



Chromosome-Scale Genome and Comparative Transcriptomic Analysis Reveal Transcriptional Regulators of β -Carotene Biosynthesis in Mango

Xiaowei Ma^{1†}, Xiang Luo^{2†}, Yongzan Wei¹, Tuanhui Bai³, Jiangli Shi³, Bing Zheng¹, Wentian Xu¹, Li Li¹, Songbiao Wang^{1*}, Jisen Zhang^{4*} and Hongxia Wu^{1*}

OPEN ACCESS

Edited by:

Jian Li Yang,
Zhejiang University, China

Reviewed by:

Tao Wu,
Hunan Agricultural University, China
Carmen A. Contreras-Vergara,
Consejo Nacional de Ciencia y
Tecnología (CONACYT), Mexico
Zhongshan Gao,
Zhejiang University, China

*Correspondence:

Songbiao Wang
songbiaowang501@163.com
Jisen Zhang
zjsen@126.com
Hongxia Wu
whx1106@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Biotechnology,
a section of the journal
Frontiers in Plant Science

Received: 29 July 2021

Accepted: 07 September 2021

Published: 12 October 2021

Citation:

Ma X, Luo X, Wei Y, Bai T, Shi J,
Zheng B, Xu W, Li L, Wang S, Zhang J
and Wu H (2021) Chromosome-Scale
Genome and Comparative
Transcriptomic Analysis Reveal
Transcriptional Regulators of
 β -Carotene Biosynthesis in Mango.
Front. Plant Sci. 12:749108.
doi: 10.3389/fpls.2021.749108

¹ Key Laboratory for Postharvest Physiology and Technology of Tropical Horticultural Products of Hainan Province, South Subtropical Crops Research Institute, Chinese Academy of Tropical Agricultural Sciences, Zhanjiang, China, ² State Key Laboratory of Crop Stress Adaptation and Improvement, Henan University, Kaifeng, China, ³ College of Horticulture, Henan Agricultural University, Zhengzhou, China, ⁴ Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, China

Mango ($2n = 2x = 40$) is an important tropical/subtropical evergreen fruit tree grown worldwide and yields nutritionally rich and high-value fruits. Here, a high-quality mango genome (396 Mb, contig N50 = 1.03 Mb) was assembled using the cultivar “Irwin” from Florida, USA. A total of 97.19% of the sequences were anchored to 20 chromosomes, including 36,756 protein-coding genes. We compared the β -carotene content, in two different cultivars (“Irwin” and “Baixiangya”) and growth periods. The variation in β -carotene content mainly affected fruit flesh color. Additionally, transcriptome analysis identified genes related to β -carotene biosynthesis. *MiPSY1* was proved to be a key gene regulating β -carotene biosynthesis. Weighted gene co-expression network analysis, dual luciferase, and yeast one-hybrid assays confirmed that transcription factors (TFs) MibZIP66 and MibHLH45 activate *MiPSY1* transcription by directly binding to the CACGTG motif of the *MiPSY1* promoter. However, the two TFs showed no significant synergistic effect on promoter activity. The results of the current study provide a genomic platform for studying the molecular basis of the flesh color of mango fruit.

Keywords: mango, genome, β -Carotene, transcriptional regulation 2, fruit ripening

INTRODUCTION

Mango (*Mangifera indica* L.) fruit is one of the most nutritious and highly-priced edible fruit worldwide. It possesses a diploid genome ($2n = 2x = 40$ chromosomes) and belongs to the Sapindales order, Anacardiaceae family, and *Mangifera* genus. Mango is grown on 2.5 million hectares in tropical and subtropical areas of the world, with an annual production of ~46.6 million (Wang et al., 2020). It is the fifth most economically important fruit crop, followed by bananas, grapes, apples, and oranges (Deshpande et al., 2017). Its yellow flesh is rich in carotenoids, an important natural pigment involved in plant growth and adaptation that the immune functions of humans and reduces the occurrence of diseases such as cancer (Sun et al., 2017). Thus, understanding the molecular basis of flesh color is the main goal for the genetic improvement of mangoes. The plastidial 2-C-methyl-D-erythritol 4-phosphate pathway,

which is responsible for carotenoid biosynthesis, is highly conserved in plants (Liu et al., 2015). However, the regulatory mechanisms governing the expression of genes of this pathway are complex and vary among plant species. To date, only a few plant transcription factors (TFs), such as phytochrome-interacting factor 1 (PIF1) in *Arabidopsis* (Toledo-Ortiz et al., 2010), SISGR1 (encoding the protein STAY GREEN) in tomato (Luo et al., 2013), and *CsMADS6* (Mcm1-Agamous-Deficiens-Srf) and *CsMADS5* in citrus (Lu et al., 2018), have been shown to directly regulate the expression of carotenoid biosynthesis genes. However, some TFs, including *SINAC4*, *TAGL1*, *ERF6*, *GLK2*, and *CubHLH1*, affect carotenoid metabolism through other pathways, such as fruit development, ethylene biosynthesis, and photomorphogenesis (Chung et al., 2010; Karlova et al., 2011). It is still unclear whether such TFs directly mediate carotenoid metabolism (Sagawa et al., 2015; Su et al., 2015). Previous studies have shown that although carotenoid composition is relatively stable in different mango cultivars, carotenoid content varies greatly among cultivars (Mercadante and Rodriguez-Amaya, 1998; Ma et al., 2018). β -carotene is the predominant carotenoid that affects the flesh color of mature mango fruit (Ma et al., 2018). Nevertheless, little is known about the transcriptional mechanisms of that regulated β -carotene biosynthesis and hence influence the flesh color of mango fruit.

Owing to the long history of cultivation, wide geographical distribution, and intense selection, there are more than a thousand mango cultivars worldwide. Genetic analysis of mango germplasm diversity has shown that the whole mango group can be divided into two subgroups: Indian and Southeast Asia groups (Warschefsky and Wettberg, 2019). Recently, the genomes of Indian cultivars (variety “Alphonso”) (Wang et al., 2020) and Southeast Asian cultivars (variety “Hong Xiang Ya”) (Li et al., 2020) have been sequenced, expanding our understanding of mango genome evolution. The Florida mango cultivars are generally considered hybrids between Indian cultivars and Southeast Asian cultivars. The Florida variety “Irwin” has excellent production stability and environmental adaptability (Olano et al., 2005). Additionally, “Irwin” mangoes have higher economic value compared to “Alphonso” and “Hong Xiang Ya” owing to their red peel and low fiber-containing flesh. Thus, “Irwin” is considered an elite maternal parent widely used for mango breeding in recent decades and is genetically much more diverse than other cultivars. However, a high-quality reference genome of Florida cultivar mangoes had not been published until now.

In this study, we sequenced and analyzed the genome of the mango cultivar “Irwin” using single-molecule real-time (SMRT) sequencing and high-throughput chromosome conformation capture (Hi-C) techniques. We subsequently performed genome and comprehensive transcriptome analyses of genes involved in carotenoid biosynthesis to explore the molecular mechanism of β -carotene biosynthesis and its regulation during fruit development and ripening. Additionally, we identified important TFs controlling the expression of a key structural gene involved in the β -carotene synthesis. Accordingly, we described a high-quality cultivar genome and elucidated the genetic basis for the flesh color of mango fruit.

MATERIALS AND METHODS

Materials

Two *M. indica* cultivars “Irwin” (IR, deep yellow-fleshed) and “Baixiangya” (BY, pale yellow-fleshed), widely cultivated in China, were chosen from the orchard of the South Subtropical Crops Research Institute, Zhanjiang (110°4'E, 21°12'N), Guangdong Province, China, for this study.

Genome Sequencing

Genomic DNA was extracted from the fresh young leaves of mango cultivar “Irwin” according to the modified CTAB method, and used for both Illumina and PacBio sequencing. DNA quantity and quality were assessed using a NanoDrop 2000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and electrophoresis on a 0.8% agarose gel. For Illumina sequencing, we constructed a short-read (300 bp) library using the NEBNext Ultra DNA Library Prep Kit (New England BioLabs, Ipswich, MA, USA). For PacBio sequencing, 20-kb SMRTbell libraries were prepared according to the protocol released by PacBio, and then the library was sequenced on the PacBio Sequel System (Biomarker Technologies Corp., Beijing, China) with one SMRT cell, yielding more than 20 Gb of raw data.

K-Mer Analysis and Genome Assembly

A total of 18.87 Gb high-quality-filtered data were used to perform a 17-kmer analysis. The genome size, heterozygous rate, and repeat rate were estimated using the program Jellyfish (Marçais and Kingsford, 2011). MECAT (version 1.2) (Xiao et al., 2017) was used for the correction and assembly of PacBio data, and Polish (version 1.22) (Walker et al., 2014) was used for the assembly correction with Illumina short reads. To check the completeness of the assembly, the RNA-seq data from fruits were mapped to the genome assembly using BLASTN (version 2.6.0+) (Altschul et al., 1990); various sequence homology and coverage parameters were determined. Furthermore, we evaluated genome completeness using the Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis (Simão et al., 2015). The Hi-C sequencing of the chromosome-level genome assembly was performed according to the previously described method (Belton et al., 2012).

Genome Annotation

The present genome was annotated with three characteristics. (i) We first used *de novo* prediction and homology-based alignment to predict repetitive sequences across the mango genome. The repeat sequence database was constructed using the LTR_FINDER (Xu and Wang, 2007), PILER-DF (Edgar and Myers, 2005), and RepeatScout (Price et al., 2005). RepeatMasker and RepeatProteinMask software (Tamura, 1992) were used to identify repeat sequences. (ii) We then annotated the predicted protein-coding genes included in the databases, namely, National Center for Biotechnology Information non-redundant database, Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, SwissProt, and EuKaryotic Orthologous Groups. (iii) Finally, we predicted miRNA and snRNA sequences using the INFERNAL software (Nawrocki and Eddy, 2013). The rRNA was identified

using the BLAST program, and tRNAscan-SE was used to predict tRNA (Chan and Lower, 2019).

RNA Sequencing

Healthy fruits of two cultivars (“Irwin” and “Baixiangya”) in three distinct developmental stages were sampled. The developmental stages were: S1, early development stage, 40 days after full boom (DAFB); S2, harvest stage, 86 DAFB; and S3, fully ripe stage, 96 DAFB. The flesh of the collected fruit was separated from the peel and frozen in liquid nitrogen for β -carotene measurement and RNA sequencing. All experiments were performed in triplicate.

Total RNA was extracted from the frozen flesh tissue of each sampling point and each cultivar using a Quick RNA Isolation Kit (Invitrogen, Carlsbad, CA, USA). The cDNA libraries were assembled according to the protocol of the manufacturer of the NEBNext Ultra RNA Library Prep Kit for Illumina (E7530) and NEBNext Multiplex Oligos for Illumina (E7500), and the libraries were sequenced on the Illumina HiSeq XTen platform (Biomarker Technologies Corp., Beijing, China). The DESeq2 software (Robinson et al., 2010) was employed to identify differentially expressed genes, which were filtered according to $p \leq 0.05$, and expression value >0 .

Analysis of β -Carotene Content

The β -carotene in the flesh was extracted as described by Ma et al. (2018). The carotenoids were analyzed using the Waters Ultra-high-performance liquid chromatography (Milford, MA, USA) with a photodiode array detector and a single quadrupole mass spectrometer detector in series (6120 Quadrupole, Agilent, Santa Clara, CA, USA). The column used was a YMC carotenoid (C30) column (Wilmington, NC, USA). The eluent phases were as follows: mobile phase A, acetonitrile:methanol = 3:1 (v/v); and phase B, 100% methyl tert-butyl ether (MTBE). Each eluent contained 0.01% butylated hydroxytoluene (BHT). Gradient elution was performed as follows: 0–2 min, 85:15 A:B; 2–4 min, 75:25 A:B; 4–7 min, 40:60 A:B; 7–10 min, 40:60 A:B; 10–13 min, 5:95 A:B; 13–23 min, 85:15 A:B. The flow rate was 0.8 ml/min, and the injection volume was 5 μ l.

RNA Extraction and Quantitative Reverse Transcription PCR Analysis

Total RNA was extracted from frozen flesh using TRIzol reagent (TaKaRa, Dalian, China) according to the instructions of the manufacturer. The primers used for real-time PCR were designed using Primer 5 and are listed in **Supplementary Table 10**. The PCR mixture (20 μ l total volume) comprised 10 μ l of LightCycler 480 SYBR Green I Master Mix (Roche, Mannheim, Germany), 2 μ l of each primer, 2 μ l of diluted cDNA, and 6 μ l DEPC-H₂O. Reverse transcription (RT)-PCR conditions were as follows: pre-incubation at 95°C for 5 min, followed by 45 cycles at 95°C for 5 s, 58°C for 15 s, and 72°C for 10 s. The expression levels of genes of interest were calculated according to the formula $2^{-\Delta\Delta Ct}$.

Yeast One-Hybrid Assay

Yeast one-hybrid (Y1H) assay was performed according to the Gold Yeast One-Hybrid Library Screening System User Manual

(Clontech, Palo Alto, CA, USA). Either the *MiPSY1* or *MiPSY1*-mut promoter was inserted into the Phis2 vector to generate the baits. *MiPSY1*-mut is a mutated version of *MiPSY1* in which the core sequence (CACGTG) of two E-box elements has been altered and is “TCTAGC.” The full-length coding sequences (CDS) of *MibZIP66* and *MibHLH45* were cloned into the pGADT7 vector to generate the prey and were transformed into the yeast strain Y187 containing the bait. The TF-promoter interaction was determined according to the growth ability of the co-transformed yeast cells on synthetic defined -Trp-Leu-His media with 3-amino-1,2,4-triazole.

Dual-Luciferase Transient Expression Assay

The full-length CDS of *MibZIP66* and *MibHLH45* was cloned into the vector pGreenII0029 62-SK as effectors. Meanwhile, the *MiPSY1* promoter was recombined with pGreenII0800-LUC to create the reporter vector. All vectors were transformed into *Agrobacterium tumefaciens* (GV3101), and the cultures were adjusted to an optical density at OD 600 nm of 0.1 with an infiltration medium containing 100 μ M acetosyringone, 0.5 M MES, and 10 mM MgCl₂. *Agrobacterium* with TFs or promoters was injected into *Nicotiana benthamiana* leaves using a 5-ml injection syringe. Transgenic plants were grown in a glasshouse condition for 3 days. Firefly luciferase (LUC) and Renilla luciferase (REN) were determined using the Dual-Luciferase Reporter Assay System (Promega Corp., Madison, WI, USA). Transcriptional activation was calculated as the ratio of LUC to REN.

Subcellular Localization Analysis

The full-length coding sequences of *MibZIP66* and *MibHLH45* without the stop codon were amplified using PCR and cloned into the GFP vector, under the control of the CaMV 35S promoter. The vector constructs of 35S-*MibZIP66*-GFP and 35S-*MibHLH45*-GFP were transformed into *Agrobacterium tumefaciens* strain GV3101 through electroporation and then injected into tobacco leaves. *MiPSY1* or the control vector (pYBA1132) plasmid was co-transformed with a plasmid coding for a mitochondrial marker red fluorescent protein (RFP) into *Arabidopsis* protoplasts by polyethylene glycol (PEG) transformation method according to Lister et al. (2007). Localizations of *MiPSY1*, *MibZIP66*, and *MibHLH45* were examined using a fluorescence microscope (Zeiss Axioskop 2 Plus; Oberkochen, Germany).

RESULTS

Genome Sequencing and Assembly

The genome of the mango cultivar “Irwin” was initially estimated to be 327 Mb with a 1.37% heterozygosity rate and 47.28% of the repeat sequence ratio based on the 17-mer sequence (**Supplementary Figure 1** and **Supplementary Table 2**). In total, 20.26 Gb of sequencing data were produced with an average coverage depth of 61.96 \times using the PacBio single molecular long-read method (**Supplementary Table 1**). After heterozygosity was accounted for, a draft genome of 396 Mb was assembled,

with N50 of 1.03 Mb (Table 1). The assembly of our genome was comparable to that of the recently published genome of the traditional Indian cultivar “Alphonso” (392.9 Mb) (Wang et al., 2020), but better than that of the Southeast Asian variety “Hong Xiang Ya” (~372 Mb) (Li et al., 2020). Then, Hi-C contact frequency derived from Hi-C sequencing was used to order and orient the polished contigs into Hi-C scaffolds. As a result, a total of 97.19% (375 Mb) sequences in the mango genome were anchored to 20 chromosomes, and the chromosome size ranged from 12.8 to 29.8 Mb (Supplementary Tables 3, 4, and Figure 1A). These results are consistent with those of studies by Li et al. (2020) and Wang et al. (2020). Two independent methods were used to evaluate assembly quality. The alignment rates of the RNA-seq reads from the three fruit developmental stages of the two cultivars were ~90.29 and 95.87% (Supplementary Table 5). In contrast, the BUSCO assay showed that ~94% of the core eukaryotic genes (1,991) were retrieved in the assembly, further confirming the continuity and high quality of the assembled genome (Supplementary Table 6). Furthermore, we performed the whole-genome alignment of our assembly and previous mango assembly. One-to-one syntenic blocks showed ~79.61% of our genome sequence matched 84.75% of the “Hong Xiang Ya” genome sequence (Figure 1B), indicating our genome sequence was substantially more complete.

Repetitive Elements and Gene Annotation

We identified a total of 138.29 Mb of repeated sequences, representing 34.92% of the assembled mango genome. Among these repetitive elements, 12.77% were unknown repetitive sequences. The most abundant repeat type was long terminal repeats (LTRs), accounting for 15.49% of the genome. Additionally, we also found long interspersed nuclear elements (LINEs, 0.57%) and DNA transposons (6.25%) (Supplementary Table 7). Combined *de novo* prediction,

protein-based homology searches, and RNA-seq analysis identified a total of 36,756 protein-coding genes with an average coding sequence length of 1,142 bp and 5.5 exons per gene (Supplementary Table 8). Moreover, a total of 109 rRNAs, 75 snRNAs, 483 tRNAs, 279 snoRNA, and 131 miRNAs were identified.

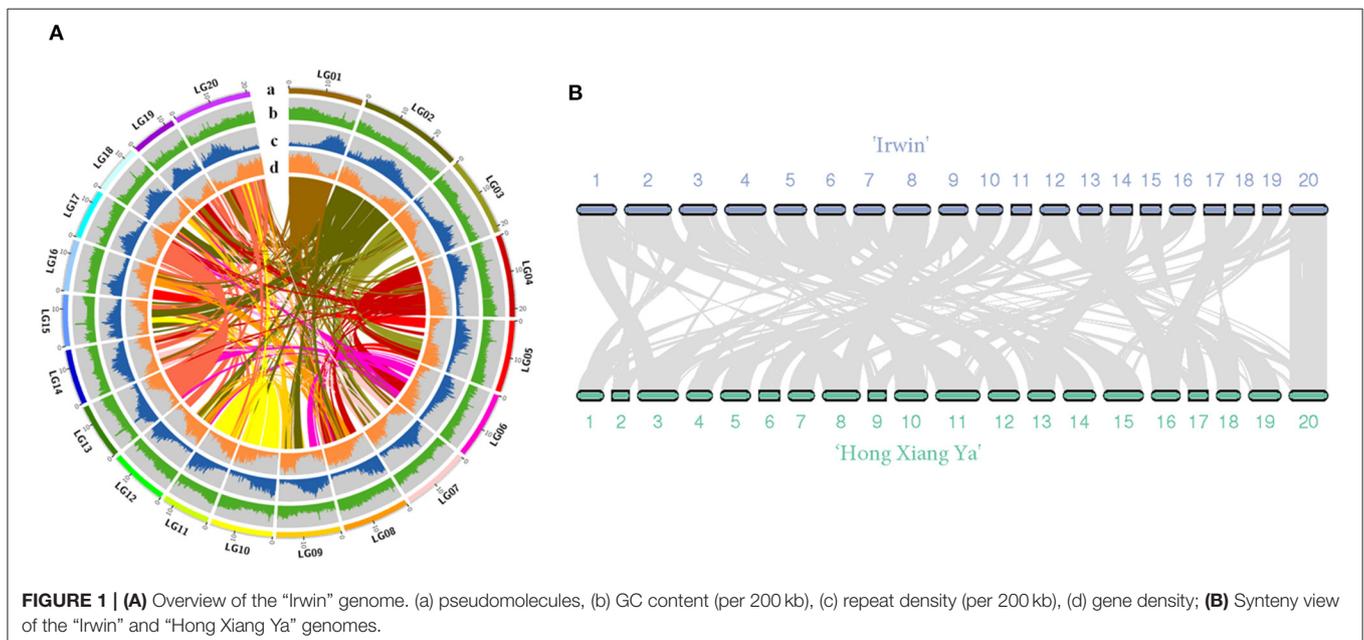
Expression of Carotenoid Genes During Fruit Development and Ripening

Studying the changes in β -carotene and the regulation of its biosynthesis will provide insight into the genetic basis of mango fruit flesh color. During fruit development and ripening, β -carotene content increased in “Irwin” mangoes, but remained low in “Baixiangya” mangoes. At the fully ripe stage, β -carotene concentrations in “Irwin” flesh was up to 50.06 $\mu\text{g/g}$ of fresh weight, but was only 2.64 μg of fresh weight in “Baixiangya” flesh (Figure 2B). This difference in β -carotene concentration was also evident visually by the distinct color of the fruits (Figure 2A).

Comparative RNA-seq analysis was performed between “Irwin” and “Baixiangya” mangoes to study the expression profiles of β -carotene biosynthesis-related genes. A total of

TABLE 1 | Statistics of mango genome assembly.

	With heterozygosis	Without heterozygosis
Number of contigs	3,022	1,305
Length of contigs (bp)	471,122,396	396,192,158
Minimum length of contig (bp)	11,830	12,183
Maximum length of contig (bp)	13,145,087	13,145,087
Average length of contig (bp)	155,897	303,595
N50 (bp)	788,704	1,034,750



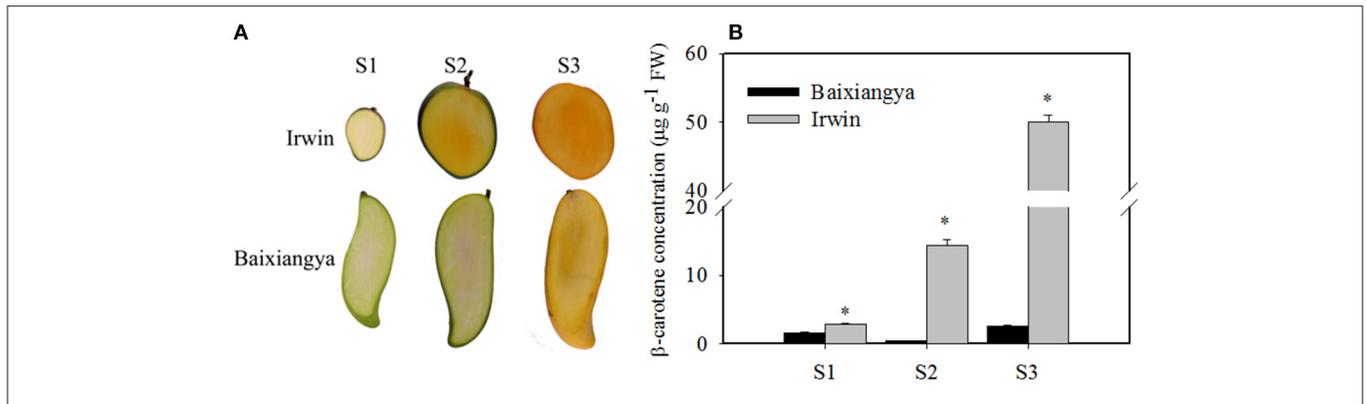


FIGURE 2 | Changes in the color (A) and β -carotene concentration (B) in the flesh of “Baixiangya” and “Irwin” during fruit development and ripening. Fruit development and ripening were divided into three stages: S1, early development stage; S2, harvest stage; and S3, fully ripe stage. A black asterisk indicates a significant change ($p \leq 0.01$) between two cultivars.

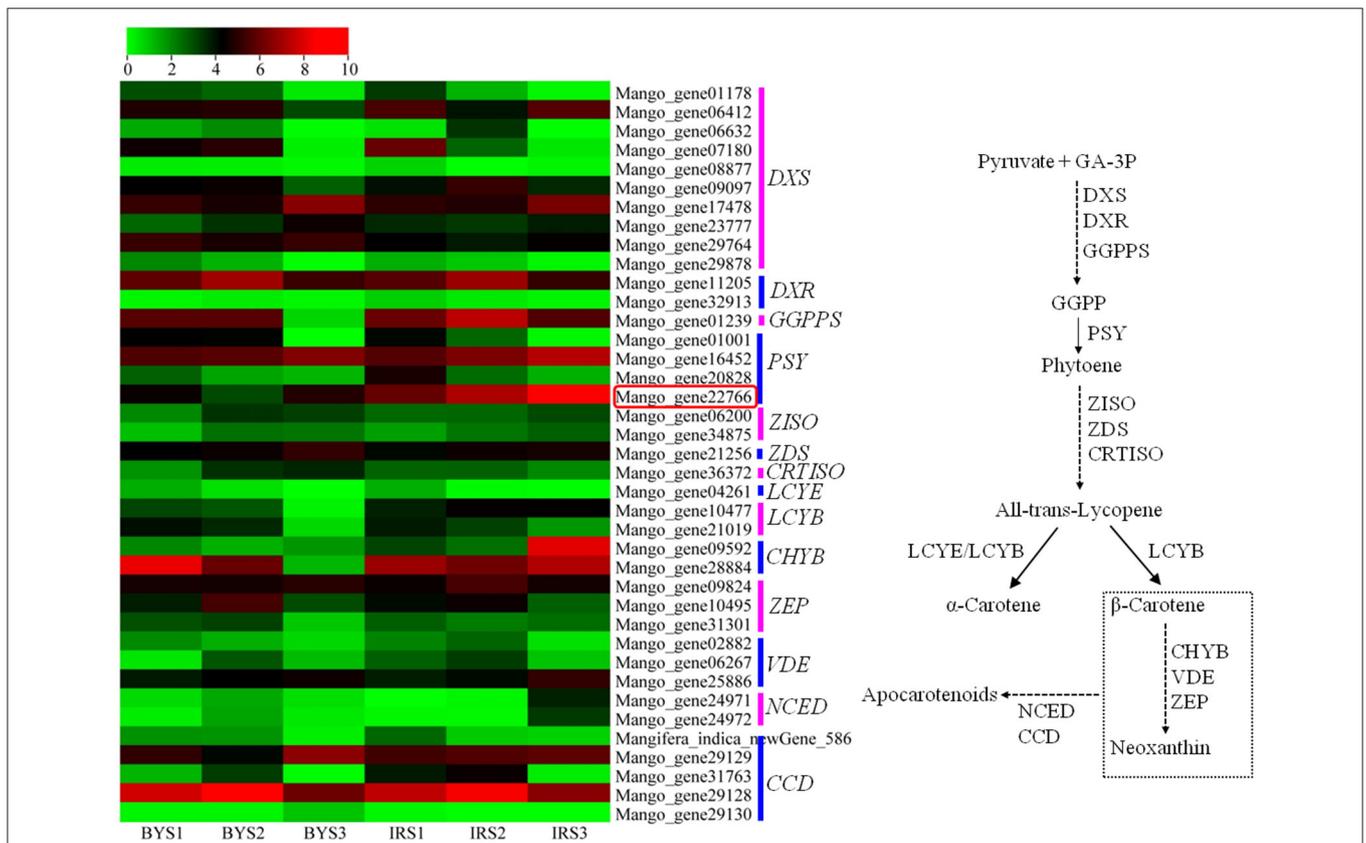


FIGURE 3 | Heatmap of the expression levels of genes related to carotenoid biosynthesis in the flesh of “Baixiangya” (BY) and “Irwin” (IR) during fruit development and ripening. Gene expression was indicated by reads per kilobase per million mapped reads (RPKM) values. Scaled log2 transcript values are shown from green to red, indicating low to high expression, respectively. A schematic of the carotenoid biosynthetic pathway in plants. DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; GGPPS, geranylgeranyl diphosphate synthase; PSY, phytoene synthase; PDS, phytoene desaturase; ZISO, ζ -carotene isomerase; ZDS, ζ -carotene desaturase; CRTISO, carotene isomerase; LCYE, lycopene ϵ -cyclase; LCYB, lycopene β -cyclase; CHYB, β -carotene hydroxylase; ZEP, zeaxanthin epoxidase; VDE, violaxanthin de-epoxidase; NCED, 9-*cis*-epoxycarotenoid dioxygenase; CCD, carotenoid cleavage dioxygenase.

52 genes were found to be associated with the carotenoid biosynthesis pathway, and 39 genes were differentially expressed between three developmental stages in both cultivars (Figure 3).

Among these differentially expressed genes, the expression levels of *MiDXS* (*Mango_gene17478* and *Mango_gene23777*, encoding 1-deoxy-D-xylulose-5-phosphatesynthase) and *MiPSY*

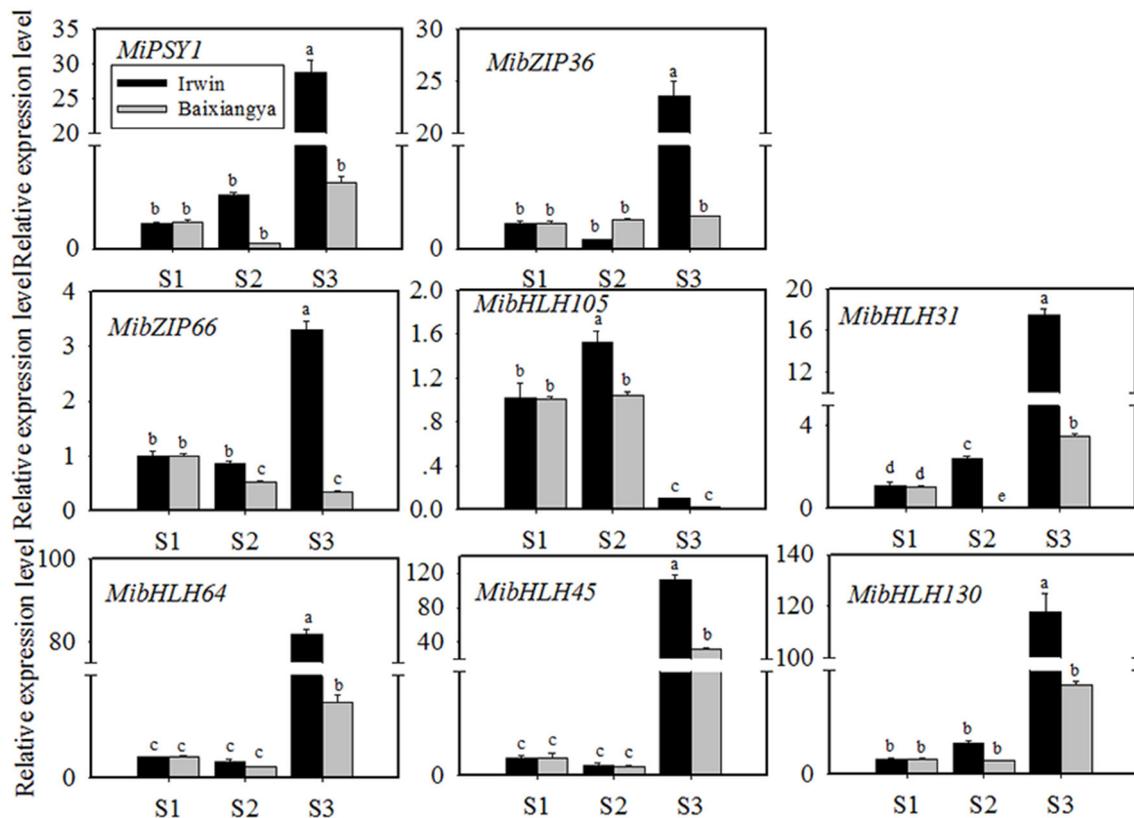


FIGURE 4 | Relative expression levels of *MiPSY1* and seven selected transcription factors (TFs) in the flesh of “Baixiangya” and “Irwin” mangoes during fruit development and ripening. Gene expression levels were analyzed by quantitative reverse transcription analysis. Different letters show significant differences in expression level as calculated by Duncan statistical analysis ($p < 0.01$).

(*Mango_gene16452* and *Mango_gene22766*) were relatively low during the early fruit development stage but increased greatly with continued fruit ripening in both cultivars. *LCYB* (*Mango_gene10477*) showed a considerably higher expression level in “Irwin” than in “Baixiangya” mangoes at the early stage of fruit development. During the fruit ripening period, its expression level remained nearly unchanged in “Irwin,” but decreased in “Baixiangya” mangoes. Similarly, *Mango_gene22766* on chromosome 7 showed a markedly high level of expression in “Irwin” during fruit development and ripening but a very low level in “Baixiangya” mangoes. At the fully ripened stage (S3), the *Mango_gene22766* expression level was 14-fold higher in “Irwin” than in “Baixiangya” based on qRT-PCR data (Figure 4). Hence, the high expression level of these genes is likely to lead to the accumulation of β -carotene in the Mango fruit. In contrast, the transcript levels of *MiCCD* (*Mango_gene31763*, *Mango_gene29128*, and *Mango_indica_newGene_586*) exhibited a significant decline during the fruit maturation phase, indicating a likely negative role in the development of Mango fruit flesh color. According to the annotation of the genome, *Mango_gene22766* plays an important role in driving the metabolic flux toward the β -carotene formation. Therefore, it can be inferred that *Mango_gene22766* may affect color

formation by regulating β -carotene production in the fruit flesh. We sequenced two coding sequences of *Mango_gene22766* from the flesh of “Irwin” and “Baixiangya,” and found they were both 1,299 bp long and encoded a peptide of 432 amino acids long (Supplementary Tables 11, 12). Except for the 400th and 417th amino acid, the rest of deduced amino acids were identical between the two cultivars (Supplementary Figure 2).

Isolation and Transactivation of Candidate TFs on *MiPSY1*

We named the cloned full-length sequence of *Mango_gene22766* as *MiPSY1* (Supplementary Table 11). *MiPSY1* was suggested to be distributed in the mitochondria by the WoLF PSORT Prediction online software (<http://www.genscript.com/wolf-psort.html>). To examine its subcellular location, *MiPSY1* was fused to RFP and transiently expressed in *Arabidopsis* protoplasts. Fluorescence microscopy analysis showed that the *MiPSY1* protein was localized in the mitochondria (Figure 5). In summary, *MiPSY1* might be a critical candidate gene controlling β -carotene formation in the mango fruit flesh.

To further investigate the potential TFs regulating *MiPSY1* expression, a weighted gene co-expression network

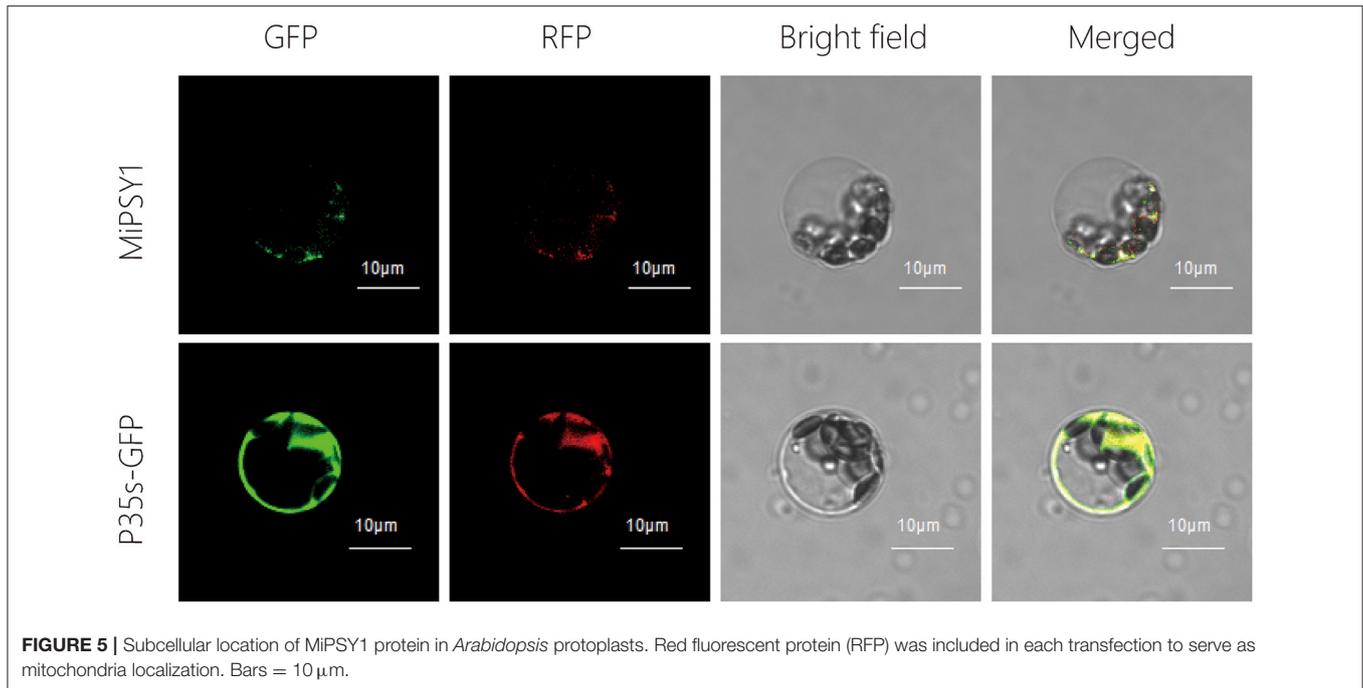


FIGURE 5 | Subcellular location of MiPSY1 protein in *Arabidopsis* protoplasts. Red fluorescent protein (RFP) was included in each transfection to serve as mitochondria localization. Bars = 10 μ m.

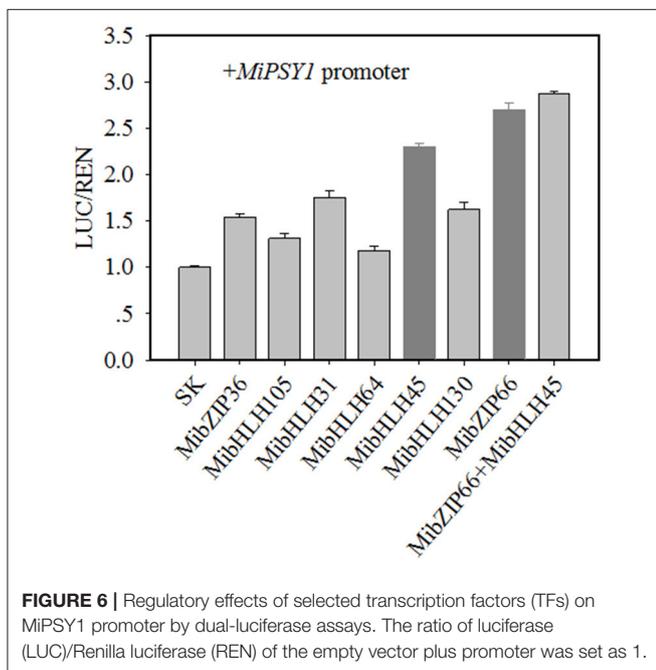


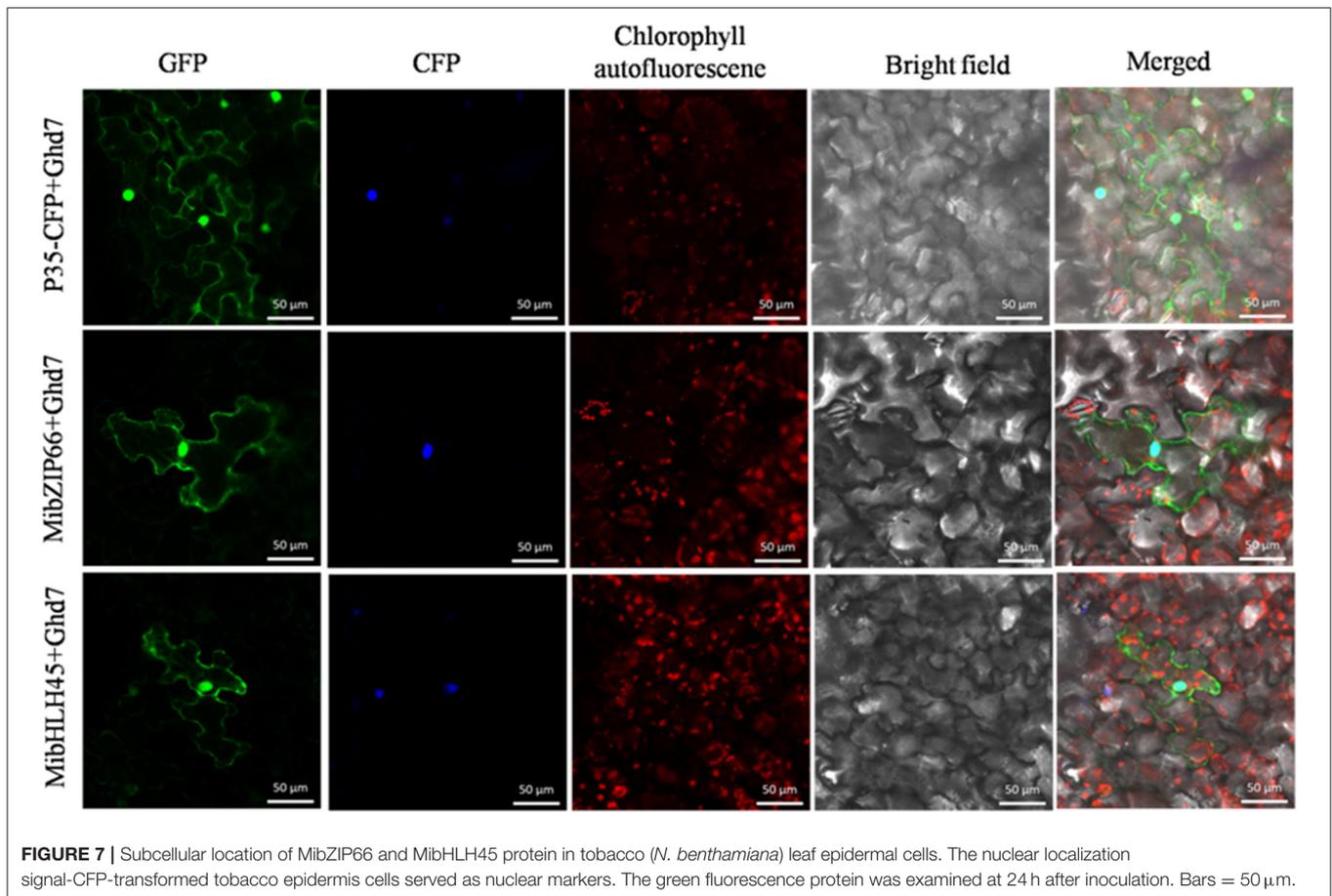
FIGURE 6 | Regulatory effects of selected transcription factors (TFs) on MiPSY1 promoter by dual-luciferase assays. The ratio of luciferase (LUC)/Renilla luciferase (REN) of the empty vector plus promoter was set as 1.

analysis of differentially expressed genes was performed (**Supplementary Figure 3A**). A total of 14 modules were identified (**Supplementary Figure 3B**). Analysis of module-trait relationships showed that the “MElightcyan” module was significantly correlated with β -carotene content ($r = 0.86$, $p = 0.00$) (**Supplementary Figure 3B**). Moreover, 49 TFs (including bHLH, bZIP, and NAC) were gathered in this module and were predicted to potentially interact with

MiPSY1 (**Supplementary Table 9**). According to PLACE and Plant-CARE databases, seven TFs, namely MibZIP36 (Mango_gene12953), MibZIP66 (Mango_gene35447), MibHLH105 (Mango_gene17414), MibHLH31 (Mango_gene17464), MibHLH64 (Mango_gene20675), MibHLH45 (Mango_gene30145), and MibHLH130 (Mango_gene30794), were predicted to be regulatory *cis*-acting elements of the *MiPSY1* promoter (**Supplementary Table 13**). Thus, we cloned these seven TFs and analyzed their sequences. Further phylogenetic analysis indicated that the seven identified TFs were grouped with *Arabidopsis thaliana* bZIP or bHLH proteins. To confirm the RNA-seq results, the seven TFs identified were further analyzed using qRT-PCR, which validated the RNA-seq results. The transcript levels of the seven TFs were significantly correlated with *MiPSY1* expression and β -carotene content during fruit development and ripening (**Figure 4**). In summary, these results suggest that these seven TFs might share the regulatory function of *A. thaliana* bZIP or bHLH and act as candidate regulators of β -carotene biosynthesis.

Dual-luciferase transient expression assays were performed to test the potential regulatory effects of the key candidate TFs on the *MiPSY1* promoter. The LUC/REN assays indicated that MibZIP66 and MibHLH45 could significantly activate the *MiPSY1* promoter, with transcription activation increments of 2.7- and 2.3-fold, respectively. The effect of the other five TFs on the *MiPSY1* promoter was limited. The action of MibZIP66 and MibHLH45 in combination had no significant additive effect on the *MiPSY1* promoter activity (**Figure 6**). MibZIP66 and MibHLH45 were localized exclusively in the nucleus (**Figure 7**).

The PlantTFDB analysis suggested that MibZIP66 and MibHLH45 could bind to the CACGTG sequence in the *MiPSY1* promoter, which was consistent with previous studies



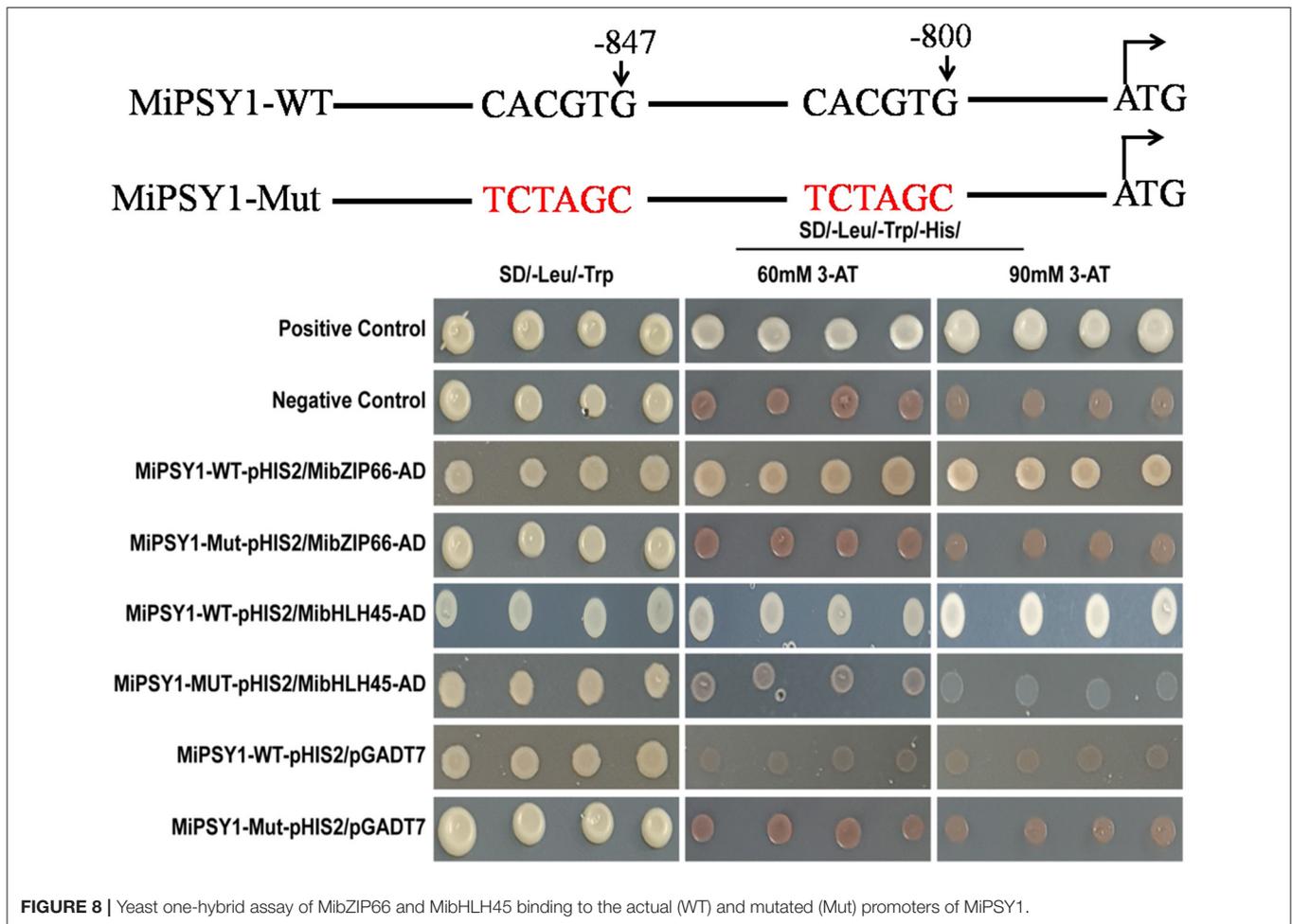
reporting that many bZIP and bHLH TFs specifically bind to the ACGT-containing elements, such as G-boxes (Pires and Dolan, 2010; Wolfgang et al., 2018). In our study, two CACGTG sequences were found to be located between the positions 852 and 800 bp, upstream of the transcriptional start codon in *MiPSY1* (Supplementary Table 13). YIH experiments and *cis*-element mutagenesis were performed. As shown in Figure 8, all of the yeast cells could grow normally on two dropout minimal media (SD/-Leu/-Trp). Nevertheless, only the yeast cells co-transformed with the positive control, pGADT7-MibZIP66+MibPSY1-WT promoter, or pGADT7-MibHLH45+MibPSY1-WT promoter grew well on selective medium SD/-His/Leu/-Trp supplemented with 60 or 90 mM 3-amino-1,2,4-triazole. The pGADT7-MibZIP66+MibPSY1-Mut promoter and pGADT7-MibHLH45+MibPSY1-Mut promoter could not grow on the selective medium when the two G-box elements CACGTG of the *MiPSY1* promoter were both mutated to TCTAGC (Figure 8). These results suggest that MibZIP66 and MibHLH45 could specifically bind to the G-box motif (CACGTG) of the *MiPSY1* promoter to regulate the expression of *MiPSY1* in the mango flesh.

DISCUSSION

Ever since the *Vitis vinifera* genome sequencing was completed in 2007 (Jaillon et al., 2007), the whole genomes of many fruit

crops, including fruit trees have been fully sequenced. However, cultivars have experienced frequent recombination and exchange of genetic fragments, and the contraction and expansion of gene families during domestication and genetic improvement through breeding and intense artificial selection. Hence, a single genome may not fully display the genetic diversity of the species. Recently, genome assemblies for the two mango cultivars “Hong Xiang Ya” (Li et al., 2020) and “Alphonso” (Wang et al., 2020) have been published, providing insights on mango genome evolution. In the current study, the whole genome of the mango cultivar “Irwin” was sequenced using the third-generation PacBio technique, which generated a high-quality genome with contig N50 and 1.03 Mb long. The “Irwin” genome provides a valuable new resource to broaden the currently available mango genome data, permitting us to deepen our knowledge on comparative genomics of mango.

Mango fruit flesh comes in different colors, including milk-white, yellow, and deep yellow. Our results showed that the β -carotene content was significantly higher in “Irwin” flesh than in “Baixiangya” flesh during the fruit ripening process, suggesting that the accumulation levels of β -carotene were related to the deepness of flesh yellow color (Vásquez-Cañedo et al., 2005). The transcriptome analysis revealed that upstream genes of the β -carotene biosynthesis pathway, including *MiDXS*, *MiPSY*, and carotenoid decomposition genes (*MiCCD*) were expressed differentially between “Irwin” and “Baixiangya” flesh, leading



us to speculate that their expression levels were contributed to the differential flesh β -carotene concentration between the two cultivars. *PSY* activity is generally accepted as the first committed step of carotenoid biosynthesis in fruit such as loquat (Fu et al., 2012) and apricot (Marty et al., 2005). We found that *MiPSY1* was expressed at significantly higher levels in “Irwin” than in “Baixiangya” during fruit development and ripening and that it was positively related to β -carotene production, suggesting the importance of its role. *LCYB* has been reported to be the rate-limiting enzyme for β -carotene biosyntheses, such as in durian fruits (Wisutiamonkul et al., 2017) and citrus fruits (Lu et al., 2018). However, in the present study, *MiLCYB* exhibited expression patterns uncorrelated with β -carotene accumulation during fruit development and ripening. Thus, we infer that *MiPSY1* acts as a key control point for β -carotene biosynthesis in the mango fruit flesh.

Transcription factors play a critical role in regulating functional gene expression to control the development of important traits in plants. A few TFs have been reported to directly regulate *PSY* expressions, such as RIN in tomato (Martel et al., 2011) and PIF1 in *Arabidopsis* (Toledo-Ortiz et al., 2010). In the present study, we found that the expression of *MibZIP66*

and *MibHLH45* correlated with both *MiPSY1* expression and β -carotene content, suggesting that the two TFs might be associated with β -carotene regulation. bHLH TFs are known to regulate carotenoid biosynthesis. In papaya, CpbHLH1/2 transactivates the expression of *CpCYC-B* and *CpLCY-B* during fruit ripening (Zhou et al., 2019). Moreover, bZIP family TFs have been shown to be involved in the regulation of anthocyanin biosyntheses, such as *Arabidopsis* HY5 (Nawkar et al., 2017), tomato SIHY5 (Liu et al., 2018), and pear PybZIPa (Liu et al., 2019). However, bZIP TFs have rarely been proven to directly regulate the structural genes involved in carotenoid production. Our results verified that MibZIP66 and MibHLH45 could specifically bind and transactivate the *MiPSY1* promoter. A recent study has identified that TF-TF interactions can participate in carotenoid biosyntheses, such as CpEIN3a interacting with CpNAC2 in papaya (Fu et al., 2017), and SINAC4 interacts with RIN in tomato (Zhu et al., 2014). However, the dual-luciferase assay showed that there was no significant synergistic effect of MibZIP66 and MibHLH45 on the *MiPSY1* promoter. Thus, MibZIP66 and MibHLH45 regulate the expression of *MiPSY1* to vary the content of β -carotene, thereby affecting the color formation in the mango flesh.

CONCLUSIONS

In summary, the “Irwin” genome and the transcriptional regulation of *MiPSY1* during mango fruit ripening are valuable to expand our understanding of the genetic basis of flesh color development. These genomic resources are expected to help accelerate the genetic improvement of mangoes.

DATA AVAILABILITY STATEMENT

The raw sequences data reported in this paper have been deposited in the Genome Sequence Archive (GSA) in national genomics data center (<https://ngdc.cnpc.ac.cn>), under accession number CRA004336.

AUTHOR CONTRIBUTIONS

XM, XL, YW, and JZ designed the study and performed the genome assembly and genome annotation. TB and JS edited the

manuscript. BZ, WX, and LL collected the samples. SW and HW provided advice on the experimental design. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Science Foundation of China (Grant No. 31672103), Guangdong Provincial National Science Foundation (No. 2021A1515010966), and Guangdong Provincial Special Fund for Modern Agriculture Industry Technology Innovation Teams (2019KJ108). We thank Dr. Mingjun Li (College of Horticulture, Northwest A & F University, Shanxi, China) for English editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.749108/full#supplementary-material>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* 58, 268–276. doi: 10.1016/j.ymeth.2012.05.001
- Chan, P. P., and Lower, T. M. (2019). tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* 1962, 1–14. doi: 10.1007/978-1-4939-9173-0_1
- Chung, M. Y., Vrebalov, J., Alba, R., Lee, J., McQuinn, R., Chuang, J. D., et al. (2010). A tomato (*Solanum lycopersicum*) APETALA2/ERF gene, SLAP2a, is a negative regulator of fruit ripening. *Plant J.* 64, 936–947. doi: 10.1111/j.1365-3113X.2010.04384.x
- Deshpande, A. B., Anamika, K., Jha, V., Chidley, H. G., Oak, P. S., Kadoo, N. Y., et al. (2017). Transcriptional transitions in Alphonso mango (*Mangifera indica* L.) during fruit development and ripening explain its distinct aroma and shelf life characteristics. *Sci. Rep.* 7:8711. doi: 10.1038/s41598-017-08499-5
- Edgar, R. C., and Myers, E. W. (2005). PILER: identification and classification of genomic repeats. *Bioinformatics* 21, 152–158. doi: 10.1093/bioinformatics/bti1003
- Fu, C. C., Han, Y. C., Kuang, J. F., Chen, J. Y., Lu, W. J. (2017). Papaya CpEIN3a and CpNAC2 Co-operatively regulate carotenoid biosynthesis-related genes CpPDS2/4, CpLCY-e and CpCHY-b during fruit ripening. *Plant Cell Physiol.* 58, 2155–2165. doi: 10.1093/pcp/pcx149
- Fu, X., Kong, W., Peng, G., Zhou, J., Azam, M., Xu, C., et al. (2012). Plastid structure and carotenogenic gene expression in red- and white-fleshed loquat (*Eriobotrya japonica*) fruits. *J. Exp. Bot.* 63, 341–354. doi: 10.1093/jxb/err284
- Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463. doi: 10.1038/nature06148
- Karlova, R., Rosin, F. M., Busscher-Lange, J., Parapunova, V., Do, P. T., Fernie, A. R., et al. (2011). Transcriptome and metabolite profiling show that APETALA2a is a major regulator of tomato fruit ripening. *Plant Cell* 23, 923–941. doi: 10.1105/tpc.110.081273
- Li, W., Zhu, X. G., Zhang, Q. J., Li, K., Zhang, D., Shi, C., et al. (2020). SMAT sequencing generates the chromosome-scale reference genome of tropical fruit mango, *Mangifera indica*. *BioRxiv* 2:960880. doi: 10.1101/2020.02.22.960880
- Lister, R., Carrie, C., Duncan, O., Ho, L. H. M., Howell, K. A., Murcha, M. W., et al. (2007). Functional definition of outer membrane proteins involved in preprotein import into mitochondria. *Plant Cell* 19, 3739–3759. doi: 10.1105/tpc.107.050534
- Liu, C. C., Chi, C., Jin, L. J., Zhu, J., Yu, J. Q., Zhou, Y. H. (2018). The b-ZIP transcription factor HY5 mediates CRY1a-induced anthocyanin biosynthesis in tomato. *Plant Cell Environ.* 41, 1762–1775. doi: 10.1111/pce.13171
- Liu, H., Su, J., Zhu, Y., Yao, G., Allan, A. C., Ampomah-Dwanmen, C., et al. (2019). The involvement of PybZIPa in light-induced anthocyanin accumulation via the activation of PyUFGT through binding to tandem G-box in its promoter. *Hortic. Res.* 6:134. doi: 10.1038/s41438-019-0217-4
- Liu, L. H., Shao, Z. Y., Zhang, M., and Wang, Q. M. (2015). Regulation of carotenoid metabolism in tomato. *Mol. Plant* 8, 28–39. doi: 10.1016/j.molp.2014.11.006
- Lu, S. W., Zhang, Y., Zhu, K., Yang, W., Ye, J. L., Chai, L., et al. (2018). The citrus transcription factor CsMADS6 modulates carotenoid metabolism by directly regulating carotenogenic genes. *Plant Physiol.* 176, 2657–2676. doi: 10.1104/pp.17.01830
- Luo, Z., Zhang, J., Li, J., Yang, C., Wang, T., Ouyang, B., et al. (2013). A STAY-GREEN protein SISR1 regulates lycopene and β -carotene accumulation by interacting directly with SLPSY1 during ripening processes in tomato. *N. Phytol.* 198, 442–452. doi: 10.1111/nph.12175
- Ma, X., Zheng, B., Ma, Y., Xu, W., Wu, H., and Wang, S. (2018). Carotenoid accumulation and expression of carotenoid biosynthesis genes in mango flesh during fruit development and ripening. *Sci. Hortic.* 237, 201–206. doi: 10.1016/j.scienta.2018.04.009
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of K-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011
- Martel, C., Vrebalov, J., Tafelmeyer, P., and Giovannoni, J. J. (2011). The tomato MADS-Box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening process in a COLORLESS NONRIPENING-dependent manner. *Plant Physiol.* 157, 1568–1579. doi: 10.1104/pp.111.181107
- Marty, I., Bureau, S., Sarkissian, G., Gouble, B., Audergon, J. M., and Albagnac, G. (2005). Ethylene regulation of carotenoid accumulation and carotenogenic gene expression in colour-contrasted apricot varieties (*Prunus armeniaca*). *J. Exp. Bot.* 417, 1877–1886. doi: 10.1093/jxb/eri177
- Mercadante, A. Z., and Rodriguez-Amaya, D. B. (1998). Effects of ripening, cultivar differences, and processing on the carotenoid composition of mango. *J. Agric. Food Chem.* 46, 128–130. doi: 10.1021/jf9702860
- Nawkar, G. M., Kang, C. H., Maibam, P., Park, J. H., Jung, Y. J., Chae, H. B., et al. (2017). HY5, a positive regulator of light signaling, negatively controls

- the unfolded protein response in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 114:2084. doi: 10.1073/pnas.1609844114
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509
- Olano, C. T., Schnell, R. J., Quintanilla, W. E., and Campbell, R. J. (2005). Pedigree analysis of Florida mango cultivars. *Proc. Fla. State Hort. Soc.* 118, 192–197.
- Pires, N., and Dolan, L. (2010). Origin and diversification of bHLH proteins in plants. *Mol. Biol. Evol.* 27, 862–874. doi: 10.1093/molbev/msp288
- Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). *De novo* identification of repeat families in large genomes. *Bioinformatic* 21, 351–358. doi: 10.1093/bioinformatics/bti1018
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Sagawa, J. M., Stanley, L. E., Lafountain, A. M., Frank, H. A., and Yao, W. Y. (2015). An R2R3-MYB transcription factors regulates carotenoid in *Mimulus lewisii* flowers. *New Phytol.* 3, 1049–1057. doi: 10.1111/nph.13647
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 1119–1125. doi: 10.1093/bioinformatics/btv351
- Su, L., Diretto, G., Purgatto, E., Danoun, S., Zouine, M., Li, Z., et al. (2015). Carotenoid accumulation during fruit ripening is modulated by the auxin-ethylene balance. *BMC Plant Biol.* 15:114. doi: 10.1186/s12870-015-0495-4
- Sun, T., Yuan, H., Cao, H., Yazdani, M., Tadmor, Y., and Li, L. (2017). Carotenoid metabolism in plants: the role of plastids. *Mol. Plant* 11, 58–74. doi: 10.1016/j.molp.2017.09.010
- Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol.* 9, 678–687. doi: 10.1093/oxfordjournals.molbev.a040752
- Toledo-Ortiz, G., Huq, E., and Rodriguez-Concepcion, M. (2010). Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11626–11631. doi: 10.1073/pnas.0914428107
- Vásquez-Cañedo, A. L., Sruamsiri, P., Carle, R., and Neidhart, S. (2005). Accumulation of All-trans- β -carotene and Its 9-cis and 13-cis Stereoisomers during postharvest ripening of nine Thai mango cultivars. *J. Agric. Food Chem.* 53, 4827–4835. doi: 10.1021/jf048168h
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wang, P., Luo, Y., Huang, J., Gao, S., Zhu, G., Dang, Z., et al. (2020). The genome evolution and domestication of tropical fruit mango. *Genome Biol.* 21:60. doi: 10.1186/s13059-020-01959-8
- Warschafsky, E. J., and Wettberg, E. (2019). Population genomic analysis of mango (*Mangifera indica*) suggests a complex history of domestication. *N. Phytol.* 222, 2023–2037. doi: 10.1111/nph.15731
- Wisutiamonkul, A., Ampomah-Dwamena, C., Allan, A. C., and Ketsa, S. (2017). Carotenoid accumulation and gene expression during durian (*Durio zibethinus*) fruit growth and ripening. *Sci. Hortic.* 220, 233–242. doi: 10.1016/j.scienta.2017.03.048
- Wolfgang, D. L., Snoek, B. L., Berend, S., and Christoph, W. (2018). The Arabidopsis bZIP transcription factor family—an update. *Curr. Opin. Plant Biol.* 45, 36–49. doi: 10.1016/j.pbi.2018.05.001
- Xiao, C. L., Chen, Y., Xie, S. Q., Chen, K. N., Wang, Y., Han, Y., et al. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* 14, 1072–1074. doi: 10.1038/nmeth.4432
- Xu, Z., and Wang, H. (2007). LTR-FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, 265–268. doi: 10.1093/nar/gkm286
- Zhou, D., Shen, Y. H., Zhou, P., Fatima, M., Lin, J., Yue, J., et al. (2019). Papaya CpbHLH1/2 regulate carotenoid biosynthesis-related genes during papaya fruit ripening. *Hortic. Res.* 6:80. doi: 10.1038/s41438-019-0162-2
- Zhu, M. K., Zhou, S. A., Dong, T. T., Chen, G. P., Wang, Y., and Hu, Z. L. (2014). A new tomato NAC (NAM/ATAF1/2/CUC2) transcription factor, SINAC4, functions as a positive regulator of fruit ripening and carotenoid accumulation. *Plant Cell Physiol.* 55, 119–135. doi: 10.1093/pcp/pct162

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ma, Luo, Wei, Bai, Shi, Zheng, Xu, Li, Wang, Zhang and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.