Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2025) **23**, pp. 2182–2193 doi: 10.1111/pbi.70034

An intronic SNP in the *Carotenoid Cleavage Dioxygenase 1* (*CsCCD1*) controls yellow flesh formation in cucumber fruit (*Cucumis sativus* L.)

Zhuonan Dai[1,†], Jiantao Guan[1,†], Han Miao[1,†], Diane M. Beckles[2], Xiaoping Liu[1], Xingfang Gu[1], Shaoyun Dong[1,*] and Shengping Zhang[1,*]

[1]*State Key Laboratory of Vegetable Biobreeding, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China*
[2]*Department of Plant Sciences, University of California, Davis, CA, USA*

Received 9 October 2024;
revised 9 January 2025;
accepted 4 February 2025.
**Correspondence* (Tel 86-010-82105952; fax 86-010-82105952; email zhangshengping@caas.cn (S.Z.); Tel 86-10-82106745; fax 86-10-82106745; email dongshaoyun@caas.cn (S.D.))
[†]These authors contributed equally to this work.

Keywords: cucumber (*Cucumis sativus* L.), yellow flesh, intronic SNP, splicing, carotenoid cleavage dioxygenase 1 (CCD1).

Summary

Vitamin A is a crucial yet scarce vitamin essential for maintaining normal metabolism and bodily functions in humans and can only be obtained from food. Carotenoids represent a diverse group of functional pigments that act as precursors for vitamins, hormones, aroma volatiles and antioxidants. As a vital vegetable in the world, elevated carotenoid levels in cucumber fruit produce yellow flesh, enhancing both visual appeal and nutritional value. However, the genetic mechanisms and regulatory networks governing yellow flesh in cucumbers remain inadequately characterized. In this study, we employed map-based cloning to identify a *Carotenoid Cleavage Dioxygenase 1* (*CsCCD1*) as a key genetic factor influencing yellow flesh in cucumbers. A causal single nucleotide polymorphism (SNP) in the eighth intron of *CsCCD1* led to aberrant splicing, resulting in a truncated transcript. The truncated protein has significantly decreased enzyme activity and increased carotenoid accumulation in the fruit. CRISPR/Cas9-generated *CsCCD1* knockout mutants exhibited yellow flesh and significantly higher carotenoid content compared to wild-type cucumbers. Metabolic profiling indicated a marked accumulation of β-cryptoxanthin in the flesh of these knockout mutants. The intronic SNP was shown to perfectly segregate with yellow flesh in 159 diverse cucumber germplasms, particularly within the semi-wild ecotype Xishuangbanna, known for its substantial carotenoid accumulation. Furthermore, transient overexpression of *CsCCD1* in yellow-fleshed Xishuangbanna cucumbers restored a white flesh phenotype, underscoring the critical role of *CsCCD1* in determining flesh colour in both cultivated and semi-wild cucumbers. These findings lay a theoretical foundation for breeding high-nutrient yellow-fleshed cucumber varieties.

Introduction

Carotenoids are essential for optimal human health and are exclusively obtainable through dietary sources. β-carotene is the primary carotenoid and precursor to vitamin A (Nisar *et al*., 2015), and lack of Vitamin A consumption can result in xerophthalmia, increased morbidity and mortality among infants and reduced immune response (Underwood, 2004). Enhancing the carotene content in crop varieties via breeding is an important way to improve human health. Cucumber (*Cucumis sativus* L.) is an important vegetable crop used widely as a fresh food, but it may also be processed and used in various culinary preparations (Ugwu and Suru, 2021). Flesh colour is an important commodity characteristic of cucumber fruit, and although white is the dominant flesh colour, fruit with green, yellow and orange endocarp and mesocarps can be found (Che and Zhang, 2019; Gebretsadik *et al*., 2021). The accumulation of carotenoids (Bo *et al*., 2012) underlie the yellow- (Kooistra, 1971; Lu *et al*., 2015) and orange- (Simon and Navazio, 1997) fleshed fruit, but their concentration is negligible in the white-fleshed fruit. Breeding new yellow-fleshed varieties could potentially enhance the nutritional quality of cucumbers, increasing the dietary source of Vitamin A, with a significant implications for global diets.

Few studies on flesh colour have been reported in cultivated cucumber, and most have been in the semi-wild ecotype Xishuangbanna (XIS; *Cucumis sativus* L. *var. xishuangbannanesis*). Qi *et al*. (1983) first described that the mature fruit flesh colour of XIS cucumber was orange. Genetic analysis of this ecotype showed that the orange mesocarp was controlled by two recessive genes, while the orange endocarp was controlled by one (Cuevas *et al*., 2010). Bo *et al*. (2012) later found that the orange flesh of the XIS cucumber genotype was due to the accumulation of high levels of β-carotene, and mapped the orange endocarp gene *ore* to chromosome 3. Through GWAS analysis, Qi *et al*. (2013) found that *CsaBCH1* in XIS cucumber can control the accumulation of β-carotene in mature fruits, resulting in orange flesh. Kishor *et al*. (2021) found that *CsOr* underscores β-carotene content in orange-fleshed fruit based on genetic mapping and whole genome sequencing. In addition, yellow flesh in 'PI 200815', an Indian ecotype (Kooistra, 1971), was found to be controlled by a recessive gene (*yf*) that mapped to chromosome 7 (Lu *et al*., 2015). Wang *et al*. (2023b) recently discovered a European greenhouse cucumber mutant with yellow flesh and showed that abscisic acid 8′-hydroxylase *Csyf2* regulates fruit colour by modulating carotenoid biosynthesis.

The accumulation of carotenoids is determined by biochemical pathways, which have largely been elucidated (Booker *et al*., 2004; Schwartz *et al*., 2001; Zheng *et al*., 2019). Carotenoid cleavage dioxygenases are a large class of carotenoid metabolizing enzymes that influence the accumulation of carotenoids and the multiple downstream processes they control. In Arabidopsis, there are four carotenoid cleavage dioxygenase genes (*CCD1*, *CCD4*, *CCD7* and *CCD8*), among which *CCD7* and *CCD8* are related to lateral shoot growth, and CCD1 leads to the accumulation of carotenoids in seeds (Auldridge *et al*., 2006). Similarly, in chrysanthemums, *CmCCD4* affects the colour of chrysanthemum petals by controlling carotenoids content (Yoshioka *et al*., 2012). In citrus, *CCD4b* is a key gene affecting the evolution of peel colour (Zheng *et al*., 2019). In strawberries, Amaya *et al*. (2024) found that CCD4 (4B) regulated the yellow flesh and carotenoid content of strawberry fruits. In tomatoes, *CCD1* has been reported to be related to the formation of flavour-volatile substances such as beta-ionone, pseudoionone and geranylacetone (Simkin *et al*., 2004). He *et al*. (2022) found that *SiCCD1* could catalyse lutein degradation in millet and affect the accumulation of carotenoids and colour development in grains. Ilg *et al*. (2010) also studied the effects of rice *CCD1* (*OsCCD1*) on the pigmentation of *Golden Rice 2* (*GR2*) and found that apocarotenoids were the substrates of *OsCCD1*. However, the roles of CCDs have not been reported yet in cucumber.

Here, through map-based cloning, we identified *CsCCD1*, which encodes a *carotenoid cleavage dioxygenase 1*, as the causal gene. A SNP located in the *CsCCD1* intron affects its normal splicing, leading to the decreased enzyme activity and yellow flesh. Metabolic profiling revealed that β-cryptoxanthin is the primary compound accumulated in yellow flesh. We also verified that the overexpression of *CsCCD1* could transform the yellow flesh of XIS cucumbers to white. These findings enhance our understanding of the genetic regulation of yellow flesh in both semi-wild and cultivated cucumbers, offering a theoretical foundation for breeding cucumbers with yellow flesh.

Result

Fine mapping of *yf*

In our previous research, based on a F_2 population, we have found that cucumber yellow flesh was controlled by a recessive single gene named *yf*, which was located on Chromosome 7 (Lu *et al*., 2015). To further elucidate the relationship between flesh colour and pigment content, we measured chlorophyll and carotenoid levels of both parental lines 'PI 2001815' with yellow flesh and '931' with white flesh at 0, 15, 30 and 45 days after flowering (DAF) (Figure 1a). The results showed that chlorophyll (Figure 1b) and carotenoid contents (Figure 1c) decreased during fruit ripening. Notably, the carotenoid content in 'PI 2001815' was significantly greater than that of '931' throughout the entire fruit development. The degradation rate of carotenoids in the fruit of 'PI 200815' was slower than that in '931', resulting in a more than fourfold higher carotenoid content in 'PI 200815' compared to '931' at 30 days after flowering (DAF) (Figure 1c).

To fine map *yf*, we screened 3000 F_2 individuals using previously reported flanking markers yfIndel3 and SSR05628 (Lu *et al*., 2015). Then, we further developed six more markers in the positioning interval. Based on the phenotype of fruit flesh colour at 45 DAF in the $F_{2:3}$ population, *yf* was mapped to a 101.76 kb genomic region (19552648-19 653 412 bp; Chinese Long v3.0 genome) between markers yfSNP1 and yfSNP2 (Figure 1d). This interval contains 15 protein-coding genes (Figure 1e; Table S2).

Candidate gene analysis

To identify potential candidate genes associated with cucumber yellow flesh, we examined the coding sequences (CDS) of 15 candidate genes, along with their 3000 bp promoter regions. Among the 15 genes analysed, two single nucleotide polymorphisms (SNPs) were detected in *CsaV3_7G031000* and *CsaV3_7G031010*, located at nucleotide positions 61 and 3735, respectively (Figure 2a; Figure S1a). The candidate gene *CsaV3_7G031000* was initially excluded from further validation due to its negligible expression in the fruit of both parental lines (FPKM <1.0), as indicated by our RNA-seq data (Figure S1b) (Abbas *et al*., 2024). In contrast, Real-Time PCR analysis indicated the expression level of *CsaV3_7G031010* in 'PI 200815' was significantly lower than that in '931' across both fruit and other tissues (Figure 2b,c). We subsequently analysed the A61G SNP of *CsaV3_7G031000* and the A3735G SNP of *CsaV3_7G031010* within a population of 159 diverse cucumber accessions. In contrast to the perfect segregation observed with the A3735G SNP for yellow flesh, the A61G SNP did not consistently associate with this trait (Figure S1c). These findings suggest that *CsaV3_7G031010* was much more likely to be implicated in the formation of yellow flesh in cucumber than *CsaV3_7G031000*. *CsaV3_7G031010* encodes an isoform of carotenoid cleavage dioxygenase (CCD).

In order to further understand the role of CCD proteases in cucumber fruit development, four homologous proteins (CCD1, CCD4, CCD7 and CCD8) were obtained from the cucumber genome according to the BlastP program (*e*-value <0.01) using the reported 4 *Arabidopsis* CCD subfamily proteins. Then, we conducted a phylogenetic analysis of CCD subfamily proteins in cucumber alongside those in *Arabidopsis*, maize, rice, sorghum, watermelon and melon (Qin *et al*., 2016). The results showed that *CsaV3_7G031010* belongs to the CCD1 subfamily, and therefore, we designated it as *CsCCD1* (Figure S2a).

An intronic SNP on the candidate gene *CsCCD1* affects its splicing

The coding sequence of eukaryotic genes is usually separated by introns. Intron and exon boundaries and the surrounding sequences are special nucleotide sequences that are conserved within the precursor mRNA (Sharp, 1994). Typically, the 5′ splice site of an intron starts with GT, while the 3′ splice site ends with AG (Yao *et al*., 2020). The SNP (A-G) of *CsCCD1* is positioned at the third base of the 5′ end of eighth intron, leading to the repetition of two GT elements (GTGT) in the 5′ GT splicing site (Figure 3a), which likely disrupts the normal recognition of the intron by the spliceosome. To evaluate the potential impact of this mutation on expression levels, we analysed the coding regions at both the 5′ (encompassing the first and second exons) and 3′ (comprising the eighth and ninth exons, as well as the thirteenth and fourteenth exons) ends of *CsCCD1* in fruits at 0, 15, 30 and 45 DAF. The results showed that the expression of 'PI 200815' was significantly lower than that of '931' in the 3′ end region of *CsCCD1*, while no significant difference was observed in the 5′ end region (Figure 3b). These qPCR findings suggest that the SNP influences transcription at the 3′ end of *CsCCD1*, which aligns with the transcriptome data (Figure 3c).

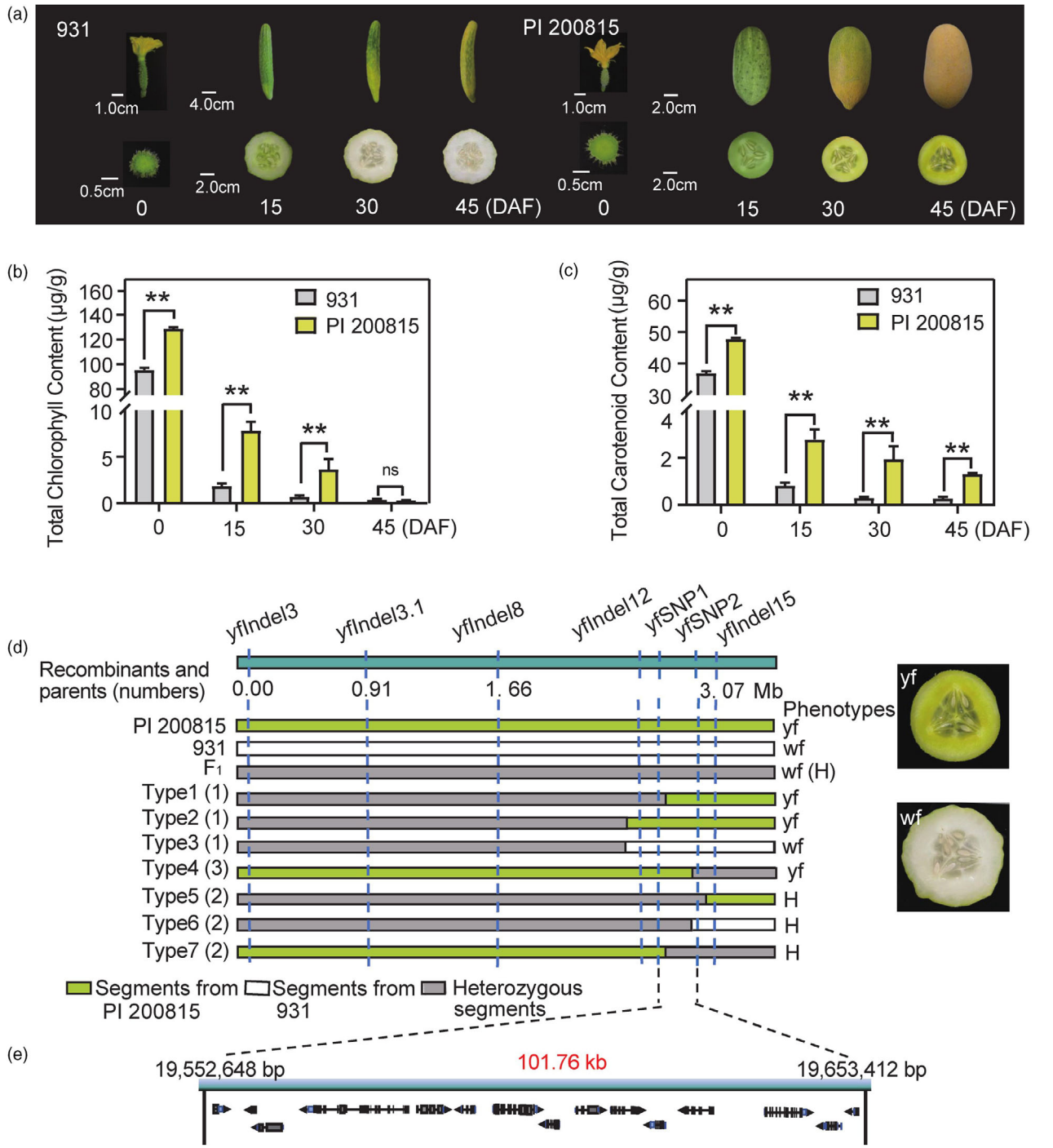


Figure 1 Map-based cloning and expression analysis of the *yf* gene. (a) Fruit appearance and flesh colour, (b) Chlorophyll contents and (c) Carotenoid contents of the two cucumber parents at 0, 15, 30, 45 days after flowering (DAF). (d) Genotyping of recombinant plants from 3000 F_2 individuals of the 'PI 200815' × '931' cross. yf, yellow flesh; wf, white flesh; H, heterozygous phenotype. (e) Annotation of the genes in the 101.81 kb interval. Blue boxes indicate UTRs; black boxes indicate exons; and black lines indicate introns. Arrow direction indicates gene orientation. ns, $P > 0.05$, $^{**}P < 0.01$ using two-sided Student's *t*-tests.

Based on the transcriptome data, two distinct isoforms were identified in the fruit of 'PI 200815': isoform1, which retained all 14 exons of the gene, and isoform2, which retained only the first eight exons and a portion of the eighth intron (Figure 3c). Notably, 'PI 200815' expresses both isoform1 and isoform2, whereas '931' only expresses isoform1 (Figure 3d). We then calculated the junction inclusion ratio (J_I) of normalized read counts for isoform2 relative to the total normalized reads for both isoforms. The J_I was significantly higher in 'PI 200815' compared to '931' (Figure 3e), indicating a dominant presence of isoform2

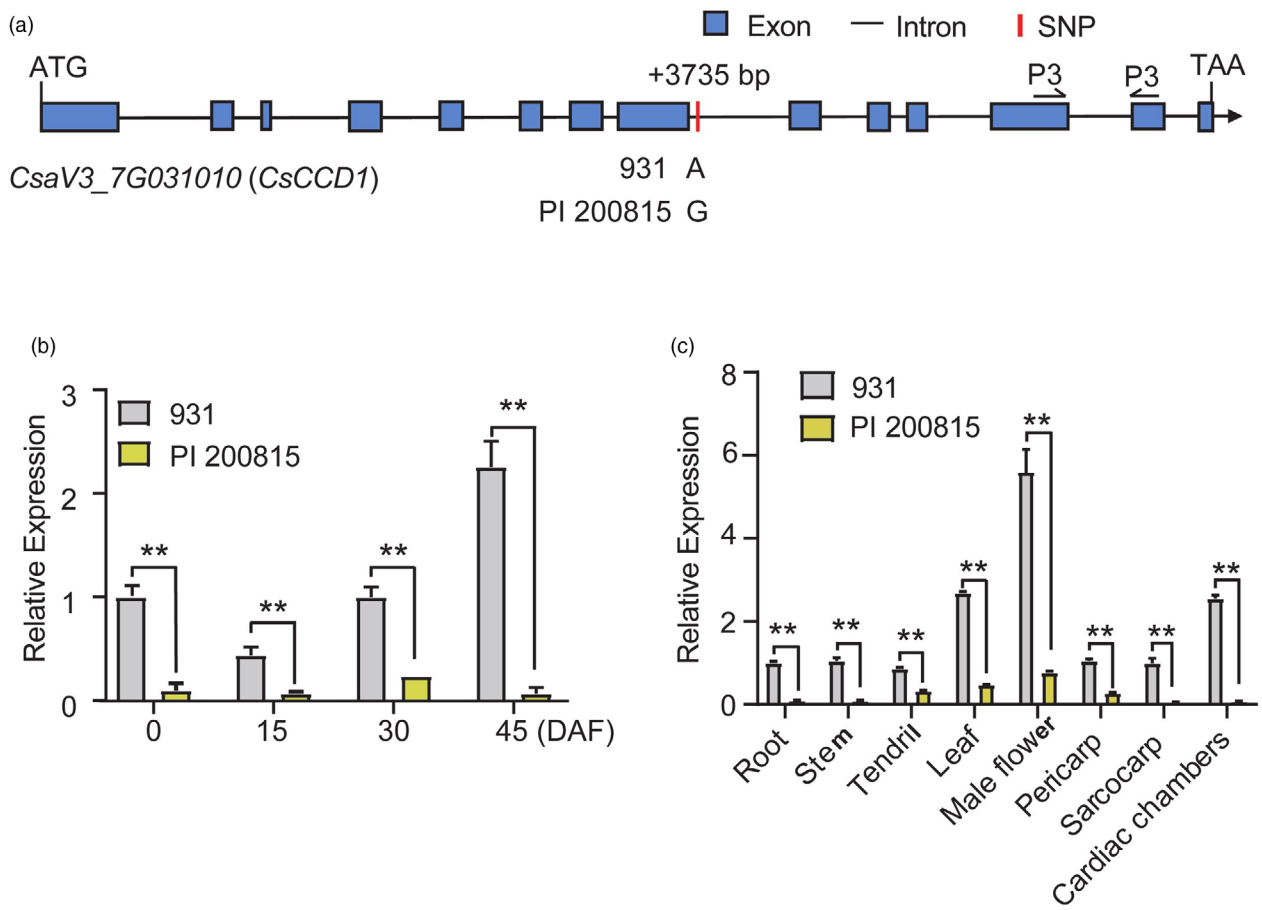


Figure 2 Sequence variation and expression analysis of candidate genes positioned in the 101.81 kb interval identified by map-based cloning. (a) Gene structure and SNP mutation of *CsaV3_7G031010* (*CsCCD1*). P3 were the primers designed for qRT-PCR. (b) Quantitative PCR expression analysis of *CsCCD1* in fruits of '931' and 'PI 200815' at 0, 15, 30 and 45 DAF. (c) Quantitative PCR expression analysis of *CsCCD1* in different tissues of '931' and 'PI 200815'.$^{**}P < 0.01$ using two-sided Student's *t*-tests.

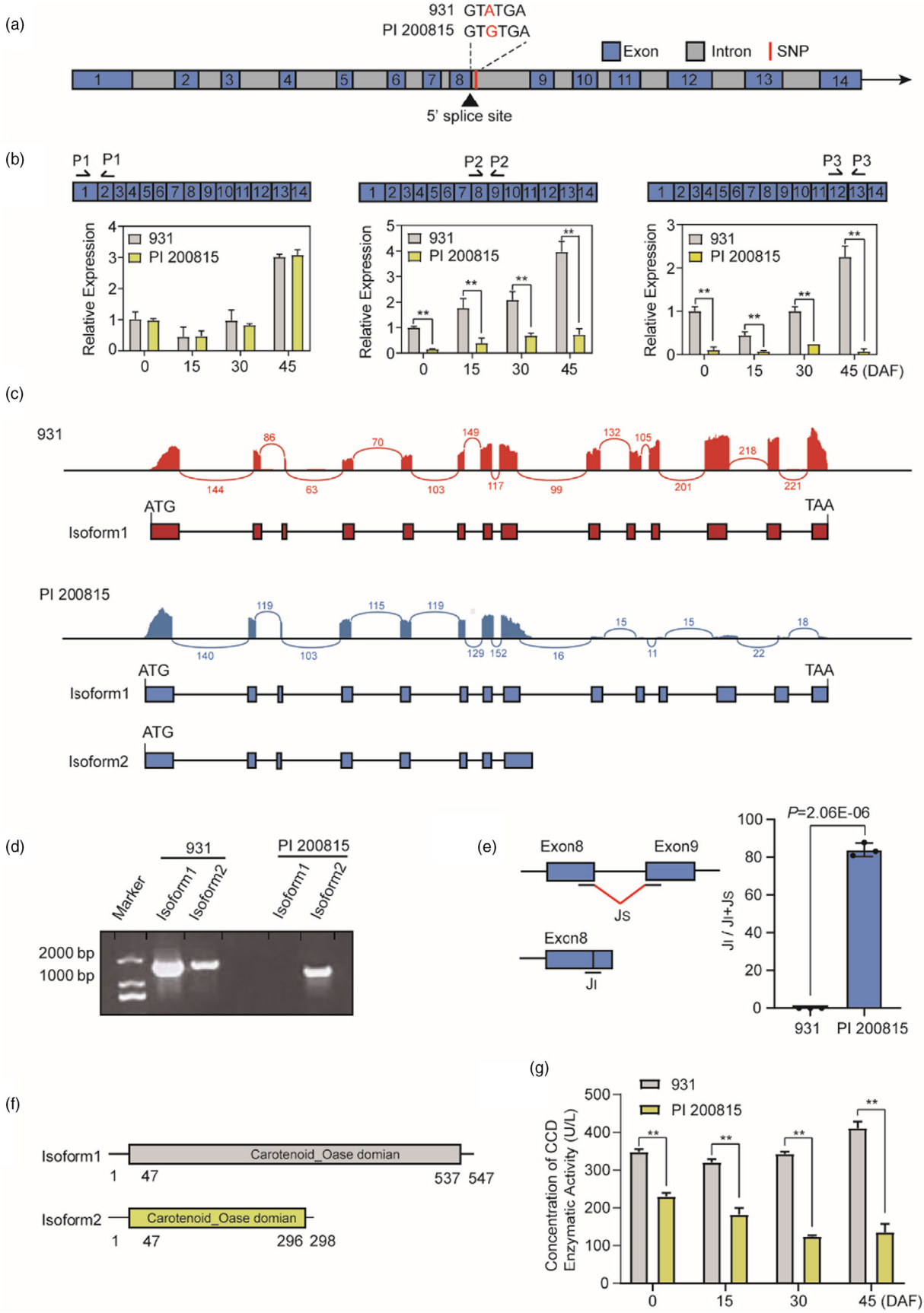
in 'PI 200815'. Furthermore, we predicted the protein domains of the two isomers, revealing that the predicted length of the Carotenoid_Oase domain was markedly shorter in isoform2 than in isoform1 (Figure 3f). To further support the influence of shortened Carotenoid_Oase domain, we predicted the three-dimensional structure of isoform1 and isoform2 and found that compared with isoform1, isoform2 was missing a variety of protein structures, including α helix, β fold, β corner, Ω ring and random curl (Figure S4).

Consistently, the CsCCD1 enzyme activity concentration in 'PI 200815' was significantly lower than in '931' at 0, 15, 30 and 45 DAF fruits (Figure 3g). All these findings suggest that the dominant truncated transcript in 'PI 200815' leads to a decrease in the enzyme activity of the CsCCD1 protein.

The *yf* mutant produced by CRISPR/Cas9 showed yellow flesh

To verify the role of *CsCCD1* in regulating cucumber yellow flesh, we employed CRISPR/Cas9 to knock out *CsCCD1* in the white flesh inbred line 'CU2', resulting in the creation of yellow-flesh (yf) mutants. Two T_2 homozygous mutants, designated *yf-1* and *yf-2*, were isolated; *yf-1* exhibits a 2-bp deletion, while *yf-2* has a 5-bp deletion in the *CsCCD1* gene, both of which are anticipated to result in premature termination of the amino acid sequence

Figure 3 Isoform analysis of CsCCD*1* in parental lines '931' and 'PI 200815'. (a) The gene structure of *CsCCD1*. SNP mutations are shown in red text. P1, P2, P3 were the primers designed for qRT-PCR. (b) The expression of different coding regions of the *CsCCD1* gene in fruit at 0, 15, 30, 45 DAF. P1, P2, P3 are the primers used for expression analysis. The black head indicates the position of the primers on the mRNA. $^{**}P < 0.01$ using two-sided Student's *t*-tests. (c) Transcription and isoform of *CsCCD1* in '931' and 'PI 200815'. Numbers meaning the number of reads supporting the junction. (d) Isoform fragments of '931' and 'PI 200815' were amplified by PCR. (e) A splicing event that occurs at the 8th intron of *CsCCD1*. On the left are two different splicing types. On the right are the proportions of J_I in '931' and 'PI 200815' transcripts, respectively. J_S, Skipping Junction. J_I, Include junction. (f) Protein domain prediction of Isoform1 and Isoform2. (g) Concentration of CCD enzyme activity of '931' and 'PI 200815' fruits at 0, 15, 30 and 45 DAF.$^{**}P < 0.01$ using two-sided Student's *t*-tests.


(a)
931 GTATGA
PI 200815 GTGTGA
Exon
Intron
SNP
5' splice site
(b)
P1 P1
P2 P2
P3 P3
Relative Expression
931
PI 200815
0 15 30 45 45 (DAF)
(c)
931
ATG
TAA
Isoform1
PI 200815
Isoform2
(d)
Marker
2000 bp
1000 bp
(e)
Exon8
Exon9
JS
Excn8
JI
JI / JI+JS
P=2.06E-06
(f)
Isoform1
Carotenoid_Oase domian
Isoform2
(g)
Concentration of CCD Enzymatic Activity (U/L)

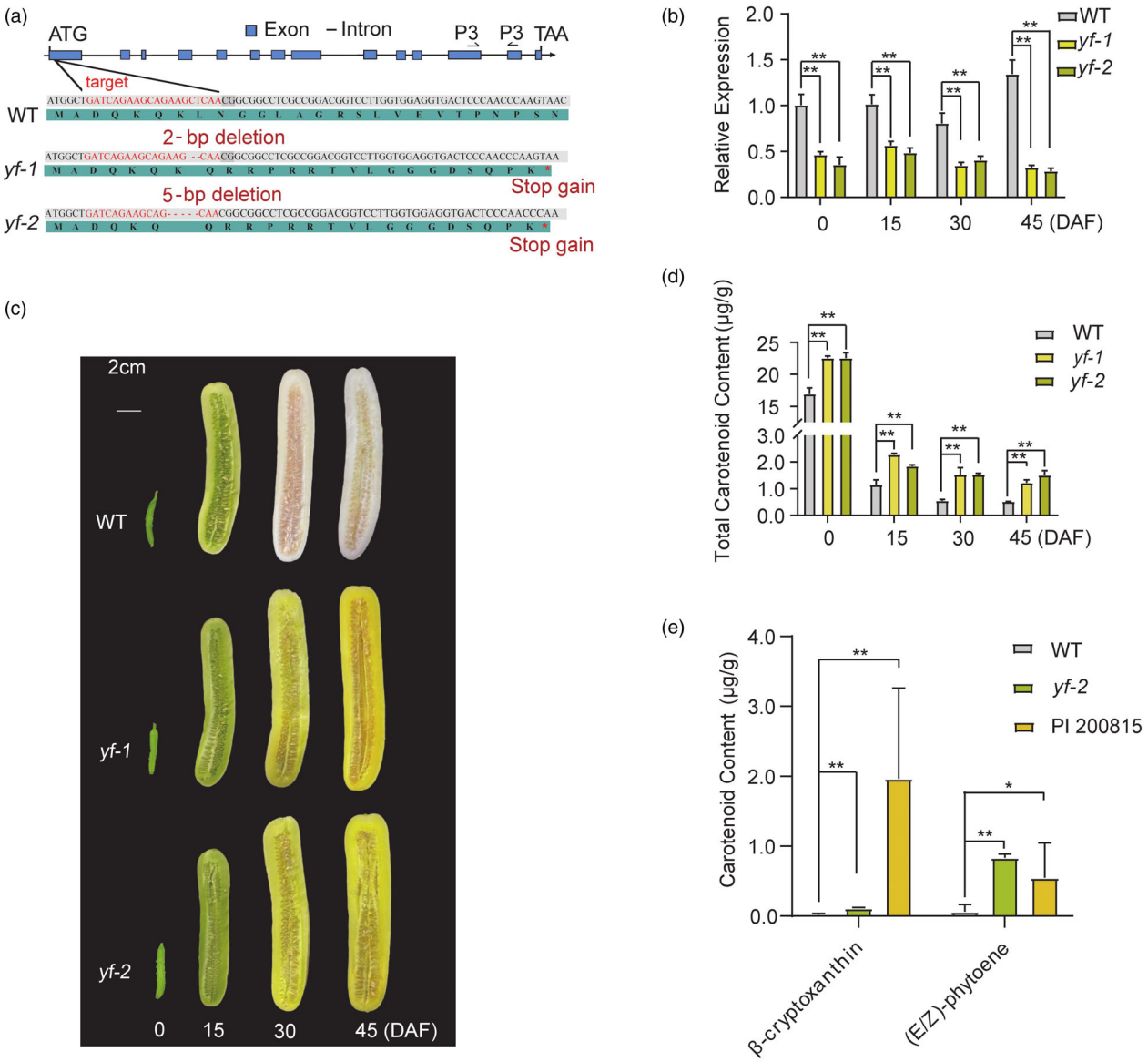


Figure 4 Phenotypic analysis of the flesh of CRISPR/Cas9 mutant of *CsCCD1*. (a) A schematic illustrating the gene structure of *CsCCD1* in the WT and the two gene-edited mutants *yf-1* and *yf-2*. P3 were the primers designed for qRT-PCR. The red text indicates the region targeted by the single guide RNAs. *yf-1* and *yf-2* are homozygous mutants with 2- and 5-base deletions, respectively, resulting in early termination of their coding amino acid. (b) *CsCCD1* expression analysis of WT, *yf-1*, *yf-2* in flesh at 0, 15, 30 and 45 DAF. (c) Flesh colour of WT, *yf-1*, *yf-2* at 0, 15, 30 and 45 DAF. (d) Carotenoid content in the flesh of WT (inbred line 'CU2'), *yf-1*, *yf-2* at 0, 15, 30and 45 DAF. (e) Carotenoid constituents in the flesh of WT, *yf-2*, 'PI 200815' fruits. Only compounds that accumulate significantly with WT were listed here, the remaining constituents were listed in Table S4. $*P < 0.05$, $**P < 0.01$ using two-sided Student's *t*-tests.

(Figure 4a). Moreover, the expression levels of the *CsCCD1* gene were significantly reduced ($P < 0.01$) in the mutants relative to the wild-type (WT) controls (Figure 4b). We assessed the changes in flesh colour among WT, *yf-1* and *yf-2*, and measured carotenoid content in the developing fruits. At 30 DAF, both *yf-1* and *yf-2* exhibited yellow flesh and demonstrated significantly higher carotenoid levels ($P < 0.01$) throughout fruit development compared to the white flesh of the WT (Figure 4c,d).

To assess potential differences in metabolites within the carotenoid metabolic pathway in yellow-fleshed cucumbers, we employed liquid chromatography-mass spectrometry (LC–MS) to analyse the flesh composition of the parental varieties '931' and 'PI 200815'. The findings revealed that '931' (white-fleshed) contained lower levels of xanthophyll and β-cryptoxanthin compared to 'PI 200815' (yellow-fleshed) (Table S3). Additionally, we identified 68 carotenoids in the WT, *yf-2* and 'PI 200815' fruit, with significant accumulation observed exclusively in (E/Z)-phytoene and β-cryptoxanthin in the yellow-fleshed fruit (*yf-2* and 'PI 200815'). Notably, β-cryptoxanthin demonstrated the most pronounced increase, with a 6.23-fold accumulation in *yf-2* relative to WT (Figure 4e). Previous studies have shown that β-cryptoxanthin can be used as substrate of CCD enzyme (Zhang *et al.*, 2024). We utilized RNA-seq data from the parental lines to examine the expression of β-cryptoxanthin synthetic gene. The

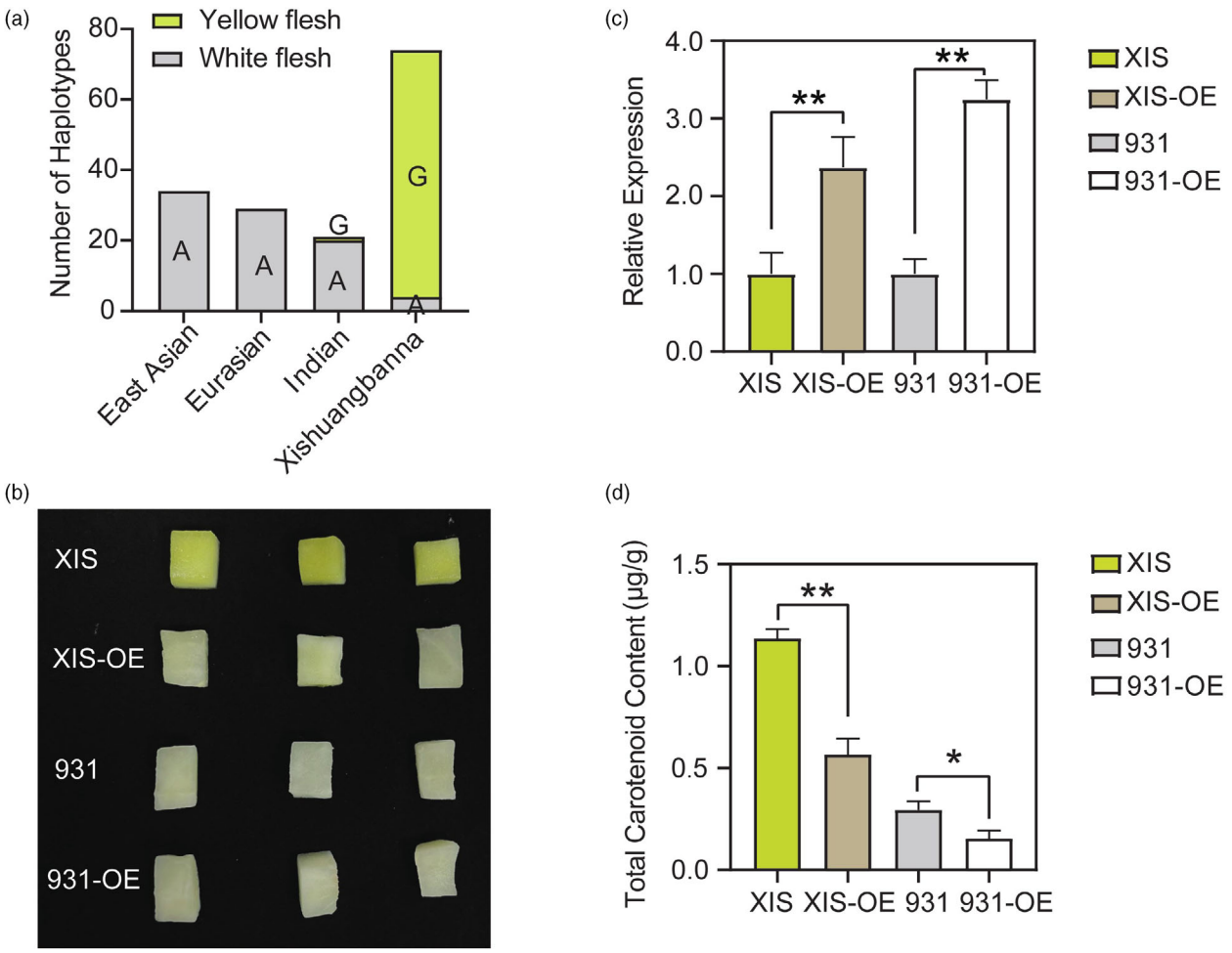


Figure 5 Transient overexpression of *CsCCD1* in flesh of XIS cucumber. (a) Flesh colour and haplotype distribution of *CsCCD1* in natural cucumber populations. (b) The representative flesh of XIS and its transient transgenic mutant XIS-OE, as well as '931' and its transient transgenic mutant 931-OE. A total of 15 samples were analysed for each genotype. (c) Quantitative PCR expression analysis of *CsCCD1* in XIS/931 and its transient transgenic mutant XIS-OE/931-OE. (d) Carotenoid content in XIS/931 and its transient transgenic mutant XIS-OE/931-OE.*$P < 0.05$, **$P < 0.01$ using two-sided Student's *t*-tests.

findings indicated no significant difference in the expression levels of the β-cryptoxanthin biosynthetic gene *BCH* between the parental lines, whereas a marked difference ($P = 1.00E-02$) was observed in the expression of the degradative gene *CCD1* (Figure S3). These results suggest that β-cryptoxanthin may be the primary metabolite accumulated in the yellow-fleshed cucumber fruit analysed.

CsCCD1 contributes to the formation of yellow flesh in XIS cucumber

XIS cucumber, a semi-wild variety cultivated by the Hani people in Xishuangbanna, Yunnan Province, China, is characterized by its high β-carotene content, which imparts a distinct yellow/orange hue (Qi *et al*., 1983; Chen *et al*., 1994). We analysed the A3735G SNP in the *CsCCD1* gene across a natural population of 159 cucumber accessions. The results revealed a perfect segregation between this SNP and the presence of yellow flesh; individuals with haplotype A exhibited white flesh, whereas those with haplotype G exhibited yellow flesh (Figure 5a). Notably, this causal SNP variation in *CsCCD1* was mostly found in XIS cucumbers. To explore the role of *CsCCD1* in the yellow pigmentation of XIS cucumber flesh, we constructed a 35S::*CsCCD1* transient overexpression recombinant plasmid by inserting the *CsCCD1* coding sequence into pGreen0029 62-SK. Transient expression of *CsCCD1* was conducted in '931' and XIS, with qRT-PCR analysis confirming significantly higher expression levels in the transgenic lines compared to wild-type (Figure 5c). In contrast to the wild-type XIS, the transgenic line XIS-OE exhibited white flesh and a marked reduction in carotenoid content (Figure 5b,d). These findings elucidate the genetic factors influencing flesh colour in XIS cucumber and underscore the role of *CsCCD1* haplotypes in flesh colour variation.

Discussion

Inheritance and QTL analysis of the cucumber flesh colour

Flesh colour is an economically and nutritionally valuable trait of cucumber fruit, and as a result, the genetic basis of flesh colour has been studied extensively. In wild cucumbers, Bo *et al*. (2019) used two parents (P_1: 'PI 183967'; P_2: '9110Gt'), and their genetic segregation populations to examine the inheritance of cucumber green flesh and concluded that green flesh was

controlled by a major QTL. In XIS cucumber, genetic analysis showed that the orange pericarp was controlled by two recessive genes and that the orange endocarp was controlled by one recessive gene (Cuevas *et al*., 2010). Lu *et al*. (2015) found that the yellow flesh in an Indian ecotype cucumber was controlled by a recessive single gene, *yf*, based on their investigation of the parental varieties (P_1: 'PI 200815'; P_2: '931'), as well as the F_1 and F_2 populations. In addition, some genes that control flesh colour have been successfully excavated, that is *CsaBCH1*, *CsOr*, *Csyf2* (Kishor *et al*., 2021; Qi *et al*., 2013; Wang *et al*., 2023a, 2023b). In this study, we further utilized an $F_{2:3}$ population derived from the two parental lines (P_1: 'PI 200815'; P_2: '931') to conduct fine mapping and successfully identified the gene responsible for yellow fruit flesh in cucumber as the *CsCCD1* locus. *CsCCD1* was different from the flesh colour genes or loci previous reported. In conclusion, the identification of *CsCCD1* for cucumber flesh colour has added a new theoretical basis for the breeding of cucumber with high nutritional quality.

Role of *CCD1* in regulating β-carotene accumulation and colour change in plants

Chlorophyll, carotenoids, flavonoids and betaine are the four primary pigment classes synthesized in plants. Among them, carotenoids are the key pigments responsible for generating natural colours of yellow, orange and red (Rosas-Saavedra and Stange, 2016). The catalytic decomposition of carotenoids is mainly accomplished by the carotenoid cleavage dioxygenase family of genes. Members of the CCD enzyme family are generally divided into two groups based on studies in *Arabidopsis*: four CCDs (*CCD1*, *4*,*7* and *8*) and five 9-cis-epoxy-carotenoid dioxygenases (*NCED 2*, *3*, *5*, *6* and *9*) (TanÃ *et al*., 2003). *CCD1* and *CCD4* possess broad substrate specificity and can cleave various types of carotenoids (Dhar *et al*., 2020). Additionally, they regulate colour changes in plants (Mi and Al-Babili, 2019). In chrysanthemums, *CmCCD4a* is highly expressed in white petals, but almost undetectable in yellow petals (Ohmiya *et al*., 2006). In peach, the gene *PpCCD4* has been identified as a critical regulator of white/yellow flesh. Its expression leads to the degradation of carotenoids in the white-fleshed genotype, while its inactivation results in the development of the yellow colour (Adami *et al*., 2013). CCD1 has been found to play a significant role in producing carotenoid-derived flavour and aroma compounds in flowers and fruits. It cleaves β-carotene and other carotenoids, leading to the formation of a range of volatile compounds (Ilg *et al*., 2016). In this study, *CsCCD1* has been identified as a crucial gene involved in regulating the white/yellow flesh phenotype in cucumber. Metabolite profiling revealed a significant accumulation of β-cryptoxanthin in yellow-fleshed fruit, suggesting that β-cryptoxanthin may serve as a substrate for *CsCCD1*. These findings provide valuable insights into the molecular mechanisms underlying colour variation in cucumber flesh and the potential involvement of *CsCCD1* in carotenoid metabolism. Indeed, while numerous studies have demonstrated an association between *CCD1* or *CCD4* and carotenoid content, the exact regulatory mechanisms governing CCD expression remain unclear. The precise factors and molecular pathways that control the transcriptional and post-transcriptional regulation of CCD genes are still being investigated. Further research is needed to elucidate the regulatory networks and signalling pathways involved in modulating CCD expression and activity, providing a more comprehensive understanding of carotenoid metabolism and its regulation in plants.

SNPs localized to introns in the *CsCCD1* gene affect its splicing

Introns are non-coding DNA sequences within genes that are removed by RNA splicing during the maturation of RNA products. Specifically, the 5′ splicing site of an intron begins with GT and the 3′ splicing site ends with AG. In general, in intron splicing reactions, U1snRNA first binds to the 5′ splicing site of the intron, the spliceosome begins to assemble, form a complex with several other factors, and then the intron forms a lasso structure and is spliced out (Valadkhan, 2005). Because introns are functionally meaningless to the structure of the translated product and are not subjected to natural selection pressure, they accumulate more mutations compared to exons. How variations in the intronic regions are regulated and how they affect phenotypic traits are important questions that need to be addressed in genetic studies. In recent years, the mechanisms by which intronic variation affect mRNA alternative splicing and transcription have gradually been elucidated. However, in plants, research on how intron variation impact their role as non-coding elements and how they affect agronomic traits remains very limited. In apples, Wang *et al*. (2023a) found that a 209 bp insertion in the fourth intron of the kinase *MdMMK2* can influence flesh colouring. In eggplants, Satterlee *et al*. (2024) discovered two independent mutations in the *prickleless* gene (*pl*) at the 5′ and 3′ splicing sites, resulting in lower expression levels and multiple mis-spliced transcripts, thereby affecting the development of fruit prickles. In this study, a SNP (A3735G) located at the third base pair of the eighth intron of *CsCCD1* in the yellow flesh genotype resulted in the generation of two consecutive repeated 5′ splice sites (GTGT). Similarly, in studies on human leukocyte antigens, researchers have found that selective splicing of an intronic SNP in *HMSD* results in a novel minor histocompatibility antigen lacked exon 2. Specifically, a C-to-A substitution at the fourth nucleotide and an A-to-G substitution at the fifth nucleotide in intron 2 (changing 5′-GUACAU-3′ to 5′-GUAAGU-3′) prevent stable binding of U1snRNA to the 5′ end of intron 2 in the mRNA (Kawase *et al*., 2007). Therefore, the SNP (A-G) in *CsCCD1* results in the repetition of two GT elements (GTGT) at the 5′ GT splice site, which may disrupt the normal recognition of introns by the spliceosome, leading to premature termination or degradation of the mRNA in the latter part. However, the specific mechanism of the alternative splicing of the mRNA precursor and its subsequent transcription still requires further research for clarification.

CsCCD1 participates in the establishment of flesh colour in XIS cucumber

The content of β-carotene in XIS cucumber is significantly higher, approximately 14 times more than ordinary cucumber varieties. Most XIS cucumbers have yellow or orange flesh. In this study, we analysed flesh colour and their *CsCCD1* haplotype from four diverse cucumber ecotypes, derived from 159 natural populations. The results showed that flesh colour was consistent with the haplotype, which in turn, was mostly found in XIS (Figure 5a). We transiently overexpressed *CsCCD1* in XIS. As a result, the flesh colour of the transgenic line XIS-OE was white, and the carotenoid content also decreased significantly compared with that of the yellow wild-type XIS line. The above study indicates that flesh colour in XIS is controlled by multiple genes, but that *CsCCD1* is likely one of the more crucial ones. The *CsCCD1*

regulatory genes and networks, however, have not been fully defined and need to be further studied.

Conclusions

In summary, we fine-mapped the yellow flesh gene (*yf*) in cucumber utilizing an $F_{2:3}$ population derived from the parental lines 'PI 200815' and '931'. The *yf* locus was confined to a 101.76 kb segment on chromosome 7, spanning from nucleotide 19 552 648 to 19 653 412 in the Chinese Long v3.0 genome. This segment contains 15 candidate genes; however, through sequence comparison, gene expression analysis and CRISPR/Cas9 gene knockout verification, we determined that *CsCCD1* is the gene responsible for controlling yellow flesh in cucumber. A significant SNP in the eighth intron of *CsCCD1* disrupts its splicing, leading to a truncated transcript and reduced enzyme activity. Carotenoid content analysis of the parental lines, wild-type (WT) and the *yf-2* mutant indicates that β-cryptoxanthin is the predominant compound accumulating in yellow-fleshed cucumber fruit, likely acting as the substrate for CsCCD1. Furthermore, transient overexpression of *CsCCD1* in the yellow-fleshed XIS cucumber restored the fruit flesh colour to white, further implicating *CsCCD1* in the determination of flesh colour in semi-wild ecotype XIS cucumber. Collectively, our study provides robust evidence supporting *CsCCD1* as the key gene regulating yellow flesh in cucumber.

Materials and methods

Plant materials and phenotype

Genotype '931' is an inbred line of East Asian ecotype, distinguished by its elongated fruit, dense white thorns and white flesh (Figure 1a). In contrast, genotype 'PI 200815' is an inbred line of Indian ecotype, first characterized in 1971 for its yellow flesh (Kooistra, 1971), featuring oval-shaped fruits with scattered black spines (Figure 1a). A cross was made between 'PI 200815' and '931' to create an F_1 population, which was self-pollinated to generate the F_2 progeny. On this basis, we selected 3000 F_2 and 53 $F_{2:3}$ segregating populations for fine mapping. All materials were grown in fall 2021 at the Nankou farm of the Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China. All plants were arranged in a complete randomized block design with three replications. The flesh colour of each fruit 30 days after flowering were observed. In addition, the contents of chlorophyll and carotenoid in fruits at 0, 15, 30 and 45 days after flowering were also determined (Wellburn and Lichtenthaler, 1984).

Genomic DNA extraction and candidate genes cloning

The genomic DNA of two parents and F_2 plants was extracted from cotyledons using the modified CTAB protocol (Saghai-Maroof *et al.*, 1984). The PCR amplification mixture with a total volume of 20 μL contains 15 ng genomic DNA, 50 ng forward and reverse primers, 10 μL of 2 × Phanta Max Master Mix (Vazyme, Nanjing, China) and 6 μL ddH_2O. The PCR reaction procedure was as follows: 95°C for 3 min, 35 cycles of (95°C denaturation for 15 s, 55°C annealing for 15 s, and 72°C extension for 1 min/kb), and a final 72°C extension for 5 min. PCR products were separated by 1% (w/v) agarose gel electrophoresis and sequenced by Sangon Biotech (Beijing, China) (https://www.sangon.com/). SSR and InDel primers were designed by Primer3.0 (https://primer3.ut.ee/); SNP primers were designed by dCAPS Finder 2.0 (http://helix.wustl.edu/dcaps/dcaps.html). All PCR primers are listed in Table S1. Sequence alignment was performed using the online software Multiple sequence alignment DNAstar (Swindell and Plasterer, 1997).

Expression levels analysis of candidate genes

The flesh samples from the parent inbred lines 'PI 200815' and '931', as well as from WT, *yf-1* and *yf-2*, were collected at 0, 15, 30 and 45 days after flowering. Additionally, various tissues including roots, stems, tendrils, leaves, male flowers, pericarps, sarcocarps and cardiac chambers from both parents were harvested for spatiotemporal expression analysis. The samples were rapidly frozen in liquid nitrogen. Total RNAs were extracted using Plant RNA Extraction Kit (TaKaRa MiniBEST), and first-strand cDNA was obtained by reverse transcription using HiScripRIII RT SuperMix for qPCR (Vazyme Biotech, Nanjing, China). ChamQ Universal SYBR qPCR Master Mix (Vazyme Biotech) was used for Real-Time PCR. *Actin1* (*CsaV3_3G038800*) was used as reference gene (Dai *et al.*, 2022), and $2^{-\Delta\Delta Ct}$ algorithm was used to calculate gene expression levels (Livak and Schmittgen, 2001). Fruits from three individual plants at the same segment were used as three biological replicates, and three technical replicates were designed for each biological replicate. Information of primers used in this study is listed in Table S1.

Identification of DEGs in the fruit flesh between parent genotypes

Fruit flesh tissues from '931' and 'PI 200815' plants, collected 45 days post-flowering, were flash-frozen in liquid nitrogen and stored at −80°C for RNA sequencing. Three biological replicates were prepared from distinct plants for each genotype. Total RNA was extracted using TRIzol reagent (Invitrogen, Burlington, ON, Canada). Strand-specific RNA-seq libraries were constructed to create DNBSEQ sequencing libraries following the manufacturer's protocol. Paired-end sequencing was executed on a DNBSEQ-T7 sequencer, yielding over 6 Gb of RNA-seq data per sample.

Raw reads underwent filtration using fastp software with default settings (Chen *et al.*, 2018). Clean reads were aligned to the reference genome (CLv3.0) via the STAR program (Dobin *et al.*, 2013), retaining unique alignments with fewer than two mismatches for subsequent differential gene expression analysis. Differentially expressed genes (DEGs) were identified utilizing the R DESeq2 package (Love *et al.*, 2014), with DEGs defined as genes exhibiting an adjusted *P*-value <0.05 and an absolute $\log_2$ fold change ≥ 1.

Determination of total pigment content and CCD enzyme activity

The contents of total chlorophyll and carotenoid were determined by light absorption method (Watada *et al.*, 1976; Wellburn and Lichtenthaler, 1984). In brief, 2 g sample was soaked in a 40 mL acetone: deionized water (4:1, v/v) mixture for 6 h, and the absorbance of the reaction mixture was measured at 663 nm (A_{663}), 646 nm (A_{646}) and 470 nm (A_{470}) using an ultraviolet–visible spectrophotometer (YOKE Co., Ltd., Shanghai, China).

$$C_a\ (\mathrm{mg/g\ FW}) = (12.21 \times A_{663} - 2.81 \times A_{646}) \times V/(1000 \times W)$$

$$C_b\ (\mathrm{mg/g\ FW}) = (20.13 \times A_{646} - 5.03 \times A_{663}) \times V/(1000 \times W)$$

$$\mathrm{Chlorophyll content (mg/g\ FW)} = (17.32 \times A_{646} + 7.18 \times A_{663}) \times V/(1000 \times W)$$

$$\text{Carotenoid content(mg/g FW)} = [(1000 \times A_{470} - 3.27 \times C_a - 104 \times C_b)/229] \times V/(1000 \times W)$$

Here, C_a, C_b, FW, V and W represent chlorophyll-a content, chlorophyll-b content, fresh weight, sample volume (mL) and sample weight (g), respectively.

To assess the enzyme activity of CCD, 1 g of flesh was homogenized in 9 mL of Phosphate Buffered Saline, and the supernatant was collected. Plant carotenoid cracking dioxygenase (CCD) enzyme-linked immunoassay kit (Mibio, Shanghai, China) was used to detect CCD enzyme activity. For details, please refer to the instruction manual of the kit.

Cucumber transformation of inbred line 'CU2' with white flesh

Two targets for guide RNA on the *yf* gene were obtained using the CRISPR-P v2.0 tool and were assembled into the CRISPR/Cas9 vector pBSE402 using *Bsa*I enzyme and T_4 Ligase (Hu *et al*., 2017). Then, the vector was transformed into the cotyledons of cucumber line 'CU2' using *Agrobacterium tumesfaciens* EHA105 as previously described (Xin *et al*., 2022). Positive plants were screened through GFP fluorescence and gene sequencing.

Determination of fruit carotenoid composition

Carotenoids contents were detected by MetWare (http://www.metware.cn/) based on the AB Sciex QTRAP 6500 LC–MS/MS platform. The flesh of the '931', 'PI 200815', WT and *yf-2* fruit collected 45 days after flowering were stored at −80°C for future use. To prepare the samples, the flesh was ground into a fine powder and then dried in a freeze-vacuum dryer. Fifty mg flesh powder was extracted with 0.5 mL mixed solution of n-hexane: acetone: ethanol (1:1:1, v/v/v) and an internal standard was added. The extract was collected by two centrifuges, then evaporated under nitrogen to dry and reconstructed in methylene chloride. It was filtered through a 0.22 μm membrane filter for LC–MS/MS analysis (Inbaraj *et al*., 2008). The sample extracts were analysed using an UPLC-APCI-MS/MS system (UPLC, ExionLC™ AD, https://sciex.com.cn/; MS, Applied Biosystems 6500 Triple Quadrupole, https://sciex.com.cn/). The samples were eluted using a gradient elution method. Acetonitrile (1:3, v/v) with 0.01% BHT and 0.1% formic acid (A), methyl tert-butyl ether with 0.01% BHT (B); gradient program, started at 0% B (0–3 min), increased to 70% B (3–5 min), then increased to 95% B (5–9 min), finally ramped back to 0% B (10–11 min). The analysis was conducted at a temperature of 28°C with a flow rate of 0.8 mL/min. For MS analysis, the API 6500 Q TRAP LC/MS/MS System was utilized, which is equipped with an APCI Heated Nebulizer, operating in positive ion mode and controlled by Analyst 1.6.3 software (Sciex). Carotenoid levels were analysed utilizing the AB Sciex QTRAP6500 LC–MS/MS platform by MetWare (http://www.metware.cn/).

Transient overexpression of *CsCCD1* in XIS cucumber

Transient overexpression analysis was performed using the procedure described in peach peel (Yu *et al*., 2021). The full-length coding sequence of *CsCCD1* was amplified using gene specific primers (Table S1) and was inserted into the pGreen0029 62-SK vector. The recombinant plasmid and the empty vector were transformed into *Agrobacterium tumefaciens* GV3101 (pSoup) using the heat shock method (Cao *et al*., 2019). XIS cucumber fruits (accession '5602') were soaked in 75% (375 mL ethanol/125 mL ddH_2O) ethanol (1 min) and 2% (10 mL sodium hypochlorite/490 mL ddH_2O) sodium hypochlorite solution (20 min) for disinfection and then washed with sterile water three times. The flesh was obtained and pre-cultured on MS medium at 24°C for 24 h. The flesh slices were soaked in the *Agrobacterium tumefaciens* suspension and placed in a vacuum (−70 kPa). The flesh slices were rinsed with sterile water and cultured on MS medium in the growth chamber (24°C, RH 85%). After three days, the flesh colour was observed, and samples were used for carotenoid content determination and gene expression analysis.

Statistical analysis

All statistical analysis were done using SAS9.4 software (Statistical Analysis System, NC, USA), and significance tests were assessed using one-way ANOVA and Student's *t*-test. * indicates $P < 0.05$, ** indicates $P < 0.01$. Histograms are represented by a bar chart using mean ± standard deviation.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (2023YFE0206900-03), the National Natural Science Foundation of China (32172571), the Earmarked Fund for Modern Agro-industry Technology Research System (Grant No. CARS-23), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2021-IVF), and the Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, PR China. .

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

Zhuonan Dai: Methodology, Investigation, Formal analysis, Writing-original draft, Writing -review & editing. Jiantao Guan: Conceptualization, Writing-review & editing. Han Miao: Conceptualization. Diane M. Beckles: Writing-review & editing. Xiaoping Liu: Conceptualization. Xingfang Gu: Conceptualization, Resources, Supervision, Project administration, Funding acquisition. Shaoyun Dong: Methodology, Writing-review & editing. Shengping Zhang: Conceptualization, Writing-review & editing, Supervision, Project administration, Funding acquisition.

Ethical approval

The authors declare that this study complies with the current laws of the countries in which the experiments were performed.

Data availability statement

All data generated or analysed during this study are included in this published article and its supplementary information files. RNA-seq data used in this study have been deposited in the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa/) of the National Genomics Data Center, part of the China National Center for Bioinformation, under accession number CRA018641.

References

Abbas, W., Shalmani, A., Zhang, J., Sun, Q., Zhang, C., Li, W., Cui, Y. *et al*. (2024) The GW5-WRKY53-SGW5 module regulates grain size variation in rice. *New Phytol.* **242**, 2011–2025.

Adami, M., Franceschi, P.D., Brandi, F., Liverani, A., Giovannini, D., Rosati, C., Dondini, L. *et al*. (2013) Identifying a carotenoid cleavage dioxygenase (ccd4) gene controlling yellow/white fruit flesh color of peach. *Plant Mol. Biol. Rep.* **31**(5), 1166–1175 1110.

Amaya, I., Roldan-Guerra, F.J., Ordonez, J.L., Torreblanca, R., Wagner, H., Waurich, V., Olbricht, K. *et al*. (2024) *Differential expression of CCD4 (4B) drives natural variation in fruit carotenoid content in strawberry (Fragaria* spp.). *bioRxiv* 601541.

Auldridge, M.E., Block, A., Vogel, J.T., Dabney-Smith, C., Mila, I., Bouzayen, M., Magallanes-Lundback, M. *et al*. (2006) Characterization of three members of the Arabidopsis carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. *Plant J.* **45**, 982–993.

Bo, K., Song, H., Shen, J., Qian, C., Staub, J.E., Simon, P.W., Lou, Q. *et al*. (2012) Inheritance and mapping of the ore gene controlling the quantity of β-carotene in cucumber (*Cucumis sativus* L.) endocarp. *Mol. Breed.* **30**(1), 335–344.

Bo, K., Wei, S., Wang, W., Miao, H., Dong, S., Zhang, S. and Gu, X. (2019) QTL mapping and genome-wide association study reveal two novel loci associated with green flesh color in cucumber. *BMC Plant Biol.* **19**, 1–13.

Booker, J., Auldridge, M., Wills, S., McCarty, D., Klee, H. and Leyser, O. (2004) MAX3/CCD7 is a carotenoid cleavage dioxygenase required for the synthesis of a novel plant signaling molecule. *Curr. Biol.* **14**, 1232–1238.

Cao, X., Xie, K., Duan, W., Zhu, Y., Liu, M., Chen, K., Klee, H. *et al*. (2019) Peach carboxylesterase PpCXE1 is associated with catabolism of volatile esters. *J. Agric. Food Chem.* **67**, 5189–5196.

Che, G. and Zhang, X. (2019) Molecular basis of cucumber fruit domestication. *Curr. Opin. Plant Biol.* **47**, 38–46.

Chen, J., Zhang, S. and Zhang, X. (1994) The Xishuangbanna gourd (*Cucumis sativus* var. xishuangbananesis Qi et Yuan), a traditionally cultivated plant of the Hanai people, Xishuangbanna, Yunnan, China. *Cucurbit Genetics Cooperative Report* **17**, 18–20.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890.

Cuevas, H., Song, H., Staub, J. and Simon, P. (2010) Inheritance of beta-carotene-associated flesh color in cucumber (*Cucumis sativus* L.) fruit. *Euphytica* **171**, 301–311.

Dai, Z., Dong, S., Miao, H., Liu, X., Han, J., Li, C., Gu, X. *et al*. (2022) Genome-Wide Identification of TIFY Genes and Their Response to Various Pathogen Infections in Cucumber (*Cucumis sativus* L.). *Sci. Hortic.* **295**, 110814.

Dhar, M.K., Mishra, S., Bhat, A., Chib, S. and Kaul, S. (2020) Plant carotenoid cleavage oxygenases: structure–function relationships and role in development and metabolism. *Brief. Funct. Genomics* **19**, 1–9.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P. *et al*. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.

Gebretsadik, K., Qiu, X., Dong, S., Miao, H. and Bo, K. (2021) Molecular research progress and improvement approach of fruit quality traits in cucumber. *Theor. Appl. Genet.* **134**, 3535–3552.

He, L., Cheng, L., Wang, J., Liu, J., Cheng, J., Yang, Z., Cao, R. *et al*. (2022) Carotenoid cleavage dioxygenase 1 catalyzes lutein degradation to influence carotenoid accumulation and color development in foxtail millet grains. *J. Agric. Food Chem.* **70**, 9283–9294.

Hu, B., Li, D., Xin, L., Qi, J., Gao, D., Zhao, S., Huang, S. *et al*. (2017) Engineering non-transgenic gynoecious cucumber using an improved transformation protocol and optimized CRISPR/Cas9 system. *Mol. Plant* **10**, 1575–1578.

Ilg, A., Bruno, M., Beyer, P. and Al-Babili, S. (2016) Tomato carotenoid cleavage dioxygenases 1A and 1B: Relaxed double bond specificity leads to a plenitude of dialdehydes, mono-apocarotenoids and isoprenoid volatiles. *FEBS Open Bio.* **4**, 584–593.

Ilg, A., Yu, Q., Schaub, P., Beyer, P. and Al-Babili, S. (2010) Overexpression of the rice carotenoid cleavage dioxygenase 1 gene in Gold Rice endosperm suggests apocarotenoids as substrates in planta. *Planta* **232**, 691–699.

Inbaraj, B.S., Lu, H., Hung, C.F., Wu, W.B., Lin, C.L. and Chen, B.H. (2008) Determination of carotenoids and their esters in fruits of *Lycium barbarum* Linnaeus by HPLC-DAD-APCI-MS. *J. Pharm. Biomed. Anal.* **47**, 812–818.

Kawase, T., Akatsuka, Y., Torikai, H., Morishima, S., Oka, A., Tsujimura, A., Miyazaki, M. *et al*. (2007) Alternative splicing due to an intronic SNP in HMSD generates a novel minor histocompatibility antigen. *Blood* **110**, 1055–1063.

Kishor, D., Lee, H.-Y., Alavilli, H., You, C.-R., Kim, J.-G., Lee, S.-Y., Kang, B.-C. *et al*. (2021) Identification of an allelic variant of the CsOr gene controlling fruit endocarp color in cucumber (*Cucumis sativus* L.) using genotyping-by-sequencing (GBS) and whole-genome sequencing. *Front. Plant Sci.* **12**, 802864.

Kooistra, E. (1971) Inheritance of fruit flesh and skin colours in powdery mildew resistant cucumbers (*Cucumis sativus* L.). *Euphytica* **20**(4), 521–523.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* **25**, 402–408.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Lu, H.W., Miao, H., Tian, G.L., Wehner, T.C., Gu, X.F. and Zhang, S.P. (2015) Molecular mapping and candidate gene analysis for yellow fruit flesh in cucumber. *Molec. Breed.* **35**, 64.

Mi, J. and Al-Babili, S. (2019) To color or to decolor: that is the question. *Mol. Plant* **12**, 1173–1175.

Nisar, N., Li, L., Lu, S., Khin, N.C. and Pogson, B.J. (2015) Carotenoid metabolism in plants. *Molec Plant* **8**, 68–82.

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.

Qi, C.Z., Yuan, Z.Z. and Li, Y.X. (1983) A new type of cucumber—Xishuangbanna cucumber. *Acta Hort. Sin.* **10**, 259–264.

Qi, J., Liu, X., Shen, D., Miao, H., Xie, B., Li, X., Zeng, P. *et al*. (2013) A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* **45**, 1510–1515.

Qin, X., Fischer, K., Yu, S., Dubcovsky, J. and Tian, L. (2016) Distinct expression and function of carotenoid metabolic genes andhomoeologs in developing wheat grains. *BMC Plant Biol* **16**(1), 155.

Rosas-Saavedra, C. and Stange, C. (2016) Biosynthesis of carotenoids in plants: enzymes and color. *Carotenoids Nat* **79**, 35–69.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A. and Allard, R. (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA* **81**, 8014–8018.

Satterlee, J.W., Alonso, D., Gramazio, P., Jenike, K.M., He, J., Arrones, A. and Villanueva, G. (2024) Convergent evolution of plant prickles by repeated gene co-option over deep time. *Science* **385**(6708), eado1663.

Schwartz, S.H., Qin, X. and Zeevaart, J.D. (2001) Characterization of a novel carotenoid cleavage dioxygenase from plants. *J. Biol. Chem.* **276**, 25208–25211.

Sharp, P.A. (1994) Split genes and RNA splicing. *Cell* **77**, 805–815.

Simkin, A.J., Schwartz, S.H., Auldridge, M., Taylor, M.G. and Klee, H.J. (2004) The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles β-ionone, pseudoionone, and geranylacetone. *Plant J.* **40**, 882–892.

Simon, P.W. and Navazio, J.P. (1997) Early orange mass 400, early orange mass 402, and late orange mass 404: high-carotene cucumber germplasm. *HortScience* **32**, 144–145.

Swindell, S.R. and Plasterer, T.N. (1997) SEQMAN. In *Sequence data analysis guidebook* (Swindell, S.R., ed), pp. 75–89. Totowa, NJ: Springer.

TanÃ, B.-C., Josephy, L.M., Dengy, W.T., Liu, L. and Li, Q.B. (2003) Molecular characterization of the Arabidopsis 9-cis epoxycarotenoid dioxygenase gene family. *Plant J. Cell Molec Biol.* **35**, 44–56.

Ugwu, C. and Suru, S. (2021) Cosmetic, culinary and therapeutic uses of cucumber (*Cucumis sativus* L.). In *Cucumber economic values and its cultivation and breeding*. London, United Kingdom: IntechOpen.

Underwood, B.A. (2004) Vitamin A deficiency disorders: international efforts to control a preventable "pox". *J. Nutr.* **134**, 231S–236S.

Valadkhan, S. (2005) snRNAs as the catalysts of pre-mRNA splicing. *Curr. Opin. Chem. Biol.* **9**, 603–608.

Wang, T., Duan, S., Xu, C., Wang, Y., Zhang, X., Xu, X., Chen, L. *et al*. (2023a) Pan-genome analysis of 13 Malus accessions reveals structural and sequence variations associated with fruit traits. *Nat. Commun.* **14**, 7377.

Wang, X., Jin, B., Yan, W., Wang, J., Xu, J., Cai, C., Qi, X. *et al*. (2023b) Cucumber abscisic acid 8′-hydroxylase Csyf2 regulates yellow flesh by modulating carotenoid biosynthesis. *Plant Physiol.* **193**, 1001–1015.

Watada, A.E., Norris, K.H., Jt, W. and Dr, M. (1976) Estimation of chlorophyll and carotenoid contents of whole tomato by light absorbance technique. *J. Food Sci.* **41**, 329–332.

Wellburn, A. and Lichtenthaler, H. (1984) Formulae and program to determine total carotenoids and chlorophylls a and b of leaf extracts in different solvents. In *Advances in Photosynthesis Research: Proceedings of the VIth International Congress on Photosynthesis*, pp. 9–12. Brussels, Belgium: Springer.

Xin, T., Tian, H., Ma, Y., Wang, S., Yang, L., Li, X., Zhang, M. *et al*. (2022) Targeted creation of new mutants with compact plant architecture using CRISPR/Cas9 genome editing by an optimized genetic transformation procedure in cucurbit plants. *Horticulture Res.* **9**, uhab086.

Yao, J., Wu, D.C., Nottingham, R.M. and Lambowitz, A.M. (2020) Identification of protein-protected mRNA fragments and structured excised intron RNAs in human plasma by TGIRT-seq peak calling. *eLife* **9**, e60743.

Yoshioka, S., Aida, R., Yamamizo, C., Shibata, M. and Ohmiya, A. (2012) The carotenoid cleavage dioxygenase 4 (CmCCD4a) gene family encodes a key regulator of petal color mutation in chrysanthemum. *Euphytica* **184**, 377–387.

Yu, Y., Guan, J., Xu, Y., Ren, F. and Xie, H. (2021) Population-scale peach genome analyses unravel selection patterns and biochemical basis underlying fruit flavor. *Nat. Commun.* **12**, 3604.

Zhang, Y., Jin, J., Wang, N., Sun, Q., Feng, D., Zhu, S., Wang, Z. *et al*. (2024) Cytochrome P450 CitCYP97B modulates carotenoid accumulation diversity by hydroxylating β-cryptoxanthin in Citrus. *Plant Commun.* **5**, 100847.

Zheng, X., Zhu, K., Sun, Q., Zhang, W., Wang, X., Cao, H., Tan, M. *et al*. (2019) Natural variation in CCD4 promoter underpins species-specific evolution of red coloration in citrus peel. *Mol. Plant* **12**, 1294–1307.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Sequence variation and expression analysis of candidate genes positioned in the 101.81 kb interval identified by map-based cloning.
Figure S2 Phylogenetic tree of CCD genes.
Figure S3 Expression level of β-cryptoxanthin pathway gene in parents.
Figure S4 Secondary structure prediction of isoform1 and isoform2 proteins.
Figure S5 Tertiary structure prediction of isoform1 and isoform2 proteins.
Figure S6 Concentration of CCD enzyme activity of WT, *yf-1* and *yf-2* fruits at 0, 15, 30, and 45 DAF.
Table S1 Primer information.
Table S2 Information of genes within the localization interval.
Table S3 Carotenoid content between parents.
Table S4 Carotenoid content of WT, *yf-2*, and PI200815.

