Theor Appl Genet (2007) 114:693–704
DOI 10.1007/s00122-006-0469-x

ORIGINAL PAPER

Carotenoid biosynthesis structural genes in carrot (*Daucus carota*): isolation, sequence-characterization, single nucleotide polymorphism (SNP) markers and genome mapping

B. J. Just · C. A. F. Santos · M. E. N. Fonseca · L. S. Boiteux · B. B. Oloizia · P. W. Simon

Received: 4 February 2006 / Accepted: 17 November 2006 / Published online: 22 December 2006
© Springer-Verlag 2006

Abstract Carotenoid pigments are important components of the human diet and carrots are the main dietary sources of the vitamin A precursors α- and β-carotene. Carotenoids play essential biological roles in plants and the genes coding for the carotenoid pathway enzymes are evolutionarily conserved, but little information exists about these genes for carrot. In this study, we utilized published carrot sequences as well as heterologous PCR approaches with primers derived from sequence information of other plant species to isolate 24 putative genes coding for carotenoid biosynthesis enzymes in carrot. Twenty-two of these genes were placed on the carrot genetic linkage map developed from a cross between orange-rooted and white-rooted carrot. The carotenoid genes were distributed in eight of the nine linkage groups in the carrot genome recommending their use for merging maps. Two genes co-localized with a genomic region spanning one of the most significant quantitative trait loci (QTL) for carotenoid accumulation. Carotenoid biosynthesis cDNAs linked to root color mutations and to QTL for carotenoid accumulation may suggest a functional role for them as candidate genes. RACE PCR and reverse transcriptase PCR were used to amplify the full-length transcript for twenty expressed carotenoid biosynthesis genes and sequences were submitted to GenBank. The cloning and sequence information of these genes is useful for PCR-based expression studies and may point toward transgenic approaches to manipulate carotenoid content in carrot.

Communicated by S. J. Knapp.

B. J. Just · C. A. F. Santos · M. E. N. Fonseca · L. S. Boiteux · B. B. Oloizia · P. W. Simon
Plant Breeding and Plant Genetics Program, and Department of Horticulture, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706, USA

C. A. F. Santos
Embrapa-Semi-Arido, CP 23, 56300-970 Petrolina-PE, Brazil

M. E. N. Fonseca · L. S. Boiteux
Embrapa-Hortaliças, National Center for Vegetable Crops Research (CNPH), CP 218, 70359-970 Brasília-DF, Brazil

P. W. Simon (✉)
USDA-Agricultural Research Service, Vegetable Crops Research Unit, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706, USA
e-mail: psimon@wisc.edu

Introduction

Carotenoid pigments provide red, yellow and orange colors and antioxidant protection to a wide variety of plants, animals, bacteria, and fungi. In plants, carotenoids play a protective role in photosynthesis by dissipating excess light energy absorbed by the photosynthetic mechanism (Demmig-Adams and Adams 2002). They also serve as precursors to other biologically important compounds (Millborrow 2001; Bouvier et al. 2003; Fester et al. 2002; Giuliano et al. 2003).

In human nutrition some carotenoid pigments (e.g. β-carotene, α-carotene, cryptoxanthin) are converted to vitamin A after molecular cleavage followed by reduction. Other carotenoids (e.g. lycopene, lutein, and zeaxanthin) function as powerful antioxidants and have

been linked to reductions in probabilities of developing certain types of cancer, vascular disease, and visual impairment (Tanumihardjo and Yang 2005).

Carrots (*Daucus carota* var. *sativus* L.) ($2n = 2x = 18$) are an excellent source of dietary carotenoid pigments. Orange carrots are colored primarily by α- and β-carotene and provide up to 30% of the provitamin A consumed in the United States (Simon 2000). In contrast, Queen Anne's Lace (*D. carota* var. *carota*) is a wild form of carrot generally believed to be the ancestor of the domesticated carrot and its storage root does not accumulate any detectable levels of carotenoid pigments.

The genes and enzymes involved in the biosynthesis of carotenoid pigments have been extensively reviewed (Bartley and Scolnik 1995; Cunningham and Gantt 1998; Hirschberg 2001) and the carotenoid pathway is highly conserved in plants and photosynthetic bacteria, though very few plant species accumulate detectable levels in roots. Recently two additional enzymes involved in carotenoid biosynthesis were discovered. An ε-ring hydroxylase from *Arabidopsis* was identified through positional cloning of the *Lut1* gene. The enzyme is a cytochrome P-450 monooxygenase and the authors identified putative orthologs in *Oryza sativa* and *Hordeum vulgare* (Tian et al. 2004). Another recently discovered gene of the carotenoid pathway codes for the carotenoid isomerase (CRTISO) enzyme that is involved in the conversion of poly-*cis* lycopene to *trans*-lycopene. This gene has been cloned in both *Lycopersicon* and *Arabidopsis* (Isaacson et al. 2002; Park et al. 2002). Carotenoid cleavage dioxygenases (CCDs) and 9-*cis*-epoxycarotenoid dioxygenases (NCEDs) cleave carotenoids into secondary products with a wide diversity of functions. Several CCDs and NCEDs have been cloned from plants (Schwartz 2001; Tan et al. 2003). A simplified schematic view of the major steps in the carotenoid pathway and the corresponding genes/enzymes is shown in Fig. 1.

Linkage maps have been created for genes in biosynthetic pathways in several species and these maps can be used to locate known genes of a pathway within a specific genomic region (e.g. Thorup et al. 2002; Chen et al. 2001). These "functional markers" represent allelic variants of genes with a known biochemical function being, therefore, useful in candidate gene approaches to isolate the factor(s) controlling a trait potentially related to a specific pathway (Pflieger et al. 2001; Varshney et al. 2005). A collection of naturally occurring single-locus mutations affecting carotenoid accumulation in the root tissue is available in carrot (Gabelman and Peters 1979; Simon 1996). They include dominant alleles such as *A* (α-carotene accumulation), *Io* (intense orange xylem, which may be an allelic form of *A*), L_1 and L_2 (lycopene accumulation), *O* (orange xylem, which may also be an allelic form of *A*) as well as the recessives alleles *y* (yellow xylem) and *rp* (reduced pigmentation). The *Y*, Y_1, and Y_2 loci control differential distribution of α- and β-carotene (xylem/phloem carotene levels) where the mutation Y_2 controls low carotene content of the storage root xylem ("core") in high carotene orange backgrounds (Simon 1996).

Several carrot genetic linkage maps have been published (Schulz et al. 1994; Vivek and Simon 1999; Santos 2001; Santos and Simon 2004; Westphal and Wricke 1989, 1991, 1997). These maps include several types of marker (isozymes, RFLPs, microsatellites, SAMPLs, and morphological markers). Most of these maps have no markers in common and therefore cannot be compared. A notable exception is presented in Santos and Simon (2004) where maps for six linkage groups from two populations (B493 × QAL and Brasilia × HCM) were merged using two codominant markers and 28 sequenced dominant AFLPs with high sequence similarity between the populations. Though PCR-based codominant markers in carrot have been published (Niemann et al. 1997; Boiteux et al. 2000, 2004; Yau and Simon 2003; Bradeen and Simon 1998; Vivek and Simon 1999), their numbers are small and their usefulness across unrelated populations is limited (Santos 2001; Bradeen and Simon 2006). Sequence tagged sites (STS) markers have not been developed for carrot, but they have been used in legumes and other crops for the purpose of creating linkage maps from different crosses that can be compared to one another (Brauner et al. 2002). This approach is desirable because the markers are codominant, and if SNP detection is used for genotyping, there is a high probability of finding polymorphism in many populations. They may also serve as candidate genes for traits.

The purpose of this study is to identify putative carotenoid biosynthetic gene sequences in carrot and place them as STS markers on a carrot linkage map that includes AFLPs marking carotenoid accumulation QTL. These markers can serve as codominant anchor loci for linking current, past and future carrot linkage maps. Additionally, these STS markers can be used as candidate genes for QTL for differential carotenoid accumulation or for major carrot color genes affecting carotenoid accumulation in carrot such as *Y*, Y_1, Y_2, *Rp*, L_1, and L_2. Another goal was to initiate evaluations of carotenoid gene expression in root tissue of orange carrots. Our final goal was to obtain full cDNA coding region sequences for the genes qualitatively expressed in orange carrot root tissue.

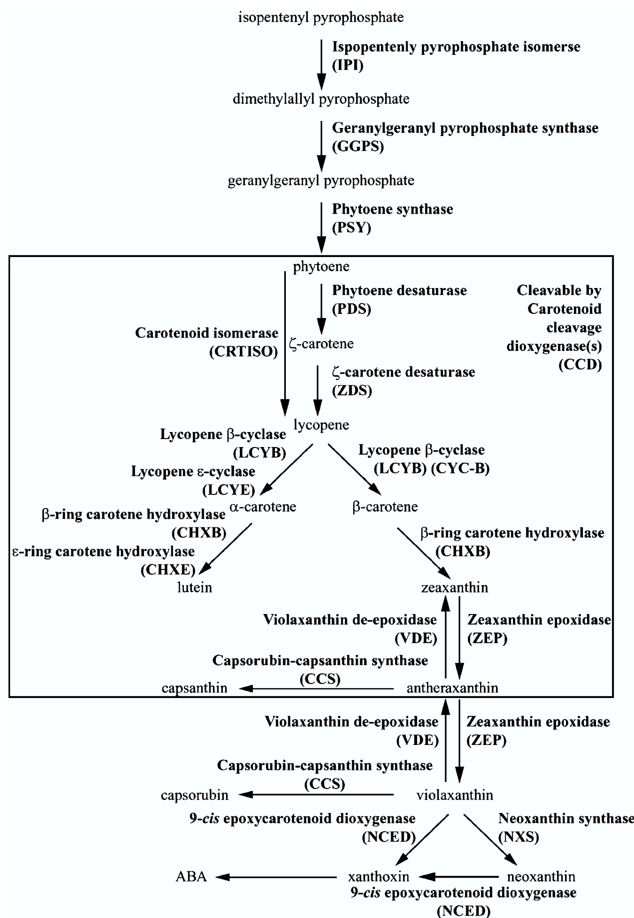

Fig. 1 A simplified diagram of the enzymes and major products in the carotenoid pathway, parts of which are common to all plants. The product names appear between the *arrows* and are written in *regular type*. The enzyme names are written in *boldface* with the abbreviation used throughout this publication below in *parentheses*. The products appearing in the *large box* can be cleaved into apocarotenoid products by a class of enzymes called carotenoid cleavage dioxygenases (*CCDs*). Figure compiled and summarized from Liu et al. (2004), Giuliano et al. (2000), Hirschberg (2001) and Cunningham and Gantt (1998)

Materials and methods

Mapping population and DNA extraction

The B493 × QAL F_2 population (Santos 2001; Santos and Simon 2002, 2004) was used for mapping STS markers. B493 is a dark orange USDA inbred carrot and QAL is a white wild carrot (*D. carota* var. *carota*). The population evaluated was derived from crossing a single B493 plant with a single QAL plant in Madison, WI, in 1989. A single F_1 plant was self-pollinated to produce the F_2 generation used for mapping. A total of 183 F_2 plants grown in field conditions in 1998 were included in the study. Total genomic DNA was extracted from freeze-dried leaf tissue following the protocol of Doyle and Doyle (1990) with modifications by Boiteux et al. (1999).

Target genes, primer design and initial PCR amplification of putative carotenoid structural gene-containing genomic segments

Genes included in the study and their abbreviations are listed in Table 1. Primers were developed directly from carrot sequence recorded in GenBank for IPI, GGPS1, GGPS2, PSY1, PSY2, CCS, LCYB, and LCYE. For genes for which no carrot sequences were available (ZDS, NCED, PDS, ZEP, VDE, CCD, CHX, CRTISO, and CHXE), degenerate primers were designed from homologous sequences from other plant species. Complementary DNA (cDNA) or mRNA sequences were gathered from the NCBI (http://www.ncbi.nlm.nih.gov) database and aligned in ClustalX, and degenerate primers were placed in highly conserved regions. Most of the primer pairs were used in PCR with genomic DNA from a population bulk of B493 × QAL DNA. Primer pairs for CRTISO and CHXE were initially run on first-strand synthesis products of root RNA extracts from B493. For these two genes, nested degenerate primer sets were used. Reactions were run on an Eppendorf MasterGradient thermocycler (Eppendorf AG, Hamburg, Germany). Each 25 μL PCR reaction mixture contained 1.5 μL RACE ready first strand synthesis product, 1 µM of each primer (for degenerate primers) or 0.5 µM of each primer (for non-degenerate primers), 0.2 mM each dNTP mix, 1.5 mM $MgCl_2$, 2.5 μL of ten times PCR buffer, and 1.25 units of *Taq* DNA polymerase. The reactions were carried out as follows: 94 °C (2 min), 35 cycles of 94 °C (30 s), variable annealing temperature (30 s), 72 °C (2 min) and a final extension step of 72 °C (5 min). If touchdown PCR was used (for CCD, CHX, CRTISO, and CHXE) reactions were begun at an initial annealing temperature above the T_m of the primer and dropped 1 °C every cycle for ten cycles followed by 25 cycles of PCR at an annealing temperature from the tenth cycle. The denaturing and extension times were the same as above. Table 1 lists the annealing temperatures and primer sequences. Products were run on 1.0% (w/v) agarose gels in one time TAE (40 mM Tris–acetate and 1 mM EDTA). A 1 kb DNA molecular weight marker (Fermentas, Hanover, MD) was run on each gel and gels were stained with ethidium bromide. Products were removed from the gel with a razor blade and purified using the Ultraclean 15 DNA purification kit (MoBio Laboratories, Carlsbad, CA). In reactions where nested primers were used, the primary reaction product

Table 1 Copy-specific primers and degenerate primers or source sequences for genes used in this study and their annealing temperatures, product size and the polymorphism type used to map each

Enzyme name	Gene symbol	Sequence of degenerate primers or source for carrot sequence	Annealing temperature for degenerate primers (°C)	Sequence of primers used to amplify fragment used for genotyping	Annealing temperature (°C)	Approximate size of fragment amplified for genotyping (kB)	Incuded on map?	Polymorphism type[a]	Full cds clone obtained?	GenBank accession number	Submitted sequence length (bp)	Accession number of closest putative ortholog and % identity[b]
Isopentenyl pyrophosphate isomerase	IPI	AF227951	–	f-cctctagtgtggactaacacttgttg r-acatacttgatctccgcgacttc	55	0.6	N[c]	NA	Y	DQ192183	1173	AB049816 89
Geranyl geranyl pyrophosphate synthase	GGPS1	AB027705	–	f-gattcaaggcgcaaatgctgtg r-cccttgtgagcccacagcc	56	0.5	N[d]	SNP, ds	Y	DQ192184	1287	AY866498 68
Geranyl geranyl pyrophosphate synthase	GGPS2	AF227950, AB027706	–	f-cttgtccaagaatttactgttcgatcatctc r-actgagttcctcctccgcctg	56	1.0	Y	SNP, ds	Y	DQ192185	1271	AY661708 72
Phytoene synthase	PSY1	AB032797	–	f-ggaggaaaatatggtaggccagtttcac r-gcatactcagcacaaacttcgcc	56	0.66, 0.67	Y	Size	Y	DQ192186	1515	AY450646 81
Phytoene synthase	PSY2	Fonseca (2000) (5′ end)	–	f-gatgaattggtagatggacctaatgcgtc r-ccatcctcatcccttcaacc	54	0.7	Y	SNP, re, *Taq*I	Y	DQ192187	1939	X68017 82
Carotenoid isomerase	CRTISO	f1-acnccnatgathaaygc[e] r1-ccnacrcartanagncc f2-taygtnaargcnccntc r2-gtrttraanggcatncc	td 52 → 42 (10) 42 (25)	f-aaccatacacacaatcaccttcctcc r-tcccagacccaatcacaatagcatc	56	0.4	Y	SNP, ds	Y	DQ192188	2269	BX813870 86
Phytoene desaturase	PDS	f-tgatwtttgcwatgccaaryaarcc r-caytgcatygamagytcrtcwg	55	f-atttgattttccggaagttctacctgc r-aacacattgcattgagagttcatcagg	56	0.6	Y	SNP, ds	Y	DQ222429	2206	AF251014 88
ζ-Carotene desaturase	ZDS1	f-gtgggatccwgttgcntaygc r-ayctcyctrcahccccacc	59	f-ggtttattgactgtgataacatgagtgctc r-atttttgttagttccttcaactaaattaacaggtaag	55	0.4	Y	SNP, ds	Y	DQ222430	2078	AF251013 90
ζ-Carotene desaturase	ZDS2	Same as ZDS1	59	f-catgcttaaaggttctcctgatgtttatttgag r-gccactttttatttcttgatacagcgtgg	57	1.0	Y	SNP, re, *Taq*I	Y	DQ192189	2038	AB072343 88
Lycopene *β*-cyclase	LCYB1	AF208531, Fonseca (2000)	–	f-tatggtgtttgggtggatgagtttgag r-cgctggcgtcttctctctattgg	57	0.85	Y	SNP, ds	Y	DQ192190	2022	AF240787 87
Lycopene *β*-cyclase (casorubin capsanthin synthase)	LCYB2 (CCS)	AF208530	–	f-cattgcgttttgctaacgccaag r-atgtaaatcaagcttcagcaaagtctcc	56	0.5	Y	SNP, re, *Tas*I	Y	DQ192191	1730	AY679168 59
Lycopene *ε*-cyclase	LCYE	AF212130, AF229684	–	f-gcttgctactgttgcatctg r-agtgtatagtaaaccaactaaagatgg	51	0.3, 0.25	Y	Size	Y	DQ192192	1779	AF251016 84
β-Ring carotene hydroxylase	CHXB1	f-tgggcnmgntgggcnc r-crtcrtgnacraacatrtangc	td 60 → 50 (10) 50 (25)	f-tagggtggcgaaattccgtattcg r-gaccgagtattacaaatgccataacactgc	57	0.3	Y	SNP, ds	Y	DQ192193	1231	Y14810 84
β-Ring carotene hydroxylase	CHXB2	Same as CHXB1	td 60 → 50 (10) 50 (25)	f-cacgaggtatgttccgagaggc r-gaacaccgtgattccaagtcc	54	0.3	Y	SNP, ds	Y	DQ192194	1137	Y09722 86

Table 1 Continued

Enzyme name	Gene symbol	Sequence of degenerate primers or source for carrot sequence	Annealing temperature for degenerate primers (°C)	Sequence of primers used to amplify fragment used for genotyping	Annealing temperature (°C)	Approximate size of fragment amplified for genotyping (kB)	Incuded on map?	Polymorphism type[a]	Full cds clone obtained?	GenBank accession number	Submitted sequence length (bp)	Accession number of closest putative ortholog and % identity[b]
β-Ring carotene hydroxylase	CHXB3	Same as CHXB1	td 60 → 50 (10) 50 (25)	f-gagctgaacgatgttttcgctataataaacgtg r-gaacaccgtgattccaagtccc	57	0.25	N[c]	SNP, ds	N	DQ192195	581	AF251018 80
ε-Ring carotene hydroxylase	CHXE	f1-atgacnctngaygtnathgg r1-gcraaytgrtcnccnacrc f2-carctnagrgaygayctnc r2-ctcatngaytcrttdatrcayc	td 52 → 42 (10) 42 (25)	f-accaaactcccaaacatgccttac r-tagcaatagctggatcactgaccac	56	0.4	Y	SNP, ds	Y	DQ192196	1803	BT012891 82
Zeaxanthin epoxidase	ZEP	f-caaacartacttygtttcttcdgatg r-atcctcwatkgccatgcawc	58	f-caaacartacttygtttcttcdgatg r-atcctcwatkgccatgcawc	58	0.65	Y	SNP, re, *Dra*I	Y	DQ192197	2431	X95732 80
Violaxanthin de-epoxidase	VDE	f-cngtwgatgcwctyaaracttg r-atrwaccayttcccrytraartc	55	f-tacttgcttacttaaggagtgcaggtac r-catgttgaagctttttacaagcacatcagc	57	0.3	Y	SNP, re, *Taq*I	N	DQ192198 DQ192199	612 854	AJ250433 100 LSU31462 84
9-*cis*-Epoxycarotenoid dioxygenase	NCED1	f-sgagctycacggccaytc r-tggaarcagaarcartccgg	56	f-cttcttgatcataagcatggcatggg r-acggaagcttaatcgaagcctctag	56	1.0	Y	SNP, re, *Taq*I	Y	DQ192200	2029	AB030293 77
9-*cis*-Epoxycarotenoid dioxygenase	NCED2	Same as NCED1	56	f-ggtggatcaccagtggtgtatgac r-gcattaacgatctgaagctccgac	56	0.6	Y	SNP, ds	Y	DQ192201	2048	AB030293 75
9-*cis*-Epoxycarotenoid dioxygenase	NCED3	NA[f]	–	f-caagtgggttgaggttccagac r-gcattaactatttgaagctctgatttccatgc	56	0.5	Y	SNP, ds	Y	DQ192202	2211	Z97215 81
Carotenoid cleavage dioxygenase	CCD1	f-caygayttygcnathacngar r-avrtanccrtcrtcytcytc	td 60 → 50 (10) 50 (25)	f-ttggtgtccttccacgatatgc r-tgttaacaggctgcagtaatttgttg	54	1.0	Y	SNP, ds	Y	DQ192203	2084	DQ003599 83
Carotenoid cleavage dioxygenase	CCD2	Same as CCD1	td 60 → 50 (10) 50 (25)	f-gacaagcggctaacacattcattc r-taaatatgactacagaaaatccccttaccttgg	54	0.35	Y	SNP, ds	N	DQ192204	1771	AY856353 71
Carotenoid cleavage dioxygenase	CCD3	Same as CCD1	td 60 → 50 (10) 50 (25)	f-tgaggctccaagttctgagatg r-tctgaaccgaatctcccaactcc	54	0.65	Y	SNP, ds	N	DQ192205	1055	DQ100347 76

[a] Polymorphism is either size or single nucleotide polymorphism (*SNP*). SNPs were genotyped with either restriction enzymes (*re*, enzyme used) or direct sequencing (*ds*)

[b] Sequence similarity at protein level based on tblastx hits. Percent similarity is shown for longest matching stretch of best hit sequence labelled as an ortholog to the carotenoid biosynthesis gene

[c] IPI and CHXB3 could not be included on the map because no polymorphic sequences were identified

[d] GGPPS1 could not be included on the map because of severe segregation distortion

[e] For CRTISO, and CHXE, nested degenerate primers were used on cDNA template from B0493 root extracts for preliminary sequence. Products of amplifications using primers with f1/r1 prefixes were used as a template in a PCR using primers with f2/r2 primers

[f] For NCED3 sequence was initially obtained from a splinkerette reaction used to obtain more sequence from NCED2

was diluted 100-fold, to be used as a template at 1,000-fold final dilution in the secondary reaction.

Cloning and sequencing

PCR products were cloned into pGEM-T vector according to the manufacturer's instructions. Plasmids were extracted using Wizard SV miniprep kits (Promega, Madison, WI) and sequenced using Big Dye terminator chemistry (Applied Biosystems, Foster City, CA). Sequencing reactions were run on an ABI 3700 capillary sequencer (Applied Biosystems, Foster City, CA). Sequences were compared to other species homologs on the NCBI Basic Local Alignment Search Tool (BLAST) web site using tblastx and blastn options (http://www.ncbi.nlm.nih.gov/blast/). Putative genes were declared if the BLAST results contained matches in regions not covered by the original degenerate primers.

Copy-specific primer design and polymorphism identification

When multiple copies of putative genes were identified (GGPS, PSY, ZDS, LCYB, CHXB, NCED, and CCD), copy-specific non-degenerate primers were developed for each putative gene and were used to amplify PCR products from a population bulk of B493 × QAL. These products were treated with 10 units of Exonuclease I (Fermentas Inc., Hanover, MD) and 1 unit of shrimp alkaline phosphatase (UBS, Cleveland, OH). Products were incubated with Exonuclease I and shrimp alkaline phosphatase for 20 min at 37 °C followed by 15 min at 80 °C. Each product was directly sequenced using one of the two primers originally developed to amplify it. Sequences were examined for polymorphism, as indicated by superimposed peaks in a chromatogram. In some cases a new primer set flanking the polymorphism was developed so the size of the fragment to be amplified was smaller.

Genotyping the population at each putative carotenoid biosynthetic genes and Y2mark

Both size polymorphisms, which were evaluated on 2% agarose gels, and single nucleotide polymorphisms (SNPs) were observed. SNPs were evaluated in one of two ways. If the polymorphism created a novel restriction site then the reaction products were treated with an excess of appropriate restriction enzyme for 2–10 h at the appropriate temperature. The digested products were run on 2% agarose gels. Other SNPs were genotyped by amplifying the fragment in each individual followed by direct sequencing. Heterozygotes were evident by the presence of a double peak at the SNP site in the sequence chromatogram (Fig. 2). An additional marker, Y2mark was also added to the map. This SCAR marker is linked to the Y_2 gene and was developed by Bradeen and Simon (1998). This marker was genotyped by scoring a SNP in direct sequenced products amplified by published primers in each individual in the population.

Linkage map construction

Genotype data for the putative carotenoid structural genes was added to the map data generated by Santos (2001), which consisted mostly of AFLP markers. MAPMAKER 3.0 (Lander et al. 1987) was used for map construction. Dominant markers from a single parent linked in coupling were used in conjunction with all codominant markers. The two-point command was used to establish linkage groups at a LOD of 4.0. One or two codominant markers per group were assigned as anchor loci in the initialization file. Markers were assigned to linkage groups using the assign command. Three-point analysis was then performed for each linkage group followed by the order command to develop a framework map. Remaining markers were added using the place command. Markers that still remained unplaced were added using the try command. The linkage group numbering of Santos (2001) and Santos and Simon (2002, 2004) was retained.

RNA extraction

RNA was extracted using the TRIZOL reagent (Invitrogen, Carlsbad, CA) following the manufacturers instructions. Fresh root tissue samples weighing 0.8 g were used. All samples were treated with RQ1 RNAse free DNAse (Promega, Madison, WI) to remove any possible remaining DNA. DNAse treated samples were quantified using a spectrophotometer. Extractions were tested for DNA contamination by using primers in the PDS gene that amplify a product that contains a small intron. Amplification of a single product of the expected size without the intron was taken as evidence of a lack of DNA contamination.

RACE PCR and full-length cDNA clone amplification

Root tissue from inbred B493 was used for RACE PCR and full coding sequence cDNA amplification. For RACE PCR, the SMART RACE kit from Invitrogen (Carlsbad, CA) was used according to the

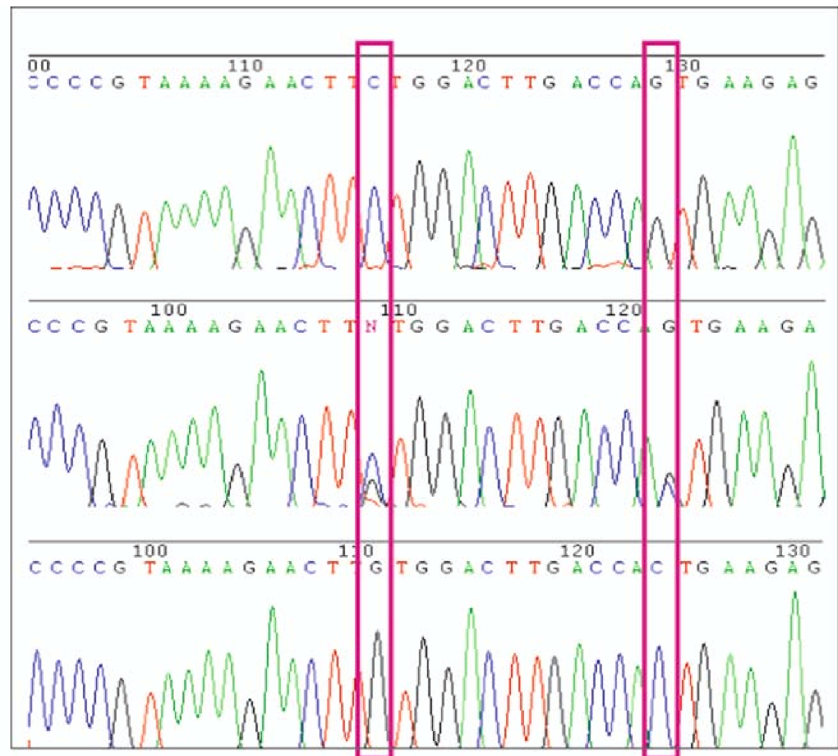

Fig. 2 An example of sequencing chromatographs showing linked SNPs that were used to map 9-*cis*-epoxycarotenoid dioxygenase 2 (*NCED2*). The *upper chromatogram* shows a portion of the sequence of the QAL allele; the *lower chromatogram* shows a portion of the sequence of the B493 allele; the *middle* is a chromatogram from a heterozygote. The *rectangles* outline the positions of the SNPs

manufacturers instructions. Each reaction was carried out using 3′ and 5′ RACE ready reverse transcribed mRNA from roots of B493. In some cases, nested primers were used to obtain RACE PCR products. A touchdown PCR program was used as suggested by the manufacturer. The program was carried out in an Eppendorf (Hamburg, Germany) Mastercycler thermalcycler as follows: five cycles of 94 °C 30 s, 70 °C 30 s, 72 °C 3 min; five cycles of 94 °C 30 s, 68 °C 30 s, 72 °C 3 min; 27 cycles of 94 °C 30 s, 65 °C 30 s, 72 °C 3 min. Each 25 μL PCR reaction mixture contained 1.5 μL RACE ready first strand synthesis product, 1 μM primer, 0.2 mM each dNTP mix, 1.5 mM $MgCl_2$, 2.5 μL of ten times PCR buffer, and 1.25 units of *Taq* DNA polymerase. Electrophoresis and PCR product purification was performed as described above for genomic sequences.

If a smear or many products were observed, a nested primer was developed for better amplification of the intended product. The conditions for the secondary reactions were identical to the primary reactions except that the 2.5 μL of 100-fold diluted product from the primary reaction was used per 25 μL reaction as a template for PCR instead of RACE ready first strand synthesis products.

Sequences from the ends of the sequences of RACE PCR products were used to develop primers that would amplify the entire coding region of the cDNA. These primers were then used on 5′RACE ready reverse transcribed mRNA from B493 to amplify the complete coding sequence of genes from which RACE products were obtained. GGPS1, GGPS2, and PSY1 have full coding sequences for carrot in the GenBank sequence database. For these genes, primers were developed from the 5′ and 3′ ends of the database sequences. All RACE PCR and full-length cDNA amplification products were separated on 1% TAE agarose gels. Amplified fragments were removed from the gels with a razor blade and purified using the UltraClean 15 gel extraction kit (MoBio Laboratories, Carlsbad, CA). RACE PCR and full-length cDNA amplification products were ligated into pGEM-T (Promega, Madison, WI) vector for sequencing.

Results

A total of 24 putative carrot carotenoid biosynthetic structural genes (Table 1) were placed on the carrot linkage map and full-length transcript was sequenced for 22 of these genes. Of these, sequence data for carrot was previously available in the GenBank database for eight (IPI, GGPS1, GGPS2, PSY1, LCYB, LCYE, and CCS) and partial cDNA sequence data for PSY2 was obtained from Fonseca (2000). Fifteen new putative genes were identified in this study (CRTISO,

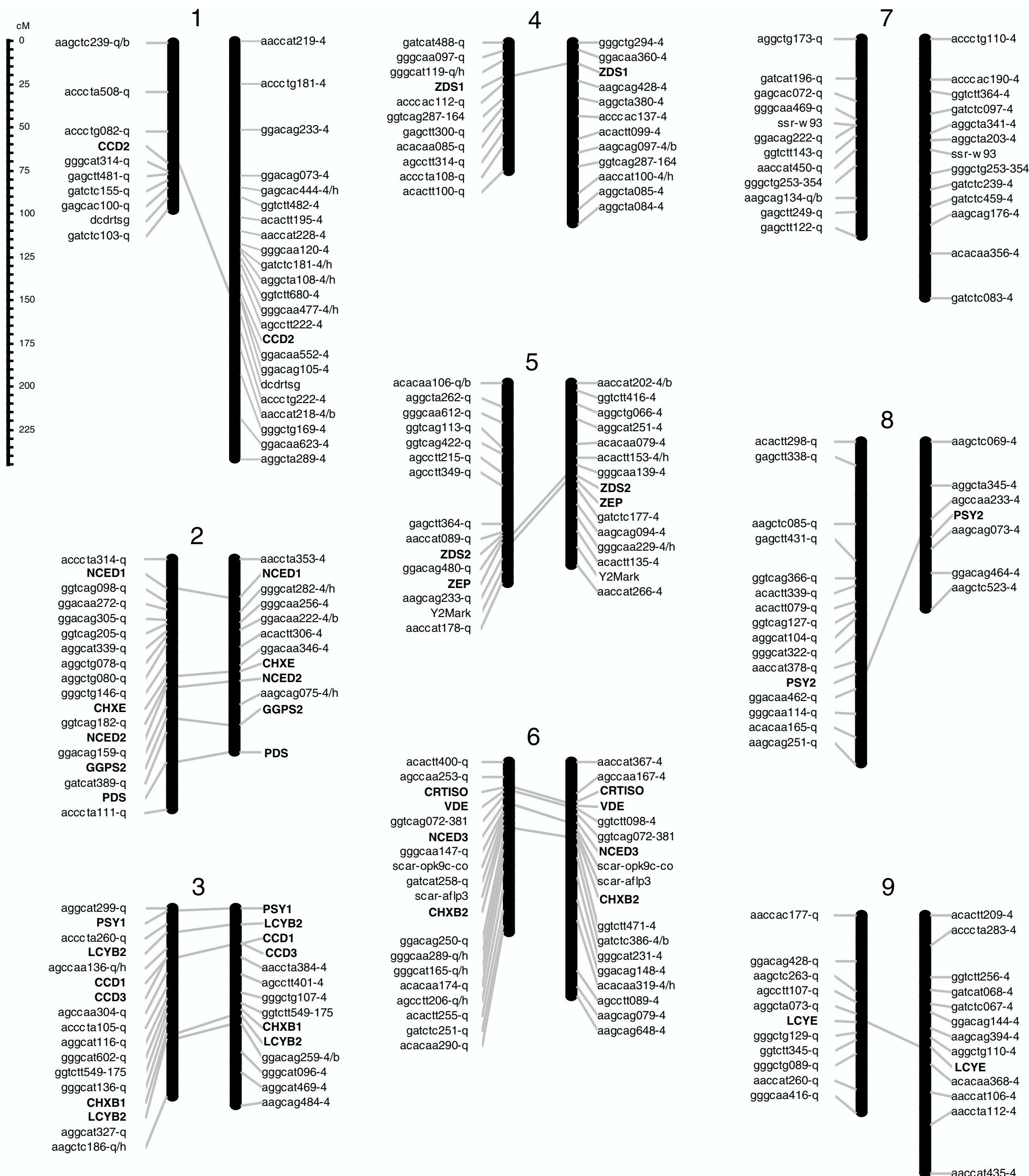

Fig. 3 B493 × QAL linkage map showing the map positions of putative carotenoid biosynthetic structural genes (abbreviations as in Table 1). All codominant markers including the carotenoid biosynthetic genes, Y2mark (linkage group 5), codominant AFLP fragments ggtctt549-175 (linkage group 3), ggtcag287-154 (linkage group 4), ggtcag072-381 (linkage group 6), microsatellite markers dcdrstg (linkage group 1) and microsatellite marker ssr-w93 (linkage group 7) are connected with *dotted lines* between the two maps. All other markers are dominant AFLP fragments as described by Santos (2001). On linkage groups where more than one codominant marker is available, they are presented in the same orientation. Since linkage groups 1, 8 and 9 each only have one codominant marker, the orientations of the two coupling maps relative to one another are ambiguous

PDS, ZDS1, ZDS2, CHXB1, CHXB2, CHXB3, ZEP, VDE, NCED1, NCED2, NCED3, CCD1, CCD2, and CCD3). Twenty-two genes and putative genes were added to the carrot genetic linkage map. For all markers we added to the map at least 90% of the population were sequenced for all markers where direct sequencing was used for genotypes except CCD1 where 86% were direct sequenced.

One gene (IPI) was omitted from the map because of lack of polymorphism and one (GGPS1) was omitted because severe segregation distortion was observed. Separate maps for each parent of the cross

were generated consisting of a mixture of only codominant and coupling phase dominant markers (Knapp et al. 1995; Mester et al. 2003).

The 24 carrot genes studied are distributed over eight of the nine carrot linkage groups (Fig. 3). The numbering of the linkage groups is the same as used by Santos (2001) and Santos and Simon (2002, 2004). CCD2 was the only structural gene to map to linkage group 1. Linkage group 2 has five genes: NCED1, CHXE, NCED2, GGPS2, and PDS. Six genes: PSY1, CCS, CCD1, CCD3, CHXB1, and LCYB all mapped to linkage group 3. ZDS1 mapped to linkage group 4. ZDS2 and ZEP mapped to linkage group 5. Five genes: CRTISO, VDE, NCED3, CHXB2, and CHXB3 were placed on linkage group 6. Linkage group 7 did not contain any of the carotenoid structural genes we evaluated. PSY2 and LCYE mapped to linkage groups 8 and 9, respectively. Figure 3 compares the QAL and B493 coupling linkage groups side by side. All codominant markers are in bold and are connected between the two maps. On linkage groups where more than one codominant marker is available, they are presented in the same orientation. Since linkage groups 1, 8 and 9 each only have one codominant marker, the orientations of the two coupling maps relative to one another are ambiguous.

Additionally, one SCAR marker (Y2mark) was added to the linkage map. This marker was developed by Bradeen and Simon (1998). While the size polymorphism described by Bradeen was not segregating in this population, a SNP was identified in the amplified fragment which was used to place the marker on the map by direct sequencing. The marker mapped to linkage group 5.

RACE PCR was used to determine qualitative expression at the mRNA level. Transcripts for twenty genes in the carotenoid pathway were detected in fresh root tissue of the orange carrot inbred line B493. Full-length cDNA sequences for these genes were determined using 5′ and 3′ RACE PCR. Primers developed on the ends of RACE PCR products allowed for amplification and cloning of full coding sequences for each of these genes. Sequence length and percent similarity to the nearest putative orthologs are presented for each GenBank submission in Table 1.

Discussion

Development of reliable codominant PCR-based markers for carrot

Codominant PCR-based markers are preferable to RFLPs in carrot because the size of plants is generally small, making it difficult to obtain the large quantities of DNA required for repeated analysis using RFLP, whereas PCR-based markers can be used even when the DNA quantity is very low. The markers developed in this study could be used in any carrot mapping population and we have found them to be useful in preliminary evaluation of unrelated populations (data not presented). Because we have generated sequence data, more PCR primers could be developed from the sequences of these genes to amplify a region known to contain a SNP, even if the polymorphisms mapped in this study do not occur in the population being studied.

These mapped genes will be used as anchor loci for future carrot genetic linkage maps. Since there is a relative paucity of conserved, PCR-based codominant markers in carrot, the markers described here should be a great aid in identifying homologous groups across studies, in addition to the PCR-based markers described previously describe for carrot (Bradeen and Simon 1998, 2006; Santos and Simon 2004).

Mapping carotenoid biosynthesis genes and Y2mark

Mapped biosynthetic pathway genes with known biochemical function can serve as candidate genes for traits. Since carotenoid pigments in the roots of carrot plants are such an important trait, it has been the subject of several mapping studies (Vivek and Simon 1997; Bradeen and Simon 1998; Santos and Simon 2002). Santos and Simon (2002) observed numerous QTL for major carotenoid pigments using single marker analysis. One highly notable region of QTL clustering was at one end of linkage group 5. AFLP markers spanning the range from agcctt349-q to aaccat178-q were detected as among the most significant QTLs for all pigments analyzed. We mapped ZDS2, ZEP, and Y2mark to that region. The data presented by Santos (2001) and Santos and Simon (2002) are inconclusive as to whether any of those linked genes directly affect the trait or they are closely linked to the gene conditioning the Y_2 phenotype. Either way, ZDS2 and ZEP may be considered tentatively as candidate genes for some the QTLs. Interestingly, Y2mark, a SNP-based marker derived from the marker linked to the Y_2 gene, described in Bradeen and Simon (1998) also maps to this region. Y_2 controls the difference between the yellow (xanthophyll) root phenotype and the orange (α- and β-carotene) root phenotype (Buishand and Gabelman 1979). Given that many yellow segregants were observed in the F_2 population derived from B493 × QAL, it is likely that the Y_2 gene is segregating in this population. Furthermore, since it maps

to this important region, Y_2 may be responsible for at least some of the QTL effect observed by Santos and Simon (1998). QTL analyses with the new maps described in this study are underway to further elucidate the answers to these questions.

Since the carotenoid pathway is so highly conserved among flowering plants (Hirschberg 2001), map positions of these genes could serve as a starting point for comparative mapping studies with other species in genera closely related to carrot. For example, this approach has been used on carotenoid biosynthesis genes in *Capsicum* and *Lycopersicon* (Thorup et al. 2000) and carbohydrate metabolism genes in *Solanum* and *Lycopersicon* (Chen et al. 2001). Gene linkage and order for the carotenoid structural genes in carrot was compared with gene linkage and order in *Arabidopsis*, pepper (Thorup et al. 2000), and tomato (Liu et al. 2003). No obvious patterns of conservation were observed among these species. This is not surprising because initial results in comparative mapping between highly diverged species such as tomato and *Arabidopsis* generally show only microsynteny, that is, conservation of gene orders within rather small sections of the genome (>10 cM) (Fulton et al. 2002). However, once more low copy number genes are mapped in carrot, it may be possible to identify regions of microsynteny with plants that have better characterized genomes. Conserved orthologous sets (COS markers) have been developed and mapped in tomato and *Arabidopsis* (Fulton et al. 2002) so it would be worthwhile to also begin mapping these genes in carrot using some of the techniques described here. We have discovered that several of the carrot carotenoid genes described here exhibit evidence for single copy hybridization using DNA from other *Daucus* species and other Apiaceaeous genera.

cDNA sequencing and detection of mRNA for carotenoid biosynthesis genes in roots

The fact that mRNA for all of these genes was present in the roots of an orange carrot is interesting, since genes both before and after beta and alpha carotene in the pathway appear to be expressed. Future research will elucidate the extent of pathway regulation at the transcription level.

It is also interesting to note that based upon the GenBank sequence, we detected mRNA for carrot capsanthin–capsorubin synthase (CCS) (AF208530), in orange carrot roots. CCS is a member of a gene family including lycopene β-cyclase. In tomato, there is a chromoplast-specific lycopene beta cyclase (CYC-B). This enzyme is very similar in amino acid composition to capsanthin–capsorubin synthase (CCS) from pepper and since CCS can perform the cyclase reaction, Ronen et al. (2000) proposed that these genes are orthologs, with CCS originating from a duplication of LCYB. Additionally, carrots have never been shown to produce either capsanthin or capsorubin. Therefore, we named the homolog we derived using primers designed from AF208530 as a lycopene β-cyclase (LCYB2), since it is more likely that the gene product is performing the cyclase reaction and is not catalyzing conversion of zeaxanthin and violaxanthin to capsanthin and capsorubin.

Messenger RNAs for genes involved in the cleavage of carotenoid molecules (NCED1, NCED2, NCED3, and CCD1) were detected. Therefore, it is likely that to a greater or lesser extent, carotenoids are being cleaved into secondary metabolites in carrot roots. Cleaved carotenoid products can act as plant hormones, provide flavor and aroma, and may be involved in the maintenance of mycorrhizal relationships (Millborrow 2001; Bouvier et al. 2003; Fester et al. 2002). Perhaps differential cleavage of carotenoids in carrot roots is responsible for some of the differences observed in carrot root color and flavor across *Daucus* germplasm. Future studies involving transcript or protein quantification will help to answer these questions.

The set of 20 carotenoid biosynthetic structural genes that we have cloned and sequence characterized is a powerful tool for the future study of carotenoid biosynthesis in carrot and other species. This will provide future researchers with a tool to study the functionality of these genes by actually producing the protein products of the genes (e.g., Gallagher et. al. 2003). This sequence information will also provide a foundation for primer development in PCR-based expression studies, since it will be possible to develop primers in less conserved regions of the genes. Finally, carotenoid gene sequence will be useful fundamental information in devising transgenic approaches to manipulating carotenoid content in carrots or other related plants (Fraser et al. 1999; Rosati el al. 2000).

Acknowledgments This research was supported by Initiative for Future Agriculture Food Systems Grant number 2000-4258 from the USDA Cooperative Research, Education, and Extension Service. The senior author acknowledges the generous support of the Gabelman-Shippo Distinguished Graduate Fellowship in the Plant Breeding and Plant Genetics Program of the University of Wisconsin-Madison. We are grateful for the proficient technical assistance of Douglas Senalik.

References

Bartley GE, Scolnik PA (1995) Plant carotenoids: pigments for photoprotection, attraction, and human health. Plant Cell 7:1027–1038

Boiteux LS, Belter JG, Roberts PA, Simon PW (2000) RAPD linkage map of the genomic region encompassing the root-knot nematode (*Meloidogyne javanica*) resistance locus in carrot. Theor Appl Genet 100:439–446

Boiteux LS, Fonseca MEN, Simon PW (1999) Effects of plant tissue and DNA purification methods on randomly amplified polymorphic DNA-based genetic fingerprinting analysis in carrot. J Am Soc Hortic Sci 124:32–38

Boiteux LS, Hyman JR, Bach IC, Fonseca MEN, Matthews WC, Roberts PA, Simon PW (2004) Employment of flanking codominant STS markers to estimate allelic substitution effects of a nematode resistance locus in carrot. Euphytica 136:37–44

Bouvier F, Suire C, Mutterer J, Camara B (2003) Oxidative remodeling of chromoplast carotenoids: identification of the carotenoid dioxygenase CsCCD and CsZCD genes involved in crocus secondary metabolite biogenesis. Plant Cell 15:47–62

Bradeen JM, Simon PW (1998) Conversion of an AFLP fragment linked to the carrot Y_2 locus to a simple, codominant, PCR-based marker form. Theor Appl Genet 97:960–967

Bradeen JM, Simon PW (2006) Mapping, markers, and molecular breeding in carrots. In: Cole CR (ed) Genome mapping and molecular breeding, vol 5. Vegetable crops. Science Publishers Inc./IBH Publishing Co. Pvt. Ltd., New Hampshire Oxford/New Delhi

Brauner S, Murphy RL, Walling JG, Przyborowski J, Weeden NF (2002) STS markers for comparative mapping in legumes. J Am Soc Hortic Sci 127:616–622

Buishand JG, Gabelman WH (1979) Investigations on the inheritance of color and carotenoid content in phloem and xylem of carrot roots (*Daucus carota* L.). Euphytica 28:611–632

Chen X, Salamini F, Gebhardt C (2001) A potato molecular function map for carbohydrate metabolism and transport. Theor Appl Genet 102:284–295

Cunningham FX, Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. Annu Rev Plant Physiol Plant Mol Biol 49:557–583

Demmig-Adams B, Adams WW III (2002) Antioxidants in photosynthesis and human nutrition. Science 298:2149–2153

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Fester T, Hause B, Schmidt D, Halfmann K, Schmidt J, Wray V, Hause G, Strack D (2002) Occurrence and localization of apocarotenoids in arbuscular mycorrhizal plant roots. Plant Cell Physiol 43:256–265

Fonseca MEN (2000) Cloning and expression of carrot (*Daucus carota* L.) cDNAs coding for enzymes of the carotenoid biosynthetic pathway in roots accumulating different types and amounts of carotenoids. PhD Thesis. University of Wisconsin, Madison, WI

Fraser PD, Kiano JW, Truesdale MR, Schuch, Bramley PM (1999) Phytoene synthase-2 enzyme activity in tomato does not contribute to carotenoid synthesis in ripening fruit. Plant Mol Biol 40:687–698

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467

Gabelman WH, Peters S (1979) Genetical and plant breeding possibilities for improving the quality of vegetables. Acta Hortic 93:243–259

Gallagher CE, Cervantes-Cervantes M, Wurtzel ET (2003) Surrogate biochemistry: use of *Escherichia coli* to identify plant cDNAs that impact metabolic engineering of carotenoid accumulation. Appl Microbiol Biotechnol 60:713–719

Giuliano G, Al-Babili S, von Lintig J (2003) Carotenoid oxygenases: cleave it or leave it. Trends Plant Sci 8:145–149

Hirschberg J (2001) Carotenoid biosynthesis in flowering plants. Curr Opin Plant Biol 4:210–218

Isaacson T, Ronen G, Zamir D, Hirschberg J (2002). Cloning of *tangerine* from tomato reveals a carotenoid isomerase essential for the production of β-carotene and xanthophylls in plants. Plant Cell 14:333–342

Knapp SJ, Holloway JL, Bridges WC, Liu BH (1995) Mapping dominant markers using F_2 matings. Theor Appl Genet 91:74–81

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Liu Y-S, Gur A, Ronen G, Causse M, Damidaux R, Buret M, Hirschberg J, Zamir D (2004) There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotechnol J 1:195–207

Mester DI, Ronin YI, Hu Y, Nevo E, Korol AB (2003) Efficient multipoint mapping: making use of dominant repulsion-phase markers. Theor Appl Genet 107:1102–1112

Millborrow BV (2001) The pathway of biosynthesis of abscisic acid in vascular plants: a review of the present state of knowledge of ABA biosynthesis. J Exper Bot 52:1145–1164

Niemann M, Westphal L, Wricke G (1997) Analysis of microsatellite markers in carrot (*Daucus carota* L. *sativus*). J Appl Genet 38A:20–27

Park H, Kreunen SS, Cuttriss AJ, DellaPenna D, Pogson BJ (2002) Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. Plant Cell 14:321–332

Pflieger S, Lefebvre V, Causse M (2000) The candidate gene approach in plant genetics: a review. Mol Breed 7:275–291

Rosati C, Aquilani R, Dharmapuri S, Pallara P, Marusic C, Tavazza R, Bouvier F, Camara B, Giuliano G (2000) Metabolic engineering of beta-carotene and lycopene content in tomato fruit. Plant J 24:413–419

Santos CAF (2001) Biometrical studies and quantitative trait loci associated with major products of the carotenoid pathway of carrot (*Daucus carota* L.). PhD Thesis. University of Wisconsin, Madison, WI

Santos CAF, Simon PW (2002) QTL analyses reveal clustered loci for accumulation of major provitamin A carotenes and lycopene in carrot roots. Mol Genet Genom 268:122–129

Santos CAF, Simon PW (2004) Merging carrot linkage groups based on conserved dominant AFLP markers in F2 populations. J Am Soc Hortic Sci 129:211–217

Schulz B, Westphal L, Wricke G (1994) Linkage groups of isozymes, RFLP and RAPD markers in carrot (*Daucus carota* L. *sativus*). Euphytica 74:67–76

Schwartz SH, Qin X, Zeevart JAD (2001) Characterization of a novel carotenoid cleavage dioxygenase from plants. J Biol Chem 276:25208–25211

Simon PW (2000) Domestication, historical development, and modern breeding of carrot. Plant Breed Rev 19:157–190

Tan B-C, Joseph LM, Deng W-T, Liu L, Li Q-B, Cline K, McCarty DR (2003) Molecular characterization of the *Arabidopsis* 9-*cis* epoxycarotenoid dioxygenase gene family. Plant J 35:44–56

Tanumihardjo SA, Yang Z (2005) Carotenoids: Epidemiology of health effects. In: Caballero B, Allen L, Prentice A (eds) Encyclopedia of human nutrition, 2nd edn. Elsevier Ltd., Oxford, pp 339–345

Thorup TA, Tanyolac B, Livingstone KD, Popovsky S, Paran I, Jahn M (2000) Candidate gene analysis of organ pigmentation loci in the Solanaceae. Proc Natl Acad Sci USA 97:11192–11197

Tian L, Musetti V, Kim J, Magallanes-Lundback M, DellaPenna D (2004) The Arabidopsis *LUT1* locus encodes a member of the cytochrome P450 family that is required for ε-ring hydroxylation activity. Proc Natl Acad Sci USA 101:402–407

Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci (in press)

Vivek BS, Simon PW (1999) Linkage relationships among molecular markers and storage root traits of carrot (*Daucus carota* L. ssp. *sativus*). Theor Appl Genet 99:58–64

Westphal L, Wricke G (1989) Genetic analysis of DIA, GOT, and PGI isozyme loci in *Daucus carota* L. ssp. *sativus*. Plant Breed 102:51–57

Westphal L, Wricke G (1991) Genetic and linkage analysis of isozyme loci in *Daucus carota* L. Euphytica 56:259–267

Westphal L, Wricke G (1997) Construction of a linkage map of *Daucus carota* L. *sativus* and its application for mapping disease resistance and restorer genes. J Appl Genet 38A:13–19

Yau YY, Simon PW (2003) A 2.5-kb insert eliminates acid soluble invertase isozyme II transcript in carrot (*Daucus carota* L.) roots, causing high sucrose accumulation. Plant Mol Biol 53:151–162