotene) is an important germplasm resource. In this study, a genetic analysis of gf segregating populations indicated that gf was controlled by a single recessive gene. BSA-seq and an initial linkage analysis placed the *gf* locus in a 290-Kb region on watermelon chromosome 1. Further fine mapping in a large population with over 1,000 F_2 plants narrowed this region to 39.08 Kb harboring two genes, *Cla97C01G008760* and *Cla97C01G008770*, which encode phytoene synthase (*ClPsy1*) and GATA zinc finger domain-containing protein, respectively. Gene sequence alignment and expression analysis between parental lines revealed *Cla97C01G008760* as the best possible candidate gene for gf trait. Nonsynonymous SNP mutations in the first exon of *ClPsy1* between parental lines cosegregated with the gf trait only among individuals in the genetic population but were not related to flesh color in natural watermelon panels. Promoter sequence analysis of 26 watermelon accessions revealed two SNPs in the *cis*-acting element sequences corresponding to MYB and MYC2 transcription factors. RNA-seq data and qRT-PCR verification showed that two MYBs and one MYC2 exhibited expression trends similar as *ClPsy1* in the parental lines, which may thus play roles in the regulation of *ClPsy1* expression. Our research findings indicate that the gf trait is determined not only by *ClPsy1* but also by *ClLCYB*, *ClCRTISO* and *ClNCED7*, which play important roles in β-carotene accumulation in watermelon flesh.

Key Messages

A gene controlling the golden flesh trait in watermelon was firstly discovered and fine-mapped to a 39.08-Kb region on chromosome 1 through a forward genetic strategy, and *Cla97C01G008760* (annotated as phytoene synthase protein, *ClPsy1*) was recognized as the most likely candidate gene.

Introduction

Watermelon (*Citrullus lanatus* L.) is a commercial Cucurbitaceae crop that is cultivated worldwide and consumed fresh. The flesh color of watermelon is an important trait based on consumer demands and market requirements. The genetic diversity of multiple flesh colors (red, pink, orange, canary yellow, pale yellow, white and pale green) in watermelon is caused by genome domestication and human breeding selection (Guo et al., 2019). Carotenoids are the main pigment responsible for formation of the watermelon flesh color. Specifically, lycopene is the key pigment responsible for the red and pink flesh colors (Sun et al., 2018), whereas β-carotene, tetra-*cis*-lycopene, and ξ-carotene are responsible for the orange flesh color (Branham et al., 2017). In addition, violaxanthin and lutein are the two main pigments responsible for the formation of canary and pale yellow flesh, and violaxanthin is significantly more abundant in canary yellow than in pale yellow watermelon (Yuan et al., 2015 and Fang et al., 2020). White-fleshed watermelon contains only small amounts of lutein and violaxanthin (Lv et al., 2015). Carotenoids, which form part of the human diet, serve as precursors of vitamin A and substrates for many nutrients and also exhibit antioxidant activity. Although these compounds are indispensable for

health, humans are incapable of *de novo* carotenoid synthesis (Rodriguez-Concepcion et al., 2018). The World Health Organization (WHO) estimates that approximately 190 million preschool children are deficient in vitamin A, and vitamin A deficiency is one of the major public health problems in the world. β-Carotene is the precursor for vitamin A synthesis, and the consumption of horticulture crops with high levels of β-carotene (golden flesh), which constitute an important germplasm resource, is an effective strategy for preventing vitamin A deficiency.

Scholars began studying the inheritance of the watermelon flesh color as early as 1937(Porter, 1937). Canary yellow (*C*) is dominant to other colors (*c*) except white (*Wf*), which is epistatic to canary yellow (*B*) (Wehner, 2007). Coral red (*Y*) is more common than orange (y^O) and salmon yellow (*y*), and orange (y^O) dominates salmon yellow (*y*). In addition, canary yellow (*C*) is epistatic to *Y* and influenced by inhibitor of canary yellow (*i-C*), and a homozygous gene results in red flesh even in the presence of *C* (Henderson et al., 1998). Bang et al. (2010) reported that the *py* gene controls the formation of pale yellow-fleshed watermelon. To date, many scientists have summarized a series of QTLs or genes related to watermelon flesh color. In 2003, two red flesh-related QTLs, which were detected in groups 2 and 8, were first reported (Hashizume et al., 2003). The biosynthesis of carotenoids has been well studied in plants (Sun et al., 2018) and reported in watermelon. *qFC.1* is reportedly related to β-carotene in a 2.4-Mb region on chromosome 1 that contains *ClPsy1* (Branham et al., 2017). The pale green flesh color is also regulated by the major effective QTL *qfc10.1* located on chromosome 10 spanning an approximately 519-Kb region, and *Cla97C10G185970*, annotated as plastid lipid-associated protein, was identified as the most candidate gene (Pei et al., 2021). The red flesh color gene *ClLCYB* (*lycopene beta-cyclase*) has been fine mapped to chromosome 4 (Liu et al., 2015; Wang et al., 2019). Zhang and colleagues found that the ClLCYB protein abundance, instead of the *ClLCYB* transcript level, is negatively correlated with lycopene accumulation (Zhang et al., 2020). Genes located in the carotenoid metabolic pathway are not the only factors responsible for flesh color formation. Y^{scr} (a single dominant gene), which has been fine mapped to an approximately 40-Kb region on chromosome 6, generates the scarlet red flesh color rather than the common red flesh color in watermelon. Four genes encoding glycine-rich cell wall structural proteins are regarded as candidate genes (Li et al., 2020). *ClPHT4;2* regulates chromoplast development, and the expression level of *ClPHT4;2* in red-fleshed accessions is approximately 10-fold higher than that in white-fleshed accessions. In addition, the overexpression of *ClPHT4;2* lead to a high accumulation of carotenoids (Zhang et al., 2017).

The gf trait has been reported in some plants, such as cauliflower (Li et al., 2001), cassava (Welsch et al., 2010), melon (Tzuri et al., 2015), rice (Bai et al., 2016), wheat (Wang et al., 2014), potato (Mortimer et al., 2016) and banana (Paul et al., 2017), and the *Psy* and *Or* genes are involved in the formation of this trait. *SlPsy1* is related to carotenoid accumulation in ripening tomato fruits (Fantini et al., 2013), and the silencing of *SlPsy1* in tomato using CRISPR/Cas9 changes the flesh color from red to yellow (Dahan-Meir et al., 2018). In watermelon and citrus, the types and expression patterns of the *Psy* genes are similar to those found in tomato (Peng et al., 2013; Lv et al., 2015). Together with the *DXS*, *LCYB*, *CHY* and *ZEP*

genes, *Psy* always affects the gf trait at the transcriptional level (Pons et al., 2014, Zeng et al., 2015, Farre et al., 2016).

The chromoplast is the main location for carotenoid accumulation, and enhancing the metabolic capacity would be beneficial for gf trait formation. The *Or* gene is known to play a role in carotenoid generation and chromoplast differentiation and exhibits chaperone activity for regulating *Psy* at the posttranscriptional level (Welsch et al., 2018). *Or* can also increase carotenoid accumulation by inhibiting hydroxylase and degradative reactions (Chayut et al., 2017), and overexpression of the *Or* gene in potato and white maize endosperm could enhance carotenoid synthesis by approximately 6- and 32-fold, respectively (Li et al., 2012, Berman et al., 2017). The carotenoid content cannot be enhanced by overexpression of the *Or* gene in tissues with abundant β-carotenoid accumulation (Shumskaya et al., 2012).

Although the *Psy* gene plays an important role in carotenoid synthesis in many plants, molecular markers developed based on forward genetics results have more effective applications in breeding. In watermelon, the gf trait remains poorly understood. The objectives of our study were to understand the genetic basis of the golden flesh color trait in watermelon and identify the genes responsible for this important trait in segregating populations. Suggestions for marker-assisted selection (MAS) for the golden trait are also provided based on the results.

Materials And Methods

Plant materials and phenotype evaluation

“Cream of Saskatchewan” (COS) with a pale-yellow flesh color and the watermelon accession PI 192938 with the gf trait (orange flesh) were used as the parental materials for obtaining the F_1, and F_2 and backcross populations. COS was kindly provided by Angela R. Davis at the U.S. Department of Agriculture, Agricultural Research Service, South Central Agricultural Research Laboratory, and PI 192938 was obtained from the U.S. National Plant Germplasm System. All the experimental materials were grown in a greenhouse at the Xiangyang Experimental Agricultural Farm of Northeast Agricultural University, Harbin, China, from 2018 to 2020. The F_2 populations were planted in 2018 and 2019 (93 and 297 plants, respectively) to verify the genetic segregation ratio and initial mapping, and BC_1P_1 (37 plants) and BC_1P_2 (84 plants) populations were planted in 2019 to further confirm their genetic inheritance. A total of 1,003 F_2 individuals were planted in 2020 for recombinant selection and fine mapping. Fifteen days after planting, each plant was numbered with a tag, and young and disease-free leaves were sampled and freshly frozen for DNA extraction using the improved hexadecyl trimethyl ammonium bromide (CTAB) method. All the plants were artificially pollinated, and this date was recorded; and 35 to 40 days after pollination, the fruits were collected, cross-cut and photographed for flesh evaluation.

BSA-seq and initial mapping

Equal amounts of genomic DNA from 20 pale yellow and 20 gf individuals were selected from 297 F_2 plants (in 2019) for flesh color gene pool construction. The two gene pools and genomic DNA of the parental lines were resequenced at the BGI Research Institute using the Illumina HiSeq Xten platform (at least 20× coverage genome sequencing depth for each sample) for BSA-seq analysis. The sequence data were aligned to the reference genome 97103 v2 (ftp://cucurbitgenomics.org/pub/cucurbit/genome/watermelon/97103/v2/) using Burrows-Wheeler Aligner (BWA, Li and Durbin, 2009) to obtain the snp.vcf file. The SNP sites were identified with SAMtools software (Li et al., 2009). Different homozygous sites in the two gene pools of the F_2 population were extracted to calculate the SNP variation frequency and analyze the regions significantly associated with the gf trait. Subsequently, 500 bp before and after the SNP site sequences were extracted from the resequencing data for molecular marker exploitation.

Candidate cleaved amplified polymorphic sequence (CAPS) loci were detected using SNP2CAPS software (Thomas et al., 2004) with the snp.vcf file. According to the BSA-seq results, CAPS loci evenly distributed in the initial chromosome region were designed with nine restriction endonucleases (*Hinc*II, *Msp*I, *Bsr*I, *Mbo*I, *Taq*I, *Alu*I, *Xho*I, *Dra*I and *BsaH*I) using Primer Premier v6.24. The protocols used for PCR amplification, mixing and enzyme digestion were previously described. SNP sites that could not be converted into CAPS markers were designed as Kompetitive Allele Specific PCR (KASP) markers and genotyped at the Vegetable Research Center of the Beijing Academy of Agricultural and Forestry Sciences. Individuals with recessive traits in the F_2 population planted in 2019 were selected and genotyped for initial mapping.

Fine mapping and candidate gene sequence analysis

A total of 1,003 F_2 individuals were sown in a greenhouse in the spring of 2020, and the two flanking markers in the initial mapping region were used for recombinant selection. All recombinants were transplanted to the greenhouse at the Xiangyang Experimental Agricultural Station of Northeast Agricultural University, Harbin China, and self-pollinated to obtain $F_{2:3}$ families. The flesh color of mature fruit was recorded, photographed and genotyped using new markers in the initial mapping segment to detect recombinant events for narrowing down the target region. The $F_{2:3}$ seeds of recombinants with dominant traits were planted with 30 plants of each family in the autumn of 2020. The genotype of the recombinants with dominant traits was confirmed according to the flesh color segregation in their $F_{2:3}$ families. Recombinants from the initial mapping panel were also selected to facilitate fine mapping.

Candidate gene annotation in the fine-mapping region was performed with the reference genome (97103 v2). A coding sequence comparison was first performed with the resequencing data, and the results were further confirmed with the sequences from the gene cloning results to detect nonsynonymous SNPs and gene structure variations between COS and PI 192938. To examine the allele diversity of *ClPsy1* among natural watermelon populations, resequencing data from 24 watermelon accessions (5 varieties resequenced in our previous study and 19 varieties resequencing data acquired from Guo et al., (2014) were used to extract the sequences of *ClPsy1* for comparison.

RNA-seq

Flesh samples of COS and PI 192938 were collected from the center area of four parental fruits at 10, 18, 26, 34 and 42 days after pollination (DAP). The flesh tissues were immediately frozen in liquid nitrogen and stored at -80°C for RNA extraction. Three out of five watermelon fruits with similar growth conditions were collected per sample. Flesh samples with three replicates at each stage were sent to Biomarker Technologies for RNA-seq with an Illumina HiSeq 2000 system. RNA-seq library construction and data analysis were performed according to the protocols described by Zhong et al. (2011) and Guo et al. (2013). Bowtie and Trapnell (Langmead et al. 2009) software were used for identification of fragment mismatches and read alignment to the watermelon genome (Guo et al. 2013). The number of reads mapped to each watermelon gene model was obtained and then standardized to the number of reads per kilobase of transcript per million mapped reads (RPKM). The RNA-seq data of COS were sent to the NCBI database with the SRA number PRJNA587316, whereas the SRA number of the PI 192938 flesh RNA-seq data was PRJNA733842. The flesh tissue samples of COS and PI 192938 were collected at the same time.

Gene expression analysis

The RNA Simple Total RNA Kit (Tiangen, China) and ReverTra Ace qPCR RT Kit (TOYOBO, Osaka, Japan) were used for total RNA extraction and cDNA synthesis, respectively. Specific primers for candidate genes (*Cla97C01G008760* and *196920*), transcription factors (TFs; *Cla97C10G196920*, *Cla97C02G046390* and *Cla97C06G112130*) and carotenoid metabolic pathway genes [*ClZDS* (*Cla97C06G118930*), *ClNCED-7* (*Cla97C07G137260*), *ClCRTISO* (*Cla97C10G200950*), *ClCHYB* (*Cla97C05G090480*), *ClLCYB* (*Cla97C04G070940*), *ClPDS* (*Cla97C07G142100*), *ClNCED-1* (*Cla97C01G024630*), and *ClCHXE* (*Cla97C01G002480*)] were designed for gene expression analysis through quantitative real-time polymerase chain reaction (qRT-PCR) (Table 3). SYBR Green Master Mix (Novogene, Beijing) was used to perform the qRT-PCRs in the QTOWER Real-Time PCR System (Analytik Jena, Germany) according to the manufacturer's instructions. qRT-PCR amplification and mixing were performed as previously described. Each experiment was performed with three biological repetitions and three technical repetitions, and relative gene expression levels were determined using the $2^{\Delta\Delta CT}$ method (Bustin et al 2009).

Promoter region cloning and candidate transcription factor prediction

The promoter region of *ClPsy1* was cloned with genomic DNA from five watermelon accessions with different flesh colors, namely, COS, PI 192938, LSW-177 (red flesh), PI 635597 (canary yellow flesh) and PI 186490 (white flesh), using the primers listed in Table 2. *Cis*-elements in each promoter were predicted using PlantCARE online software (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), and the five promoter sequences were compared to detect the variation. SNPs or structural alterations located in *cis*-elements were selected as important loci. The resequencing data of 24 other watermelon accessions were also used to extract the sequences of the *ClPsy1* promoter and thus detect the variation diversity.

The RNA-seq data of COS and PI 192938 flesh tissues were used to preliminarily view the expression patterns of the TFs that may bind with the *cis*-elements located in the variation region. TFs exhibiting expression patterns similar to that of *ClPsy1* between COS and PI 192938 may be regarded as important factors for further verification.

Statistical analysis

The genetic analysis and evaluations of the differences in gene expression were performed using SPSS v.21.0 software (SPSS Inc., Chicago, IL, USA). Prism 7.0 software was used (GraphPad Inc., La Jolla, CA, USA) for illustration preparation.

Results

The gf trait in watermelon is controlled by a simply inherited gene

The carotenoid composition and content in the mature flesh of COS and PI 192938 were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) in our previous study (Fang et al. 2020). β-carotene and violaxanthin appeared to be the two pigments showing major differences in content between the two parental materials, and the contents in PI 192938 (16.133 ± 0.952 μg/g) were approximately 53.2-fold higher than those in COS (0.033 ± 0.004 μg/g). Based on the flesh color associated with the two pigments, we speculated that the high accumulation of β-carotene in PI 192938 may be the main reason for the gf trait (Fig. 1a). The flesh color of F_1 was canary yellow, similar to that of COS, and four flesh categories were segregated in the F_2 generation: gf, pale yellow, canary yellow and gf mixed with canary yellow (Fig. 1b). F_2 individuals can also be divided into gf and non-gf groups (pale yellow, canary yellow and gf mixed with canary yellow, Fig. 1c to e). According to these classification criteria, the COS×PI 192938-F_2 population in 2018 (84 individuals in total) consisted of 17 gf and 67 non-gf plants and exhibited a 1:3 genetic ratio ($\chi^2 = 1.016$, $p = 0.313$), whereas in 2019, 279 F_2 individuals consisted of 74 gf and 205 non-gf plants, which was also consistent with the 1:3 genetic ratio ($\chi^2 = 0.345$, $p = 0.557$). For generation of a backcross population between F_1 and COS, none of the individuals exhibited the gf color. In the backcross population derived from F_1 and PI 192938, 39 plants had the gf color, and 45 plants had a non-gf color, which corresponded to a ratio of 1:1 ($\chi^2 = 0.429$, $p = 0.513$, Table 1). Based on the above results, we conclude that the *gf* trait in watermelon flesh is controlled by a simply inherited gene and that pale yellow flesh is partly dominant to the gf color.

BSA-seq and recombinant deletion of the gf locus in a 39-Kb region identified ClPsy1 as a candidate gene

After filtering low-quality and short reads, 125,549,810 and 60,739,330 clean read pairs were obtained for COS and PI 192938 with approximately 10.01 (28.76× depth coverage) and 8.56 (22.87× depth coverage) Gbp clean bases, respectively, and Q30 values above 89.89% were found. A total of 86.50% and 92.48% of these clean reads of COS and PI 192938, respectively, were successfully mapped to the reference genome, and 68,585,634 and 68,541,696 clean read pairs were generated from the gf pool (25.7× depth

coverage and 92.35% properly mapped ratio) and pale-yellow flesh pool (25.88× depth coverage and 93.16% properly mapped ratio), respectively, through the Illumina high-throughput sequencing platform. A total of 366,358 and 374,417 SNPs were identified between the reference genome and the two gene pools, respectively. The SNP index for each identified SNP was calculated, and the average SNP index was computed in a 1-Mb interval using a 10-Kb sliding window. By combining the SNP index information from the gf color pool and pale-yellow flesh pool, the ΔSNP index was calculated and plotted against the genome positions. According to the ΔSNP index value, an obvious signal related to the gf color was detected on chromosome 1, spanning approximately 2.99 Mb (from 8,912,000 bp to 11,900,000 bp, Fig. 2a).

A total of 30 CAPS and two KASP markers evenly distributed in BSA-seq chromosome segments were developed based on parental line resequencing data, and 10 markers (eight CAPS markers and two KASP markers) were used for initial mapping after polymorphism detection among COS, PI 192938 and their F_1 generation. Individuals with a recessive phenotype (gf trait) from 2019 were selected for genotyping with the 10 polymorphic markers. The candidate region was narrowed to a physical distance of 290.214 Kb (from 9,272,322 bp to 9,562,536 bp) using 11 recessive-trait plants (including nine recombinants) between the CAPS markers *Chr01_9272322* and *Chr01_9562536* with one and two recombinants (Fig. 2b). To further narrow down the initial mapping region precisely, a larger COS×PI 192938-F_2 segregating population including 1,003 individuals was subjected to genotyping of the primary flanking markers *Chr01_9242322* and *Chr01_9562536* in the spring of 2020. A total of 20 recombinants were screened for further fine mapping of the *gf* gene. Another nine polymorphic markers were developed to genotype the 20 recombinants. The target trait genotype of the dormant recombinants was confirmed based on phenotypic segregation in their F_3 families. Finally, the *gf* locus was delimited between the CAPS markers *Chr01_9440282* and *Chr01_9479366* (physical distance of approximately 39.08 Kb) with two and nine recombinants, respectively (Fig. 2c).

According to the watermelon reference genome, the 39.08-Kb region contained only two annotated candidate genes, *Cla97C01G008760* and *Cla97C01G008770*. *Cla97C01G008760* encodes a phytoene synthase protein (*ClPsy1*), and *Cla97C01G008770* was annotated as a GATA zinc finger domain-containing protein. To identify the candidate gene for the *gf* locus, we first analyzed the genomic variations in the two candidate genes between the parental lines with resequencing data. The results identified no polymorphic sites in *Cla97C01G008770*, whereas one nonsynonymous SNP mutation, $SNP^{9,448,870}$ (A→G, located in the first exon at the 9,448,870th bp position), was detected in the coding region of *Cla97C01G008760* between COS and PI 192938. In COS, base A encodes glutamic acid (Glu), whereas in PI 192938, this base is mutated to base G, resulting in an amino acid change from Glu to lysine (Lys). To further confirm the sequence variation, we cloned the coding regions of the two candidate genes in COS and PI 192938. This SNP mutation was still found between the two parental lines. We further developed this nonsynonymous SNP into the KASP marker *Chr01_9448870* and genotyped F_2 individuals from 2018 to 2019. As a result, $ClPsy1^{A:A}$ exhibited the gf color, whereas $ClPsy1^{G:A/G:G}$ showed

a non-gf color, which indicated that *Chr01_9448870* cosegregated with the phenotype in all the plants (Fig. 3a).

Although no variation in the *Cla97C01G008770* gene sequence was found between COS and PI 192938, we found some polymorphic sites in the promoter region. To further confirm this hypothesis, we analyzed the gene expression patterns of the two candidate genes in COS and PI 192938 flesh tissues collected from different stages of flesh color formation. The results showed that the two parental lines exhibited similar expression trends across the five developmental stages (10, 18, 26, 34 and 42 DAP), and no significant difference in *Cla97C01G008770* was found (Fig. 3b). For *Cla97C01G008760* (*ClPsy1*), the two parental lines also showed similar expression patterns (the expression level was upregulated gradually during flesh maturation), but at 26 DAP, the allele of *ClPsy1* in PI 192938 exhibited significantly higher expression than that in COS, and this increase continued to be observed until 42 DPA (mature stage, Fig. 3c). Our previous research also showed that 26 DAP may be an important developmental stage for flesh color formation (Fang et al., 2020). At this stage, the colored COS and PI 192938 watermelon flesh started to abundantly accumulate carotenoids. Hence, we hypothesized that *ClPsy1* is the most likely candidate gene for the *gf* locus and is responsible for high β-carotene accumulation in watermelon.

Nucleotide variation in the ClPsy1 **gene structure among natural watermelon accessions**

To examine the allelic diversity of the *ClPsy1* gene in natural watermelon groups, we examined the nucleotide variation of the *ClPsy1* locus in 26 resequenced accessions with different flesh colors (red, orange, canary yellow, pale yellow, light green and white), including 18 *C. lanatus*, four *C. mucosospermus* and four *C. amarus* accessions (Fig. 4). $SNP^{9,448,870}$ was still present, but this mutation was not correlated with flesh color among the different watermelon accessions and exhibited no obvious difference between the cultivated and wild-type watermelon groups. These results indicated that this site may not affect carotenoid accumulation. In addition to $SNP^{9,448,870}$, another nonsynonymous SNP mutation, $SNP^{9,448,438}$ (C→T, located in the first exon at the 9,448,438th bp position), was also detected. Interestingly, $SNP^{9,448,438}$ existed only in the *C. amarus* group, resulting in an amino acid substitution from proline (Pro in the *C. lanatus* and *C. mucosospermus* groups) to serine (Ser in the *C. amarus* group).

To analyze the reason for the low *ClPsy1* expression in COS, we cloned the 1,996-bp promoter sequence from four cultivated watermelon varieties: COS, PI 192938, LSW-177 (red flesh) and PI 635597 (canary yellow). The promoter sequences of 14 other cultivated watermelon accessions were extracted from their genome resequencing data. Interestingly, a total of six SNPs (SNP^{342}, SNP^{598}, SNP^{898}, $SNP^{1,257}$, $SNP^{1,634}$ and $SNP^{1,694}$) were detected in the COS promoter region compared with the other 17 watermelon accessions, which exhibited consistent promoter sequences. SNP^{598} and $SNP^{1,257}$ were located in the MYC- and MYB-binding sites, respectively, whereas the other four SNPs were not located in the sequence of any *cis*-acting element (Fig. 5a).

MYB and MYC2 may be important transcription factors regulating the expression level of ClPsy1

We then used the RNA-seq data of COS and PI 192938 flesh tissues (collected at 18, 26 and 42 DAP, data not shown in this manuscript) to obtain an overview of the expression patterns of all MYB and MYC TFs. A total of 65 MYB TFs with read per kilobase per million mapped reads (RPKM) values were detected, and only two MYB TFs (*Cla97C10G196920* and *Cla97C02G046390*) exhibited expression tendencies similar to that of *ClPsy1* in COS and PI 192938. An obvious significant difference in expression began to be observed at 26 DAP in PI 192938 compared with COS and continued to be observed at 42 DAP (mature stage). For all 22 MYC TFs, *Cla97C06G112130* (annotated as a MYC2 transcription factor) also showed an obvious significant difference in expression between COS and PI 192938 at 26 DAP. We further examined the expression levels of *Cla97C10G196920*, *Cla97C02G046390* and *Cla97C06G112130* in COS and PI 192938 flesh tissues collected at five developmental stages by qRT-PCR to verify the expression pattern. The results showed the same tendency as the RNA-seq data (Fig. 5b to d). LSW-177 was a red flesh-colored watermelon accession with the same genotype in the gene sequence and promoter region as PI 192938. The RNA-seq data between COS and LSW-177 (red flesh) flesh tissues (BioProject number PRJNA338036) were also used for analyzing the expression levels of *Cla97C10G196920*, *Cla97C02G046390* and *Cla97C06G112130*. The results showed that the three TFs also presented a higher expression level in LSW-177 than in COS (Fig. 8a to c). In watermelon fruit rinds, the expression levels of the three TFs were also clearly lower than those in the flesh of 97103 (red flesh), which indicated that these TFs may be expressed in tissues with high carotenoid accumulation according to the RNA-seq data of SRP012849 (Fig. 8d to f).

The conserved domains of *Cla97C10G196920* (148 aa) and *Cla97C02G046390* (110 aa) were extracted and compared to the Arabidopsis Information Resource (TAIR). The results showed that AtMYB21/AtMYB3, AtMYB24, AtMYB57, AtMYB59, and AtMYB48 were the first three (or two) homologs. *Cla97C06G112130* has two conserved domains: the N-terminus of the bHLH-MYC and R2R3-MYB TFs and the N-terminus of a family of MYB and MYC transcription factors (156 aa). The other superfamily is the bHLH domain superfamily (70 aa), and AtMYC2 exhibited the highest homology. We speculated that ClMYB and ClMYC2 may be two important TFs regulating *ClPsy1* expression due to variations in the binding sites in their promoter region between COS and PI 192938. Although MYB and MYC TFs have many functions, they have not been reported to play a role in flesh color formation in watermelon.

The gene expression and genotype variations in carotenoid pathway genes between COS and PI 192938 provide insight into gf trait formation in watermelon flesh

We examined the transcript abundances of *ClPDS*, *ClZDS*, *ClCRTISO*, *ClLCYB*, *ClCHYB* and *ClNCED7* in COS and PI 192938 flesh tissues collected at five developmental stages (10, 18, 26, 34 and 42 DAP) by qRT-PCR (Fig. 6a to f). *ClPDS*, *ClZDS* and *ClCRTISO* exhibited the same expression trend as *ClPsy1* between the two parental lines. The transcript abundance of PI 192938 was always higher than that of COS throughout all developmental stages, particularly at 26 DAP. It has been reported that *ClLCYB* regulates lycopene accumulation at the protein level (Zhang et al., 2020), and COS and PI 192938 have the same single-nucleotide mutation as red-fleshed watermelon accessions. The mutation of G^{676th} to T^{676th} altered the 226th amino acid from valine (Val) to phenylalanine (Phe), whereas the mutation of

$G^{1,305th}$ to $C^{1,305th}$ altered the 435th amino acid from lysine (Lys) to asparagine (Asp). This finding indicated that the ClLCYB protein may have the same function in COS and PI 192938. The lycopene content in the mature flesh of the two parental lines was quite low compared with that in a red-fleshed variety identified in our previous study (Fang et al., 2020). The expression level of *ClNCED*-7 in both parental lines showed an increasing trend over time after pollination, and COS presented a significantly higher expression level than PI 192938, except at 26 DAP.

The sequence variations in genes encoding enzymes at each step of the carotenoid pathway were also analyzed between COS and PI 192938 using resequencing data. Three SNPs were found in the coding region of *ClZDS* in PI 192938 compared with that in COS, and two of these SNPs led to amino acid substitutions. Mutation of the 161st base (G→A) resulted in a change in the 54th amino acid from serine (Ser) to asparagine (Asp), and mutation of the 480th amino acid (G→T) resulted in a change in the 160th amino acid from lysine (Lys) to asparagine (Asp). Only one nonsynonymous residue was detected in *ClCRTISO*, and mutation of the 526th base (T→C) changed the 176th amino acid from tyrosine (Tyr) to histidine (His). No variations were detected in the coding region of *ClPDS* or *ClNCED-7*.

According to the above-described results, we speculated the cause of the high β-carotene content shown in Fig. 7. High expression of *ClPsy1* in PI 192938 contributed to increased phytoene accumulation, and the abundance of phytoene may upregulate the expression of *ClPDS*, *ClZDS* and *ClCRTISO* at each step of the carotenoid metabolism pathway to result in the synthesis of higher amounts of zeta-carotene and tetra-*cis*-lycopene in PI 192938. Tetra-*cis*-lycopene could be isomerized through *ClCRTISO* to generate lycopene, which is the carotenoid upstream of β-carotene. The same genotype of the ClLCYB protein may have a similar cyclization effect in PI 192938 and COS, and nearly all lycopene can be cyclized into β-carotene. High expression of *ClCHYB* increased violaxanthin accumulation, whereas low expression of *ClNCED-7* may prevent β-carotene metabolism in PI 192938. These factors may be the main reasons for the high accumulation of violaxanthin and β-carotene in PI 192938 and thus its orange flesh color.

Discussion

Various patterns of *ClPsy1* regulation may exist among different watermelon accessions

The first committed and rate-controlling step is mediated by the phytoene synthase (Psy) protein, which catalyzes the conversion of two molecules of GGPP to phytoene (colorless) as the first carotenoid (Cazzonelli and Pogson, 2010; Welsch et al., 2010; Giuliano et al., 2017). As an important gene in the carotenoid metabolic pathway, *Psy1* has been reported to function in carrot (Maass et al., 2009 and Massimo et al., 2016), banana (Paul et al., 2017), citrus (Lu et al., 2018), tomato (Xiong et al., 2019) and pepper (Jeong et al., 2019). Compared with white and bitter wild-type watermelon, modern cultivated watermelon (*C. lanatus*) exhibits various flesh colors. *ClPsy1*, located in a selective chromosome segment, may play an important role in watermelon flesh color formation during the domestication process (Guo et al., 2020). In this study, the sequence variations in both coding and promoter regions were analyzed in 26 watermelon accessions belonging to three subspecies with different flesh colors. No

representative mutations were found between the *C. mucosospermus* and *C. lanatus* groups. However, multiple variations were detected in the *C. amarus* group compared with the *C. mucosospermus* and *C. lanatus* groups. A low gene expression level of *ClPsy1* was also found in *C. amarus* (such as PI 296341-FR, Guo et al., 2015). $SNP^{9,448,438}$ was detected only in the *C. amarus* group, but whether this SNP mutation alerts *ClPsy1* function needs further verification. Based on published RNA-seq data or previous research, white-fleshed *C. mucosospermus* (such as PI 186490, Fang et al., 2020 and Wang et al., 2021) showed a significantly lower expression level than colored cultivated watermelon varieties, although the former presented the same genotype in coding and promoter regions as the colored accessions. This feature may be regulated by TFs or epigenetic factors, but further research is needed. Compared with other watermelon accessions, two SNP mutations in *cis*-acting element sites were only found in COS. Further qRT-PCR results implied that these two SNP mutations may be the main reason for the low *ClPsy1* expression level.

Enzyme activity is also an important factor affecting *ClPsy1* gene function in carotenoid accumulation. The reduction in total carotenoids was consistent with the requirement of galactolipids for PSY protein activity in etiolated seedlings of *Arabidopsis* mutants (Fujii et al., 2018). In loquat (*Eriobotrya japonica* Lindl.), a mutant *EjPSY2Ad* lacking the C-terminal region (694-bp segment) in the fifth exon and thus no corresponding catalytic activity resulted in the absence of carotenoids and thus an inability to form white flesh (Fu et al., 2014). The aspartic acid-rich region (DXXXD) is an important functional domain of the PSY protein (Zhai et al., 2016). We also compared the mutation locations among watermelon accessions with different flesh colors, but none of the SNPs were located in this domain region. Among the varieties belonging to the *C. mucosospermus* and *C. lanatus* groups, $SNP^{9,448,870}$ was also not correlated with flesh color in the natural watermelon panel. These two lines of evidence may imply that the function of *ClPsy1* in flesh color formation is dependent on expression rather than a difference in enzymatic activity. These results showed that the expression pattern of *ClPsy1* was regulated in different ways in different watermelon subspecies.

Regulatory factors in carotenoid accumulation and flesh color formation

The function of *Psy* is affected by many regulatory factors at both the transcriptional and protein translation levels. In tomato, ripening inhibitors (RINs), through their interaction with the *SlPsy1* promoter, have been shown to regulate fruit carotenoid concentrations (Martel et al., 2011). B-box zinc-finger transcription factor (SlBBX20) can activate the expression of *SlPsy1*, leading to dark-green fruits and leaves and higher levels of carotenoids, by directly binding to a G-box motif in its promoter (Xiong et al., 2019). A transcriptomic analysis of AdMYB7-infiltrated *N. benthamiana* leaves showed upregulation of *NbPSY* (Ampomah-Dwamena et al., 2019). The overexpression of CsMADS6 in citrus calli can increase the carotenoid content by directly binding to the *CsPSY*, *CsPDS* and *CsCCD1* promoters (Lu et al., 2018). Two ClMYB and one ClMYC2 TFs exhibited expression trends similar to that of *ClPsy1* in COS and PI 192938 and were thus hypothesized to be regulatory factors of *ClPsy1*. Compared with the *Arabidopsis* genome, AtMYB21/AtMYB3, AtMYB24, AtMYB57, AtMYB48 and AtMYB59 showed high identity in conserved domains with the two ClMYBs (*Cla97C10G196920* and

Cla97C02G046390), whereas bHLH-MYC/R2R3-MYB and AtMYC2 showed the highest homology to ClMYC2 (*Cla97C06G112130*). With the exception of the two mutation sites, some other MYB- and MYC2-binding sites were located in the promoter regions of COS and PI 192938, which could partly explain why obvious expression levels of *Cla97C10G196920*, *Cla97C02G046390* and *Cla97C06G112130* in COS could be detected at 26 DAP, even though the binding sites participate in the variation.

The MYB TF family is very large and functionally diverse in plants. MYB members have been reported to participate in carotenoid metabolism. Chili pepper fruits synthesize and accumulate carotenoid pigments, which are responsible for the yellow, orange and red colors. *CaCCS* (capsanthin-capsorubin synthase), *CaBCH* (β-carotene hydroxylase), and *CaPSY* were clustered with six MYB-related genes (CaDIV1, CaDIV3, CaMYBR13, CaTRF2, CaMYBC1, and CaPHR9) and an atypical MYB (CaMYB5R) based on a coexpression analysis (Arce-Rodriguez et al., 2021). The R2R3-MYB transcription factor CrMYB68 directly regulates the α- and β-branches in the carotenoid pathway in *Citrus reticulata*. Reduced expression of *CrBCH2* and *CrNCED5* is responsible for the delay in α- and β-carotene synthesis. The expression of these genes is negatively correlated with the expression of CrMYB68 in green fruit (Feng et al., 2017). A coexpression network and transient expression analysis suggested a potential direct link between flavonoid and carotenoid biosynthesis pathway genes (*PSY*, *ZDS* and *CYP97C*) through MYB TF regulation in *Primula vulgaris* (Li et al., 2020). SlMYB72 directly binds to the *SlPSY*, *SlZISO*, and *SlLCYB* genes and regulates carotenoid biosynthesis in tomato. The downregulation of SlMYB72 decreases the lycopene content and promotes β-carotene production and chromoplast development (Wu et al., 2020). CpMYB1 (MYB44-like) and CpMYB2 are transcriptional repressors that can bind to the *CpPDS2*, *CpPDS4*, and *CpCHY-b* promoters and suppress the activities of these genes in papaya (Fu et al., 2020).

In *Medicago truncatula*, an R2R3-MYB protein (WHITE PETAL1, WP1) functions as a transcriptional activator that modulates floral carotenoid pigmentation by directly regulating the expression of multiple carotenoid biosynthesis-related genes (Meng et al., 2019). The silencing of strigolactone (SL, carotenoid-derived phytohormone) biosynthesis-related genes results in the upregulation of MYC2 (Xu et al., 2019). To date, the regulation of flesh color by MYB or MYC2 TFs in watermelon has not been studied. The binding effect and regulatory mechanism of *Cla97C10G196920*, *Cla97C02G046390* and *Cla97C06G112130* need further verification and research.

The *Or* gene (encoding a plastid-targeted protein containing a cysteine-rich zinc finger domain) is also another important gene for golden trait formation. In melon (Tzuri et al., 2015) and cauliflower (Li et al., 2001), the *Or* gene reportedly regulates β-carotene accumulation based on forward genetic research. *CmOr* exerts only a slight effect on *CmPsy1* expression but strongly affects the CmPsy1 protein levels and enzymatic activity in melon (Chayut et al., 2017). In *Arabidopsis* and sweet potato, the PSY protein is affected by OR/OR-like proteins (Zhou et al., 2015; Park et al., 2016). The AtOR protein enhances AtPSY protein stability and increases the enzymatically active proportion of AtPSY in Atclpc1 in *Arabidopsis* (Welsch et al., 2018).

Molecular breeding suggestions for golden-fleshed watermelon

This study provides the first report of the *gf* gene in watermelon. Although the *Psy* gene reportedly functions in gf trait formation in some plants, genetic evidence remains lacking. According to the results, abundant *ClPsy1* transcripts may lead to a rich β-carotene content in PI 192938. Red-colored flesh always has a higher *ClPsy1* expression level than flesh with other colors (Wang et al., 2016), starting at the color turning stage and ending at the mature stage. Interestingly, the *ClPsy1* content in the red-fleshed watermelon accession LSW-177 was higher than that in PI 192938 (Fang et al., 2020), but the β-carotene content in LSW-177 flesh was significantly lower than that in mature PI 192938 flesh (2.605±0.375 versus 16.133±0.952 μg/g), whereas the lycopene content in LSW-177 was significantly higher than that in PI 192938 (1.523±0.199 versus 25.950±0.390 μg/g). This finding indicated that the high expression of *ClPsy1* was not the only main reason for β-carotene accumulation. The upstream gene of β-carotene is *ClLCYB*, which encodes a cyclase that can catalyze the conversion from lycopene to β-carotene. The downregulation of *ClLCYB* causes lycopene accumulation, whereas the overexpression of *ClLCYB* in the red-fleshed line causes the flesh color to change to orange (β-carotene) (Zhang et al., 2020). Among watermelon accessions, two nonsynonymous SNP sites generate three haplotypes (*ClLCYB*^{red}, *ClLCYB*^{white}, and *ClLCYB*^{yellow}). COS and PI 192938 contained the same haplotype of *ClLCYB*^{yellow}, which indicated that the ClLCYB protein had the same function in the two lines. In PI 192938, nearly all lycopene was cyclized into β-carotene to form a gold flesh. This finding may indicate that *ClLCYB* is also an important gene for β-carotene accumulation and partly explains the accumulation of β-carotene observed in LSW-177, COS and PI 192938.

In tomato, the accumulation of tetra-*cis*-lycopene (the upstream carotenoid of lycopene) can also lead to the formation of orange flesh by *SlCRTISO* (Jayaraj et al., 2021, Dahan-Meir et al., 2018). In watermelon, the genotype of *ClCRTISO* also cosegregates with an orange flesh color (Jin et al., 2019). Visual observation alone does not allow differentiation of the two types of carotenoids. Although they have the same color, tetra-*cis*-lycopene is not converted into the precursor of vitamin A without β-cyclization. The golden trait not only focuses on flesh color but also considers the physiological function of carotenoids. For the gf trait, the MAS process depends only on the expression level and the genotype of *ClPsy1*, and visual observation is not sufficient or accurate.
The genotypes of *ClPsy1*, *ClCRTISO* and *ClLCYB* should also be considered at the same time to confirm the carotenoid composition.

“Push” and “block” are the two main strategies for gf trait molecular breeding and biofortification (Watkins et al., 2020). The “push” strategy involves enhancing the metabolic flux of carotenoids synthesized upstream of β-carotene and is regarded as the most effective method. The overexpression of *Psy* or 1-deoxy-D-xylulose-5-phosphate synthase (DXS) increases the synthetic amounts of total carotenoids or β-carotene. Cassava (Welsch et al., 2010), rice (Bai et al., 2016), wheat (Wang et al., 2014), potato (Mortimer et al., 2016) and banana (Paul et al., 2017) show up to 100- to 1,000-fold higher contents of total carotenoids. Silencing the expression of genes downstream of β-carotene to prevent the degradation of target products is another strategy for regulating the gf trait, and this method

is called “block”. In potato, the inhibition of zeaxanthin epoxidase (*ZEP*), lycopene epsilon-cyclase (*LCYE*) and carotene hydroxylase (*CHY*) could lead to approximately 10-fold increases in the zeaxanthin and β-carotene contents (Pons et al., 2014). “Push” and “Block” have always been used together to obtain better results: the overexpression of *CrtB* (a homologous gene of *Psy* in bacteria) accompanied by the silencing of *CHY* could produce a large amount of β-carotene in the endosperm of wheat (Zeng et al., 2015). Significantly high accumulation of zeaxanthin could be induced by the overexpression of *Psy1* and the silencing of *LCYE* in the maize endosperm (Farre et al., 2016).

Declarations

Acknowledgments This work was supported by funding from the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province [Grant# UNPYSCT-201816].

Author contribution statement Shi Liu designed the experiment, analyzed the data and wrote the manuscript. Zhongqi Gao performed the major experiments with the help of Shi Liu. Xuezheng Wang and Feishi Luan revised the manuscript. Zuyun Dai, Zhongzhou Yang, Qian Zhang provided part of materials used in this study and participated in phenotypic data collection. Shi Liu and Zhongqi Gao contributed equally to this work. Feishi Luan, Xuezheng Wang and Shi Liu were as the co-corresponding authors.

Data availability All data pertinent to the reported work have been provided in the manuscript or in the supplemental online materials.

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Ampomah-Dwamena C, Thrimawithana AH, Dejnoprat S, Lewis D, Espley RV, Allan AC (2019) A kiwifruit (*Actinidia deliciosa*) R2R3-MYB transcription factor modulates chlorophyll and carotenoid accumulation. New Phytol 221(1):309–325
2. Arce-Rodriguez ML, Martinez O, Ochoa-Alejo N (2021) Genome-wide identification and analysis of the MYB transcription factor gene family in chili pepper (*Capsicum* spp.). Int J Mol Sci 22(5):2229
3. Bai C, Capell T, Berman J, Medina V, Sandmann G, Christou P et al (2016) Bottlenecks in carotenoid biosynthesis and accumulation in rice endosperm are influenced by the precursor product balance. Plant Biotech J 14(1):195–205
4. Bang H, Davis AR, Kim S, Leskovar DI, King S (2010) Flesh color inheritance and gene interactions among canary yellow, pale yellow, and red watermelon. J Am Soc for Hortic Sci 135:362–368
5. Berman J, Zorrilla-López U, Medina V, Farre G, Sandmann G, Capell T et al (2017) The Arabidopsis *ORANGE* (*AtOR*) gene promotes carotenoid accumulation in transgenic corn hybrids derived from parental lines with limited carotenoid pools. Plant Cell Rep 36:933–945
6. Branham S, Vexler L, Meir A, Tzuri G, Frieman Z, Levi A et al (2017) Genetic mapping of a major codominant QTL associated with β-carotene accumulation in watermelon. Mol Breed 37:146

7. Bustin SA, Vandesompele J, Pfaffl MW (2009) Standardization of qPCR and RT-qPCR. Genet Eng Biotech N 29(14):40–43
8. Chayut N, Yuan H, Ohali S, Meir A, Sa'ar U, Tzuri G et al (2017) Distinct mechanisms of the ORANGE protein in controlling carotenoid flux. Plant Physiol 173:376–389
9. Dahan-Meir T, Filler-Hayut S, Melamed-Bessudo C, Bocobza S, Czosnek H, Aharoni A et al (2018) Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. Plant J 95(1):5–16
10. Fang XF, Liu S, Gao P, Liu HY, Wang XZ, Luan FS et al (2020) Expression of *ClPAP* and *ClPsy1* in watermelon correlates with chromoplast differentiation, carotenoid accumulation, and flesh color formation. Sci Hortic 270 (2020): 109437
11. Fantini E, Falcone G, Frusciante S, Giliberto L, Giuliano G et al (2013) Dissection of tomato lycopene biosynthesis through virus-induced gene silencing. Plant Physiol 163:986–998
12. Farre G, Perez-Fons L, Decourcelle M, Breitenbach J, Hem S, Zhu CF et al (2016) Metabolic engineering of astaxanthin biosynthesis in maize endosperm and characterization of a prototype high oil hybrid. Transgenic Res 25:477–489
13. Fu CC, Chen HJ, Gao HY, Lu Y, Han C, Han YC (2020) Two papaya MYB proteins function in fruit ripening by regulating some genes involved in cell-wall degradation and carotenoid biosynthesis. J Sci Food Agr 100(12):4442–4448
14. Fu XM, Feng C, Wang CY, Yin XR, Lu PJ, Grierson D et al (2014) Involvement of multiple phytoene synthase genes in tissue and cultivar specific accumulation of carotenoids in loquat. J Exp Bot 65:4679–4689
15. Fujii S, Kobayashi K, Nagata N, Masuda T, Wada H ((2018) Digalactosyldiacylglycerol is essential for organization of the membrane structure in etioplasts. Plant Physiol 177:1487–1497
16. Giuliano G (2017) Provitamin A biofortification of crop plants: a gold rush with many miners. Curr Opin Biotech 44:169–180
17. Guo SG, Sun HH, Zhang HY, Liu JA, Ren Y, Gong GY et al (2015) Comparative transcriptome analysis of cultivated and wild watermelon during fruit development. PLoS One 10(6):e0130267
18. Guo SG, Zhang JG, Sun HH, Salse J, Lucas WJ, Zhang HY et al (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nat Genet 45(1):51–58
19. Guo SG, Zhao SJ, Sun HH, Wang X, Wu S, Lin T et al (2019) Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. Nat Genet 51(11):1616–1623
20. Hashizume T, Shimamoto I, Hirai M (2003) Construction of a linkage map and QTL analysis of horticultural traits for watermelon [*Citrullus lanatus* (THUNB.) MATSUM & NAKAI] using RAPD, RFLP and ISSR markers. Theor Appl Genet 106:779–785
21. Henderson W, Scott G, Wehner T (1998) Interaction of flesh color genes in watermelon. J Hered 89:50–53

22. Iorizzo M, Ellison S, Senalik D, Zeng P, Satapoomin P, Huang JY et al (2016) A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat Genet 48(6):657–666
23. Jayaraj KL, Thulasidharan N, Antony A, John M, Augustine R, Chakravartty N et al (2021) Targeted editing of tomato carotenoid isomerase reveals the role of 5' UTR region in gene expression regulation. Plant Cell Rep 40(4):621–635
24. Jeong HB, Kang MY, Jung A, Han K, Lee JH, Jo J et al (2019) Single-molecule real-time sequencing reveals diverse allelic variations in carotenoid biosynthetic genes in pepper (*Capsicum spp*.). Plant Biotech J 17(6):1081–1793
25. Jin B, Lee J, Kweon S, Cho Y, Choi Y, Lee SJ et al (2019) Analysis of flesh color-related carotenoids and development of a *CRTISO* gene-based DNA marker for prolycopene accumulation in watermelon. Hortic Environ Biote 60(3):399–410
26. Langmead B, Schatz MC, Lin J, Pop M, Salzberg SL (2009) Searching for SNPs with cloud computing. Genome Biol 10(11):R134
27. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25
28. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760
29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
30. Li L, Paolillo DJ, Parthasarathy MV, DiMuzio EM, Garvin DF (2001) A novel gene mutation that confers abnormal patterns of β-carotene accumulation in cauliflower (*Brassica oleracea* var. *botrytis*). Plant J 26:59–67
31. Li L, Yang Y, Xu Q, Owsiany K, Welsch R, Chitchumroonchokchai C et al (2012) The *Or* gene enhances carotenoid accumulation and stability during post-harvest storage of potato tubers. Mol Plant 5:339–352
32. Li L, Ye J, Li HH, Shi QQ (2020) Characterization of metabolites and transcripts involved in flower pigmentation in *Primula vulgaris*. Front Plant Sci 11:572517
33. Li N, Shang J, Wang JM, Zhou D, Li NN, Ma SW et al (2019) Discovery of the genomic region and candidate genes of the scarlet red flesh color (*Yscr*) locus in watermelon (*Citrullus lanatus* L.). Front Plant Sci 11:116
34. Liu S, Gao P, Wang XZ, Davis AR, Baloch A, Luan FS (2015) Mapping of quantitative trait loci for lycopene content and fruit traits in *Citrullus lanatus*. Euphytica 202:411–426
35. Lu SW, Zhang Y, Zhu KJ, Yang W, Ye JL, Chai LJ et al (2018) The citrus transcription factor CsMADS6 modulates carotenoid metabolism by directly regulating carotenogenic genes. Plant Physiol 176:2657–2676
36. Lv P, Li N, Liu H, Gu HH, Zhao WE (2015) Changes in carotenoid profiles and in the expression pattern of the genes in carotenoid metabolisms during fruit development and ripening in four watermelon

cultivars. Food Chem 174:52–59

37. Maass D, Arango J, Wust F, Beyer P, Welsch R et al (2009) Carotenoid crystal formation in *Arabidopsis* and carrot roots caused by increased phytoene synthase protein levels. PLoS One 4(7):e6373
38. Martel C, Vrebalov J, Tafelmeyer P, Giovannoni JJ (2011) The tomato MADS box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-dependent manner. Plant Physiol 157:1568–1579
39. Mortimer CL, Misawa N, Ducreux L, Campbell R, Bramley PM, Taylor M et al (2016) Product stability and sequestration mechanisms in *Solanum tuberosum* engineered to biosynthesize high value ketocarotenoids. Plant Biotech J 14(1):140–152
40. Park S, Kim HS, Jung YJ, Kim SH, Ji CY, Wang Z et al (2016) Orange protein has a role in phytoene synthase stabilization in sweetpotato. Sci Rep-UK 6:33563
41. Paul JY, Khanna H, Kleidon J, Hoang P, Geijskes J, Daniells J et al (2017) Golden bananas in the field: elevated fruit pro-vitamin A from the expression of a single banana transgene. Plant Biotechnol J 15(4):520–532
42. Pei S, Liu Z, Wang XZ, Luan FS, Dai ZY, Yang ZZ et al (2021) Quantitative trait loci and candidate genes responsible for pale green flesh color in watermelon (*Citrullus lanantus*). Plant Breed 140(2):349–359
43. Peng G, Wang CY, Song S, Fu XM, Azam M, Grierson D et al (2013) The role of 1-deoxy-D-xylulose-5-phosphate synthase and phytoene synthase gene family in citrus carotenoid accumulation. Plant Physiol Biochem 71:67–76
44. Pons E, Aquezar B, Rodriguez A, Martorell P, Genoves S, Ramon D et al (2014) Metabolic engineering of beta-carotene in orange fruit increases its in vivo antioxidant properties. Plant Biotech J 12:17–27
45. Porter D (1937) Inheritance of certain fruit and seed characters in watermelons. Hilgardia 10:489–509
46. Rodriguez-Concepcion M, Avalos J, Bonet M (2018) A global perspective on carotenoids: Metabolism, biotechnology, and benefits for nutrition and health. Prog Lipid Res 70:62–93
47. Shumskaya M, Bradbury LMT, Monaco RR, Wurtzel ET (2012) Plastid localization of the key carotenoid enzyme phytoene synthase is altered by isozyme, allelic variation, and activity. Plant Cell 24:3725–3741
48. Sun TH. Yuan H, Cao HB, Yazdani M, Tadmor Y, Li L (2018) Carotenoid metabolism in plants: the role of plastids. Mol Plant 11:58–74
49. Tzuri G, Zhou XJ, Chayut N, Yuan H, Portnoy V, Meir A et al (2015) A 'golden' SNP in *CmOr* governs the fruit flesh color of melon (*Cucumis melo*). Plant J 82:267–279
50. Wang C, Zeng J, Li Y, Hu W, Chen L, Miao YJ et al (2014) Enrichment of provitamin A content in wheat (*Triticum aestivum* L.) by introduction of the bacterial carotenoid biosynthetic genes *CrtB* and *CrtI*. Journal Exp Bot 65(9):2545–2556

51. Wang CN, Luan FS, Liu HY, Davis AR, Zhang QA, Dai ZY et al (2021) Mapping and predicting a candidate gene for flesh color in watermelon. Journal Integr Agr DOI. 10.1016/S2095-3119(20)63487-6
52. Wang CN, Qiao AH, Fang XF, Sun L, Gao P, Angela AR et al (2019) Fine mapping of lycopene content and flesh color related gene and development of molecular marker-assisted selection for flesh color in watermelon (*Citrullus lanantus*). Front Plant Sci 10:1240
53. Wang N, Liu S, Gao P, Luan FS, Davis AR (2016) Developmental changes in gene expression drive accumulation of lycopene and beta-carotene in watermelon. J Am Soc Horti Sci 141:1–10
54. Watkins JL, Pogson BJ (2020) Prospects for carotenoid biofortification targeting retention and catabolism. Trends in Plant Sci 25(5):501–512
55. Wehner T (2007) Gene list for watermelon. Cucurbit Genet Coop Rep 30:9–120
56. Welsch R, Arango J, Bar C, Salazar B, Al-Babili S, Beltran J et al (2010) Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell 22:3348–3356
57. Welsch R, Zhou XJ, Yuan H, Alvarez D, Sun T, Schlossarek D et al (2018) Clp protease and OR directly control the proteostasis of phytoene synthase, the crucial enzyme for carotenoid biosynthesis in *Arabidopsis*. Mol Plant 11:149–162
58. Wu MB, Xu X, Hu XW, Liu YD, Cao HH, Chan HE et al (2020) SlMYB72 regulates the metabolism of chlorophylls, carotenoids, and flavonoids in tomato fruit. New Phytol 183(3):854–868
59. Xiong C, Luo D, Lin AH, Zhang CL, Shan LB, He P et al (2019) A tomato B-box protein SlBBX20 modulates carotenoid biosynthesis by directly activating *PHYTOENE SYNTHASE 1*, and is targeted for 26S proteasome-mediated degradation. 221(1): 279–294, New Phytol
60. Xu XC, Fang PP, Zhang H, Chi C, Song LX, Xia XJ et al (2019) Strigolactones positively regulate defense against root-knot nematodes in tomato. J Exp Bot 70(4):1325–1337
61. Yuan H, Zhang JX, Nageswaran D, Li L (2015) Carotenoid metabolism and regulation in horticultural crops. Hortic Res 2:15036
62. Zeng J, Wang XT, Miao YJ, Wang C, Zang ML, Chen X et al (2015) Metabolic engineering of wheat provitamin A by simultaneously overexpressing *CrtB* and silencing carotenoid hydroxylase (*TaHyd*). J Agr Food Chem 63:9083–9092
63. Zhai SN, Li GY, Sun YW, Song JM, Li JH, Song GQ et al (2016) Genetic analysis of phytoene synthase 1 (*Psy1*) gene function and regulation in common wheat. BMC Plant Biol 16:228
64. Zhang J, Guo SG, Ren Y, Zhang HY, Gong GY, Zhou M et al (2017) High-level expression of a novel chromoplast phosphate transporter *ClPHT4;2* is required for flesh color development in watermelon. New Phytol 213:1208–1221
65. Zhang J, Sun HH, Guo SG, Ren Y, Li MY, Wang JF et al (2020) Decreased protein abundance of lycopene β-cyclase contributes to red flesh in domesticated watermelon. Plant Physiol 183(3):1171–1183

66. Zhong S, Joung JG, Zheng Y, Chen YR, Liu B, Shao Y et al (2011) High-throughput illumine strand-specific RNA sequencing library preparation. Cold spring harbor protocols, (8), pdb-prot5652
67. Zhou XJ, Welsch R, Yang Y, Alvarez D, Riediger M, Yuan H et al (2015) Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. P Natl Acad Sci USA112:3558–3563
68. Zhu F, Luo T, Liu CY, Wang Y, Yang HB, Yang W et al (2017) An R2R3-MYB transcription factor represses the transformation of alpha- and beta-branch carotenoids by negatively regulating expression of *CrBCH2* and *CrNCED5* in flavedo of *Citrus reticulate*. New Phytol 216(1):178–192

Tables

Table 1. Segregation of golden flesh color in six generations from COS × PI 192938 cross over 2 years

Populations	Plants	Orange	Non-orange	Expected segregation ratio	P value of Chisquare tests
PI 192938	15	15	0	N/A	N/A
COS	15	0	15	N/A	N/A
F_1	15	0	15	N/A	N/A
2018 F_2	84	17	67	1:3	0.313
2019 F_2	279	74	205	1:3	0.557
2019 BC_1P_1	37	0	37	0:1	0.513
2019 BC_1P_2	84	39	45	1:1	0.563

N/A: Not Applicable

Table 2 and 3 are not available with this version.

Figures

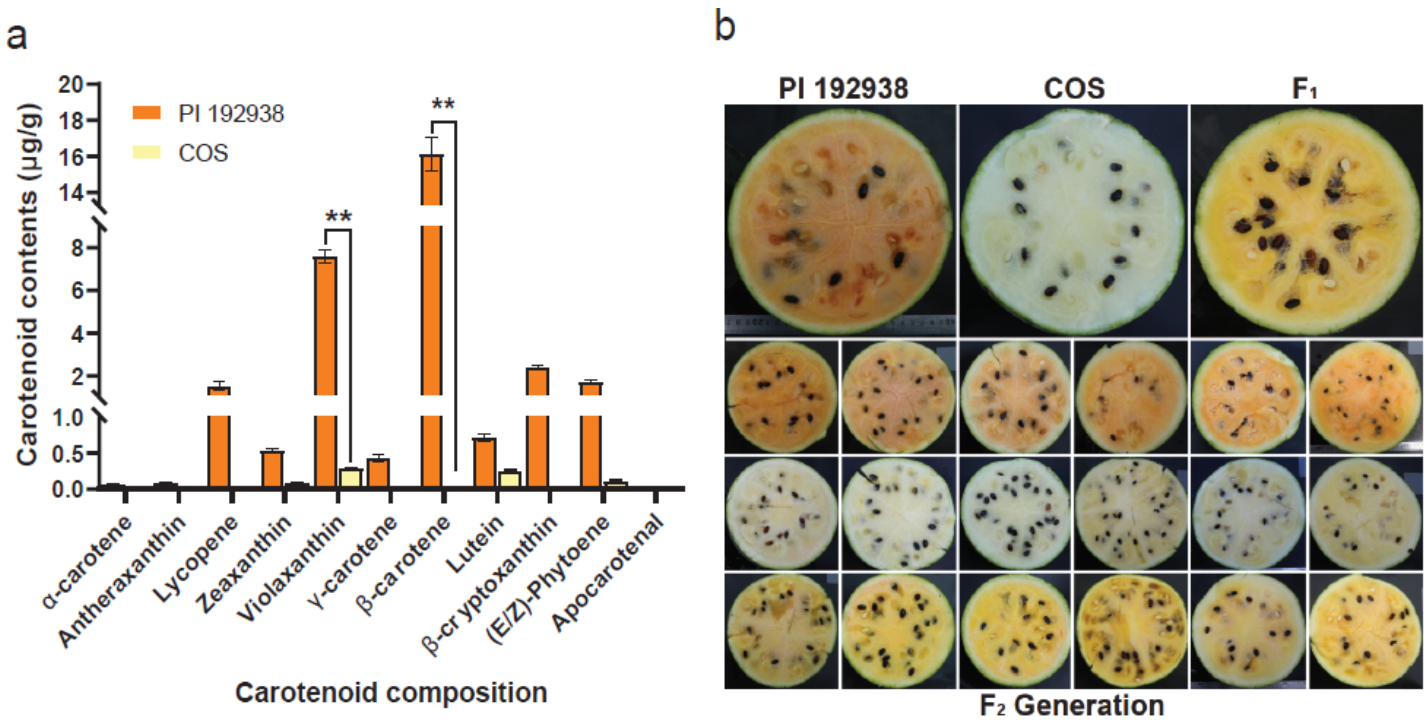


Figure 1

Flesh color segregated in the genetic populations and carotenoid composition of the parental lines. a. Carotenoid composition of COS and PI 192938. The carotenoid data were obtained from our previous publication (Fang et al., 2020). b. From left to right, flesh color of mature fruits of COS, PI 192938 and their F1 generation. c to e. Three flesh colors in the F2 generation derived from COS and PI 192938.

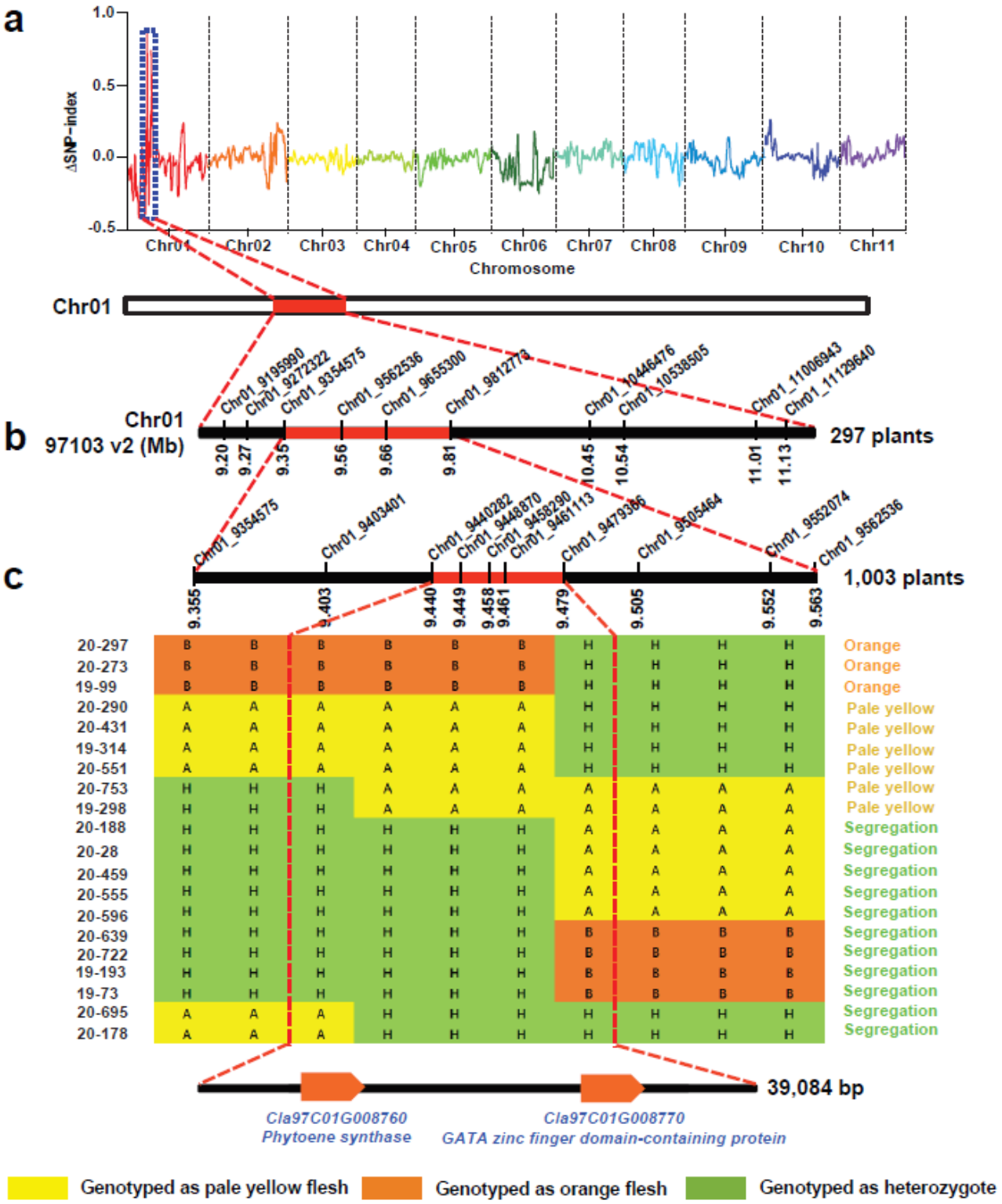


Figure 2

BSA-seq analysis and fine mapping of the golden flesh color trait. a. BSA-seq results and chromosome region related to the golden flesh color trait. b. Initial mapping of the golden flesh color trait. c. Fine mapping of the golden flesh color trait and candidate genes in the fine-mapping region.

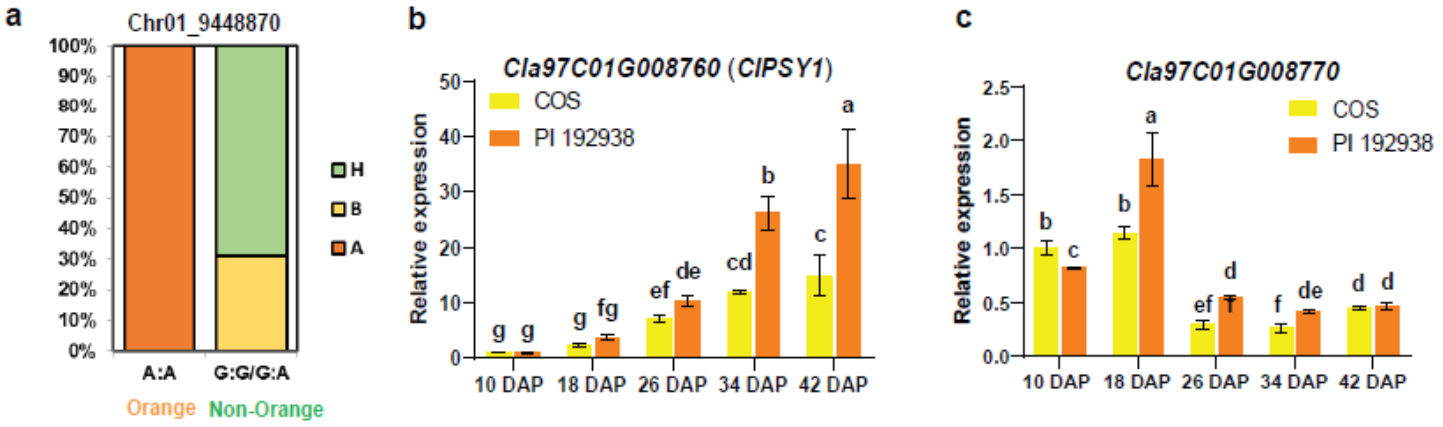
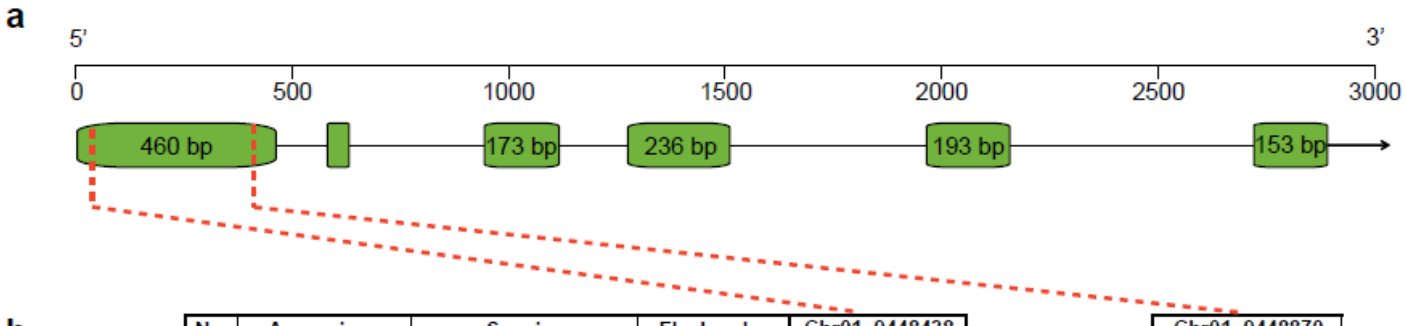


Figure 3

Genotyping of the individuals in the F2 generation and expression patterns of two candidate genes in COS and PI 192938. Genotyping results of the SNP at the 9,448,870th bp in the F2 generation. Gene expression patterns of the two parental materials at different flesh color formation stages.



b

No.	Accessions	Species	Flesh color	Chr01_9448438		Chr01_9448870	
1	PI 192938	*C. lanatus*	Orange	C	Pro	G	Lys
2	Huayuan Muben		Red	C		G	
3	Yinyu Jingxin		Red	C		G	
4	JX-2		Red	C		G	
5	97103		Red	C		A	Glu
6	LSW-177		Red	C		A	
7	JXF		Red	C		A	
8	RZ-901		Red	C		A	
9	XHBFGM		Red	C		A	
10	Black Diamond		Red	C		A	
11	Calhoun Gray		Red	C		A	
12	Sugarlee		Red	C		A	
13	Sy-904304		Red	C		A	
14	RZ-900		Red	C		A	
15	Yellow Crimson		Orange	C		A	
16	PI 635597		Canary yellow	C		A	
17	JLM		Canary yellow	C		A	
18	COS		Pale yellwo	C		A	
19	PI 249010	*C. mucosospermus*	White	C		A	
20	PI 248178		White	C		A	
21	PI 595203		White	C		A	
22	PI 186490		White	C		A	
23	PI 482326	*C. amarus*	White	T	Ser	A	
24	PI 482303		Yellow green	T		A	
25	PI 296341-FR		Yellow green	T		A	
26	PI 482276		Yellow green	T		A	

Figure 4

Nucleotide variation in the ClPsy1 gene structure among natural watermelon accessions. a. Gene structure of ClPsy1. b. Nucleotide variations and amino acid substitutions in the ClPsy1 gene among 26 watermelon accessions.

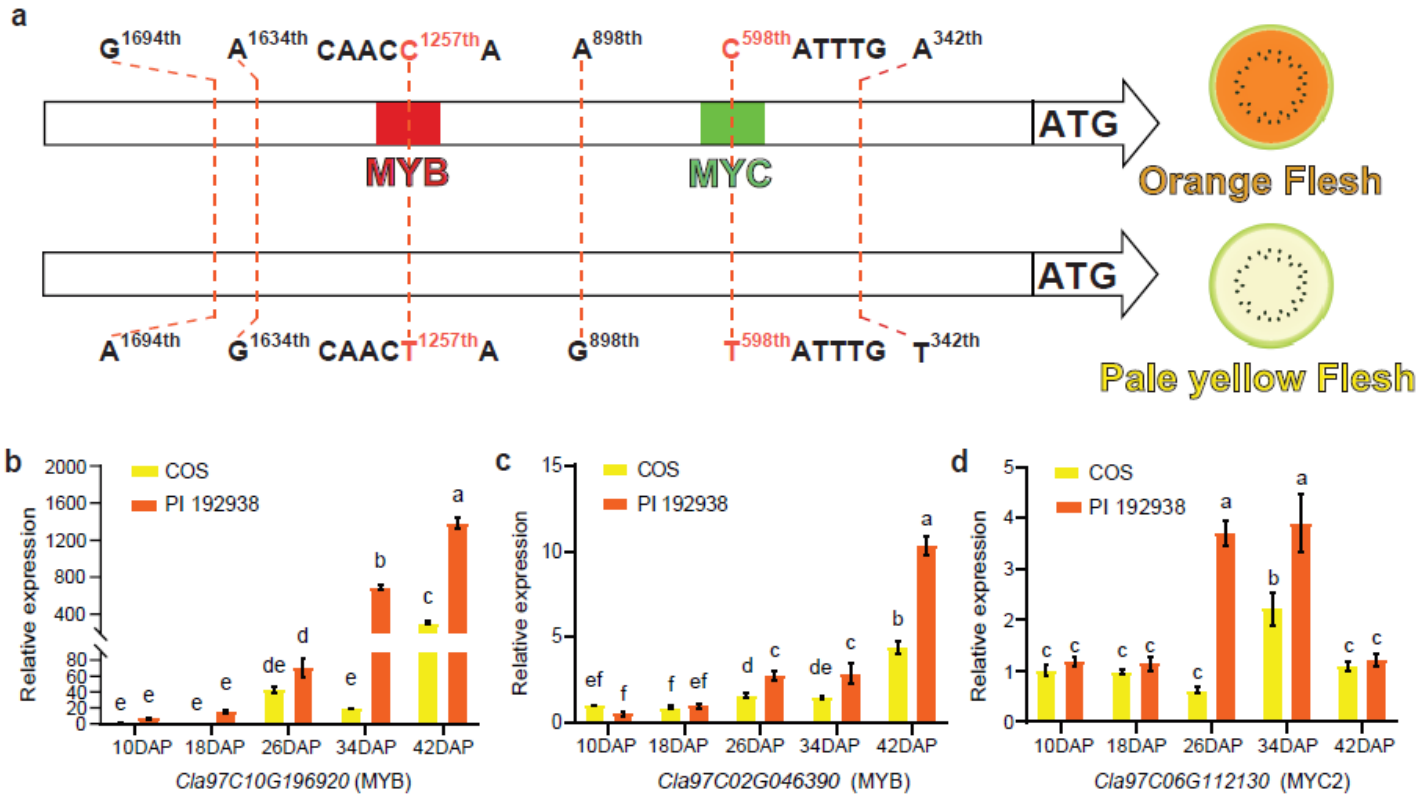


Figure 5

SNP variations in the promoter region and expression of candidate transcription factors. a. SNPs located in the promoter regions of COS and PI 192938. b to d. Gene expression patterns of the three candidate transcription factors in flesh tissues at different flesh color formation stages.

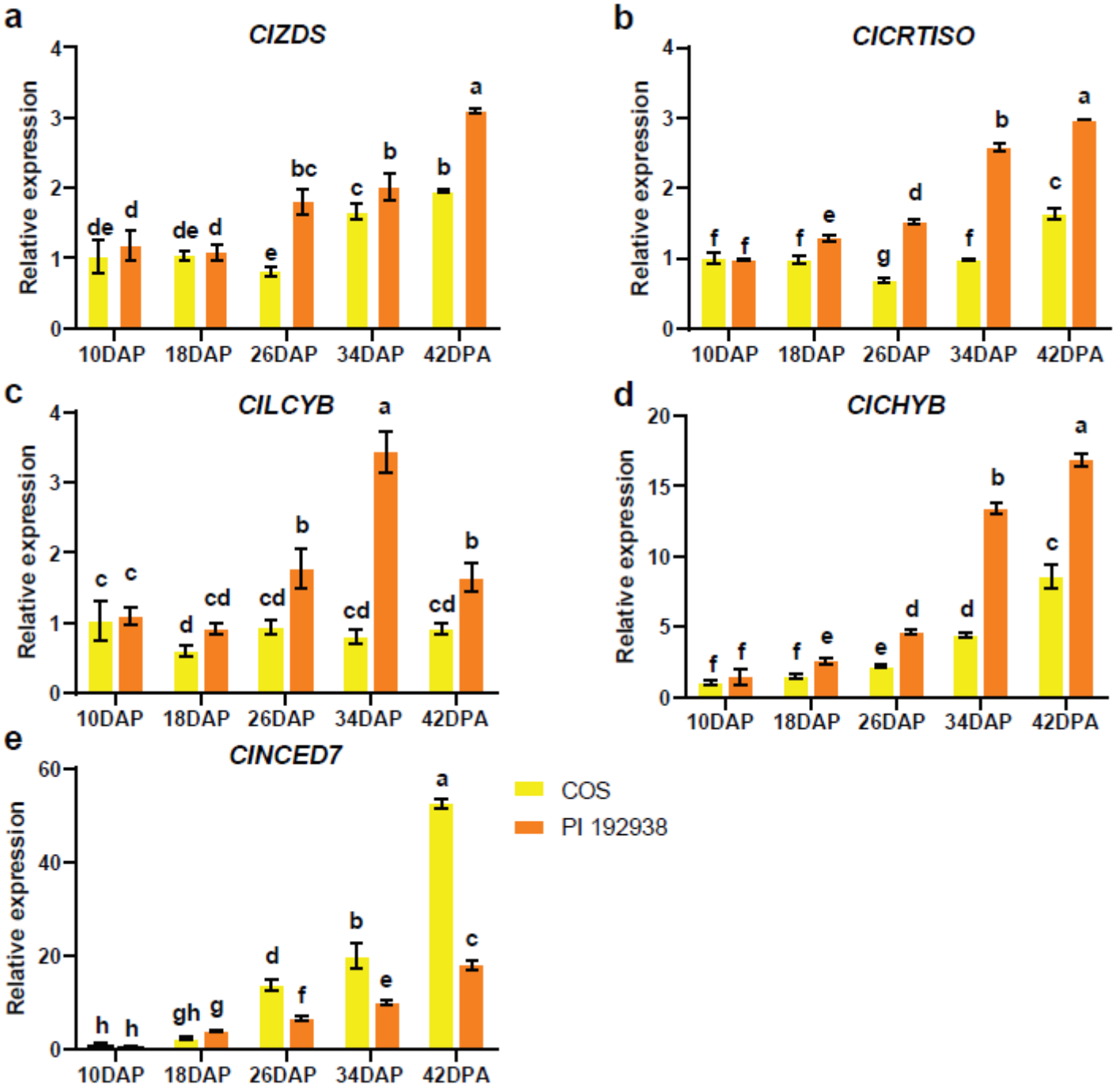


Figure 6

Differentially expressed genes involved in the carotenoid metabolic pathway.

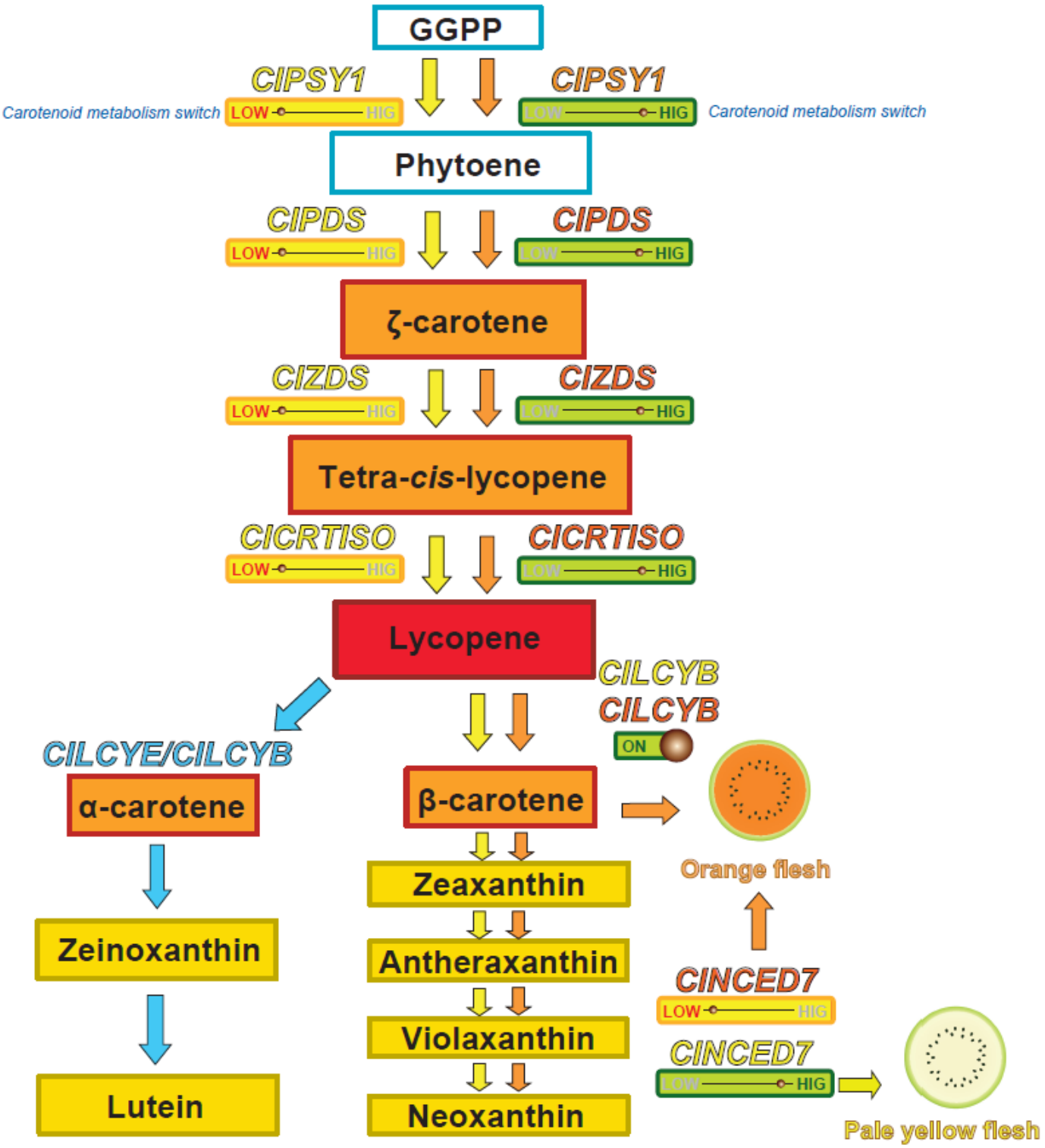


Figure 7

Inferred golden flesh trait formation based on gene expression and gene structure variations in genes involved in the carotenoid metabolic pathway. The yellow and orange arrows indicate pale yellow flesh color and gf color formation, respectively. The blue arrows indicate that genes in the carotenoid metabolic pathway located in the α-band did not show any gene expression difference between the two parental lines.

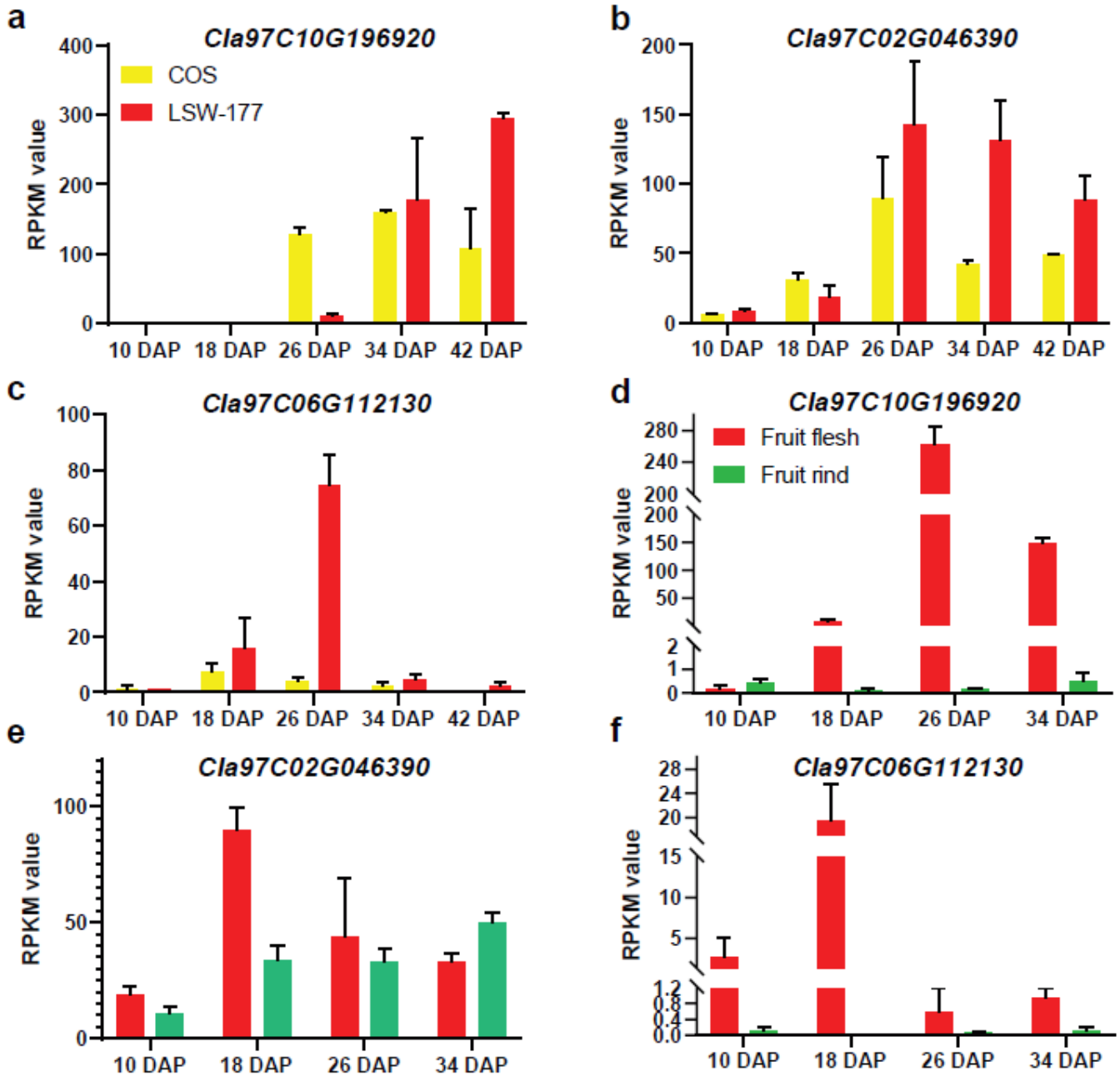


Figure 8

Gene expression patterns of the three candidate transcription factors based on published RNA-seq data. a to c, BioProject number PRJNA338036. d to f, BioProject number SRP012849.

Supplementary Files

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

- Tables.xlsx