**Natural Variation in *CCD4* Promoter Underpins Species-specific Evolution of Red Coloration in Citrus Peel**

Xiongjie Zheng1, Kaijie Zhu1, Quan Sun1, Weiyi Zhang1, Xia Wang1, Hongbo Cao1, a, Meilian Tan1, b, Zongzhou Xie1, Yunliu Zeng1, Junli Ye1, Lijun Chai1, Qiang Xu1, Zhiyong Pan1, \*, Shunyuan Xiao1, 3, Paul D. Fraser2, Xiuxin Deng1, \*

1Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, China

2 School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK

3Department of Plant Science and Landscape Architecture, Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD, USA

a H.C. present address is College of Horticulture, Agricultural university of Hebei, Baoding Hebei 071000, China

b M.T. present address is Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan 430070, China.

\* To whom correspondence should be addressed. xxdeng@mail.hzau.edu.cn; zypan@mail.hzau.edu.cn

Running Title: Genetic basis of apocarotenoids pigments diversity

**ABSTRACT**

Carotenoids and apocarotenoids can act as phytohormone and volatile precursors key for plant development and confer aesthetic and nutritional value critical to consumer preference. Citrus fruits display considerable natural variation for carotenoids and apocarotenoids pigments. A multifaceted genetic approach revealed that a 5´*cis*-regulatory change at *CCD4b* encoding CAROTENOID CLEAVAGE DIOXYGENASE 4bis a major genetic determinant for the natural variation of C30 apocarotenoids responsible for the red coloration of citrus peel. Functional analyses demonstrated that besides the known catalytic role in synthesizing red β-citraurin, *CCD4b* is also responsible for the production of another important C30 apocarotenoid pigment, β-citraurinene. Furthermore, promoter and transcript analyses for *CCD4b* among citrus germplasm accessions established a tight correlation between presence of a putative 5´ *cis*-regulatory enhancer within a MITE transposon and the enhanced allelic expression of *CCD4b* in C30 apocarotenoid-rich red-peeled accessions. Phylogenetic evidence suggests that functional diversification and naturally occurring promoter variation of *CCD4b* constituted the stepwise evolution of the red-peel trait in mandarins and their hybrids. Findings herein not only provide a new insight into genetic and evolutionary basis of apocarotenoid diversity in plant, but also may facilitate breeding efforts towards improving nutritional and aesthetic value of citrus and perhaps other fruit crops.

**Key words:** Citrus, apocarotenoid, natural variation, carotenoid cleavage dioxygenase (CCD), promoter, transposon

**INTRODUCTION**

Carotenoids are present in most fruits and vegetables. Because of their potent antioxidant properties and/or provitamin A activity, dietary acquired carotenoids are associated with health promoting benefits (Fraser and Bramley, 2004; Krinsky and Johnson, 2005). In many cases, the levels and types of carotenoids present in fresh fruits and vegetables can also determine the intensity of its color (Yuan et al., 2015), thereby contributing to the aesthetic properties and consumer quality of the product. Industrially, carotenoids are used across the cosmetic, food, health, feed and pharma sectors (Bouvier et al., 2003; Noviendri et al., 2011).

Citrus is one of the most important fruit tree crops in the world with an annual production of about 120 million tons since 2011 (FAO, 2017). Citrus fruits contain a diverse array of carotenoids in their flesh and peel (Fanciullino et al., 2006; Ikoma et al., 2016; Rodrigo et al., 2013). Compared with other crops such as tomato and carrot known to contain a high content of carotenes (α, β carotene and lycopene), citrus fruits display a greater level of carotenoid diversity. For example, while the red flesh of citrus fruits has a high content of lycopene or β-cryptoxanthin, the red peel is rich in citrus-specific C30 apocarotenoids (β-citraurin and β-citraurinene) (Supplemental Figure 1) (Alquezar et al., 2008; Alquézar et al., 2008; Farin et al., 1983; Kato et al., 2004). These carotenoids, besides their nutritional value, also contribute to the red color of the peel and/or flesh of citrus fruits, thereby enhancing consumer acceptance (Carmona et al., 2012). Additionally, these carotenoids and/or their catabolites have been shown to have protective roles against various disease states (Frusciante et al., 2014; G Gutheil et al., 2012; Matsumoto, 2013). The abundant citrus germplasm with diverse flesh and peel colors are excellent genetic resources for dissection of new biochemical and molecular genetic mechanisms controlling carotenoid content and diversity.

Apocarotenoids are enzymatic and non-enzymatic products derivative from carotenoids. Some apocarotenoids act as precursors for phytohormones such as abscisic acid and strigolactones that affect plant development and adaption (Alder et al., 2012; Tan et al., 2003). In plants, the biochemical pathways involved in the biosynthesis of carotenoids and apocarotenoids have been largely elucidated (Supplemental Figure 1). Through a series of biochemical reactions, colorless phytoene is converted to pink-colored lycopene, and then to α and β-carotenes and various xanthophylls. It has been known that CAROTENOID CLEAVAGE DIOXYGENASE 4 (CCD4) acts as a negative regulator of carotenoid content in various plant tissues by cleaving carotenoids into C27apocarotenoids which will be subsequently converted to colorless compounds (Campbell et al., 2010; Bruno et al., 2015; Bruno et al., 2016; Falchi et al., 2013; Gonzalez-Jorge et al., 2013; Zhang et al., 2015). Interestingly, one duplicated copy of the *CCD4* gene in citrus has undergone neofunctionalization to produce a CCD4b isozyme that catalyzes the cleavage of C40 β- xanthophylls (β-cryptoxanthin and zeaxanthin), leading to the formation of a red C30 apocarotenoid, β-citraurin (Ma et al., 2013; Rodrigo et al., 2013). The other major C30 apocarotenoid pigment responsible for red coloration of citrus fruits is β-citraurinene (Farin et al., 1983). However, how C30 β-citraurinene pigment is produced still awaits elucidation.

Currently, little is known about the genetic basis underlying citrus fruit color, owing to the difficulty in genetic transformation and its long juvenility. Nevertheless, it has been found that peel color of citrus fruits appears to observe simple Mendelian inheritance of two major genes, with red peel color being dominant over yellow (Chen et al., 1993). In the present study, using a combinatory genetic approach, we demonstrate that a 5´ *cis*-regulatory change due to an insertion of a MITE transposon caused enhanced expression of the *CCD4b* gene, which in turn resulted in increased production of citrus-specific C30 β-citraurin and β-citraurinene that are attributable to the red coloration of citrus fruit peel. Thus, our results provide new insights into the biochemical, genetic and evolutionary basis of fruit peel coloration in *Citrus* and should facilitate breeding and engineering efforts towards improving citrus fruit quality.

**Results**

**C30 apocarotenoids are strongly associated with red-peel color** **trait in a citrus F1 population**.

To determine the genetic basis of the red peel phenotype, we examined the segregation of peel color within a F1 pseudo-testcross population, derived from a cross between a red-peeled Red tangerine *(Citrus reticulata)* and a yellow-peeled Trifoliate orange (*Poncirus trifoliata*) as previously reported (Tan et al., 2007). Given the discontinuous distribution of the color index, e.g. the H value among the F1 individuals, we were able to group them into either the red peel or the yellow peel type (Figure 1A, 1B and Supplemental Figure 2A). The segregation ratio (47: 34) of red peel plants vs. yellow peel plants fits an approximate 1:1 ratio expected for segregation of a monogenic dominant gene in a testcross (Supplemental Figure 2B).

Red peel of citrus fruits is known to contain a high content of C30 apocarotenoids (Farin et al., 1983). To determine if this holds true for the F1 population, the carotenoid profiles for 81 F1 plants (Supplemental Dataset 1 and 2) were determined. Chromatographic separation and HPLC-DAD-APCI-MS/MS were used to identify C30 β-Citraurinene (m/z 419.3 [M+H]+) and C30 β-Citraurin (m/z 433.3 [M+H]+) previously found in citrus species with red peel (Supplemental Figure 3 and 4) (CURL, 1965; Leuenberger and Stewart, 1976). The F1 plants with red and yellow peel colorations showed distinct difference in carotenoid composition and content, especially for the two C30 apocarotenoids (β-citraurin and β-citraurinene) (Figure 1C). Principal component analysis (PCA) of carotenoids profiles showed that there are two distinct clusters, consisting of individual F1 plants with either red-peel or yellow-peel (Supplemental Figure 5A and 5B). Combining the results of the quantitative levels displayed in the heat map and Pearson correlation analysis, it was observed that β-citraurin, β-citraurinene, phytoene, and phytofluene all positively correlated, whereas lutein and zeaxanthin negatively correlated with the red color of fruit peel among the F1 plants (Figure 1D, Supplemental Figure 5C and D). Particularly, the ratio of C30 apocarotenoids to total carotenoids showed the highest correlation (r > 0.9, P-value = 2.2E-16) with the red-peel color index in the F1 population (Supplemental Figure 5D). Taken together, the above results suggest that the red color trait may be controlled by a dominant gene that promotes the production of C30 apocarotenoids.

**A** ***cis*-acting eQTL correlates the variations in *CCD4b* expression and co-segregates with the red peel** **color.**

To map the gene controlling peel color, we performed Bulked Segregant RNA Sequencing (BSR-seq) using two extreme pools containing equal amount of RNA from each of 22 plants that yielded either red-peel or yellow-peel fruits. As inferred from the BSR-seq average Δ (Single Nucleotide Polymorphism (SNP) index) scans, a dramatic bimodal distribution was displayed for Chromosome 8 (Figure 2A and Supplemental Dataset 3), suggesting that a major genetic determinant(s) of the red-peel trait is located in this chromosome. Then, fine mapping using recombinants were used to narrow the genetic interval down to a 1.3Mb physical region (scaffold8: 16.7M~18M) between M-PCR14 and M-PCR3 (Figure 2B). This 1.3Mb genomic interval contains 62 predicted protein encoding genes (Supplemental Dataset 4), of which only two exhibited significant difference in expression between the two pools (Supplemental Figure 6A). The first gene is *CCD4b* whose expression was ~30-fold higher in the red peel pool, and the other is a defense-related gene with higher expression in the yellow peel pool (Supplemental Figure 6A, B and Supplemental Dataset 5). Quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR) analysis also confirmed that the red-peel individuals of the F1 population showed higher levels of *CCD4b* expression than yellow-peel individuals (Figure 2C and Supplemental Dataset 6). Thus, it appears that higher expression of *CCD4b* is responsible for the red peel coloration. There existed a second CCD4-like gene Ciclev10030384m.g in the genetic interval of red-peel trait. However, BSRNA-seq together with Q-RT-PCR results found it showed very low expression in citrus tissues (Supplemental Figure 6B and C). Consistently, the expressed sequence tag (EST) of Ciclev10030384m.g was also absent in the Citrus EST database (http://harvest.ucr.edu/, software HarvEST 1.32, assembly C52) from Citrus cDNA libraries. The results indicated the Ciclev10030384m.g gene is not expressed in any citrus tissue, which suggests that it might be a pseudogene (Rodrigo et al., 2013). Interestingly, while no polymorphism was detected in the *CCD4b* coding sequences among the F1 plants, as reported for the two parents (Zheng et al., 2015), there are a few SNPs segregating among the F1 individuals, two of which [SNP1G and SNP2G] are located within a 158bp Harbinger Miniature Inverted-repeat Transposon (MITE) at 1537 bp 5´ of the ATG start codon of *CCD4b* that is present only in the red-peel parent (HJ) and co-segregating with the red-peel phenotype in the F1 population (Figure 2D and Supplemental Figure 7). These results imply that a MITE-introduced 5´regulatory change in the promoter region might cause enhanced expression of *CCD4b*.

To test the above hypothesis, expression Quantitative Trait Loci (eQTL) analysis were carried out to search for genomic loci that can explain the variation in expression levels of *CCD4b* in the F1 population. As shown in Supplemental Table 1, four eQTLs were detected, among which one *cis*-acting eQTL (Scaffold 8: 16946945~18572910, P<0.001) having the highest LOD (13.06) and explaining 55.2% of the variance of *CCD4b* expression was found to be located in the same genomic region where the red peel phenotype mapped (Scaffold 8: 16707807~18021439) (Figure 2E). These results suggest that this *cis*-acting eQTL of *CCD4b* gene is associated with the red peel trait, which supports our initial hypothesis.

***CCD4b* is located in a mQTL hotspot associated with variations in carotenoids and C30 apocarotenoids.**

Investigating the genetic control of intermediate phenotypes such as metabolites would contribute to bridging genotype–phenotype gap and may provide useful genetic and biochemical interpretation of mechanisms underlying the complex phenotypic traits (Luo, 2015). Thus, we performed metabolite (m) QTL analysis for variations of 12 “traits” related to levels of 11 individual carotenoids and the ratio of C30 apocarotenoids to total carotenoids (C30/Total) (Supplemental Table 2). Nineteen significant overlapping mQTLs were identified for these 12 “traits” assessed over two years (Supplemental Table 3), and 8 of which (β-citraurin, β-citraurinene, Zeaxanthin, Lutein, luteoxanthin, phytoene, phytofluene and C30/Total) were co-localized on a mQTL hotspot region (66.663-68.663 CM on Chromosome 8) (Figure 3A and Supplemental Table 3). Of these 8 traits, C30/Total showed the highest percentage of phenotypic variance (~91%) explained by the mQTL and has the highest LOD values (~36) (Supplemental Table 2). This mQTL hotspot is located on Chr 8: 17664105~18572910 based on the Clementine genome sequences (Wu et al., 2014) and overlaps with the genomic region (Chr 8: 16707807~18021439) where the red-peel trait and *CCD4b* were mapped (Figure 3A). Thus, this result suggests that the *CCD4b* locus may be a major genetic determinant for variation of C30-apocarotenoids (β-citraurin and β-citraurinene) attributable to red-peel coloration of citrus fruits, which would also control the variation of the C40 carotenoids (i.e. Lutein and Zeaxanthin). Giving the CCD4b was responsible for biosynthesis of C30 β-citraurin (Ma et al., 2013; Rodrigo et al., 2013), the mQTL result suggested that CCD4b might be also involved in the unknown pathway for biosynthesis of C30 β-citraurinene. Consistent with this speculation, spatiotemporal expression of *CCD4b* showed a positive correlation with the levels of peel-specific C30 β-citraurin and β-citraurinene within peel of citrus fruits during fruit ripening (Figure 3B-D).

**Overexpression of *CCD4b* leads to production of peel-specific C30 β-citraurinene pigment in** **citrus callus.**

Citrus callus tissues have a very low-level *CCD4b* expression (Cao et al., 2012), and have been used for functional analysis of candidate genes implicated in carotenoid biosynthesis (Cao et al., 2012; Lu et al., 2018). To directly assess the impact of elevated *CCD4b* expression in production of C30 apocarotenoids, we overexpressed *CCD4b* in the callus of two previously established multigene Engineered Cell Models (ECMs) (Cao et al., 2012). ECM1 callus produces abundant precursors of carotenes (Supplemental Table 4) due to overexpression of *CrtB* (a bacterial phytoene synthase gene). The overexpression of *CCD4b* in ECM1 (Ox-4b-ECM1) led to the production of a new C30 apocarotenoid (i.e.β-citraurinene) that should be only detected in the peels of citrus fruits (Figure 4 A-C, Supplemental Figure 8A and Supplemental Table 4). QRT-PCR and Western Blot analyses suggested that the expression levels of CCD4bmRNA and protein positively correlated with the ratio of C30 apocarotenoids to total carotenoids, but negatively correlated with the ratios of α-derived xanthophylls (i.e. lutein) to total carotenoids and α/ β-carotene to total carotenoids among different Ox-4b-ECM1 lines (Figure 4A-C and Supplemental Figure 8C-E).

ECM2 is abundant in xanthophyll precursors due to the overexpression of both *CrtB* and Os*HYD* with a carotenoid composition similar to that of citrus fruit peel. Overexpression of *CCD4b* in ECM2 also resulted in production of β-citraurinene, along with changes in C40 carotenoids similar to that in ECM1 overexpressing *CCD4b* (Supplemental Table 4). The CCD4b expression levels in different ECM2 lines transgenic for *CCD4b* also showed a positive correlation with the ratio of C30 apocarotenoids to total carotenoids (Supplemental Figure 8C-E). Together, the above results from the analysis of transgenic ECM clearly demonstrate a prominent role of enhanced expression of *CCD4b* in synthesizing C30 β-citraurinene while also impacting C40 carotenoid composition.

To determine the subcellular localization of CCD4b, we transiently co-expressed CCD4b-YFP with AtFBN1a-CFP, a well-accepted marker for plastoglobules (PG) of plastids (Rottet et al., 2016), in protoplasts of Arabidopsis leaves. CCD4b-YFP was found as yellow punctate fluorescence inside the chloroplasts with red autofluorescence. The yellow fluorescence, in most cases, colocalized with cyan fluorescence of FBN1a-CFP, indicating that CCD4b could mainly localizes in plastoglobules (Supplemental Figure 9). Consistent with this, TEM analysis revealed that overexpression of *CCD4b* in ECM1 resulted in increased numbers of PG within the chromoplast, and increased total PG area per plastid (Figure 4D and 4E). These observations imply structural changes may occur in plastids for callus cells to adapt or respond to CCD4b-induced increase in synthesis of C30 xanthophylls (as reported by increased C30 β-citraurinene content) from carotenes (as reported by reduced β-carotene content).

**A 5´ *cis*-regulatory change accounts for *CCD4b* allele-specific expression and production of C30 apocarotenoids**.

Our efforts in genetic mapping collectively suggest that a 5´ *cis*-regulatory change may underscore the red-peel color (Figure 2). To determine the precise DNA polymorphism at the 5´*cis*-regulatory region of *CCD4b* that influences its transcription, we sequenced 1900 bp 5´ of the ATG start codon of *CCD4b* for 115 citrus germplasm accessions, including three basic species, mandarin (*C. reticulata*), pummelo (*C. maxima* or *C. grandis*) and citron (*C. medica*) and other citrus species of hybrid origin. As shown in Supplemental Table 5-7, only some mandarin species and their hybrid contain the MITE transposon at 5´ cis-regulatory region of CCD4b. Those accessions containing MITE(AG) or MITE(GG) allele with SNP2G in MITE (i.e. the MITE\_AG or MITE\_GG genotype) produce fruits with red-peel when fruits mature on the tree or during postharvest storage (such as oranges) (Supplemental Figure 11). By contrast, the other accessions containing MITE(AA) allele with SNP2A (MITE\_AA genotype) produce yellow-peel fruits. In addition, the accessions lacking this MITE insertion (i.e. the no-MITE genotype) also produce fruits with yellow-peel or greenish-yellow-peel. Analysis of peel carotenoids in 95 of the 115 citrus germplasm accessions also established a strong positive correlation between the MITE\_AG or MITE\_GG genotype and the levels of C30 β-citraurin and β-citraurinene (Figure 5A-C and Supplemental dataset 7). Because the *CCD4b* promoter sequences with MITE from the MITE(AA) allele in yellow-peel species differ from that of the MITE(AG) allele in red-peel species only in one nucleotide (i.e. SNP2A/G (…CAACTA/GTGG…) in the MITE sequence (Supplemental Table 6). Based on these data it could be speculated that SNP2G in the MITE sequence alone results in enhanced transcription of *CCD4b*.

To test this hypothesis, we first measured the expression levels of *CCD4b* among different citrus accessions using QRT-PCR and found a positive correlation between enhanced *CCD4b* expression and the presence of SNP2G (in either the MITE\_AG or MITE\_GG genotype) in the promoter of *CCD4b* (Figure 5C). We then used one SNP(C/T) within the cDNA of *CCD4b* that corresponds to distinct promoter sequences and performed high-resolution melting (HRM) analysis to detect allele-specific expression of *CCD4b* in the hybrid citrus species. it was found that almost all *CCD4b* transcripts detected were derived from alleles that contain an upstream sequence of either MITE(AG) or MITE(GG) (Figure 6A and 6B). Consistent with these findings, it was found that when cDNA from red-peel citrus species was used as a template for amplification of *CCD4b*, only the allele carrying SNP(C) *cis*-linked with MITE(AG) and/or MITE(GG) allele was detected in the PCR products, this was despite both alleles (SNP(C/T) being amplified when genomic DNA was used as the PCR template (Figure 6C). Together, these results demonstrate that the *CCD4b* alleles with the 5´ SNP2G sequence are preferentially expressed relative to the other *CCD4b* alleles.

To directly verify that the*CCD4b* promoters containing the 5´ SNP2G sequence have higher promoter activity, promoter-GUS analyses in *N. benthamiana* were carried out. It was found that the promoters containing MITE(AG) or MITE(GG) allele indeed had higher activities than those containing MITE(AA) (Supplemental Figure 10). It has been reported that the expression of CCD4b in peel of mandarin and orange can be induced by ethylene treatment (Ma et al., 2013; Rodrigo et al., 2013). Additionally, we also found that transcription of *CCD4b* with MITE\_GG from red peel F1 plants, but not that of *CCD4b* with MITE\_AA from yellow peel F1 plants was highly induced by ethylene treatment (Figure 6D). This suggests that the 5´ SNP2G change within the MITE may enhance the transcription of *CCD4b* via interaction with an ethylene-sensitive transcription factor. Based on the above results, we conclude that the 5´ SNP2G regulatory change in the *CCD4b* promoter is responsible for the enhanced expression of *CCD4b*, which in turn results in increased production of C30 apocarotenoids, leading to red coloration of fruit peel.

**The 5´ SNP2G*-CCD4b* alleles probably evolved after the speciation of mandarin.**

To track the evolution of the 5´ SNP2G*-CCD4b* allele in citrus, we conducted a phylogenetic analysis of the promoter sequences of 38 representative *CCD4b* alleles from diverse citrus accessions and related genera. The alleles from Pummelo, Ichang Papeda, Eremocitrus, Poncirus, Papeda, Microcitrus and Citron, which lack the MITE transposon in the promoter regions (promoter Type I and II type), are distant from the alleles in the mandarin cluster (Figure 7A). Notably, based on this phylogenetic tree, the alleles derived from a primitive yellow peel mandarin species “Mangshan mandarin” (Wang et al., 2018) and other wild mandarins, which have no transposon insertion in the same promoter region (promoter Type II), are grouped in a cluster that is distant from the large cluster containing the semi-domesticated and domesticated mandarins (Figure 7A). Within this large cluster, the *CCD4b* promoter alleles are grouped into three clades: clade I with MITE(AG) (Type V, red peel accessions), clade II with MITE(GG) (Type VI, red peel accessions), clade III with MITE(AA) (Type IV, yellow peel accessions) and clade IV without MITE (No-MITE, type III, yellow peel accessions) (Figure 7A and 7B). Thus, it appears likely that the insertion of the MITE transposon into the 5´ regulatory regions of *CCD4b* (MITE(AG) and MITE(GG) alleles) occurred after the speciation of mandarin. Combining these results, it can be speculated that the MITE(AA) and MITE(GG) alleles might have evolved from a single MITE transposition event in a more primitive pure mandarin species (No-MITE allele) followed by domestication and hybridization of mandarin species. The MITE(AG) allele occurred only in the hybrids derived from crosses between mandarin and other species (Supplemental Table 8), and phylogenetic analysis showed that the cluster of alleles containing MITE(AG) was phylogenetically closer to MITE(AA) cluster than to the MITE (GG) cluster (Figure 7A). These results suggest that both MITE(GG) and MITE(AG) allele (5´ SNP2G-CCD4b allele) probably evolved after the speciation of mandarin and that the MITE(AG) allele might have derived from the MITE(AA) allele through a single nucleotide substitution (Figure 7B).

A phylogenetic analysis of various CCD4 homologs at the protein level showed that CCD4b from citrus (mandarin) is distinct from its counterparts from all other plant species that produce C27 apocarotenoid (Supplemental Figure 12). Furthermore, an alignment of nine *CCD4b* alleles from different primitive, wild and cultivated citrus species showed that *CCD4b* is indeed highly conserved among all citrus species examined (Supplemental Figure 13) and there are 9 highly conserved/identical motifs known to be required for its enzymatic activity (Ma et al., 2013; Rodrigo et al., 2013; Zheng et al., 2015). Therefore, it seems likely that during the course of citrus evolution, a progenitor *CCD4b* genemight have undergone functional diversification in a primitive citrus species, producing an enzyme that is capable of catalyzing the production of colored Citrus-specific C30-apocarotenoids. The transcription of *CCD4b* in a primitive pure mandarin was then enhanced due to a MITE-introduced 5´ regulatory change, resulting in the formation of red-peel mandarins and their hybrids.

**DISCUSSION**

Red colored peel or is an important trait of aesthetic value for citrus fruits, and hence an important breeding objective for improved consumer preference. Identification of genetic determinant controlling red-peel trait would greatly facilitate conventional breeding through marker-assisted selection and/or genetic engineering via CRISPR-based base editing. In this study, we elucidated the genetic/molecular basis and evolutionary origin of red-peel coloration with biochemical insight into the biosynthesis of C30-apocarotenoids especially abundant in the red-peel citrus fruit. The new information should help design novel strategies for improving citrus fruit quality with enhanced aesthetic value.

**An integrated approach for resolving the genetic basis of a complex fruit color trait in a woody plant.**

Fruit color in many cases is determined by the composition and content of various carotenoids and thus considered to be a complex trait. While several genetic determinants of carotenoid formation have been identified in herbaceous plants such as Arabidopsis, tomato and maize (Gonzalez-Jorge et al., 2013; Gonzalez-Jorge et al., 2016; Harjes et al., 2008; Yan et al., 2010; Zhu et al., 2018), little is known about the mechanisms underlying fruit coloration in woody plants due to spatiotemporal limitations. The present study used an integrated approach that combines genetic, transcriptomic, and metabolic analyses across a small segregating population concurrently with a collection of diverse germplasm. Functional characterization through ectopic expression of the candidate gene in callus to identify a major genetic determinant that controls the red coloration of citrus peels. Collectively, these strategies could be effective in the identification of genes controlling other important fruit quality traits in perennial trees.

Firstly, our success can be attributed to the employment of the BSR-seq for the initial mapping of red-peel coloration of mandarin to a genetic interval in Chromosome 8 where *CCD4b* is located (Figure 2A and B). Although a CCD4-like gene Ciclev10030384m.g is also located in this genetic interval, BSR-seq together with Q-RT-PCR results found it showed very low expression in citrus tissues (Supplemental Figure 6B and C). Consistently, the expressed sequence tag (EST) of Ciclev10030384m.g was also absent in the Citrus EST database (http://harvest.ucr.edu/, software HarvEST 1.32, assembly C52) from 141 Citrus cDNA libraries, suggesting the Ciclev10030384m.g gene might be a pseudogene (Rodrigo et al., 2013). Therefore, we selected the first CCD4 gene (Ciclev10028113m.g, CCD4b) rather than CCD4-like (Ciclev10030384m.g) as a candidate and did subsequent analyses. Thanks to the prior knowledge that *CCD4b* encodes an enzyme involved in carotenoid metabolism (Ma et al., 2013; Rodrigo et al., 2013), we were able to infer that enhanced expression of *CCD4b* may contribute to red-peel coloration based on the subsequent eQTL (Figure 2E). Because the cDNA sequence for *CCD4b* is identical between the two citrus parents, one may overlook the MITE-insertion caused DNA polymorphism 5´ to *CCD4b* if DNA instead of RNA is for the bulked sergeant analysis because the population size is too small. Thus, RNA-seq appears to be a better choice for BSA with a small segregating population, because it can reveal differences at level of both the DNA ((polymorphism in exons) and mRNA (transcript levels) between the two pools. Secondly, our ability in rapid and effective validation of results from the initial genetic mapping also played a significant role in this process. In this regard, our possession of a large collection of citrus germplasm made it possible for the verification of the tight association between the presence of the MITE-introduced 5´ regulatory change and red-peel color among a wide variety of mandarins and their hybrids. The subsequent promoter-activity analyses via transient expression in *N. benthamiana* further validated the results from allelic expression analysis (Figure 6A-D and Supplemental Figure 10). Lastly, in order to further establish genotype–phenotype relationship underlying thepeel trait, we use mQTL to investigate the genetic determinant of intermediate phenotypes such as carotenoids and link these results to substantiate fine-mapping result of the trait (Figure 3A). Furthermore, our deployment of the callus cell-based expression system (i.e. the ECMs system) greatly accelerated the final confirmation of the biological function of *CCD4b* with additional genetic and biochemical evidence (Figure 4 and supplemental Figure 8). This is particularly important for gene identification and characterization in perennial trees whose genetic studies are inherently restricted by spatiotemporal constraints such large tree size and long juvenility.

***CCD4b* is responsible for the production of red-peel specific C30 β-citraurinene pigment in citrus.**

We detected two C30-apocarotenoids (β-citraurin and β-citraurinene) that predominantly accumulated in red peel of the F1 hybrids and their content was associated with red-peel coloration (Figure 1). The only difference in chemical structure between β-citraurinene and β-citraurin is that the latter has an aldehyde group in the 8'-position whereas the former has a methyl group (Supplemental Figure 3 and 4). Like many aldehydes, the aldehyde groups of apocarotenoids are highly reactive in plant cells (Demurtas et al., 2018), and might be directly reduced to methyl by non-specific chemical reactions (Gevorgyan et al., 2001). Interestingly, previous studies based on enzymatic activity assays of CCD4bexpressed in prokaryotic cells showed that CCD4b exclusively produced β-citraurin by cleaving zeaxanthin and/or β-cryptoxanthin at the positions 7, 8 or 7´, 8´ *in vitro* (Ma et al., 2013; Rodrigo et al., 2013); By contrast, red peel of mandarin contains both β-citraurin and β-citraurinene and citrus callus overexpressing *CCD4b* (Ox-4b-ECM1 and Ox-4b-ECM2) predominantly accumulated β-citraurinene (Figure 4A-C and Supplemental Figure 8). Hence, we speculate that the enzymatic activities of CCD4b produced by prokaryotic cells may not fully reflect those of the native CCD4b protein produced in citrus callus cells and fruit peel tissues. Alternatively, β-citraurin in citrus fruit peel is converted to β-citraurinene automatically or catalyzed by a yet unidentified enzyme in citrus cells, which does not occur in bacterial cells.

Consistent with the above reasoning, we found that the decrease of β-citraurin was accompanied with an increase of β-citraurinene in citrus peel from 26 to 90 days after postharvest storage (Supplemental Figure 14) and that CCD4b-YFP is localized in the plastoglobuli of plastids (Supplemental Figure 8), which is probably very different from the cellular environment of the prokaryotic cells. A similar phenomenon has also been observed for Saffron CCD2, which cleaved Zeaxanthin to yield crocetin dialdehyde in *E.coli* but predominantly produced crocetin in transgenic rice callus overexpressing it (Ahrazem et al., 2016; Frusciante et al., 2014). It should be pointed out that the citrus callus overexpressing CCD4b had quite lower content of carotenoid/apocarotenoid compared with citrus peel, which could explain why transgenic callus did not turn red. This implies that red coloration may require accumulation of both carotenoids, β-citraurin and β-citraurinene above a threshold level as in the case of peel tissues of red mandarins.

**Stepwise evolution of red peeled mandarins.**

CCD4s from other plant species have been shown to cleave carotenoids at (9, 10) 9’ 10’ to form C27 apocarotenoids (Bruno et al., 2015; Bruno et al., 2016; Huang et al., 2009), they are thought to play a negative role in coloration (Campbell et al., 2010; Falchi et al., 2013; Gonzalez-Jorge et al., 2013; Ohmiya et al., 2006; Zhang et al., 2015). Apparently, CCD4b-induced red-peel coloration is an evolutionary innovation in citrus, which is not only supported by the genetic and molecular evidenced from this study, but also by the fact that the two C30 apocarotenoids β-citraurin β-citraurinene are rarely found in other plants (CURL, 1965; Leuenberger and Stewart, 1976). Based on the results from this study, the evolution of red-peeled mandarins appears to consist of two major steps. As step one, the evolution of the red peel trait in citrus probably began with duplication of a *CCD4* progenitor gene followed by sequence diversification and neofunctionalization of *CCD4b* whose product could cleave carotenoids at (7, 8) 7´ 8´ position to produce colorful β-citraurin and β-citraurinene pigment (Ma et al., 2013; Rodrigo et al., 2013). Indeed, citrusCCD4b is divergent among the five paralogous CCD4 family members in citrus (Zheng et al., 2015) and also among the likely orthologs from 10 plant species (Supplemental Figure 12). How CCD4b achieves the new enzymatic function awaits future investigation.

Perhaps the most interesting finding of this study is that enhanced expression of *CCD4b* caused by a transposon insertion-induced 5´ *cis* regulatory change underscores the red-peel coloration in mandarins and their hybrids. Clearly, the insertion of MITE to the promoter region of *CCD4b* in a primitive mandarin constituted the second step of the red-peel trait evolution. Given that enhanced gene expression resulting from transposon introduced *cis*-elements to the promoter regions of targeted genes has been reported to be associated with color variation in horticultural plants (Butelli et al., 2012; Chiu et al., 2010; Fernandez et al., 2013; Li et al., 2015), the transposon-caused “yellow peel-to-red peel” mutation in citrus is not entirely surprisingly. However, it is remarkable that one SNP in the MITE-introduced sequence (i.e. MITE (AG) vs. MITE (AA)) underlies the difference in *CCD4b* expression and the peel color. Because one primitive type of mandarin, the yellow-peeled ‘Mangshan mandarin’, has been thought to be the progenitor of the domesticated yellow-peeled and red-peeled mandarins in the southern side and northern side of Nanling Mountain, respectively (Wang et al., 2018), it is possible that the original transposon insertion event might have happened in a ‘Mangshan mandarin’ descendant. Furthermore, all the 25 mandarins and their hybrids from the southern Nanling surveyed are yellow-peeled and contain the MITE (AA) and/or the no-MITE alleles (five haplotypes in total), whereas the 48 red-peeled mandarins and their hybrids from northern Nanling had either the MITE (AG) and/or the MITE (GG) allele (thus two haplotypes only) (Supplemental Table 8). Hence, one may speculate that ‘no-MITE to MITE (AA) or MITE (GG)” and then “MITE (AA) to MITE (AG)” mutations might have occurred in ‘Mangshan mandarin’ descendants in the northern Nanling (Figure 7B). Moreover, the presence of more red-peeled mandarins in the northern Nanling is perhaps indicative of strong selection for red-peeled mandarins during the domestication of mandarins in that area. How the SNP2A/G (…CAACTA/GTGG…) between MITE (AA) to MITE (AG) or MITE (GG) specifies the remarkable transcriptional difference of *CCD4b* is currently unclear. It is possible that the DNA sequence in the MITE (AG) or MITE (GG) alleles creates a potential enhancer element (TGTGG) (Manimaran et al., 2015) for an unknown transcription factor, or a new binding site (CAACTG) for MYB /MYC transcription factor known to be responsible for variation of secondary metabolites in other plants (Chiu et al., 2010; Van Moerkercke et al., 2011) in the promoter region of *CCD4b*, thereby enhancing *CCD4b* transcription.

**Materials and Methods**

**Plant Materials**

A total of 117 citrus species from *P. trifoliata,* *C. reticulata*, *C. grandis*, *C. sinensis* and otherrelated hybrid citrus specieswere used for genotype-phenotype association analysis in this study (Supplemental Table 5). For most orange species, the orange-yellow peel will turn red after postharvest storage. Therefore, the red color trait of citrus fruit was determined by the observation in the mature stage as well as after postharvest storage. The HJ x ZK F1 population was obtained by crossing red-peeled Red tangerine (*Citrus reticulata* Blanco，HJ, female parent)with a yellow-peeled trifoliate orange(*Poncirus trifoliata*, ZK, male parent) in 2003 (Tan et al., 2007). Ethylene (25PPM) treatment with fruits of the F1 progeny was performed as previously reported (Rodrigo et al., 2013).

**Pigment Extraction and Analysis**

The n-hexane: acetone: ethanol (2:1:1 [v/v/v]) and two mobile phases were used for carotenoid extraction as previously described (Cao et al., 2012; Zheng et al., 2015). Mobile phase A (Methanol: H2O, 9:1, v/v) and mobile phase B (100 %MTBE containing 0.01 % (w/v) BHT) were used to perform HPLC at a flow rate of 1 ml/min. The elution conditions were as follows: 0-35 min 8-25% B; 35-50min 25-50% B; 50-60 min 50-70% B; 60-70 min 70% B; 70-75 min 8% B. Carotenoid identification was performed by comparison of characteristic spectral properties and typical retention time based on previous literatures (Cao et al., 2012; Zheng et al., 2015). The carotenoid quantification was achieved by comparison with calibration curves of authentic standards from CaroteNature (Lupsingen, Switzerland). For principal component analysis (PCA) of carotenoid was performed by using SIMCA- P 11.5 program (Version 11, Umetrics, Umea, Sweden).

**Chromatographic separation and mass spectrometry identification of apocarotenoids**

The chromatographic separation of individual apocarotenoids was performed by using Dionex UltiMate 3000 HPLC system equipped with a semi-preparative column (COSMOSIL，5C18-MS-Ⅱ, 10mm id. 250mm) and Ultimate RS variable wavelength detector. Mobile phase A (Methanol: H2O, 9:1, v/v) and Mobile Phase B (100 % MTBE) were used as eluent at a flow rate of 3 ml/min. The elution conditions for isolation were as follows: 0-20 min 8-18% B；20-23 min 18-50% B；23-25 min 50-60% B；

The mass spectrometry identification was performed by using a HPLC-DAD-APCI-MS/MS system (Agilent1100, Palo Alto, CA, USA) consisting of an ion-trap mass spectrometer and an APCI ionization source (Esquire 4000, Bruker Daltonics, Bremen, Germany) operated in both negative and positive modes. Nebulizer pressure was set at 60 psi, dry gas (nitrogen) flow was set at 5 L/h and source temperature was 320°C; APCI source settings were set as follows: corona discharge voltage, 4000 nA; HV capillary, 3500 V.

**Measurement of Fruit Color Index**

The color parameters L (black to white), a (green to red) and b (blue to yellow) were measured by a color analyzer (KONICA MINOLTA CR-400, Japan) with a minimum of 9 fruits. The value of hue angle (H0, arctangent (b/a)) were used to calculate the color index of fruits as previously reported (Luo et al., 2015).

**QTL Analysis**

A SNP-based high-density genetic map of HJ x ZK F1 population was constructed by specific length amplified fragment (SLAF) sequencing (Zheng et al., 2019). Metabolite and expression Quantitative Trait Loci (mQTL and eQTL) detection in this F1 segregating population was carried out using MapQTL® version 6.0 software with Interval mapping (IM) algorithm (Ooijen and Kyazma, 2009). 1,000-permutation test (PT) were used to calculate threshold values of logarithm of odds (LOD) for declaring QTL at the confidence level of 95%. The final QTL confidence intervals were displayed by using 1 LOD unit falls as previously reported (Graham, 2010).

**Bulked segregant RNA-Seq (BSR-seq) analysis**

An Illunima HiSeq 2500 platform (Illumina, Inc; San Diego, CA, U.S.) was used to do RNA-seq experiment with 10GB sequencing data for each pool peel sample from 22 red-peel and 22 yellow-peel individuals respectively. Cufflinks software was used to identify the differentially expressed genes (DEG) between the two extreme pools (Trapnell et al., 2012). The DEG with FPKM values greater than 100 in at least one extreme pool and Log2 fold change (LFC) value of FPKM higher than 2 were selected. The Heml 1.0 software was used for heatmap displaying and clustering analysis of the selected DEG data. Reads from the bulk RNA-Seq were used to calculate values of Δ (SNP index) (Takagi et al., 2013) and 95% and 99% confidence intervals (CIs) of Δ (SNP index) for BSA analysis as previously reported (Tang et al., 2017).

**CAPs Markers Development and Fine Mapping**

In order to convert the SNPs into the cleaved amplified polymorphic sequence (CAPs) makers, the SAMtools were used to identify the restrict enzyme sites in the sequence containing the SNPs (Tang et al., 2017). For the preliminary mapping, six CAPs makers developed from BSR-seq data were used to confirm the genetic region identified by BSA. Then, the PCR-based sequencing SNP makers developed from BSR-seq and SLAF-sequencing (Zheng et al., 2019) were used to narrow the locus by screening recombinants genotype.

**Sequence analyses**

For sequence analysis, the primers 4b-P-F&4b-p-R and 4b-F&4b-R were used to amplified the promoters and coding sequences of *CCD4b* from different citrus species (Supplemental Table 9). ClustalX (version 1.8) and Genedoc software were used to perform sequences alignments and editing respectively. The phylogenetic analysis was conducted in MEGA5 by using the maximum likelihood method and neighbor-joining algorithm as previously described (Tamura et al., 2011; Zheng et al., 2015).

**Quantitative Real-time PCR**

Extraction of total RNA was performed by using RN38-EASY RNA extraction kit (Aidlab Biotechnology, Beijing, PRC) following the manufacturer’s instructions. The total RNA was treated with DNaseI and the first-strand cDNA was synthesized with a HiScript II 1st Strand cDNA Synthesis Kit (Vazyme Biotech, Nanjing, PRC). Quantitative Real-time PCR was performed by using an ABI 7500 Real Time System (Applied Biosystems) as previously reported (Wang et al., 2017; Zheng et al., 2015). The E −ΔΔCt method was used for the calculation of relative expression.

**Callus Transformation**

The CDS of *CsCCD4b* was cloned into the pH7GW2D and pB2GW7 gateway vector respectively. The primers used in the gateway reaction were listed in Supplemental Table 9. Agrobacterium tumefaciens strain EHA105 containing pH7GW2D-*CsCCD4b* and pB2GW7-*CsCCD4b* were transformed into one stable transgenic callus line of ECM1 and ECM2 respectively (Cao et al., 2012). To avoid the occurrence of chimeric transgenic callus lines, a minimum 10 cycles of subculture was performed onto culture media containing appropriate antibiotics for selection of the transgene. Each independent transgenic line was cultured in dark at room temperature.

**Subcellular Localization**

The CDS of *CsCCD4b* was cloned into the p2YGW7 vector in frame with YFP; the vector p2CGW7 carrying Arabidopsis *FBN1a* in frame with CFP was used as a marker for plastoglobules colocalization (Vidi et al., 2006). Arabidopsis mesophyll protoplast were isolated by using the method “Tape-Arabidopsis Sandwich” as previously described (Wu et al., 2009) and the protoplast transfection assay was immediately performed according to a PEG-mediated method (Yoo et al., 2007). The fluorescence in protoplast was observed with a Leica TCS SP8 (Leica Microsystems) confocal laser scanning microscope.

**Western Blot**

The *CsCCD4b* coding sequence isolated from *C. reticulata* was cloned in the pET28a vector and introduced in *Escherichia coli* strain BL21 (DE3). The expressed recombinant proteins were purified and administered to rabbits to develop anti-CsCCD4b polyclonal antisera commissioned from Frdbio company (Wuhan, P.R. China). Total protein extraction and Western Blot analysis were performed according to a previous protocol (Cao et al., 2012).

**Microscopy Observation and Analysis**

To visualize the plastoglobules in the amyloplasts-abundant callus, we applied sucrose starvation treatment in the transgenic callus before Transmission Electron Microscopy (TEM) imaging analysis. The Sample preparation, staining and microscopy observation for TEM analysis were performed as described previously (Cao et al., 2012), and the size and area of plastoglobules were measured by ImageJ software (National Institutes of Health (NIH), USA).

**Promoter activity analysis**

Three fragments with MITE (AA), MITE (GG), MITE (AG) and no-MITE allele sequences were used to compare the promoter activity. Because the full-length MITE transposon sequence in the promoter resulted in very low promoter activity in tobacco leave, we use the fragments with the deletion of 95bp in the MITE sequence to assess the impact of the presence of SNP2G in the MITE on promoter activity. All these fragments were inserted into the corresponding cloning sites of pCambia-1301 vector to produce promoter-GUS vector, respectively. The GUS assays were performed according to a previous protocol (Lu et al., 2016; Lu et al., 2018). The GUS enzyme activity expressed as μmoles MU/min.mg protein was used to measure the promoter activity. Three biological replicates of the GUS assays were performed for each sample.

**High-resolution melting (HRM) analysis**

Primers used for HRM analysis were listed in Supplemental Table 9. The HRM detection was performed in a 384-well formatted program of gene scanning in the Roche (Indianapolis, IN, USA) LightCycler® 480 system with the LightCycler® 480 High Resolution Melting Master mix (Roche) by following the manufacturer’s instructions. The PCR products amplified from validated plasmid with known genotypes were used as standard genotypes references. The melting curve data were analyzed by using the LightCycler® 480 Software 1.5.

**Statistical Analysis**

For statistical analysis, mean ± standard errors (SE) was used to present data at least in triplicate. One-way ANOVA and Student’s t-tests were used to calculate significant differences between transgenic and wild type samples. Pearson correlation (r) analysis was performed by using ANOVA in the Microsoft Office 2010.

**Accession Numbers**

The citrus *CCD4b* promoter sequences used in the phylogenetic analysis have been deposited in the GenBank database (accession nos. MH705289-MH705326). The BSR-seq sequencing data have been deposited at Sequence Read Archive database in NCBI (SRA accession nos. SRR7621313-SRR7621316).

**FUNDING**

This research was financially supported by the National Natural Science Foundation of China (no. 31330066, 31521092 and 31630065), China Agriculture Research System (CARS-27), the 111 project (B13034) and China council scholarship (to X.Z.).

**AUTHOR CONTRIBUTIONS**

X.D. conceived the project and supervised the experiments; X.D. X.Z. and Z.P. designed the experiments; X.Z. performed most of the experiments and analyzed the data with contributions from K.Z. Q.S. W.Z. X.W.; H.C. M.T. and Z.X. provided vital materials; X.Z. Z.P. S.X. X.D. and P.D.F. wrote the article; Z.P. S.X. P.D.F. Y.Z. Q.X., L.J.C., and J.L.Y. provided critical comments in experiment technology and writing improvement.

**ACKNOWLEDGMENTS**

We thank Dr. Hanhui Kuang, Dr. Chunying Kang, Dr. Elliott J. Price and Dr. Rui Guan for critical reading of this manuscript and suggestions. The authors declare no conflict of interest.

**REFERENCES**

**Ahrazem, O., Rubio‐Moraga, A., Berman, J., Capell, T., Christou, P., Zhu, C., and Gómez‐Gómez, L.** (2016). The carotenoid cleavage dioxygenase CCD2 catalysing the synthesis of crocetin in spring crocuses and saffron is a plastidial enzyme. *New Phytol* 209:650-663.

**Alder, A., Jamil, M., Marzorati, M., Bruno, M., Vermathen, M., Bigler, P., Ghisla, S., Bouwmeester, H., Beyer, P., and Al-Babili, S.** (2012). The path from β-carotene to carlactone, a strigolactone-like plant hormone. *Science* 335:1348-1351.

**Alquezar, B., Rodrigo, M.J., and Zacarías, L.** (2008). Regulation of carotenoid biosynthesis during fruit maturation in the red-fleshed orange mutant Cara Cara. *Phytochemistry* 69:1997-2007.

**Alquézar, B., Rodrigo, M.J., and Zacarías, L.** (2008). Carotenoid biosynthesis and their regulation in citrus fruits. *Tree for sci biotech.* 2:23-37.

**Bouvier, F., Dogbo, O., and Camara, B.** (2003). Biosynthesis of the food and cosmetic plant pigment bixin (annatto). *Science* 300:2089-2091.

**Bruno, M., Beyer, P., and Al-Babili, S.** (2015) "The potato carotenoid cleavage dioxygenase 4 catalyzes a single cleavage of β-ionone ring-containing carotenes and non-epoxidated xanthophylls." *Arch Biochem Biophys* 572: 126-133.

**Bruno, M., Koschmieder, J., Wuest, F., Schaub, P., Fehling-Kaschek, M., Timmer, J., Beyer, P., and Al-Babili, S.** (2016). "Enzymatic study on AtCCD4 and AtCCD7 and their potential to form acyclic regulatory metabolites." *J Exp Bot* 67: 5993-6005.

**Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., and Martin, C.** (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24:1242-1255.

**Campbell, R., Ducreux, L.J., Morris, W.L., Morris, J.A., Suttle, J.C., Ramsay, G., Bryan, G.J., Hedley, P.E., and Taylor, M.A.** (2010). The metabolic and developmental roles of carotenoid cleavage dioxygenase4 from potato. *Plant Physiol* 154:656-664.

**Cao, H., Zhang, J., Xu, J., Ye, J., Yun, Z., Xu, Q., Xu, J., and Deng, X.** (2012). Comprehending crystalline β-carotene accumulation by comparing engineered cell models and the natural carotenoid-rich system of citrus. *J Exp Bot* 63:4403-4417.

**Carmona, L., Zacarías, L., and Rodrigo, M.J.** (2012). Stimulation of coloration and carotenoid biosynthesis during postharvest storage of ‘Navelina’orange fruit at 12 C. *Postharvest Biol Technol* 74:108-117.

**Chen, L., Chen, K., Li, J., and Hu, Y.** (1993). The genetic characterization of flavedo color in citrus. *Acta Horticulturae Sinica* 20:221-224.

**Chiu, L.W., Zhou, X., Burke, S., Wu, X., Prior, R.L., and Li, L.** (2010). The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiol* 154:1470-1480.

**CURL, A.L.** (1965). The Occurrence of beta‐Citraurin and of beta‐Apo‐8′‐Carotenal in the Peels of California Tangerines and Oranges. *J Food SCI* 30:13-18.

**Demurtas, O.C., Frusciante, S., Ferrante, P., Diretto, G., Azad, N.H., Pietrella, M., Aprea, G., Taddei, A.R., Romano, E., and Mi, J., et al.** (2018). Candidate enzymes for saffron crocin biosynthesis are localized in multiple cellular compartments. *Plant Physiol* 177: 990-1006.

**Falchi, R., Vendramin, E., Zanon, L., Scalabrin, S., Cipriani, G., Verde, I., Vizzotto, G., and Morgante, M.** (2013). Three distinct mutational mechanisms acting on a single gene underpin the origin of yellow flesh in peach. *Plant J* 76:175-187.

**Fanciullino, A.L., Dhuique-Mayer, C., Luro, F., Casanova, J., Morillon, R., and Ollitrault, P.** (2006). Carotenoid diversity in cultivated citrus is highly influenced by genetic factors. *J Agric Food Chem* 54:4397-4406.

**Farin, D., Ikan, R., and Gross, J.** (1983). The carotenoid pigments in the juice and flavedo of a mandarin hybrid (Citrus reticulata) cv Michal during ripening. *Phytochemistry* 22:403-408.

**Fernandez, L., Chaib, J., Martinez-Zapater, J.M., Thomas, M.R., and Torregrosa, L.** (2013). Mis-expression of a PISTILLATA-like MADS box gene prevents fruit development in grapevine. *Plant J* 73:918-928.

**Fraser, P.D., and Bramley, P.M.** (2004). The biosynthesis and nutritional uses of carotenoids. *Prog Lipid Res* 43:228-265.

**Frusciante, S., Diretto, G., Bruno, M., Ferrante, P., Pietrella, M., Prado-Cabrero, A., Rubio-Moraga, A., Beyer, P., Gomez-Gomez, L., Al-Babili, S., et al.,** (2014). Novel carotenoid cleavage dioxygenase catalyzes the first dedicated step in saffron crocin biosynthesis. *Proc Natl Acad Sci USA* 111:12246-12251.

**Gutheil, WG., Reed, G., Ray, A., Anant, S., and Dhar, A.** (2012). Crocetin: an agent derived from saffron for prevention and therapy for cancer. *Curr Pharm Biotechnol* 13:173-179.

**Gevorgyan, V., Rubin, M., Liu, J.-X., and Yamamoto, Y.** (2001). A direct reduction of aliphatic aldehyde, acyl chloride, ester, and carboxylic functions into a methyl group. *J Org Chem* 66:1672-1675.

**Gonzalez-Jorge, S., Ha, S.H., Magallanes-Lundback, M., Gilliland, L.U., Zhou, A., Lipka, A.E., Nguyen, Y.N., Angelovici, R., Lin, H., Cepela, J., et al.** (2013). Carotenoid cleavage dioxygenase4 is a negative regulator of beta-carotene content in Arabidopsis seeds. *Plant Cell* 25:4812-4826.

**Gonzalez-Jorge, S., Mehrshahi, P., Magallanes-Lundback, M., Lipka, A.E., Angelovici, R., Gore, M.A., and DellaPenna, D.** (2016). ZEAXANTHIN EPOXIDASE activity potentiates carotenoid degradation in maturing Arabidopsis seed. *Plant Physiol* 171: 1837-1851.

**Graham, I.A., Besser, K., Blumer, S., Branigan, C. A., Czechowski, T., Elias, L., ... & Larson, T. R.** (2010). The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science* 327:328-331.

**Harjes, C.E., Rocheford, T.R., Bai, L., Brutnell, T.P., Kandianis, C.B., Sowinski, S.G., Stapleton, A.E., Vallabhaneni, R., Williams, M., and Wurtzel, E.T., et al.** (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 319:330-333.

**Huang, F.C., Molnar, P., and Schwab, W.** (2009). Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. *J Exp Bot* 60:3011-3022.

**Ikoma, Y., Matsumoto, H., and Kato, M.** (2016). Diversity in the carotenoid profiles and the expression of genes related to carotenoid accumulation among citrus genotypes. *Breed Sci* 66:139-147.

**Kato, M., Ikoma, Y., Matsumoto, H., Sugiura, M., Hyodo, H., and Yano, M.** (2004). Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit. *Plant Physiol* 134:824-837.

**Krinsky, N.I., and Johnson, E.J.** (2005). Carotenoid actions and their relation to health and disease. *Mol Aspects Med* 26:459-516.

**Leuenberger, U., and Stewart, I.** (1976). β-Citraurinene, a new C30-citrus carotenoid. *Phytochemistry* 15:227-229.

**Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R.J., Ma, Z., Shang, H., Ma, X., Wu, J., et al.** (2015). Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol.* 33:524-530.

**Lu, S., Zhang, Y., Zheng, X., Zhu, K., Xu, Q., and Deng, X.** (2016). Isolation and functional characterization of a lycopene β-cyclase gene promoter from citrus. *Front Plant Sci* 7:1367.

**Lu, S., Zhang, Y., Zhu, K., Yang, W., Ye, J., Chai, L., Xu, Q., and Deng, X.** (2018). The Citrus Transcription Factor CsMADS6 Modulates Carotenoid Metabolism by Directly Regulating Carotenogenic Genes. *Plant Physiol* 176:2657-2676.

**Luo, J.** (2015). Metabolite-based genome-wide association studies in plants. *Curr Opin Plant Biol.* 24:31-38.

**Luo, T., Xu, K., Luo, Y., Chen, J., Sheng, L., Wang, J., Han, J., Zeng, Y., Xu, J., Chen, J., et al.** (2015). Distinct Carotenoid and Flavonoid Accumulation in a Spontaneous Mutant of Ponkan (Citrus reticulata Blanco) Results in Yellowish Fruit and Enhanced Postharvest Resistance. *J Agric Food Chem* 63:8601-8614.

**Ma, G., Zhang, L., Matsuta, A., Matsutani, K., Yamawaki, K., Yahata, M., Wahyudi, A., Motohashi, R., and Kato, M.** (2013). Enzymatic formation of beta-citraurin from beta-cryptoxanthin and Zeaxanthin by carotenoid cleavage dioxygenase4 in the flavedo of citrus fruit. *Plant Physiol* 163:682-695.

**Ma, G., Zhang, L., Iida, K., Madono, Y., Yungyuen, W., Yahata, M., Yahata, M., Yamawaki, K., and Kato, M.** (2017). Identification and quantitative analysis of β-cryptoxanthin and β-citraurin esters in Satsuma mandarin fruit during the ripening process. Food chemistry 234: 356-364.

**Manimaran, P., Raghurami Reddy, M., Bhaskar Rao, T., Mangrauthia, S.K., Sundaram, R.M., and Balachandran, S.M.** (2015). Identification of cis-elements and evaluation of upstream regulatory region of a rice anther-specific gene, OSIPP3, conferring pollen-specific expression in *Oryza sativa* (L.) ssp. indica. *Plant Reprod* 28:133-142.

**Matsumoto, C., Ashida, N., Yokoyama, S., Tominari, T., Hirata, M., Ogawa, K., Sugiura M., et al.** (2013). The protective effects of β-cryptoxanthin on inflammatory bone resorption in a mouse experimental model of periodontitis. *Biosci Biotechnol Biochem.* 77:860-862.

**Noviendri, D., Hasrini, R.F., and Octavianti, F.** (2011). Carotenoids: Sources, medicinal properties and their application in food and nutraceutical industry. *J Med Plant Res* 5:7119-7131.

**Ohmiya, A., Kishimoto, S., Aida, R., Yoshioka, S., and Sumitomo, K.** (2006). Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. *Plant Physiol* 142:1193-1201.

**Ooijen, J., and Kyazma, B.** (2009). MapQTL 6.0, software for the mapping of quantitative trait loci in experimental populations of dihaploid species. Netherlands: Wageningen.

**Rodrigo, M.J., Alquezar, B., Alos, E., Medina, V., Carmona, L., Bruno, M., Al-Babili, S., and Zacarias, L.** (2013). A novel carotenoid cleavage activity involved in the biosynthesis of citrus fruit-specific apocarotenoid pigments. *J Exp Bot* 64:4461-4478.

**Rottet, S., Devillers, J., Glauser, G., Douet, V., Besagni, C., and Kessler, F.** (2016). Identification of plastoglobules as a site of carotenoid cleavage. *Front Plant Sci* 7:1855.

**Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., and Takuno, S., et al.** (2013). QTL‐seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J* 74:174-183.

**Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S.** (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731-2739.

**Tan, B.-C., Joseph, L.M., Deng, W.-T., Liu, L., Li, Q.-B., Cline, K., and McCarty, D.R.** (2003). Molecular characterization of theArabidopsis9-cisepoxycarotenoid dioxygenase gene family. *Plant J* 35:44-56.

**Tan, M., Song, J., and Deng, X.** (2007). Production of two mandarin × trifoliate orange hybrid populations via embryo rescue with verification by SSR analysis. *Euphytica* 157:155-160.

**Tang, Q., Tian, M., An, G., Zhang, W., Chen, J., and Yan, C.** (2017). Rapid identification of the purple stem (Ps) gene of Chinese kale (*Brassica oleracea* var. *alboglabra*) in a segregation distortion population by bulked segregant analysis and RNA sequencing. *Mol Breed.* 37:153.

**Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L.** (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562.

**Van Moerkercke, A., Haring, M.A., and Schuurink, R.C.** (2011). The transcription factor EMISSION OF BENZENOIDS II activates the MYB ODORANT1 promoter at a MYB binding site specific for fragrant petunias. *Plant J* 67:917-928.

**Vidi, P.-A., Kanwischer, M., Baginsky, S., Austin, J.R., Csucs, G., Dörmann, P., Kessler, F., and Bréhélin, C.** (2006). Tocopherol cyclase (VTE1) localization and vitamin E accumulation in chloroplast plastoglobule lipoprotein particles. *J Biol Chem* 281:11225-11234.

**Wang, L., He, F., Huang, Y., He, J., Yang, S., Zeng, J., Deng, C., Jiang, X., Fang, Y., and Wen, S., et al**. (2018). Genome of wild mandarin and domestication history of mandarin. *Mol Plant* 11: 1024-1037.

**Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., Wu, G., Tian, S., Chen, C., Liu, Y., et al.** (2017). Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat Genet* 49:765-772.

**Wu, F.-H., Shen, S.-C., Lee, L.-Y., Lee, S.-H., Chan, M.-T., and Lin, C.-S.** (2009). Tape-Arabidopsis Sandwich-a simpler Arabidopsis protoplast isolation method. *Plant methods* 5:16.

**Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., Terol, J., et al.** (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat Biotechnol* 32:656-662.

**Yan, J., Kandianis, C.B., Harjes, C.E., Bai, L., Kim, E.-H., Yang, X., Skinner, D.J., Fu, Z., Mitchell, S., and Li, Q., et al.** (2010). Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain. *Nat genet* 42:322.

**Yoo, S.-D., Cho, Y.-H., and Sheen, J.** (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat protoc* 2:1565.

**Yuan, H., Zhang, J., Nageswaran, D., and Li, L.** (2015). Carotenoid metabolism and regulation in horticultural crops. *Hortic Res* 2:15036.

**Zhang, B., Liu, C., Wang, Y., Yao, X., Wang, F., Wu, J., King, G.J., and Liu, K.** (2015). Disruption of a CAROTENOID CLEAVAGE DIOXYGENASE 4 gene converts flower colour from white to yellow in Brassica species. *New Phytol* 206:1513-1526.

**Zheng, X., Xie, Z., Zhu, K., Xu, Q., Deng, X., and Pan, Z.** (2015). Isolation and characterization of carotenoid cleavage dioxygenase 4 genes from different citrus species. *Mol Genet Genomics* 290:1589-1603.

**Zheng, X., Tang, Y., Ye, J., Pan, Z., Tan, M., Xie, Z., Chai, L., Xu, Q., Fraser, P.D., and Deng, X.** (2019). SLAF-based construction of a high-density genetic map and its application in QTL mapping of carotenoids content in citrus fruit. *J Agric Food Chem* 67(3): 994-1002.

**Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., Lin, T., Qin, M., Peng, M., Yang, C., et al.** (2018). Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell* 172:249-261 e212.

**Figure 1. Phenotypic distribution of citrus peel color and association between carotenoids/apocarotenoids and red peel trait in the HJ x ZK F1 population.**

**(A)** The color segregation of the outside peel in the HJ x ZK F1 pseudo-testcross population derived from the cross between a red-peel tangerine (HJ, *Citrus reticulata*) and a yellow-peel trifoliate orange (ZK, *Poncirus trifoliata*). **(B)** The discontinuous distribution of color index of citrus peel among the F1 individuals in year 2015. The red and yellow column respectively indicates red- and yellow-peel F1 individuals. **(C)** The HPLC PDA profile of carotenoid from the peel of four representative individuals from the HJ x ZK F1 population. The peaks indicated by the red arrow at 28min and 43min represent β-citraurin and β-citraurinene, respectively. The UV/VIS spectra and Mass spectrometry information of β-citraurin and β-citraurinene were displayed on the right. **(D)** The heatmap of carotenoid profiles among the F1 individuals in 2015, the numbers with red and yellow color displayed on the left respectively indicate the red-peel and yellow-peel F1 individuals. Neo: Neoxanthin; Luteo: Luteoxanthin; Vio: Vioxanthin; β-cit: β-citraurin; β-C.e: β-citraurinene; Phy: Phytoene; Phytof: Phytofluene; Lut: lutein; Zea: Zeaxanthin; β-cry: β-cryptoxanthin; β-car: β-carotene.

**Fig. 2 BSR-seq analysis and positional cloning of the red peel trait using the HJ x ZK** **F1 population****.**

**(A)** BSA result of two extreme pools (red/yellow peel) was displayed by average values of SNP-index (Δ). The X-axis indicates nine chromosomes according to the clementine reference genome, and the Y- axis represents the average values of Δ. **(B)** Genetic region for citrus red-peel phenotype was identified by fine mapping using recombinants. The genome interval for candidate region is based on Clementina (*Citrus clementina*) genome. **(C)** The expression pattern of *CCD4b* in the red/yellow peel individuals from the HJ x ZK F1 population. **(D)** Schematic of the *CCD4b* 5´ *cis*-regulatory region sequences in HJ (red peel, female parent) and ZK (yellow peel, male parent). A 158-bp MITE transposon (1) and a 201-bp MITE like transposon (2) were detected in the promoter of *CCD4b* in HJ. A107-bp MITE-like (3) insertion was detected in the promoter of one *CCD4b* allele in ZK. The two SNPs (G and G) in MITE 1 showed co-segregation with the red-peel phenotype. **(E)** A *cis*-acting eQTL for the *CCD4b* expression was located in Chr8: 16946945~18572910.

**Figure 3.** **The *CCD4b* gene is located in a mQTL hotspot associated with variations in C40 carotenoids and C30 apocarotenoids.**

**(A)** An overlapping QTL detected for β-citraurin, β-citraurinene, Zeaxanthin, Lutein, luteoxanthin, phytoene, phytofluene and C30/Total in chromosome 8. **(B)** and **(C)** The contents of β-citraurin and β-citraurinene in different tissues and peels of citrus fruit during development and ripening. Peel-7, 8, 9, 10, 11, 12, represent the peel from one red-peel progeny of HZ x ZK F1 population collected at 60, 90, 120, 150, 180, and 210 days after flower, respectively. **(D)** Spatial-temporal expression levels of *CCD4b*.

**Figure 4. Functional analyses of the citrus *CCD4b* gene using multigene Engineered Cell Models (ECMs).**

The ECM1 indicates the stable transgenic ECMs with overexpression of CrtB fused with Pea rbcS transit peptide in citrus callus. The Ox-4b-ECM1 represents the ECMs with overexpression of the citrus *CCD4b* gene in ECM1 background. **(A)** The expression level of *CCD4b* in ECM1 and different transgenic lines of Ox-4b-ECM1. **(B)** The ratio of C30 apocarotenoids to total carotenoids (C30/Total), α and β-carotene to total carotenoids (α and β-car/Total) and α-xanthophylls contents to total carotenoids in ECM1 and Ox-4b-ECM1 lines. **(C)** The content of individual carotenoids in different ECM1 and Ox-4b-ECM1 lines. The abbreviation of individual carotenoids was described in notes of Figure 1. **(D)** Ultrastructure of ECM1 and Ox-4b-ECM1 transgenic callus cell. To make the plastid structure more easily visible, the transgenic callus was cultured in sucrose-free medium for 5 days to decrease the starch granule. “ch” indicates chromoplast; “p” represents plastoglobuli; “c” indicates carotenoid crystal. **(E)** The number of plastoglobules and the plastoglobules area per plastid in the ECM1 and line9 of Ox-4b-ECM1 transgenic callus.

**Figure 5. The levels of C30 apocarotenoids and *CCD4b* expression are associated with the *CCD4b* promoter genotype.**

**(A)** and **(B)** The contents of C30 apocarotenoids (β-citraurin and β-citraurin) in 95 citrus accessions. The MITE\_GG and MITE\_AG genotype contain MITE(GG) and MITE(AG) allele with SNP2G (CAACTG) in the MITE transposon respectively; MITE\_AA genotype contain the MITE(AA) allele with SNP2A (CAACTG) in the MITE. The alleles of the No-MITE genotypes have no MITE insertion. The detail of MITE\_GG, MITE\_AA, MITE\_AG and No-MITE genotypes were provided in Supplemental Table 5 and Table7. **(C)** The *CCD4b* expression levels in different *CCD4b* genotypes of 95 citrus accessions.

**Figure 6. Allelic difference in citrus *CCD4b* gene expression**

**(A)** The HRM result of the standard reference of two *CCD4b* alleles. **(B)** The HRM analysis revealed the allelic difference for *CCD4b* expression. Both DNA and cDNA samples from different citrus species with the MITE\_GG and MITE\_AG genotypes were used for HRM analysis. These selected genotypes contain two alleles, one has SNP2G (CAACTG) in the MITE (i.e. MITE(GG) or MITE(AG)), the other allele has no MITE insertion (i.e. no-MITE). The PCR fragments of the genomic DNA from the C/T genotype were represented by red curves of HRM. The fragments amplified by the same primers from cDNA of the C/C genotype were indicated by blue curves of HRM. The gray curves indicate the T/T genotype from the standard reference. **(C)** DNA sequences obtained from Sanger sequencing of the fragments amplified from genomic DNA and cDNA respectively. Sequence chromatograms indicate the C/T genotype of fragment from DNA, however the only C/C genotype was detectable from the cDNA sample . X14 indicates 14 independent clones from the fragment amplified form cDNA were selected for sequencing; X4 and X5 respectively indicates four and five independent clones selected for sequencing that belong to the C/C genotype and the C/T genotype, respectively. **(D)** *CCD4b* expression upon ethylene treatment in the peel tissues of the red-peeled and yellow-peeled individuals of the HJ x ZK F1 population.

**Figure 7. Proposed evolutionary history for the *CCD4b* promoters in Citrus**

**(A)** A phylogenetic tree constructed by using the *CCD4b* promoter sequences from different citrus and related genera species. The accession numbers of the *CCD4b* promoter sequences were described in Materials and Method. **(B)** An evolutionary model proposing how MITE insertion and its polymorphism emerged in *CCD4b* locus during the speciation of citrus species. “I”, type I *CCD4b* promoter sequence from ichang papeda, which contains a 1225 bp deletion compared to other types. “II”, type II *CCD4b* promoter sequences from pummelo, eremocitrus, poncirus, papeda, microcitrus and citron, which lack the MITE transposon (red box) and Transposon-like insertion (yellow box). “III”, “IV”, “V”, and “VI” denote *CCD4b* promoter sequences from different mandarin species with red or yellow color peels.