Genome-wide Association Study and Possible Candidate Genes for Root Color and Carotenoid Contents in Japanese Orange Carrot F2 Populations

Taeko Shibaya (komai@kazusa.or.jp)
Fujii Seed Co. Ltd
Chika Kuroda
Fujii Seed Co. Ltd
Hisano Tsuruoka
Kazusa DNA Research Institute
Chiharu Minami
Kazusa DNA Research Institute
Akiko Obara
Kazusa DNA Research Institute
Shinobu Nakayama
Kazusa DNA Research Institute
Takayoshi Fujii
Fujii Seed Co. Ltd
Sachiko Isobe
Kazusa DNA Research Institute

Research Article

Keywords: Carrot , provitamin A, genome-wide association study (GWAS), single-nucleotide polymorphisms (SNPs)

Posted Date: November 10th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1014140/v1

Abstract

Carrot is a major source of provitamin A in a human diet. Two of the most important traits for carrot breeding are carotenoid contents and root color. To examine genomic regions related to these traits and develop DNA markers for carrot breeding, we performed a genome-wide association study (GWAS) using genome-wide single-nucleotide polymorphisms (SNPs) in two F2 populations, both derived from crosses of orange root carrots bred by a Japanese seed company. The GWAS revealed 21 significant associations, and the physical position of some associations suggested two possible candidate genes. An Orange (Or) gene was a possible candidate for visual color evaluation and the α- and β-carotene contents. Sanger sequencing detected a new allele of Or with an SNP which caused a non-synonymous amino acid substitution. Genotypes of this SNP corresponded to the visual evaluation of root color in another breeding line. A chromoplast-specific lycopene β-cyclase (CYC-B) gene was a possible candidate for the β/α carotene ratio. On CYC-B, five amino acid substitutions were detected between parental plants of the F2 population. The detected associations and SNPs on the possible candidate genes will contribute to carrot breeding and the understanding of carotenoid biosynthesis and accumulation in orange carrots.

Introduction

Carrot (*Daucus carota* L.), a major source of provitamin A carotenes in the human diet, is consumed worldwide [1]. Carrots accumulate abundant carotenoids in their taproots, and these carotenoids (which are responsible for the orange pigmentation in the carrot roots) are thought to provide health benefits [2]. A variety of colors has been observed in carrot taproots, including orange, white, yellow, red, and purple. Quantitative trait loci (QTL) analyses and association studies for carrots' root color and carotenoid contents have been performed in several populations, and important and useful QTLs have been reported [3–6]. These studies used populations derived from crosses between accessions showing clearly different root colors such as orange and white [3–5] and orange and dark orange [4], and other studies used inter-crossed populations derived from white, yellow, red, and orange carrots [6].

Carotenoid biosynthesis is well established, and a highly conserved carotenoid biosynthesis pathway has been characterized in many plant species (Fig. 1) [7–9]. In carrot, several carotenoid biosynthetic genes have been mapped [3], and the released carrot whole-genome sequences showed orthologous and homologous genes involved in the carotenoid biosynthesis pathway [10,11]. Several genes involved in carotenoid biosynthesis and accumulation in carrot have also been identified. An ortholog of carotene hydroxylase CYP97A3 in the carotenoid biosynthesis pathway has been identified in carrot; it controls the α-carotene, total carotenoid contents, and the α/β carotene ratio [12]. A candidate gene association study of the carotenoid biosynthesis pathway revealed associations between the total carotenoid and β-carotene contents and the genes *zeaxanthin epoxidase* (*ZEP*), *phytoene desaturase* (*PDS*), and *carotenoid isomerase* (*CRTISO*), between the α-carotene content and the genes *CRTISO* and *plastid terminal oxidase* (*PTOX*), and between color components and the gene *ZEP* [6].

It was also reported that not only genes in the carotenoid biosynthesis pathway but also genes that have other functions considerably affect carotenoid contents. *Y* and Y_2 loci account for most of the color differences of orange, yellow, and white carrot roots [13]. The *Y* gene has been identified, and this gene has been hypothesized to regulate photosystem development and functional processes, including photomorphogenesis and root de-etiolation [10]. The *Y2* locus has been mapped to an approx. 650-kb genomic region; in addition, no annotated gene involved in the carotenoid biosynthesis pathway was located within the candidate region [14]. An *Orange* (*Or*) gene, which was first identified in cauliflower and accounted for an abnormally elevated β-carotene accumulation [15], was identified in carrot and is associated with the presence of carotenoid in carrot [16]. However, the genes, polymorphisms, and genomic regions involved in carotenoid biosynthesis and the carotenoid accumulation that cause slight differences in root color and carotenoids are not fully understood, especially within the orange carrots.

In Japan, consumers prefer a bright orange root color for carrots, and a cultivar showing uniform root colors is popular. There are accessions showing slight color differences in bright orange roots, and breeders in Japan have selected the best 'bright orange' and uniform color among the accessions that have bright orange roots. DNA markers that can be used to distinguish slight differences within bright orange color have thus been sought in Japanese carrot breeding. Toward this goal, there has been no study using populations derived from a cross between orange root carrots with slight color differences, but the recent release of whole genome sequences of carrot has made it easier to analyze whole-genome constitutions with high marker density and to conduct association analyses, even in the populations derived from genetically close orange carrots [10,11].

In the present study, we developed two F_2 populations that have a common parent. Both populations were derived from crosses between orange-root parents. We performed a genome-wide association study (GWAS) to investigate the genomic regions that cause slight but important differences in the root color and carotenoid contents within carrots with orange root color.

Methods

Plant materials

We developed two F_2 populations (A and B) using orange-colored carrot plants bred by a Japanese seed company, Fujii Seed (Osaka, Japan). Population A was derived from a cross between Fs001 and Fs002, and population B was derived from a cross between Fs002 and Fs003 (Fig. 2). Fs002 was the pollen parent for F_2 population A and the seed parent for F_2 population B. Plants of F_2 populations A (n=146) and B (n=136) were cultivated from mid-February to early June 2018 in a natural field at Narashino, Chiba, Japan, and used for DNA extraction and the visual evaluation of root colors. Roots of population A were also used for the quantification of carotenoid content by high-performance liquid chromatography (HPLC) and the measurement of color components.

To examine a developed DNA marker on *Or* gene, we also used breeding line C, which was bred by Fujii Seed. This line was developed by using Fs002 as one of the breeding materials (Fig. 2). Breeding line C was cultivated from the end of March to early July 2017 in a natural field at Oirase, Aomori, Japan, and 40 plants were used for DNA extraction and the visual evaluation of root colors.

Experimental research and field studies on plant materials comply with relevant institutional, national, and international guidelines and legislation.

Visual evaluation of root colors and evaluation of color components

The visual evaluation of root colors was performed by two experienced breeders at Fujii Seed. The root colors were visually evaluated to ten grades of orange darkness in F_2 population A, and to seven grades in F_2 population B, and to three grades in breeding line C. In the F_2 population A, color components (L*, a* and b*) were measured with a spectrocolorimeter (model CM2600d, Minolta, Tokyo) equipped with a 5-mm measuring area. The surface of the middle part of washed carrot root was measured three times, and the average values were used for phenotypic data.

Quantification of carotenoid contents (α-carotene, β-carotene, and lutein) by HPLC

Carrot root surface, i.e., approx. 1–2 mm of outer parts from phloem in the middle of roots was cut and collected. The collected samples were immediately frozen in liquid nitrogen and stored at −80°C. Root surface was used for HPLC because the visual and color component evaluations were performed on the carrot root surface. The extraction for HPLC was performed as described [6] with a scale-down and some modifications. Frozen samples were crushed into a powdery status with a tube mill control (S001, IKA, Staufen, Germany). Extraction was done on approx. 50 mg (50 mg ± 5%) of crushed frozen material to which 50 µL of b-apo-8'-carotenal at 5 µg/mL was first added as an internal standard. Samples were mixed with 600 µL of $MgCO_3$ 0.57%, 3,5-di-tert-butyl-4-hydroxytoluene (BHT) 0.1% in methanol, then vortexed, and mixed with 600 µL of 0.1% BHT-containing chloroform. After 10 times of vertical mixing and incubation for 15 min in darkness at 4°C, 600 µL of ultrapure water was added, and samples were centrifuged at 236 g for 10 min. Next, 400 µL from the lower layer was concentrated under vacuum evaporation, and the dry extract was dissolved in 50 µL of acetone containing 0.1% BHT. Samples were kept at 4°C and protected from direct light during the entire procedure.

The carotenoid quantification was done on an Ultimate 3000 HPLC system coupled with a diode array detector (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instruction with slight modifications. Carotenoids were separated on an Acclaim C30 column (150 × 2.1 mm, 3 µm, Thermo Fisher Scientific). The mobile phases were acetonitrile as eluent A, methanol/acetic ether (1:1,

v/v) as eluent B, and 10 mM formic acid (pH 3.0) as eluent C. The elution program was as follows: the proportions of solvent A, B and C were 85% A, 14.5% B, and 0.5% C at 0–2 min; 85%–44.5% A, 14.5%–55% B, and 0.5% C at 2–7 min; 44.5% A, 55% B, and 0.5% C at 7–21 min; and returned to the initial conditions (85% A, 14.5% B, and 0.5% C) at 21.1–28.5 min. The flow rate was 0.4 mL/min. The injection volume of the filtered sample by a 0.22-μm PTFE membrane filter was 3.9 μL. Analytes were detected by a photodiode array detector at 450 nm. The data were analyzed using Chromeleon 7 software (Thermo Fisher Scientific) based on internal calibration using b-apo-8'-carotenal and the extraction yield.

Double-digest restriction site-associated DNA sequencing (ddRAD-seq) and GWAS

Total genomic DNA was extracted from young leaves of carrot plants with the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). A double-digest restriction site-associated DNA sequencing (ddRAD-seq) analysis was performed as described [17] with the restriction enzymes *Pst*I and *Msp*I. The ddRAD-seq libraries were constructed and sequenced on a HiSeq 4000 platform (Illumina, San Diego, CA) in paired-end 101-nucleiotide (nt) mode as described [17]. Primary data processing such as deleting low-quality bases and trimming adapters, mapping onto carrot genome Daucus carota v2.0 [10], and filtering single-nucleotide polymorphisms (SNPs) to obtain high-confidence SNPs were performed as described [18]. The association analysis between the phenotype data and the genotype data was performed using the generalized linear model (GLM) of trait analysis by association, evolution, and linkage (TASSEL) ver. 5.2.40 [19].

Sanger sequencing of candidate genes

For the comparison of the genomic sequences of possible candidate genes between parental plants in F_2 populations A and B, we performed Sanger sequencing from the start codon to the stop codon on the genes. The primers used in the Sanger sequencing are listed in Supplementary Table S1.

SNP genotyping with KASP marker

KASP marker, which genotypes an SNP on *Or* gene in this study, was developed and performed according to the manufacturer's instructions (Biosearch Technologies, Novato, CA).

Phylogenetic analysis

The sequences of LCYE, LCYB, CYC-B, NSY, and CCS in several reported plant species such as *Solanum lycopersicum*, *Carica papaya*, *Citrus sinensis*, *Capsicum annuum*, and *Lillium lancifolium* [20–23], *Arabidopsis* [7], and carrot [24] were obtained from the public databases NCBI

(http://www.ncbi.nlm.nih.gov) and Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#). We used CLUSTALW [25] to align the amino acid sequences and constructed the phylogenetic tree by using the neighbor-joining method [26] provided by MEGA X [27].

Results

GWAS for the visual evaluation of root color and evaluations of color components and carotene contents in roots of F_2 populations A and B

F_2 populations A and B both showed a normal distribution in all root color evaluations (Suppl. Fig. S1), suggesting the involvement of multiple associations in carrot root color. The ddRAD-seq analysis detected 3,106 and 1,901 high-confidence SNPs in F_2 populations A and B, respectively. The GWASs were performed using these genotypic data and values from the visual evaluation and evaluations of the color components and carotene contents in the carrot roots. In F_2 population A, significant associations were detected for the visual evaluation of root color (Fig. 3a); color components a* (Fig. 3c) and b* (Fig. 3d); α-carotene (Fig. 3e), β-carotene (Fig. 3f), and lutein contents (Fig. 3g); and the β/α-carotene ratio (Fig. 3h) in root (Table 1). No significant associations were detected for color component L* (Fig. 3b).

The associations for visual evaluation, color components a* and b*, and α- and β-carotene contents on chromosome 1 were detected at close physical positions, and the highest associations were detected at a physical position around 31 Mb (Fig. 3, Table 1), suggesting that these associations are caused by an identical locus. The physical positions of the associations for the α- and β-carotene contents on chromosome 3 were close, and the highest associations were detected at a physical position around 6 Mb (Fig. 3, Table 1). An association detected in population B for visual color evaluation on chromosome 3 showed the highest association at physical position 5.4 Mb, and this physical position was similar to those of the associations detected in population A for α- and β-carotene contents (Figs. 3, 4, Table 1). These results suggest that the associations are caused by an identical locus. Interestingly, the association detected on chromosome 5 (showing the highest association for visual evaluation in F_2 population A) was not detected in any other evaluations (Fig. 3, Table 1).

Table 1

Significant associations for carrot root color identified by GWAS

Trait	F2_population	physical position[a]		-log10P
		Chr	bp	
Visual evaluation	A	5	39,247,011	5.80
	A	1	30,782,032	5.31
	A	4	20,589,731	5.01
	A	6	6,440,506	4.72
	B	3	5,419,538	9.69
a*	A	1	32,693,618	8.42
	A	3	1,689,065	6.37
b*	A	1	31,112,534	12.49
α-carotene	A	1	30,704,558	11.08
	A	1	15,832,013	8.22
	A	3	5,849,853	7.77
β-carotene	A	1	30,704,558	13.72
	A	1	15,832,013	11.82
	A	3	6,455,250	5.79
Lutein	A	5	11,540,998	8.38
	A	2	10,281,204	5.09
	A	6	32,679,396	4.75
β/α-carotene ratio	A	6	4,640,387	11.02
	A	1	12,574,447	8.78
	A	5	29,337,124	6.63
	A	2	34,903,955	5.04

[a]physical position of the SNP showing the most significant association in the GWAS based on carrot genome Daucus carota v2.0 (10).

Correlations among visual evaluation, color components, and carotene contents in root of F_2 population A

The Pearson correlation between each phenotype showed that three color components, i.e., L*, a* and b*, the α-carotene content, and the β-carotene content were highly correlated (Table 2). The lutein content was slightly correlated with L*, a* and b* and highly correlated with the α-carotene content. As lutein is biosynthesized downstream of the α-carotene (Fig. 1), this high correlation of lutein and α-carotene is consistent with the biosynthesis pathway. The visual evaluation was not highly correlated with any other phenotypes.

Allelic effects of associations detected by GWAS on chromosomes 1 and 3 for the α-carotene and β-carotene contents in F_2 population A

We examined the allelic effects of the associations detected by the GWAS for the α- and β-carotene contents. At the median, the carrots with AA allele on the SNP showing the highest association for α-carotene (DCARV2_CHR1_30704558) had approx. 1.5-fold higher contents of α- and β-carotene than those with GG allele (Fig. 5a, b). Similarly, at the median, the carrots with GG allele on the SNP showing the highest association for α-carotene (DCARV2_CHR3_5849853) had approx. 1.3-fold higher contents of α-carotene and approx. 1.2-fold higher contents of β-carotene compared to those with AA allele (Fig. 5c, d). A clear genetic interaction such as epistasis was not observed between the associations detected on chromosomes 1 and 3 (Suppl. Fig. S2). Together with both associations detected on chromosomes 1 and 3, at the median, the carrots that had alleles showing higher carotenoid content in both associations also had approx. 2.6-fold higher α-carotene and approx. 1.8-fold higher β-carotene contents in carrot surface compared to those with alleles showing lower carotenoid contents in both associations (Suppl. Fig. S2).

Possible candidate gene for the association detected on chromosome 3 by the GWAS and sequence comparison between parents in F_2 populations A and B

Significant associations were detected around the physical position at 5–6 Mb on chromosome 3 for α-carotene and β-carotene contents in F_2 population A and for visual evaluation in F_2 population B (Figs. 3, 4, Table 1). Within this region, the reported *Or* gene (DCAR_009172), which affects carotenoid contents in carrot [16], is located at 5.2 Mb on chromosome 3. To examine the involvement of *Or*, we performed Sanger sequencing of *Or* in the parents of populations A and B. The Sanger sequencing detected only one non-synonymous amino acid substitution at the fourth amino acid from the end, which was caused by an SNP between both parents of F_2 populations A and B (Fig. 6a). A thymine which was identical to that in

the carrot reference genome [10] in Fs001 and Fs003 was changed to guanine in Fs002, which resulted in a change from Tyr309 in the Fs001 and Fs003 to aspartic acid in the Fs002.

To examine the effect of the SNP on *Or* causing the non-synonymous amino acid substitution, we developed a KASP marker which could genotype the SNP. We applied the developed KASP marker to breeding line C whose root color was segregated and that is the progeny of Fs002 (Fig. 2). The root color of breeding line C was visually evaluated into three grades (Fig. 6c). The genotype of KASP marker on *Or* was clearly correlated with the visual evaluation (Fig. 6b). All of the carrots whose root color was bright and middle orange had a heterozygote for the SNP on *Or*, and all of the carrots whose root color was slightly light orange had a TT homozygote for the SNP. We thus speculate that the associations detected in F_2 population A for α-carotene and β-carotene contents on chromosome 3 and the association detected in F_2 population B for visual evaluation were responsible for the SNP causing the non-synonymous amino acid substitution on *Or*.

Possible candidate gene for the β/α-carotene ratio in population A, and the amino acid comparison between parents of population A

In the GWAS of F_2 population A, the association for the β/α-carotene ratio was detected on chromosome 6 and showed the highest association on the physical position at around 4.6 Mb. Iorizzo *et al*. [10] summarized the carrot orthologous and homologous candidate genes involved in the plastid 2-C-methyl-D-erythritol 4-phosphate (MEP) and carotenoid pathways in a table. According to the table, DCAR_022896 (which has a lycopene cyclase domain) is located on a physical position at 4.1 Mb on chromosome 6, which is between the SNP showing the highest association for the β/α-carotene ratio and the next SNP (Suppl. Table S2). Carotenoid biosynthesis bifurcates after lycopene to produce e- and β-carotenoids by enzymatic activity of the two lycopene cyclases, lycopene e-cyclase (LCYE) and lycopene β-cyclase (LCYB) [28] (Fig. 1). In addition, it is known that the proportions of β-carotene and α-carotene are determined mostly by the comparative amounts and/or activities of the LCYB and LCYE enzymes [20, 29–32].

However, the genes annotated as carrot *LCYB* and *LCYE* are not DCAR_022896 [10, 24]. Our BLAST search (NCBI; http://www.ncbi.nlm.nih.gov) of amino acids of DCAR_022896 showed sequence homology with neoxanthin synthase (NYS) and capsanthin-capsorubin synthase (CCS). The Phytozome database annotated DCAR_022896 as NYS, CCS, and lycopene cyclase (https://phytozome.jgi.doe.gov/pz/portal.html#). It is known that LCYB, NSY (which catalyzes violaxanthin into neoxanthin), CCS (which catalyzes the conversion of antheraxanthin and violaxanthin into capsanthin and capsorubin, respectively) (Fig. 1), and chromoplast-specific lycopene β-cyclase (CYC-B) have high sequence homology and similar putative catalytic mechanisms [20, 21, 33, 34].

Our phylogenic analysis of DCAR_022896, LCYE, LCYB, NSY, CCS, and CYC-B in carrot and *Arabidopsis* as well as *Solanum lycopersicum*, *Carica papaya*, *Citrus sinensis*, *Capsicum annuum*, and *Lillium lancifolium* showed that DCAR_022896 belonged to the same clade as CYC-B in *C. sinensis* and *C. papaya* (Fig. 7a). At the amino acid level, DCAR_022896 had 76.9% identity to CYC-B in *C. sinensis* and 62.1% to CYC-B in *C. papaya*. CYC-B is a *LCYB*, and it converts lycopene to β-carotene in chromoplasts, where carotenoids are accumulated [35, 36], in a specific manner [20] (Fig. 1). Moreover, our BLAST search of primers for the reported *LCYB2* in carrot showed that *CYC-B* (DCAR_022896) in the present study is identical to *LCYB2* [5, 6, 37, 38]. We thus presume that DCAR_022896 is a possible candidate gene for the β/α-carotene ratio, and we compared the amino acid sequences between the parents of F_2 population A by Sanger sequencing. The amino acid comparison revealed five amino acid substitutions between the parents of F_2 population A (Fig. 7b). These results suggested the possibility of the involvement of CYC-B in the β/α-carotene ratio in carrot root.

Discussion

Our GWAS using the two F_2 populations derived from orange root carrots detected 21 associations for visual color evaluation, color component a* and b*, α- and β-carotene content, lutein content, and the β/α-carotene ratio (Figs. 3, 4, Table 1). Some associations were detected on close physical positions for several evaluations of root color. However, interestingly, associations for visual evaluation in F_2 population A on chromosomes 4 and 5 were not detected for any other phenotypes. The Pearson correlation also showed no high correlation between visual color evaluation and other phenotypes (Table 2). These results suggest that we could not evaluate carrot root colors as same as experienced breeders by using spectrocolorimetry and HPLC. Experienced breeders evaluate root color comprehensively including the gloss and texture of the carrot surface, and thus the detected associations only for visual evaluation might be associated for these phenotypes.

Table 2

Pearson correlation between visual evaluation, color components and carotenoid contents of carrot root in F2 population A.

	Visual evaluation	L*	a*	b*	α-carotene	β-carotene
L*	0.079					
a*	0.093	0.674***				
b*	-0.167*	0.615***	0.899***			
α-carotene	-0.176*	0.367***	0.533***	0.537**		
β-carotene	0.048	0.333***	0.547***	0.508**	0.775***	
lutein	0.065	0.266*	0.265*	0.186*	0.354***	0.284**

(p: ***<0.0001; **<0.001; *<0.05)

On the other hand, the associations detected herein on chromosome 1 were significant for visual evaluation, color components a* and b*, and the α- and β-carotene contents (Fig. 3, Table 1). However, there are no annotated genes for MEP and carotenoid pathways within 5 Mb from the physical position of 31 Mb on chromosome 1 where the highest association was detected [10]. Similarly, a highly significant association was detected in F_2 population A for lutein content on chromosome 5 (Fig. 3, Table 1), whereas there are no predicted genes annotated for MEP and carotenoid pathways [10] or for chromatin-modifying histone methyltransferase, *SDG8* (*CCR1*), which affects the lutein content in leaves [39] around this locus except for *neoxanthin synthase* (*NSY*). The *NSY* gene is located approx. 1.1 Mb away from the physical position of the highest association for lutein content. NSY has a role downstream of another branch which does not include lutein in carotenoid biosynthesis (Fig. 1), and no feedback regulation between NSY and lutein content has been reported. However, we cannot exclude the possibility that the mutation of the *NSY* of another branch affects the flow rate of each branch, resulting in an effect on the lutein content. Further analyses such as a map-based strategy is necessary to narrow down the candidate regions and identify candidate genes causing the associations detected on chromosome 1 for visual evaluation, color components a* and b*, and α- and β-carotene contents and chromosome 5 for lutein content, and for the other significant associations revealed in this study.

The GWAS in F_2 population A for the α- and β-carotene content and in F_2 population B for visual evaluation detected the association at a similar physical position around 5–6 Mb on chromosome 3 (Figs. 3, 4, Table 1), and the previously reported *Or* was located on a similar physical position (5.2 Mb). The *Or* gene is involved in carotenoid accumulation via chromoplast development and biosynthesis via *15-cis-phytoene synthase* (*PSY*) expression, and it affects the total carotenoid content [40, 41]. Carrot *Or* was recently identified and is associated with the carotenoid presence in carrot root [16]. The similarity of

physical positions and function of *Or* suggests that these associations were caused by *Or*, in addition, an SNP causing a non-synonymous mutation was detected in the present study between the parents of F_2 populations A and B by Sanger sequencing (Fig. 6a). Iorizzo *et al*. [10] re-sequenced 35 carrot accessions and released SNPs on Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). In the released SNPs, there are many SNPs on the *Or* gene, including three SNPs causing a non-synonymous mutation. However, the SNP that causes the non-synonymous mutation detected in Fs002 is not included, suggesting that this is a new allele of *Or* and that a distribution of the SNP detected in this study might be limited. Further analyses of the distribution of this SNP are needed. As breeding line C was derived from Fs002 (Fig. 2), a bright orange allele would be derived from the Fs002. Peaks of the Manhattan plots of this region were also observed for visual color evaluation (Fig. 3a), color components L* (Fig. 3b) and b* (Fig. 3d), and lutein content (Fig. 3g). Although their associations were not significant, these peaks might be caused by *Or*.

The GWAS in F_2 population A revealed the association for the β/α-carotene ratio on chromosome 6; *CYC-B* is located on the associated region (Fig. 3h, Suppl. Table S2). In several model plants such as *Arabidopsis thaliana* [42], rice (*Oryza sativa*) [43], and maize (*Zea mays*) [29, 44], LCYB is encoded by a single gene. However, LCYB is encoded by two genes in some plant species that accumulate high levels of carotenoids in non-photosynthetic organs, such as fruits and flowers [38]. These genes are differentially expressed in photosynthetic and non-photosynthetic organs, and genes that are expressed in non-photosynthetic organs were named *CYC-B*. As named, *CYC-B* is a chromoplast-specific lycopene β-cyclase.

Carrot has two LCYBs: LCYB1 and LCYB2 [3]. Our present phylogenetic analysis demonstrated that carrot LCYB1 (LCYB) and LCYB2 (CYC-B) belong to LCYB and CYC-B clades, respectively (Fig. 7a). The *CYC-B* was first reported in tomato (*Solanum lycopersicum*) as a fruit- and flower-specific lycopene β-cyclase using two mutations named *Beta* and *old-gold*. *Beta* increases the β-carotene content in fruit, and *old-gold*, a null mutation of *CYC-B*, abolishes β-carotene and increases lycopene contents in fruits and caused tawny orange flowers [20]. The *CYC-B* has also been reported to be responsible for fruit color in papaya (*Carica papaya*) [22] and citrus (*Citrus sinensis*) and for the involvement of the null allele in the high lycopene accumulation in red grapefruits [23]. In carrot, unlike plants that have organ-specific LCYBs, *LCYB1* is expressed in both leaves and root, and the transcript level of *LCYB1* increases as the carotenoid content increases during root development [38, 45]. Since the GWAS detected an association around the *CYC-B* region in this study, we speculate that in carotenoid-accumulating carrot root, in addition to the LCYB (LCYB1), CYC-B (LCYB2) would also have a role in carotenoid biosynthesis. To the best of our knowledge, the involvement of *CYC-B* (*LCYB2*) in carotenoid contents in carrot root has not been examined. A chromoplast-specific carotenoid biosynthesis pathway including *CYC-B* such as that reported in tomato [46] might have a role in carrot roots. Further functional analyses such as an expression study of *CYC-B* (*LCYB2*) in several organs and an investigation of the subcellular location of CYC-B (LCYB2) in carrot are necessary to clarify the functions of the two LCYBs.

Visual appearance traits are important targets in carrot breeding in Japan, and the 'best bright orange color' is selected based on a comparison of minute color differences as shown in Figure 6c. The differences are actually difficult for non-specialists to detect, but the resultant selected cultivars attract consumers in Japan with their 'best bright orange color'. The present study provides the first results of a GWAS analysis for carrot root color for the selection of bright orange color in orange root populations. The developed KASP marker on *Or* as well as the SNPs showing significant associations will contribute to orange carrot breeding.

Declarations

Data availability

Nucleotide sequence data for the ddRADseq in F_2 population A and B is available in the DDBJ Sequence Read Archive under accession numbers from DRA012848 to DRA012853.

Acknowledgements

We are grateful to Dr. K. Shirasawa at the Kazusa DNA Research Institute for technical advises, and Ms. S. Sasamoto, Ms. R. Aomiya, and Ms. T. Shibazaki at the Kazusa DNA Research Institute for their technical assistance. We are also grateful to Mr. H. Hirose, and Mr. Y. Yoshida and Mr. S. Yoshida at Yoshida Seed for their carrot-field management, and to all staff members at Fujii Seed for preparing carrot materials.

Author contributions

T.S. designed the experiment, performed GWAS, and wrote the manuscript; S.I. provided direction for the study, designed the experiment, and correction to the manuscript; H.T., C.M. and A.O. performed the experiment; S.N. performed NGS data analysis; C.K. and T.F. provided plant materials and evaluated phenotypes. All authors read and approved the final manuscript.

Competing Interests:

The authors declare no competing interests.

References

1. Simon, P., Pollak, L., Clevidence, B., Holden, J. & Haytowitz, D. Plant breeding for human nutrition. *Plant Breed. Rev.* **31**, 325-392 (2009).

2. Krinsky, N. I., Johnson, E. J. Carotenoid actions and their relation to health and disease. *Mol. Aspects Med*. **26**, 459-516 (2005).
3. Just, B. J., Santos, C. A. F., Fonseca, M. E. N., Boiteux, L. S., Oloizia, B. B., & Simon, P. W. Carotenoid biosynthesis structural genes in carrot (Daucus carota): Isolation, sequence-characterization, single nucleotide polymorphism (SNP) markers and genome mapping. *Theor. Appl. Genet*. **114**, 693-704. doi: 10.1007/s00122-006-0469-x. (2007).
4. Santos, C. A. F. & Simon, P. W. QTL analyses reveal clustered loci for accumulation of major provitamin A carotenes and lycopene in carrot roots. *Mol. Genet. Genomics*. **268**, 122-129 (2002).
5. Just, B. J., Santos, C. A. F., Yandell, B. S. & Simon, P. W. Major QTL for carrot color are positionally associated with carotenoid biosynthetic genes and interact epistatically in a domesticated × wild carrot cross. *Theor. Appl. Genet*. **119**, 1155-1169 (2009).
6. Jourdan, M., *et al*. Carotenoid content and root color of cultivated carrot: A candidate-gene association study using an original broad unstructured population. *PLOS ONE.* **10**, e0116674. doi: 10.1371/journal.pone.0116674 (2015).
7. Ruiz-Sola, M. Á. & Rodríguez-Concepción M. Carotenoid biosynthesis in *Arabidopsis*: A colorful pathway. *Arabidopsis Book*. **10**, e0158. doi: 10.1199/tab.0158 (2012).
8. Stanley, L., Yuan, Y.-W. Transcriptional regulation of carotenoid biosynthesis in plants: so many regulators, so little consensus. *Front. Plant Sci*. **10**, 1017. doi: 10.3389/fpls.2019.01017 (2019).
9. Al-Babili, S. *et al*. Identification of a novel gene coding for neoxanthin synthase from *Solanum tuberosum*. *FEBS Lett*. **485**, 168-172 (2000).
10. Iorizzo, M. *et al*. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **48**, 657-666 (2016).
11. Wang, F., Wang, G.-L., Hou, X.-L., Li, M.-Y., Xu, Z.-S. & Xiong, A.-S. The genome sequence of 'Kurodagosun', a major carrot variety in Japan and China, reveals insights into biological research and carrot breeding. *Mol. Genet. Genomics*. **293**, 861-871 (2018).
12. Arango, J., Jourdan, M., Geoffriau, E., Beyer, P. & Welsch, R. Carotene hydroxylase activity determines the levels of both α-carotene and total carotenoids in orange carrots. *Plant Cell*. **26**, 2223-2233 (2014).
13. Buishand, J. & Gabelman, W. Investigations on the inheritance of color and carotenoid content in phloem and xylem of carrot roots (*Daucus carota* L.). *Euphytica* **28**, 611-632 (1979).
14. Ellison, S., Senalik, D. Bostan, H, Iorizzo, M. & Simon, P. Fine mapping, transcriptome analysis, and marker development for Y2, the gene that conditions β-carotene accumulation in carrot (*Daucus carota* L.). *G3 (Bethesda)*. **7**, 2665-2675. doi: 10.1534/g3.117.043067 (2017).
15. Li, L., Paolillo, D., Parthasarathy, M., Dimuzio, E. & Garvin, D. A novel gene mutation that confers abnormal patterns of β-carotene accumulation in cauliflower (*Brassica oleracea* var. *Botrytis*). *Plant J*. **26**, 59-67 (2001).
16. Ellison, S. L. *et al*. Carotenoid presence is associated with *Or* gene in domesticated carrot. *Genetics*. **210**, 1497-1508 (2018).

17. Shirasawa, K., Hirakawa, H., & Isobe, S. Analytical workflow of double-digest restriction site-associated DNA sequencing based on empirical and in silico optimization in tomato. *DNA Res*. **23**, 145-153 (2016).
18. Arafa, R. A. *et al*. Rapid identification of candidate genes for resistance to tomato late blight disease using next-generation sequencing technologies. *PLOS ONE.* **12**, e0189951. doi: 10.1371/journal.pone.0189951 (2017).
19. Bradbury, P. J. *et al*. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*. **23**, 2633-5 (2007).
20. Ronen, G., Carmel-Goren, L., Zamir, D., & Hirschberg J. An alternative pathway to β-carotene formation in plant chromoplasts discovered by map-based cloning of *Beta* and *old-gold* color mutations in tomato. *Proc. Natl. Acad. Sci*. **97**, 11102-11107 (2000).
21. Jeknić, Z. *et al*. Cloning and functional characterization of a gene for capsanthin-capsorubin synthase from tiger lily (*Lilium lancifolium* Thunb. 'Splendens'). *Plant Cell Physiol*. **53**, 1899-1912. doi: 10.1093/pcp/pcs128 (2012).
22. Blas, A. L. *et al*. Cloning of the papaya chromoplast-specific lycopene b-cyclase, CpCYC-b, controlling fruit flesh color reveals conserved microsynteny and a recombination hot spot. *Plant Physiol*. **152**, 2013-2022. doi: 10.1104/pp.109.152298 (2010)
23. Alquézar, B., Zacarías, L., Rodrigo, M. J. Molecular and functional characterization of a novel chromoplast-specific lycopene b-cyclase from *Citrus* and its relation to lycopene accumulation. *J. Exp. Bot*. **60**, 1783-1797 (2009).
24. Simon, P.W., Geoffriau, E., Ellison, S., & Iorizzo, M. Carrot carotenoid genetics and genomics, in *The carrot genome* (eds. Simon, P., Iorizzo, M., Grzebelus, D., Baranski, R.) 247-260 (Springer, 2019).
25. Thompson, J. D., Higgins, D. G., & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680 (1994).
26. Saitou, N., & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. **4**, 406-425 (1987).
27. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol*. **35**, 1547-1549 (2018).
28. Cazzonelli, C. I, & Pogson, B. J. Source to sink: Regulation of carotenoid biosynthesis in plants. *Trends Plant. Sci.* **15**, 266-274 (2010).
29. Bai, L., Kim, E.-H., DellaPenna, D., Brutnell, T. P. Novel lycopene epsilon cyclase activities in maize revealed through perturbation of carotenoid biosynthesis. *The Plant J.* **59**, 588-599 (2000).
30. Pogson, B., McDonald, K. A., Truong, M., Britton, G., DellaPenna, D. Arabidopsis carotenoid mutants demonstrate that lutein is not essential for photosynthesis in higher plants. *Plant Cell*. 8, 1627-1639 (1996).
31. Ronen, G., Cohen, M., Zamir, D., Hirschberg, J. Regulation of carotenoid biosynthesis during tomato fruit development: Expression of the gene for lycopene epsilon-cyclase is down-regulated during

ripening and is elevated in the mutant *Delta*. *The Plant J*. **17**, 341-351 (1999).

32. Harjes, C. E. *et al*. Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. *Science*. **319**, 330-333. doi: 10.1126/science.1150255 (2008).
33. Bouvier, F., d'Harlingue, A., Backhous, R. A., Kumagai, M. H. & Camara, B. Identification of neoxanthin synthase as a carotenoid cyclase paralog. *Eur. J. Biochem*. **267**, 6346-6352 (2000).
34. Mialoundama, A. S. *et al*. Characterization of plant carotenoid cyclases as members of the flavoprotein family functioning with no net redox change. *Plant Physiol*. **153**, 970-979 (2010).
35. Egea, I. *et al*. Chromoplast differentiation: current status and perspectives. *Plant Cell Physiol*. 51, .1601-1611 (2010).
36. Kim, J. E., Rensing, K. H., Douglas, C. J., Cheng, K. M. Chromoplasts ultrastructure and estimated carotene content in root secondary phloem of different carrot varieties. Planta. **231**, 549-558 (2010).
37. Fuentes, P., Pizarro, L., Moreno, J. C., Handford, M., Rodriguez-Concepcion, M., Stange, C. Light-dependent changes in plastid differentiation influence carotenoid gene expression and accumulation in carrot roots. *Plant Mol. Biol*. **79**, 47-59 (2012).
38. Moreno, J. C., Pizarro, L., Fuentes, P., Handford, M., Cifuentes, V., Stange, C. Levels of lycopene b-cyclase 1 modulate carotenoid gene expression and accumulation in *Daucus carota*. *PLOS ONE.* **8**, e58144. doi: 10.1371/journal.pone.0058144 (2013).
39. Cazzonelli, C. I. *et al*. Regulation of carotenoid composition and shoot branching in Arabidopsis by a chromatin modifying histone methyltransferase, SDG8. *Plant Cell*. **21**, 39-53. doi: 10.1105/tpc.108.063131 (2009).
40. Lu, S. *et al*. The cauliflower *Or* gene encodes a DnaJ cysteine-rich domain containing protein that mediates high levels of β-carotene accumulation. *The Plant Cell*. 18, 3594-3605 (2006).
41. Zhou, X. *et al*. Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. *Proc. Natl. Acad. Sci. USA.* **112**, 3558-3563 (2015),
42. Lange, B. M., Ghassemian, M. Genome organization in Arabidopsis thaliana: A survey for genes involved in isoprenoid and chlorophyll metabolism. *Plant Mol. Biol*. **51**, 925-948 (2003).
43. Fang, J. *et al*. Mutations of genes in synthesis of the carotenoid precursors of ABA lead to pre-harvest sprouting and photo-oxidation in rice. *Plant J*. **54**, 177-189 (2008).
44. Singh, M. *et al*. Activator mutagenesis of the pink scutellum1/viviparous7 locus of maize. *Plant Cell*. **15**, 874-884 (2003).
45. Fuentes. P. *et al*. Light-dependent changes in plastid differentiation influence carotenoid gene expression and accumulation in carrot roots. *Plant Mol. Biol*. **79**, 47-59. doi: 10.1007/s11103-012-9893-2 (2012).
46. Galpaz, N., Ronen, G., Khalfa, Z., Zamir, D., Hirschberg, J. A chromoplast-specific carotenoid biosynthesis pathway is revealed by cloning of the tomato *white-flower* locus. *Plant Cell*. **18**, 1947-1960 (2006).

Figures

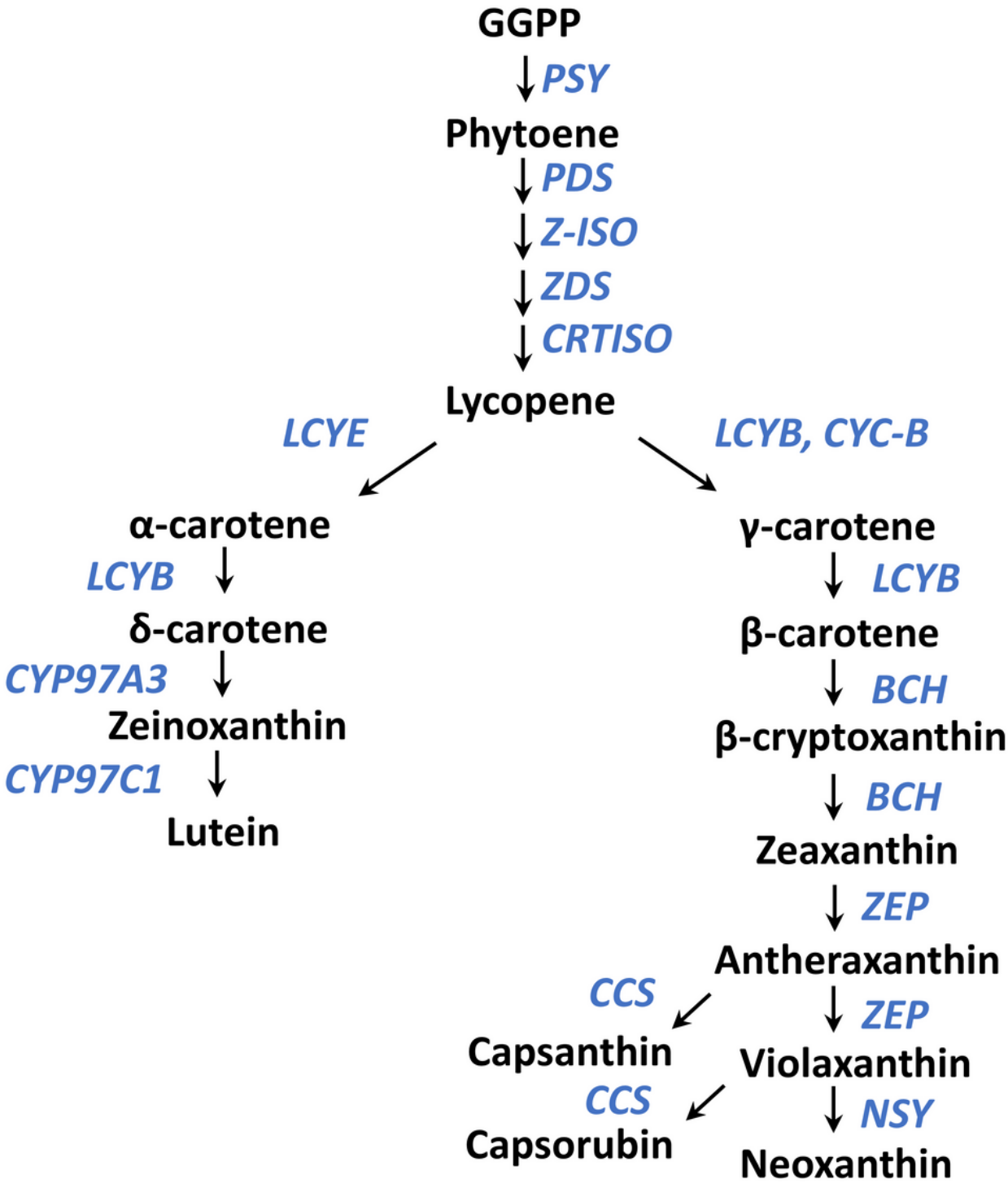

Figure 1

Carotenoid biosynthesis pathways. The carotenoid biosynthesis pathway is shown in black, with carotenoid biosynthesis genes indicated in blue. Figure compiled and summarized from Stanley et al. [8] and Al-Babili et al. [9].

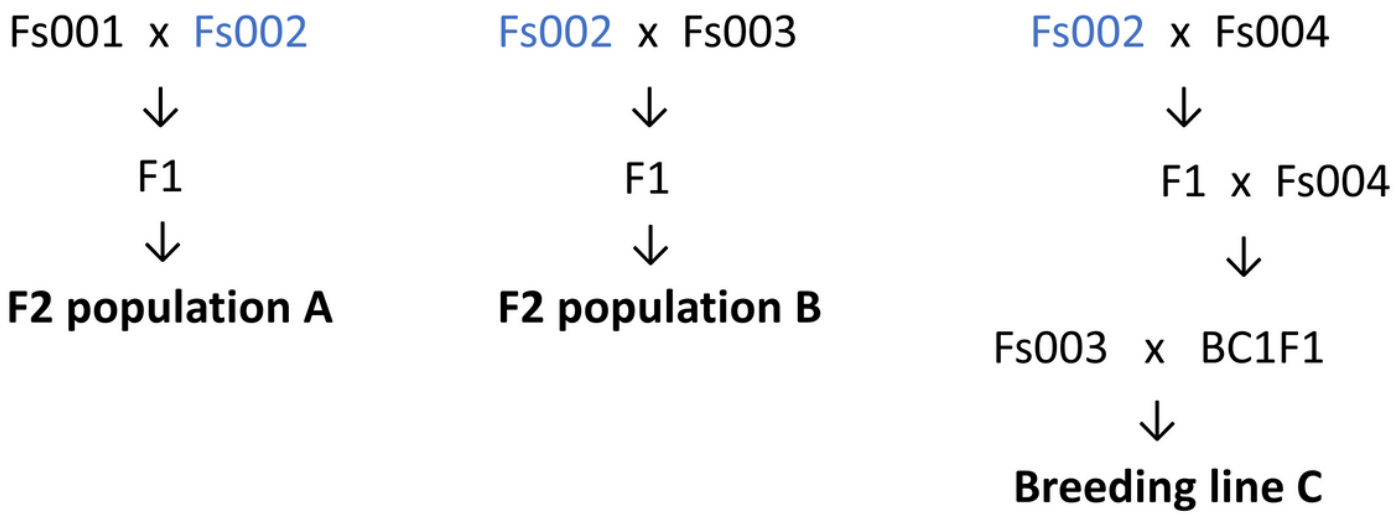

Figure 2

Lineage images of the plant materials, F2 populations A and B, and breeding line C. Fs002 was used as a common breeding material.

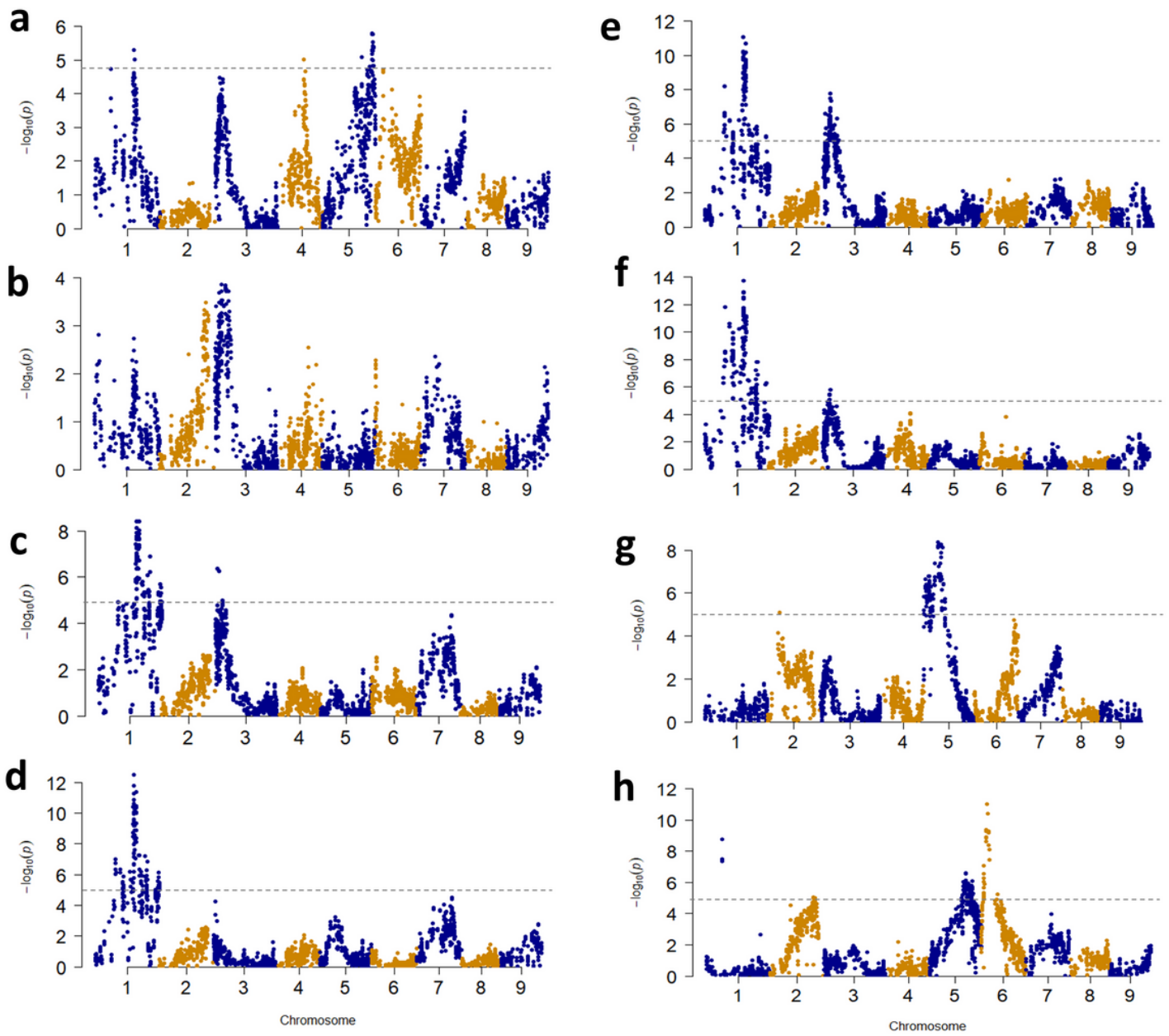

Figure 3

Manhattan plots for carrot taproot color in F2 population A. Plots for the visual evaluation (a), L* (b), a* (c), b* (d), α-carotene (e), β-carotene (f), lutein (g), and the β/α-carotene ratio (h). Horizonal line indicates the Bonferroni correction (0.05).

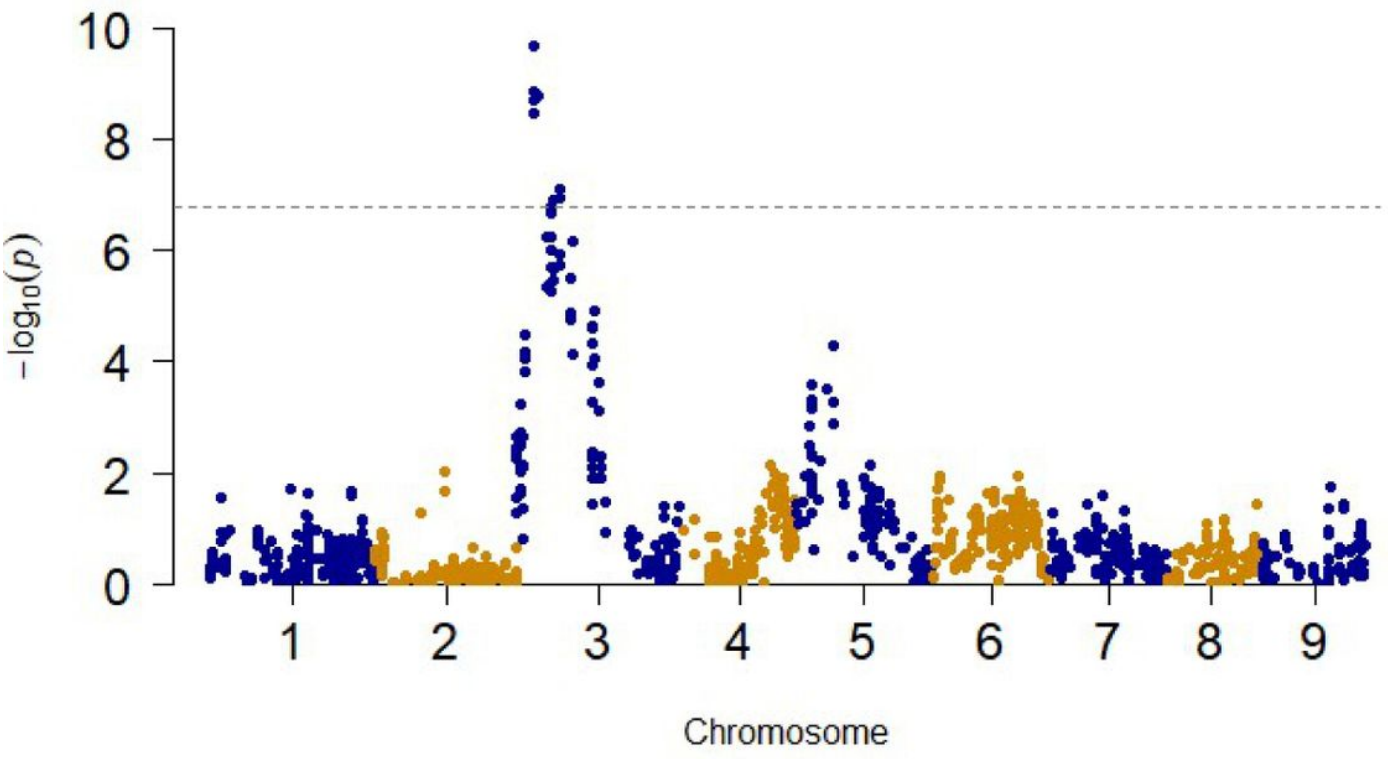

Figure 4

Manhattan plots for carrot taproot color in F2 population B. Plots for the visual evaluation of carrot root color. Horizonal line: Bonferroni correction (0.05).

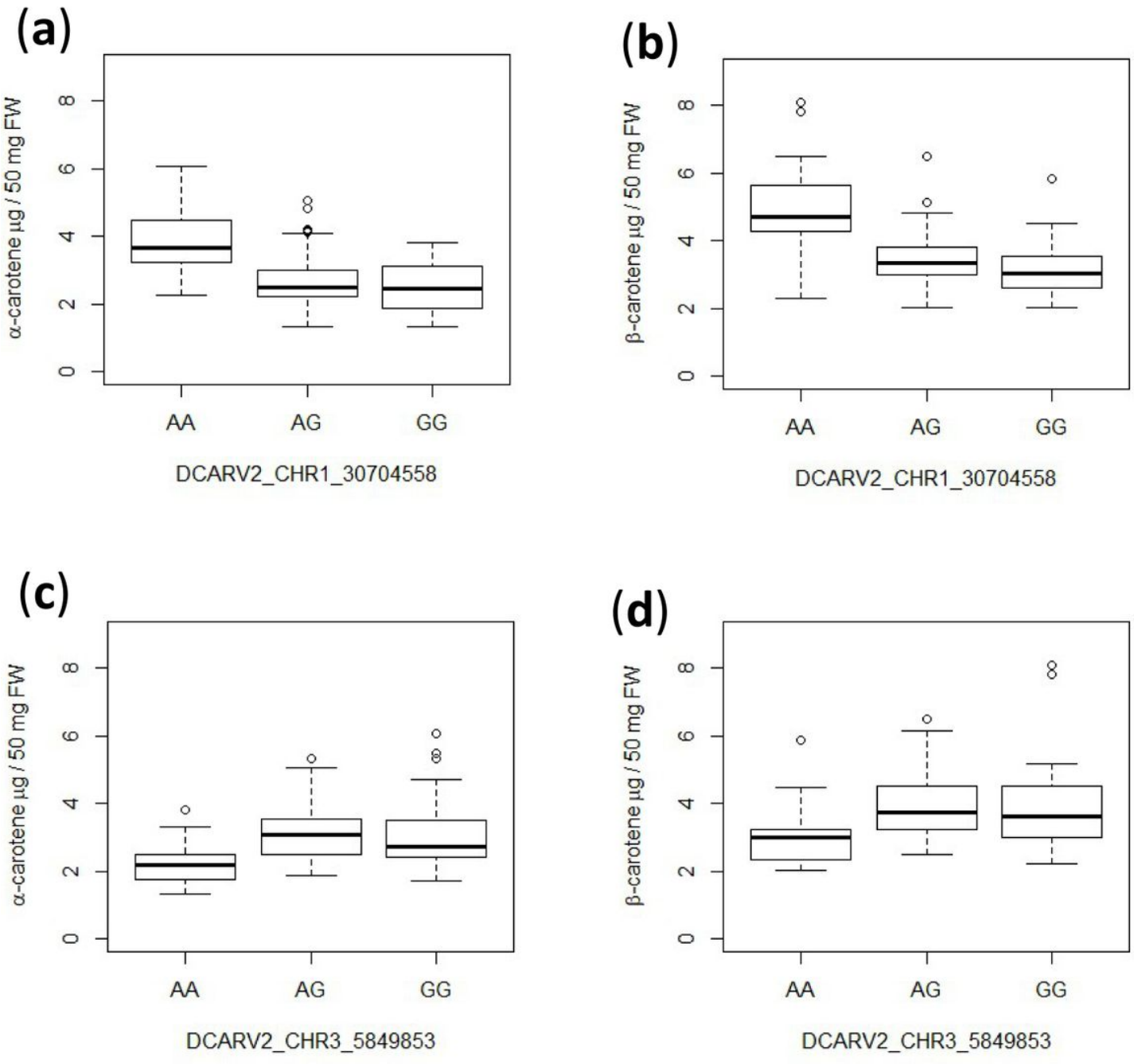

Figure 5

Allelic effects of associations detected by the GWAS on chromosome 1 (a, b) and chromosome 3 (c, d) for the α-carotene (a, c) and β-carotene (b, d) content in carrot root surface. Carotene contents were box-plotted by the SNP showing the highest –log10P in the GWAS for α-carotene content. The numbers of plants were 30, 81 and 34 for the AA, AG, and GG SNPs respectively in panels a and b, and 40, 70, 35 for the AA, AG and GG SNPs respectively in panels c and d.

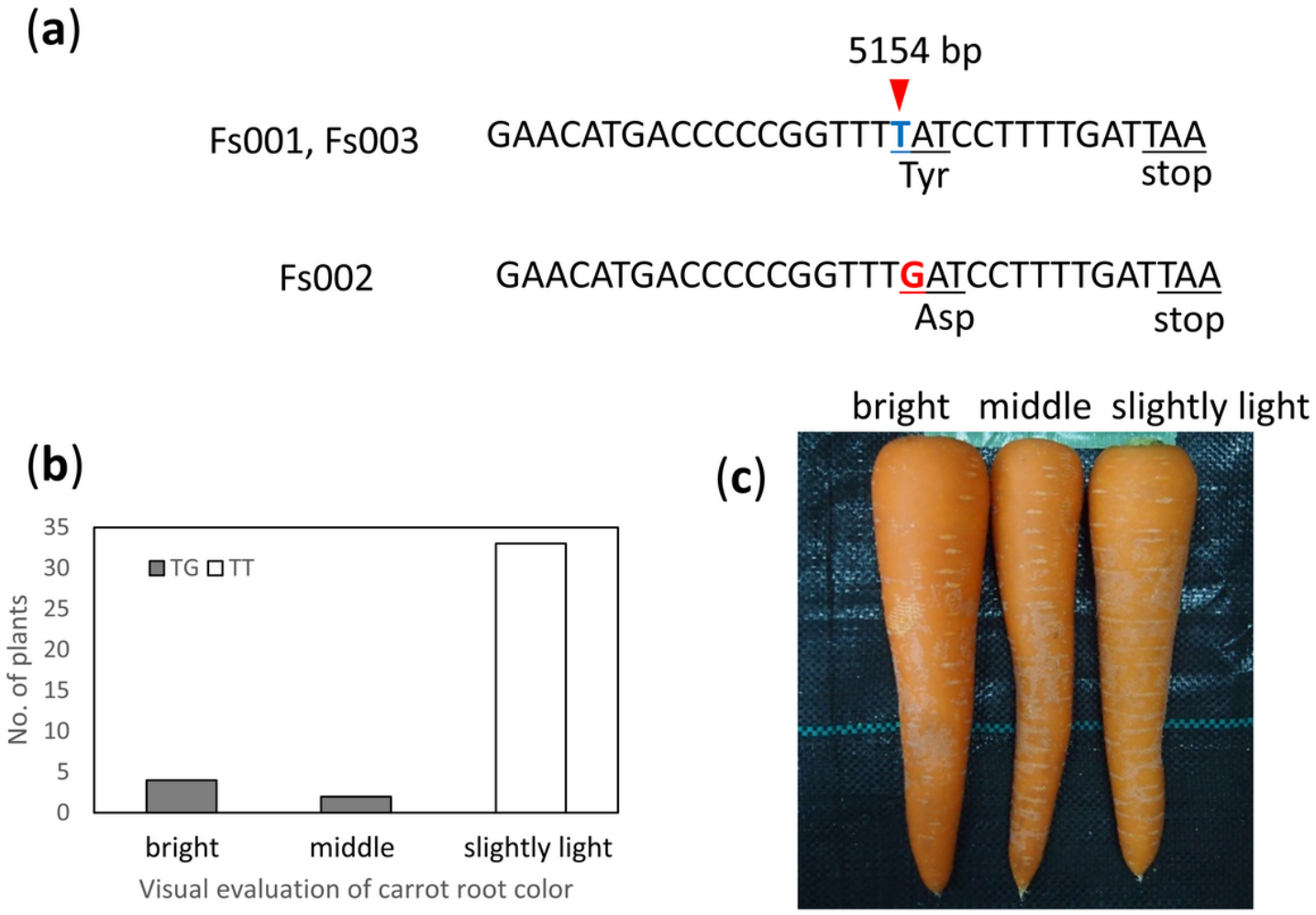

Figure 6

The SNP on Or and the examination of its effect on carrot root color in breeding line C. (a) The SNP detected on Or between parental plants in F2 populations A and B. The SNP causes an amino acid substitution. The upper sequence is identical to the reference sequence of Iorizzo et al. [10], and the lower sequence is a new allele of Or. (b) The allelic effect of the SNP on Or in another breeding line at Fujii Seed. TT and TG showed TT homozygote and heterozygote of the Or SNP, respectively. Carrot root color was visually evaluated as three grades: brilliant, middle, and pale orange color (b, c). All of the plants with bright or middle orange color roots showed a heterozygote for the Or SNP, and all plants with the slightly light orange color roots showed TT homozygote for the Or SNP. (c) Examples of carrot root color visually evaluated as three grades.

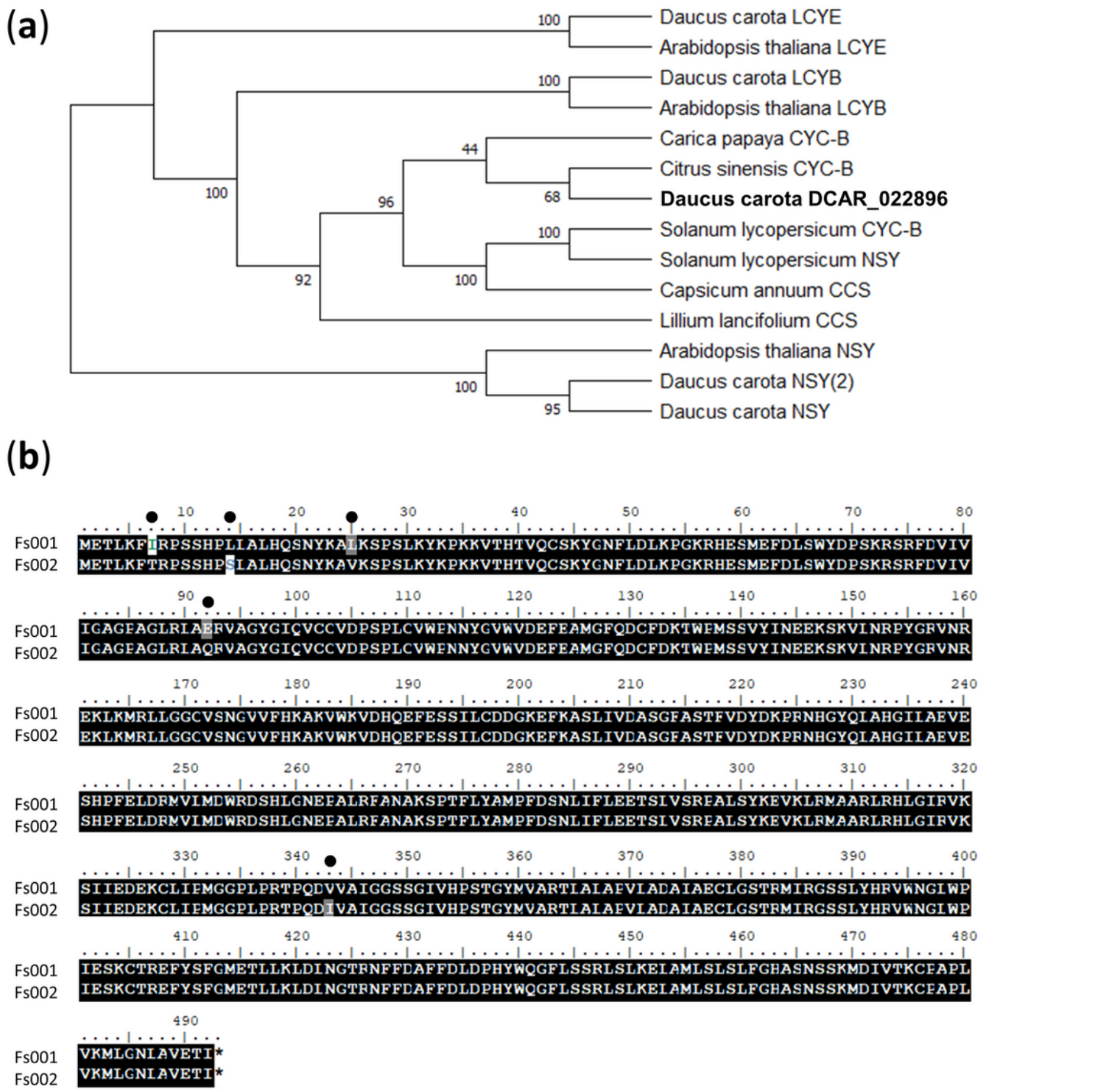

Figure 7

The phylogenetic tree based on the amino acid sequences of LCYE, LCYB, CYC-B, NSY, and CCS, and the amino acid substitutions between the parental plants of F2 population A on DCAR_022896. (a) The phylogenetic tree was drawn by the neighbor-joining method based on the amino acid sequences of Daucus carota DCAR_022896, D. carota LCYE (DCAR_028276), Arabidopsis thaliana LCYE (At5g57030), D. carota LCYB (DCAR_020544), A. thaliana LCYB (At3g10230), Solanum lycopersicum CYC-B, Carica

papaya CYC-B (evm.model.supercontig_195.16), Citrus sinensis CYC-B (orange1.1g010693m.g), D. carota NSY (DCAR_017191), D. carota NSY(2) (DCAR_025914), A. thaliana NSY (At1g67080), S. lycopersicum NSY (CAB93342.1), Capsicum annuum CCS (Q42435.1), and Lillium lancifolium CCS (JF304153). Values at the nodes indicate the percentage consensus support as calculated using a bootstrapping test with 1,000 replications. (b) Five amino acid substitutions (●) were detected between seed (Fs001) and pollen (Fs002) parents of F2 population A on DCAR_022896. Similar substitutions are shown in gray background.

Supplementary Files

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

- supplementaryfigures.pdf
- Supplementarytables.xlsx