LARGE-SCALE BIOLOGY ARTICLE

CAROTENOID CLEAVAGE DIOXYGENASE4 Is a Negative Regulator of β-Carotene Content in Arabidopsis Seeds

Sabrina Gonzalez-Jorge,a,1 Sun-Hwa Ha,b,1 Maria Magallanes-Lundback,a Laura Ullrich Gilliland,c Ailing Zhou,d Alexander E. Lipka,e Yen-Nhu Nguyen,a Ruthie Angelovici,a Haining Lin,f Jason Cepela,g Holly Little,h C. Robin Buell,g Michael A. Gore,i and Dean DellaPennaa,2

a Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824–1319
b Graduate School of Biotechnology and Crop Biotech Institute, Kyung Hee University, Yongin, Korea
c Department of Biology, Pennsylvania State University, University Park, Pennsylvania 119677
d Syngenta Biotechnology, Research Triangle Park, Durham, North Carolina 14853
e Institute for Genomic Diversity, Cornell University, Ithaca, New York 14853
f Dupont Pioneer, Johnston, Iowa 50139
g Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824
h Department of Biology, Saginaw Valley State University, University Center, Michigan 48710
i Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York 14853

Experimental approaches targeting carotenoid biosynthetic enzymes have successfully increased the seed β-carotene content of crops. However, linkage analysis of seed carotenoids in Arabidopsis thaliana recombinant inbred populations showed that only 21% of quantitative trait loci, including those for β-carotene, encode carotenoid biosynthetic enzymes in their intervals. Thus, numerous loci remain uncharacterized and underutilized in biofortification approaches. Linkage mapping and genome-wide association studies of Arabidopsis seed carotenoids identified CAROTENOID CLEAVAGE DIOXYGENASE4 (CCD4) as a major negative regulator of seed carotenoid content, especially β-carotene. Loss of CCD4 function did not affect carotenoid homeostasis during seed development but greatly reduced carotenoid degradation during seed desiccation, increasing β-carotene content 8.4-fold relative to the wild type. Allelic complementation of a ccd4 null mutant demonstrated that single-nucleotide polymorphisms and insertions and deletions at the locus affect dry seed carotenoid content, due at least partly to differences in CCD4 expression. CCD4 also plays a major role in carotenoid turnover during dark-induced leaf senescence, with β-carotene accumulation again most strongly affected in the ccd4 mutant. These results demonstrate that CCD4 plays a major role in β-carotene degradation in drying seeds and senescing leaves and suggest that CCD4 orthologs would be promising targets for stabilizing and increasing the level of provitamin A carotenoids in seeds of major food crops.

INTRODUCTION

Plants, algae, and bacteria de novo synthesize a diverse group of more than 700 carotenoids that function as photosynthetic accessory pigments, antioxidants, and photoprotectants (Ruiz-Sola and Rodriguez-Concepcion, 2012). As animals cannot synthesize carotenoids, dietary intake and processing of carotenoids with provitamin A activity (e.g., α- and β-carotenes and β-cryptoxanthin) provide the vitamin A necessary for normal tissue differentiation as well as organ, immune, and visual development (Sommer and Vyas, 2012). Fruits, vegetables, and seeds are the primary dietary sources of carotenoids. However, provitamin A carotenoid levels in seeds of the most abundantly consumed staple crops are insufficient to meet minimum nutritional requirements (Fitzpatrick et al., 2012); consequently, vitamin A deficiency remains prevalent in developing countries. Improving the provitamin A content of staple crops through molecular breeding is a crucial strategy for alleviating vitamin A deficiency.

Plant carotenoid biosynthesis has been fully elucidated and starts with the production of phytoene by the condensation of two geranylgeranyl diphosphate molecules by PHYTOENE SYNTHASE (PSY). Phytoene is subjected to a series of desaturation and isomerization reactions to yield trans-lycopene (see Supplemental Figure 1 online; Ruiz-Sola and Rodriguez-Concepcion, 2012), which is acted on by ε- and β-ring cyclases with one ε-ring and one β-ring cyclization, producing α-carotene, and two β-ring cyclizations, producing β-carotene. These cyclized
carotenoids are further oxygenated by a series of enzymes to produce the xanthophylls most commonly found in leaf tissue: lutein, zeaxanthin, antheraxanthin, violaxanthin, and neoxanthin.

In addition to functioning as photoprotectants, antioxidants, and accessory pigments in photosynthesis, carotenoids are also substrates for the synthesis of apocarotenoids, biologically active derivatives formed by oxidative cleavage and further modification of carotenoids. Apocarotenoids include retinol (vitamin A), the plant hormones abscisic acid and strigolactones, and a wide range of plant volatiles that serve as attractants for pollinators and herbivores (Bouvier et al., 2005). Apocarotenoid production is initiated by carotenoid cleavage dioxygenases (CCDs) that target specific bonds in the conjugated carotenoid polyene chain and can affect carotenoid composition and content (Ohmiya et al., 2006). In plants, CCDs are generally encoded by small gene families. For example, the Arabidopsis thaliana genome codes for nine members, five 9-cis-epoxycarotenoid dioxygenases involved in abscisic acid synthesis (encoded by 9-CIS-EPOXYPYRANOID DIOXYGENASE2 [NCED2], NCED3, NCED5, NCED6, and NCED9; Tan et al., 2003) and four CCDs with broader substrate specificities (encoded by CCD1, CCD4, CCD7, and CCD8; Auldridge et al., 2006; Huang et al., 2009). All proteins encoded by this Arabidopsis gene family are targeted to the plastid, with the exception of CCD1, which is cytoplasmic (Tan et al., 2003; Auldridge et al., 2006; Ytterberg et al., 2006). CCD1 and CCD4 enzymes have been implicated in the production of apocarotenoid-derived pigments, flavors, and aromas in various plants (Schwartz et al., 2001; Schmidt et al., 2006; Garcia-Limones et al., 2008; Ilg et al., 2009), but in Arabidopsis only CCD1 has been demonstrated to degrade pigments in vivo (Auldridge et al., 2006). CCD7 and CCD8 are required for the synthesis of strigolactones, a class of transmissible plant hormones regulating auxiliary branching and tillering (Schwartz et al., 2004; Gomez-Roldan et al., 2008; Alder et al., 2012).

In addition to biosynthesis and degradation, sequestration and storage of carotenoids can also strongly influence accumulation in tissues (Shewmaker et al., 1999; Li et al., 2001). For example, mutation of the HIGH PIGMENT2 gene of tomato (Solanum lycopersicum) disrupts a negative regulator of light signaling, resulting in larger plastid size and increased carotenoid content (Kolotilin et al., 2007). Similarly, a splicing mutation of the ORANGE gene in Brassica oleracea results in the production of mutant variants of a plastid-localized DnaJ domain protein that induce chromoplast differentiation, thus increasing carotenoid accumulation (Lu et al., 2006).

Changes in carotenoid levels have been successfully engineered by overexpressing the gene responsible for the committed step of the carotenoid pathway, PSY, alone or in combination with those for other pathway enzymes (Shewmaker et al., 1999; Fraser et al., 2007; Fitzpatrick et al., 2012), with the development of Golden Rice (Oryza sativa) being the most well-known example (Ye et al., 2000; Al-Babili and Beyer, 2005; Paine et al., 2005). However, in other cases, pathway engineering has not been as predictable (Fray et al., 1995; Fraser et al., 2007; Simkin et al., 2007). For example, constitutive overexpression of PSY in tomato resulted in dwarfism due to competition with gibberellic acid biosynthesis for geranylgeranyl diphosphate in leaf tissue. Fruit-specific overexpression was required to increase fruit lycopene levels without negative vegetative phenotypes (Fray et al., 1995; Fraser et al., 2002). Similarly, carotenoid biosynthetic mutations in Arabidopsis yield surprisingly different phenotypes in seeds and leaves (Kim et al., 2009). For example, disruption of the cytochrome P450 LUTEIN DEFICIENT1 (LUT1) leads to a loss of lutein in all tissues, with leaf tissue showing compensatory increases in β-xanthophylls, while in seeds, β-xanthophyll levels decrease threefold. These data suggest that, despite the carotenoid pathway being fully elucidated and some notable successes in engineering the pathway, our understanding of the mechanisms regulating carotenoid content and composition in specific tissues remains incomplete.

Given the dietary significance of carotenoids and the prevalence of vitamin A deficiency in developing countries, a better understanding of the mechanisms regulating plant carotenoid composition, especially in edible seeds, of staple crops, is needed (Shewmaker et al., 1999; Howitt and Pogson, 2006; Kim et al., 2012; Chandler et al., 2013). Recently, linkage and association mapping studies have assessed the contribution of allelic variation at carotenoid biosynthetic genes to natural variation in seed carotenoids. For instance, allelic variants of phytoene synthase, lycopene β-cyclase, and CrtRB-type β-carotene hydroxylases affect the content and composition of seed carotenotes and xanthophylls in maize (Zea mays; Harjes et al., 2008; Yan et al., 2010; Fu et al., 2013). Linkage mapping in crops with draft genomes has identified numerous quantitative trait loci (QTLs) intervals associated with carotenoid traits, yet more than half of these intervals are devoid of carotenoid biosynthetic genes, indicating that numerous novel loci influence carotenoid traits (Wong et al., 2004; Pozniak et al., 2007; Chandler et al., 2008, 2013; Fernandez et al., 2008). Identification and functional characterization of such loci will be key to providing alternative or synergistic means for altering the carotenoid content of specific plant tissues. To assess the genetic control of carotenoid levels in seeds, we used a combination of linkage mapping and a genome-wide association study (GWAS) to identify CCD4 as a major locus negatively regulating carotenoid levels in Arabidopsis seeds.

RESULTS

Analysis of Two Arabidopsis Recombinant Inbred Line Populations Identifies QTLs for Natural Variation in Seed Carotenoid Levels

To assess the natural variation of individual and total carotenoid contents in mature, dry Arabidopsis seeds, the parental accessions, Columbia-0 (Col-0), Landsberg erecta (Ler), and Cape Verde Islands (Cvi), for two recombinant inbred line (RIL) populations, Col-0/Ler and Cvi/Ler, were analyzed (Table 1). Compared with Col-0, total carotenoids in Ler seeds were 1.7-fold higher due to significant increases in all carotenoids, with the respective 5- and 3.8-fold increases in violaxanthin and β-carotene being particularly noteworthy. In comparison with Col-0, Cvi had higher levels of neoxanthin, violaxanthin, and antheraxanthin and lower zeaxanthin, lutein, and total carotenoid levels. Carotenoid levels of both RIL populations were normally distributed.
Table 1. Carotenoid Contents of Mature Seeds for Three Arabidopsis Accessions: Col-0, Ler, and Cvi

<table>
<thead>
<tr>
<th>Accession</th>
<th>Neoxanthin</th>
<th>Violaxanthin</th>
<th>Antheraxanthin</th>
<th>Zeaxanthin</th>
<th>Lutein</th>
<th>β-Carotene</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col-0</td>
<td>0.75 ± 0.04</td>
<td>1.23 ± 0.13</td>
<td>1.08 ± 0.17</td>
<td>1.89 ± 0.05</td>
<td>27.8 ± 1.3</td>
<td>0.18 ± 0.03</td>
<td>32.9 ± 1.6</td>
</tr>
<tr>
<td>Ler</td>
<td>1.72 ± 0.07**</td>
<td>6.11 ± 0.27**</td>
<td>3.42** ± 0.15**</td>
<td>4.51 ± 0.06**</td>
<td>43.1 ± 0.56**</td>
<td>0.69 ± 0.06**</td>
<td>56.6 ± 0.8**</td>
</tr>
<tr>
<td>Cvi</td>
<td>1.01 ± 0.08**</td>
<td>1.89 ± 0.05**</td>
<td>1.42 ± 0.15**</td>
<td>2.12 ± 0.12**</td>
<td>24.5 ± 0.63**</td>
<td>0.16 ± 0.03</td>
<td>30.2 ± 0.7**</td>
</tr>
</tbody>
</table>

Average values are shown in nmol/g with SE for n = 4. Asterisks indicate statistically significant differences (*P < 0.05 and **P < 0.01) relative to Col-0.

and on average exceeded those of the parental accessions (Figure 1). QTLs were identified for antheraxanthin, neoxanthin, violaxanthin, lutein, β-carotene, and total carotenoids and are designated as qCRT (for quantitative trait locus for carotenoids; see Supplemental Table 1 online). In the Cvi/Ler RIL population, 11 QTLs (qCRT1 to qCRT11) were identified with Ler as the superior parent (conferring higher trait levels) for all but two intervals, while in the Col-0/Ler population, a total of 21 QTLs (qCRT12 to qCRT33) were identified, of which 10 had Ler as the superior parent. The identified intervals could be broadly grouped into three categories: (1) intervals that contained carotenoid biosynthetic pathway genes; (2) intervals that contained carotenoid degradation genes; and (3) intervals lacking carotenoid biosynthetic pathway genes and carotenoid degradation genes and containing novel loci. For the Cvi/Ler population, 27, 0, and 73% of intervals were in categories 1, 2, and 3, respectively, while intervals in the Col-0/Ler population were 21, 17, and 63% in categories 1, 2, and 3, respectively (see Supplemental Table 1 online). qCRT27 affected the largest number of carotenoid traits in either population (six), with log of odds (LOD) scores ranging from 5.9 to 15 and explaining 30 to 58% of the phenotypic variation (see Supplemental Table 1 online). The qCRT27 interval on chromosome 4 (48 to 62 cM) spanned a 2.25-Mb region (8.60 to 10.85 Mb) containing 695 genes, of which, three were carotenoid related: PHYTOENE DESATURASE (PDS), NCED2, and CCD4.

Fine Mapping of the qCRT27 Seed Carotenoid Interval

A line containing chromosome 4 of Ler (the superior allele for qCRT27) introgressed into a Col-0 background, CSS4 (Kourproglou et al., 2002), was crossed to Col-0 to construct an F2 mapping population. Fifty independent homozygous recombinant lines were selected based on the delineation of informative recombination events (see Supplemental Data Set 1 and Supplemental Table 2 online), and their seed carotenoid levels were quantified and used for QTL analysis (Figure 2A; see Supplemental Figure 2 online). Five chromosome 4 intervals were identified, three of which were not resolved in the initial Col-0/Ler RIL population. Two new intervals from 74 to 84 cM and 93 to 103 cM had significant effects on β-carotene–derived xanthophylls, while a third from 41 to 46 cM had a modest effect on seed β-carotene levels. These data allowed narrowing of the 2.25-Mb, 695-gene qCRT27 interval to 559 kb (10.28 to 10.84 Mb) and 161 genes that still included CCD4 and NCED2 but not PDS, which was eliminated as a QTL candidate gene. The 559-kb qCRT27 region explained 26, 39, 49, 53, and 62% of the phenotypic variation for total carotenoids, violaxanthin, neoxanthin, antheraxanthin, and β-carotene, respectively, with LOD scores ranging from 3.5 to 10.34.

The presence of NCED2 and CCD4, two genes encoding carotenoid degradation enzymes, in the qCRT27 interval led us to perform additional fine mapping across the interval to resolve their respective contributions. A Col-0 × CSS4 F2 line homozygous for a Ler introgression from 8 to 16 Mb was crossed with Col-0, and additional F2 recombinants were selected across the 559-kb qCRT27 interval (see Supplemental Tables 3 and 4 online). On the basis of the recombination break points and seed carotenoid phenotypes of 14 informative recombinants, the 559-kb qCRT27 interval was further narrowed to 40.3 kb (10.46 to 10.50 Mb), which eliminated NCED2 (10.14 Mb) as a candidate gene (Figure 2B). This 40.3-kb interval contained 11 genes, of which CCD4 was the only one with a predicted function related...
to carotenoids and the only gene encoding a chloroplast-targeted protein (Figure 2C).

A GWAS Identifies CCD4 as a Key Locus Affecting Seed β-Carotene Content

Independently of the Cvi/Ler and Col-0/Ler QTLs and fine-mapping analysis described above, a GWAS of β-carotene levels in Arabidopsis seeds was performed. This approach utilized a 315-accession diversity panel that was designed to capture maximum genetic diversity and was genotyped with a high-density, genome-wide coverage 250K single-nucleotide polymorphism (SNP) array (see Supplemental Data Set 2 online; Baxter et al., 2010; Li et al., 2010; Platt et al., 2010). In order to reduce the incidence of false-positive signals, a unified mixed linear model that controls for population structure and familial relatedness was used (6PC+K [for 6 Principal Components and Kinship]; Yu et al., 2006) to test for associations between seed β-carotene levels and 170,679 of the 250,000 SNPs with minor allele frequency $\geq 5\%$.

Analysis of the natural variation of seed β-carotene revealed an 11-fold difference across the diversity panel coupled with a 64% (broad-sense) heritability, suggesting that the observed natural variation is largely dictated by genetic variation across the population rather than environmental factors (Harjes et al., 2008; Chandler et al., 2013). A strong GWAS association ($P = 6.76E-06$) with β-carotene was identified on chromosome 4. The most strongly associated SNP (i.e., the peak SNP), SNP147077, was located at 10,482,452 bp and within the coding region of CCD4 (Figures 3A and 3B), the most likely causative gene identified through fine mapping of qCRT27 (Figure 2B). Accessions with the thymine allele for SNP147077 exhibited on average 26% higher levels of seed β-carotene compared with accessions containing the cytosine allele. The SNP147077 thymine allele was present in only 20% of the 315 accessions (i.e., it is the minor allele) and explained 15% of the phenotypic variation for β-carotene in the 6PC+K model.

In a survey of linkage disequilibrium (LD) between pairs of SNPs within ±10 kb of SNP147077, we estimated that LD declined to nominal levels ($r^2 \approx 0.1$) within 7989 bp (see Supplemental Figure 3A online). Furthermore, LD estimates ($r^2$) for marker pairs with the peak SNP were $\leq 0.32$ over distances of $>4.4$ kb, indicating rapid LD decay at CCD4 (see Supplemental Figure 3B online). Based on the 250K SNP data set, this 7989-bp region was extended to 8.35 kb due to the physical positions of available SNPs. Given the estimated rate of LD decay, we focused on an 8.35-kb genomic region centered on the peak SNP, which encompassed three genes: a gene for an unknown...
protein, CCD4, and a gene encoding a GDA1/CD39 nucleoside phosphatase family protein (Figure 3B). Given the enzymatic function of CCD4 in plants (Ohmiya, 2009) and that the strongest associated SNP is within the coding region of CCD4, our previously identified fine linkage mapping candidate gene (Figure 2), CCD4, was considered as a very strong candidate gene to affect seed β-carotene levels in the association panel.

To resolve the association signals at the genomic region that included CCD4, a multiple-locus mixed model (MLMM) analysis was performed (Segura et al., 2012) for the genomic region that included all SNPs ±100 kb from peak SNP. The optimal MLMM contained only SNP147077 (see Supplemental Table 5 online) within this CCD4-centered region. When the GWAS was performed with SNP147077 as a covariate for the 6PC+K model, no other significant associations for seed β-carotene were identified (see Supplemental Figures 4A and 4B online). This is consistent with SNP147077 being the primary SNP tagging the association with seed β-carotene levels.

To assess the molecular basis of natural variation for seed β-carotene content, haploblock analysis of the 8.35-kb genomic region containing CCD4 was conducted and seven haploblocks were identified (HB1 to HB7; Figure 3B). Of these, only HB2 and HB4 had significantly different haplotypes for seed β-carotene levels (P = 2.33E-07 and P = 1.38E-07, respectively; Figure 3D). Across the panel, for HB2, four haplotypes (AC, AT, TA, and TT) were present, with TT having 40% higher β-carotene levels than AC. For HB4, the CT and TC haplotypes were present in 85 and 116 accessions, respectively, and had β-carotene levels 32 and 23% lower than the 105 accessions containing the CC haplotype. Triple asterisks indicate statistically significant differences (P < 0.0001).

**Figure 3.** Identification of CCD4 through Genome-Wide Association.

(A) Manhattan plot of genome-wide association for seed β-carotene levels using the 6PC+K model. The x axis shows the physical positions of SNPs across the five Arabidopsis chromosomes, which are shown in alternating colors. The y axis shows the negative log of P values, with each dot representing an individual SNP. The candidate region containing CCD4 is indicated by a dashed area, with the arrow indicating the location of the most significant genome-wide SNP, SNP147077, at 10,482,452 bp on chromosome 4.

(B) LD analysis relative to SNP147077 (black arrow) showed LD decay to be 8.35 kb encompassing three genes: At4g19160, At4g19170 (CCD4), and At4g19180. Black dashed boxes indicate seven haploblocks (HB) in the interval, with triple asterisks identifying HB2 and HB4, which have statistically (P = 2.33E-07 and P = 1.38E-07) contrasting haplotypes in the association panel. Inverted triangles denote InDels in the 5′ and 3′ regions of CCD4. Of these, only the 3′ InDel is significant (P = 5.49E-05, denoted by triple asterisks). The thick black line at the bottom indicates the region used in allelic complementation of ccd4-1.

(C) Thirteen nonsynonymous SNPs (NS SNPs; in green) were located across the 1.7-kb CCD4 coding region. Of these nonsynonymous SNPs, two were within HB4: SNP 10, a Lys-to-Thr conversion, and SNP 11, an Arg-to-Thr conversion. SNP 10 is a rare Ler polymorphism and was present in only 5 of the 307 accessions. Additional details are provided in Supplemental Data Set 3 online.

(D) β-Carotene seed levels (nmol/g) in lines with contrasting haplotypes for HB2 and HB4. For HB2, the 59 accessions with the TT haplotype showed β-carotene levels 40% higher than the 92 accessions with the AC haplotype. For HB4, the CT and TC haplotypes were present in 85 and 116 accessions, respectively, and had β-carotene levels 32 and 23% lower than the 105 accessions containing the CC haplotype. Triple asterisks indicate statistically significant differences (P < 0.0001).
were observed in 92, 1, 155, and 59 accessions, respectively. Accesessions carrying the HB2 TT haplotype had β-carotene levels 40% higher than accessions containing the AC haplotype (including Cvi). In addition, HB2 contained SNP147077, the peak SNP from GWAS (Figures 3B and 3C). HB4 also had four haplotypes (CC, CT, TC, and TT) present in 105, 85, 116, and 1 accessions, respectively. Accessions harboring the CC haplotype (including Ler) had 32 and 23% higher seed β-carotene levels than those containing the CT (including Cvi) or TC (including Col-0) haplotype, respectively. These haplotype data are in agreement with the estimated parent QTL allelic effects (see Supplemental Table 1 online).

**CCD4 Is Highly Expressed Late in Seed Development**

To determine if CCD4, identified as an important locus for seed carotenoid content by both linkage mapping and a GWAS (Figures 2 and 3), has a seed-related gene expression profile, its in silico gene expression during seed development and maturation in Col-0 was compared with those of genes encoding three other carotenoid cleavage enzymes: CCD1 and the chromosome 4 carotenoid cleavage enzyme genes NCED2 and CCD8 (http://www.weigelworld.org/resources/microarray/AtGenExpress; Figure 4A). Although CCD1 is on chromosome 3 and was not found to contribute to seed carotenoid natural variation in this study, previous data have implicated CCD1 in seed carotenoid turnover (Auldridge et al., 2006). Only CCD4 and CCD1 were expressed at detectable levels in Col-0 seeds, with CCD1 being constitutively expressed at high levels throughout seed development while CCD4 expression was below detection limits at stages 6 and 7 and strongly induced at later stages of seed development and maturation (stages 8 to 10; Figure 4A). Comparative analysis of CCD4 gene expression in developing Col-0 and Ler seeds showed a pattern similar to in silico data (Figure 4B). CCD4 transcript levels were low at 12 to 15 d after pollination (DAP) and strongly elevated 18 and 21 DAP (Figure 4B), although the 18- and 21-DAP levels in Ler seeds were markedly lower compared with Col-0. The level and timing of CCD4 expression and the loss of seed carotenoid content during desiccation were consistent with CCD4 activity and the difference in CCD4 expression levels between Col-0 and Ler playing an important role in the extent of carotenoid degradation late in seed maturation.

**Dry Seed Carotenoid Levels in nced2, ccd4, ccd1, and ccd1 ccd4 Knockout Mutants**

To functionally characterize CCD4 and conclusively determine its role in seed carotenoid homeostasis, a null mutant that disrupts CCD4 transcript accumulation during seed maturation (12 to 15 DAP) and desiccation (18 to 21 DAP) was identified (ccd4-1; Figures 2C and 4B). As controls, knockouts for CCD1 (Auldridge et al., 2006), a cytosolic enzyme previously shown to affect dry seed carotenoid levels, and NCED2 were also identified and characterized. As expected, nced2-3 did not significantly affect seed carotenoid levels, while ccd1-1 significantly increased total carotenoids, lutein, neoxanthin, and violaxanthin 170 to 210% and β-carotene 400% relative to the wild type (Figure 5; Auldridge et al., 2006). ccd4-1 had an even higher effect on all seed carotenoid levels. Total carotenoids in ccd4-1 were elevated 270%, lutein 230%, violaxanthin 590%, neoxanthin 390%, and β-carotene a remarkable 840% compared with the wild type. Combining ccd4-1 and ccd1-1 into a single background was additive for total carotenoids, anthoxanthin, and lutein levels (360, 470, and 240% of wild-type levels, respectively), while for β-carotene, violaxanthin, and neoxanthin (at 1710, 1220, and 1620% of wild-type levels, respectively), ccd4-1 and ccd1-1 were clearly synergistic (Figure 5). This suggests that plastid-localized CCD4 is an important component in the enzymatic turnover of carotenoids during seed maturation and desiccation, especially in relation to dry seed β-carotene levels. These data are consistent with the significant association of natural allelic variation at CCD4 with seed β-carotene levels (Figure 3).
Carotenoid Degradation during Seed Maturation and Desiccation

Like those of many crop plants, Arabidopsis seeds are green and photosynthetically active during development and maturation (King et al., 1998). As seeds enter desiccation, the degradation of chlorophyll and carotenoids culminates in dry, fully mature brown seeds (Mansfield and Briarty, 1992). To assess the rate and level of carotenoid degradation during Arabidopsis seed development, Col-0, ccd1-1, ccd4-1, and ccd1-1 ccd4-1 genotypes were analyzed at six time points distributed across seed development. The stages included fully developed mature green seeds (15 DAP), desiccating seeds (18 and 21 DAP), and three different stages of drying seeds (28, 35, and 42 DAP; Figure 6). In Col-0, individual and total seed carotenoid levels showed a steep decrease between 15 and 21 DAP, coincident with the induction of CCD4 expression (Figure 4B), followed by a slower decline to 42 DAP. Carotenoids in mutant lines followed one of two profiles leading to higher levels at 42 DAP: (1) an attenuated decrease with elevated levels throughout desiccation (e.g., β-carotene, zeaxanthin, and neoxanthin) or (2) a stage-specific increase followed by an attenuated decrease (e.g., lutein, antheraxanthin, violaxanthin, and total carotenoids; Figure 6; see Supplemental Figure 5 online). At 15 DAP, seed carotenoid levels in the three mutant genotypes were statistically indistinguishable from Col-0, indicating that loss of CCD1 and/or CCD4 activity does not affect carotenoid synthesis and deposition during seed development. However, at and beyond the onset of desiccation (18 DAP), ccd4-1 and ccd1-1 ccd4-1 and to a lesser extent ccd1-1 contained higher levels of individual and total carotenoids compared with the wild type (Figure 6; see Supplemental Figure 5 online). In general, ccd4-1 had an effect greater than or equal to ccd1-1, and in most cases, with the exception of β-carotene, the two mutations combined produced an additive effect. At 21 DAP, β-carotene levels in ccd4-1 were 11-fold higher than ccd1-1 and similar to ccd1-1 ccd4-1, indicating that ccd1-1 has a limited contribution to β-carotene retention. Indeed, at 42 DAP, ccd4-1 retained 21% of the β-carotene originally present at 15 DAP while ccd1-1 and Col-0 retained 2 and 1%, respectively. These data indicate that while both CCD4 and CCD1 contribute to the turnover of carotenoids, CCD4 activity is a major determinant of dry seed β-carotene retention.

CCD4 Also Facilitates Carotenoid Degradation during Leaf Senescence

Leaf senescence involves desiccation and pigment degradation and, in this regard, is similar to maturing and desiccating seeds. Leaf senescence can be artificially induced by extended dark treatment (Weaver and Amasino, 2001). Thus, we used dark-induced senescence to assess the contribution of CCD1 and CCD4 to this process. As expected, increasing numbers of days in darkness (corresponding to increasing progression of senescence) resulted in decreased chlorophyll levels, with no significant difference between Col-0, ccd1-1, and ccd4-1 until 6 d after dark treatment, when ccd1-1 and ccd4-1 contained significantly lower chlorophyll than Col-0 (Figure 7). Prior to dark treatment, β-carotene levels were slightly but not significantly higher in ccd4-1 compared with Col-0 or ccd1-1, a trend that was exacerbated and made significant by extended dark treatment. After 10 d of darkness, ccd4-1 contained 2.5-fold higher levels of β-carotene in comparison with Col-0 and 1.8-fold higher in comparison with ccd1-1 (Figure 7; see Supplemental Figure 6 online). These data indicate that, as in seeds, CCD4 is a major regulator of carotenoid degradation during dark-induced leaf senescence.

Transgenic Complementation Demonstrates That Polymorphisms at CCD4 Are Causal for the Natural Variation of β-Carotene Levels in Dry Seeds

To determine if allelic variation between Col-0 and Ler at the CCD4 locus and their respective lower and higher β-carotene phenotype predictions from haplotype analysis and GWAS (Figure 3D) are correct, allelic complementation of ccd4-1 was performed. From Col-0, Ler, and Cvi 5.4-kb genomic regions (Figure 3B) containing the CCD4 coding region, with 1.9 kb upstream and 2.1 kb downstream of the start and stop codons, respectively, were isolated and transformed into the ccd4-1 knockout mutant. The effects on dry seed carotenoid levels were assessed relative to the wild-type Col-0 background. Independent single-insert homozygous transgenic lines carrying the Col-0 CCD4 allele (11 lines), the Ler CCD4 allele (14 lines), or the Cvi CCD4 allele (6 lines) were selected, and all were found to functionally complement the ccd4-1 mutant to varying degrees (Figure 8). Consistent with Ler having the qCRT27 QTL allele that increases carotenoid levels in both the Cvi/Ler and Col-0/Ler mapping populations (see Supplemental Table 1 online), neoxanthin, violaxanthin, and β-carotene levels were significantly higher in Ler relative to the Col-0 complemented alleles, while the Cvi and Col-0 complemented alleles were indistinguishable.
for all compounds (Figure 8). These data demonstrate that allelic variation at the CCD4 locus is the genetic basis of the qCRT27 QTL (Figure 2; see Supplemental Table 1 online).

To assess the molecular basis of the observed complementation (Figure 8), the 5.4-kb regions of the Col-0, Ler, and Cvi CCD4 alleles were sequenced and 122 polymorphisms (including insertions, deletions, and substitutions) were identified (see Supplemental Table 6 online). Of these 122 polymorphic sites, Ler had a unique allele at 59 of them relative to Col-0 and Cvi. Based on the strong effect of the Ler CCD4 allele (Figure 8), it is likely that one or more of the 59 polymorphisms are causative. These 59 polymorphisms were distributed between the promoter (31), coding (1), and 3′ (27) regions of CCD4. Relative to Col-0, seven nonsynonymous SNPs were identified in the Ler and Cvi CCD4 coding regions, of which only one (Lys-445 to Thr) was unique to Ler (see Supplemental Table 6 online). Based on the presence of these nonsynonymous polymorphisms, which can result in protein changes, was observed in the RIL parental lines, their distribution and contribution to the β-carotene trait across the diversity panel was unknown. To assess their possible contribution, the sequence of the CCD4 coding region from 307 accessions in the diversity panel was obtained from the 1001 Arabidopsis genome database (180 accessions; http://signal.salk.edu/atg1001/3.0/gebrowser.php; Weigel and Mott, 2009) or direct amplification and Sanger sequencing (127 accessions). An additional six nonsynonymous SNPs were identified, bringing the total to 13 (see Supplemental Data Set 3 online). While these 13 nonsynonymous polymorphisms spanned the CCD4 coding region, only SNPs 10 and 11, which caused a Lys-to-Thr and Arg-to-Thr conversion, respectively, were present in a statistically significant haploblock, HB4 (P = 1.38E-07; Figure 3C; see Supplemental Table 7 online). SNP 10 was a rare polymorphism present in only 5 (including Ler) of the 307 accessions of the diversity panel; therefore, it was not possible to score it by association analysis. To assess the contribution of the other 12 nonsynonymous polymorphisms to seed β-carotene levels, a model accounting for population structure was fitted and all 12 were found to be nonsignificant, suggesting that these nonsynonymous SNPs alone are neither causal nor in strong LD with the causative polymorphism(s).

In addition to SNPs, insertions/deletions (InDels) were also identified across the diversity panel in the 5′ and 3′ ends of CCD4 (Figure 3B; see Supplemental Data Set 3 online). Three states were observed in the 5′ region at −470 nucleotides relative to the Col-0 start codon: a deletion (Col-0 state) and 36- and 56-bp insertions. Three InDel states were also identified in the 3′ CCD4 region +282 nucleotides from the Col-0 stop codon: a deletion (Col-0 state) and 19- and 120-bp insertions. To assess any contributions of InDels to seed β-carotene levels, a model was fitted using the CCD4 InDel states. The 19-bp 3′ InDel was significantly associated (P = 5.49E-05) with seed β-carotene levels (Figure 3B). Pairwise significance testing showed that accessions containing the 3′ 19-bp insertion had

Figure 6. Carotenoid Levels during Seed Development and Desiccation in Col-0, ccd1-1, ccd4-1, and ccd1-1 ccd4-1 Lines. Flowers were tagged at anthesis, seeds were harvested at specific DAP, and carotenoids were analyzed by HPLC. Average values (nmol/seed) are shown with SE for n = 3 of 150 seeds per biological replicate. Asterisks indicate statistically significant differences (*P < 0.01, **P < 0.001, and ***P < 0.0001) relative to Col-0. Error bars smaller than the symbols are not shown.
significantly higher levels of β-carotene compared with accessions containing the 120-bp InDel (Col-0) states (36 and 17%; \( P = 4.15 \times 10^{-5} \) and \( P = 1.39 \times 10^{-4} \), respectively; Figure 3D). Interestingly, although these analyses underscore the significant association of the 39 InDel with seed β-carotene content, the physical location of the 39 InDel places it in the nonsignificant HB5 (Figure 3B). These data, combined with those for the peak SNP, SNP147077, suggest that multiple variants collectively contribute to the seed β-carotene trait, and due to their modest effect size and low allele frequencies, they do not individually breach genome-wide significance, a phenomenon in agreement with previous observations (Yang et al., 2010). This seems likely that the same will be true for some novel loci as well. To identify such loci, a combinatorial approach integrating linkage mapping in two different biparental populations and a GWAS in an Arabidopsis diversity panel of 315 accessions was performed.

Analysis of the Cvi/Ler and Col-0/Ler RIL populations for seed carotenoid content identified a total of 33 QTLs, with 21, 9, and 70% of these intervals containing carotenoid biosynthetic genes, carotenoid degradation genes, and novel loci, respectively (Figure 7).

**DISCUSSION**

Approaches to increase the levels of provitamin A carotenoids in the edible portions of staple crops are crucial for ameliorating the widespread vitamin A deficiency that affects millions of people in developing countries, including an estimated 200 million preschool children (http://www.who.int/vmnis/database/vitamina/x/en/index.html). To date, efforts have focused on engineering or selecting natural variants of specific carotenoid biosynthetic enzymes for a desired outcome (Shewmaker et al., 1999; Ye et al., 2000; Paine et al., 2005; Fitzpatrick et al., 2012). However, the majority of QTL intervals affecting seed carotenoid traits in several crops do not contain carotenoid biosynthetic genes, suggesting that novel genes make substantial contributions to observed natural variation in carotenoid traits (Wong et al., 2004; Pozniak et al., 2007; Chander et al., 2008, 2013; Fernandez et al., 2008). This is also true in model plants like Arabidopsis, where ~80% of the QTL intervals in the Col-0/Ler and Cvi/Ler RIL populations lack carotenoid biosynthetic genes (see Supplemental Table 1 online). Given the past successes in applying the molecular and biochemical knowledge of carotenoid pathway enzymes obtained from model organisms to crops (Ye et al., 2000; Abbo et al., 2005; Ducrœux et al., 2005; Paine et al., 2005; Pozniak et al., 2007; Fernandez et al., 2008; Cong et al., 2009; Welsch et al., 2010; Chander et al., 2013), it seems likely that the same will be true for some novel loci as well.

**Figure 7.** Carotenoid Levels in Leaves of Col-0, ccd1-1, and ccd4-1 Lines Treated for the Indicated Days in Darkness to Induce Leaf Senescence. Individual leaves were covered with foil for 3, 6, and 10 d. Average values (nmol/g) are shown with se for \( n = 4 \). Asterisks indicate statistically significant differences (*\( P < 0.01 \) and **\( P < 0.001 \)) relative to Col-0. Error bars smaller than the symbols are not shown.

**Figure 8.** Seed Carotenoid Levels When ccd4-1 Is Functionally Complemented with Col-0, Ler, or Cvi CCD4 Alleles. Average values (nmol/g) are shown with se for 11, 14, and 6 ccd4-1 lines homozygous for the introduced Col-0, Ler, and Cvi CCD4 alleles, respectively, relative to wild-type Col-0 seeds. The dashed line indicates the wild-type Col-0 level for each seed carotenoid, with asterisks indicating statistically significant differences (*\( P < 0.01 \) and **\( P < 0.001 \)) compared with wild-type Col-0. Average carotenoid levels in wild-type Col-0 seeds (nmol/g) were as follows: neoxanthin = 1.21, violaxanthin = 3.06, antheraxanthin = 2.75, lutein = 38.81, zeaxanthin = 3.41, β-carotene = 0.19, and total carotenoids = 49.43.
haploblocks in the CCD4 panel lacks sufficient statistical power to resolve such smaller-effect QTLs. However, GWAS did allow the 40.3-kb CCD4-containing interval defined by linkage mapping (Figure 2B) to be reduced to an 8.35-kb LD interval that still contained CCD4 (Figure 3B). Furthermore, based on the location of significant haplodblocks in the CCD4 coding region and the 3' InDel, the associated interval could be further reduced to a 2.1-kb region that includes significant haplodblocks 2 and 4 and the 3' InDel (Figure 3B), which collectively encompasses much of CCD4.

To provide molecular confirmation of the predicted high and low trait contributions of alleles identified by haplotype and linkage mapping (Figure 3D; see Supplemental Table 1 online), 5.4-kb regions encompassing the equivalent promoter, coding, and 3' regions for CCD4 alleles from Col-0, Ler, and Cvi (low, high, and low β-carotene levels for RIL parents, respectively) were used for accession-specific functional complementation of the ccd4-1 null mutant (Figure 8). As expected, substantial complementation for seed carotenoid levels was achieved with all three CCD4 alleles. When the complementing accession-specific alleles were compared relative to the Col-0 knockout background, only the Ler allele resulted in a statistically significant increase in seed carotenoids, especially for β-carotene, neoxanthin, and violaxanthin (Figure 8). These data demonstrate that allelic variation present in the 5.4-kb region encompassing the CCD4 locus can recapitulate the observed QTL (Figure 2) and, moreover, that among the three accessions tested, Ler contained the most favorable combination of polymorphisms, especially for seed β-carotene levels.

As amino acid changes (nonsynonymous SNPs) in the CCD4 coding region were not significantly associated with seed β-carotene levels, developing seeds were assessed for differences in allele-specific CCD4 expression (Figure 4). Col-0 showed a significant negative correlation between CCD4 expression (Figure 4B) and seed carotenoid content (Figure 6) during late maturation and early desiccation. It is especially noteworthy that CCD4 expression levels were accession specific, with the Ler CCD4 allele being expressed at significantly lower levels late in development relative to Col-0 (Figure 4B). These contrasting expression patterns likely contribute to the higher carotenoid levels in mature dry seeds of Ler relative to Col-0 (Table 1; Figure 6) and are in agreement with Ler being the superior parental allele of qCRT27 (see Supplemental Table 1 online). These data suggest that allele-specific differences in CCD4 expression in developing seeds significantly contribute to natural variation in seed β-carotene levels (Figure 4B), a finding consistent with studies of carotenoid natural variation in maize kernels, where altered gene expression at loci for pathway biochemical enzymes also make large contributions to traits (Harjes et al., 2008; Vallabhaneni et al., 2009; Vallabhaneni and Wurtzel, 2009; Yan et al., 2010). Regulation of tissue carotenoid levels by alteration of CCD4 expression appears to be an evolutionarily conserved mechanism, as it also underlies the selection through breeding of elevated carotenoid levels in nonseed tissues, including Chrysanthemum morifolium petals, potato (Solanum tuberosum) tubers, and peach (Prunus persica) fruit flesh (Ohmiya et al., 2006; Campbell et al., 2010; Brandi et al., 2011; Falchi et al., 2013).

In Arabidopsis, the majority of enzyme-mediated degradation of carotenoids is catalyzed by the cytosolic CCD1 and plastid-localized CCD4 enzymes. In vitro assays performed with Escherichia coli strains engineered to accumulate different carotenoid profiles have identified several putative substrates for CCD1 and CCD4 orthologs from multiple organisms (Schwartz et al., 2001; Schmidt et al., 2006; García-Limones et al., 2008; Rubio et al., 2008; Huang et al., 2009; Ilg et al., 2009). However, as the full in vivo suite of carotenoids and xanthophyll substrates cannot be produced in such E. coli assay systems, knowledge of potential in vivo substrates for both enzymes has been fragmentary. In this regard, the in vivo activities defined by single and double ccd mutant lines are particularly informative. Loss of either CCD1 or CCD4 activity in seeds results in significant increases in all six seed carotenoids, indicating that both enzymes can utilize them as in vivo substrates (Figures 5 and 6; see
and Col-0/Ler, contained 162 and 100 lines, respectively (Lister and Dean, 1993; Alonso-Blanco et al., 1998). CSS4 is a chromosomal substitution line containing Ler chromosome 4 in a Col-0 background (Koumproglou et al., 2002). T-DNA knockout lines for CCD1 (At3g63520, SAIL_390_C01, ccd1-1), NCED2 (At4g18350, SALK_090937, nced2-3), and CCD4 (At4g19170, SALK_097984, ccd4-1) were PCR identified, as was the ccd1-1 ccd4-1 double mutant from a ccd1-1 and ccd4-1 cross. All lines were obtained from the ABRC.

DNA Isolation and Genotyping
Genomic DNA for PCR-based genotyping (see Supplemental Tables 2 and 3 online) of all plant lines was isolated as described by Gilliland et al. (2006). Fine mapping of qCRT27 on RIL populations and CSS4 recombinants was performed using derived cleaved-amplified polymorphic sequence markers http://helix.wustl.edu/dcaps/dcaps.html and primers designed based on the Cereon InDel database https://www.arabidopsis.org/browse/Cereon/.

Seed and Leaf Carotenoid Extraction and Quantification
Except where indicated, seed samples were harvested and dried for a minimum of 6 weeks prior to HPLC analysis. For analysis of seed and leaf carotenoids, extractions were performed from 10 to 12 mg of seeds and 20 to 25 mg of leaves weighed into a 1.4-mL U-bottom barcoded extraction tube (Micronic USA) containing two 5-mm glass beads. For seed carotenoid extraction, 450 μL of extraction buffer (2:1 [v/v] methanol:chloroform containing 1 mg/mL butylated hydroxytoluene and 500 ng of tocot [Matreya] as an internal recovery control) was added to samples followed by 10 min of grinding using a commercial paint shaker (PACER Industrial Mixers) set on high. Three hundred microliters of HPLC-grade water and 150 μL of chloroform were added and mixed. Samples were centrifuged for 10 min at 2500g in a Sorvall Legend RT centrifuge (Kendro Laboratory Products). From the lower organic phase, 200 μL was transferred to a 750-μL tube, SpeedVac dried, and resuspended in 100 μL of 70:30 (v/v) acetonitrile:ethyl acetate by shaking on a microplate shaker for 15 min at 2000 rpm. For leaf carotenoid extraction, 400 μL of leaf extraction buffer (60:40 [v/v] acetonitrile:ethyl acetate containing 1 mg/mL butylated hydroxytoluene and 1 mg/mL α-tocopherol acetate [Sigma-Aldrich] as an internal recovery control) was added to the sample and ground for 10 min as mentioned before. One hundred fifty microliters of HPLC-grade water was added, and following 10 min of centrifugation, 200 μL from the upper organic phase was transferred into a 750-μL tube for evaporation and resuspension prior to HPLC. From this point onward, seed and leaf extractions were identical. Tubes were spun for 5 min at 2500g in a Sorvall Legend RT centrifuge, and the supernatant was transferred to a microtiter plate for HPLC injection. Ten microliters of extract was resolved using a Prominence HPLC unit (Shimadzu Scientific) equipped with a Kinetex 2.6-μm, 100–4.6-mm C18 column (Phenomenex) at 40°C. Mobile phases were ethyl acetate (A) and acetonitrile:water:triethylamine (85:15:0.1 [v/v/v]; B) at 8 mL/min using the following gradient: 0 to 3 min, 100% buffer B; 3 to 9.2 min, buffer B reduced to 0% and buffer A increased to 100%; and held for 40 s; 10 to 12 min, 100% buffer B for column reequilibration. Carotenoids were analyzed at 450 nm. Standard curves were done using commercial standards or prepared in the laboratory.

Map Construction and QTL Analysis
Carotenoid levels from Cvi/Ler and 100 Col-0/Ler RILs were used for QTL analysis using composite interval mapping with Plant breeding and Biology QTL analysis (PLABQTL) (Utz and Melchinger, 1996). The PLABQTL program automatically chose cofactors used for calculations,
and a critical LOD score of 2.5 was preset. For each trait, a permutation procedure (Churchill and Doerge, 1994) was run 1000 times to determine the optimal \( \alpha = 0.05 \) threshold. These thresholds varied from 2.54 to 2.68 for different traits.

**Phenotypic Data Processing of Association Panel Data**

The phenotypic data were processed according to the method of Lipka et al. (2013). Seed \( \beta \)-carotene levels from three replicates of the Arabidopsis association panel were screened for outliers using Studentized deleted residuals (Kutner et al., 2004) and a model including replicate and accession as random effects. After outlier removal, the Box-Cox procedure (Box and Cox, 1964) was implemented to correct for nonnormality of error terms and unequal variances. Best linear unbiased predictors (BLUPs) for these accessions were predicted using a mixed linear model fitted across the three replicates and used as the phenotypes in subsequent analyses. The outlier screen, Box-Cox procedure, and BLUP model fitting were all conducted in SAS version 9.2 (SAS Institute).

**GWAS**

A 250K Affymetrix SNP genotyping data set (version 3.06, call method 75; Li et al., 2010; Horton et al., 2012) for the Arabidopsis association panel was downloaded from AtPolyDB (https://cynin.gmi.oaw.ac.at/home/resources/atpolydb/250k-snp-data/call_method_75.tar.gz/view). GWAS of seed \( \beta \)-carotene BLUPs was performed using 170,679 SNPs with \( \geq 5\% \) minor allele frequency for the 315 accessions. A unified mixed linear model (Yu et al., 2006) with population parameters previously determined (Zhang et al., 2010) was used to test for an association between each SNP and \( \beta \)-carotene content. This mixed linear model included principal components (Price et al., 2006) as fixed effects to account for population structure and a kinship matrix (Loisel et al., 2005) to account for familial relatedness. A backward elimination procedure utilizing the Bayesian information criterion (Schwarz, 1978) was implemented to determine the optimal number of principal components to include in the mixed model. The principal components and kinship matrix were calculated from 214,052 SNPs with minor allele frequency \( > 10\% \) among 1307 accessions in the 250K SNP data set. The Benjamin and Hochberg (1995) procedure was used to control the false discovery rate at 5%. The goodness of fit of each SNP was assessed using a likelihood ratio-based \( r^2 \) statistic (Sun et al., 2010). All analyses were conducted using the Genome Association and Prediction Integrated Tool software package (Lipka et al., 2012).

**LD**

LD for pairs of SNP loci was evaluated as described previously (Weir and Hill, 1986). The squared allele frequency correlations (\( r^2 \)) were estimated between pairs of loci that had SNPs with a minor allele frequency \( \geq 0.05 \) in TASSEL version 3.0 (Bradbury et al., 2007).

**Haploblock and Haplotype Analysis**

Haploblocks in an 8.35-kb genomic region surrounding CCD4 were created in Haplovlew version 4.2 (Barrett et al., 2005) using the confidence interval method. To assess the association between each haploblock and \( \beta \)-carotene content, the unified mixed linear model (Yu et al., 2006) was fitted with principal components and a kinship accounting for population structure and familial relatedness, respectively. This model was fitted in SAS version 9.2 (SAS Institute) in order to analyze the haplotypes as a class variable. The LSMEANS statement in PROC MIXED was used to compare \( \beta \)-carotene content between haplotypes. The Tukey-Kramer test was used to adjust for the multiple testing problem (Tukey, 1953; Kramer, 1956).

**MLMM Analysis**

The MLMM procedure (Segura et al., 2012) was implemented to resolve complex association signals involving a major-effect locus. The extended Bayesian information criterion (Chen and Chen, 2008) was used to identify the optimal model. All SNPs within \( \pm 100 \) kb of CCD4 were considered for inclusion in the optimal model.

**Plasmid Construct and Gene Transformation**

A 5464-bp genomic fragment containing the 1788-bp CCD4 coding region, 1622 bp of 5' flanking sequence, and 2054 bp of 3' flanking sequence was amplified from genomic DNA of Col-0, Cvi, and Ler accessions using High Fidelity Platinum Taq DNA polymerase (Invitrogen) and the primers 5'-TGGTTTATGACAAGTACAAG-3' and 5'-AGT- TGTCTTCTGATAGCTAGGGG-3'. Amplified PCR products were cloned into the pGEM-T EASY vector (Promega), fully sequenced, and transferred into the plant transformation vector pMLBART (Gleave, 1992). Constructs in Agrobacterium tumefaciens GV3101 were used to transform the cdd4-1 knockout (Col-0) by floral dipping (Clough and Bent, 1998). Primary transgenic plants were selected by BASTA screening, transformants segregating for a single transgene locus were selected, and homozygous T3 seeds were isolated and assayed for carotenoid composition by HPLC.

**Real-Time Quantitative RT-PCR**

Total RNA was isolated from developing seeds using the hot-borate protocol (Biricic and Kranner, 2006) followed by TURBO DNA-free (Ambion) DNase treatment. cDNA synthesis was performed from 1 \( \mu \)g of total RNA using SuperScript II reverse transcriptase (Invitrogen) and oligo(dT) primer. Full details are available in Supplemental Methods 1 online.

**Dark-Induced Leaf Senescence**

Rosette leaves 5, 6, 7, and 8 of Arabidopsis Col-0, cdd4-1, and cdd4-1 plants were individually covered with foil to initiate dark-induced senescence. Following foil covering, leaves were harvested at 0, 3, 6, and 10 d of darkness. For each time point, leaf samples were harvested in triplicate and carotenoids were quantified by HPLC.

**Carotenoid Degradation during Seed Development and Desiccation**

Arabidopsis flowers of Col-0, cdd7-1, cdd4-1, and cdd1-1ccd4-1 plants were tagged at anthesis, and developing seeds were collected at 15, 18, 21, 28, 35, and 42 DAP. For each developmental time point, a total of 450 developing seeds were collected (~150 seeds per line per time point in triplicate) and carotenoids were quantified by HPLC.

**Accession Numbers**

Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession numbers: PDS (At4g14210), NCED2 (At4g18350), CCD4 (At4g19170), CCD1 (At5g35320), CCD8 (At4g32810), gene of unknown protein (At4g19160), and GDA1/CD39 nucleoside phosphotase family protein (At4g19180).

**Supplemental Data**

The following materials are available in the online version of this article.

**Supplemental Figure 1.** Carotenoid Biosynthesis and Turnover Pathways in Arabidopsis.
Supplemental Figure 2. CSS4 × Col-0 Recombinant QTL Mapping.
Supplemental Figure 3. LD Plots of the CCD4 Region.
Supplemental Figure 4. Manhattan Plots of Covariate Analysis of Arabidopsis Seed β-Carotene GWAS.
Supplemental Figure 5. Carotenoid Levels (nmol/seed) during Seed Development and Desiccation in Col-0, ccd1-1, ccd4-1, and ccd1-1ccd4-1 Lines.
Supplemental Figure 6. Carotenoid Levels (nmol/g) in Leaves of Col-0, ccd1-1, and ccd4-1 Lines Treated for the Indicated Days in Darkness to Induce Leaf Senescence.

Supplemental Table 1. Summary of Seed Carotenoid QTLs Identified in Two Arabidopsis RIL Populations, Col-0/Ler and Cvi/Ler.
Supplemental Table 2. Thirty-One Primer Pairs Used for CSS4 × Col-0 Recombinant Mapping.
Supplemental Table 3. Mapping of (CSS4 × Col-0 F2) × Col-0 Recombinants across Chromosome 4.
Supplemental Table 4. Twenty-Three Primer Pairs Used for CSS4 Recombinant Mapping.
Supplemental Table 5. MLMM Analysis of Arabidopsis Seed β-Carotene for a ±100-kb Region of the Peak SNP, SNP147077.
Supplemental Table 6. Full List of Polymorphisms between Col-0, Ler, and Cvi.
Supplemental Table 7. Haploblock and Haplotype Analysis of the CCD4 LD Region.

Supplemental Methods 1. Real-Time Quantitative RT-PCR.
Supplemental Data Set 1. Mapping of CSS4 × Col-0 Recombinants and Carotenoid Quantification across Chromosome 4.
Supplemental Data Set 3. Profile of InDels and Synonymous and Nonsynonymous SNPs in the Arabidopsis Diversity Panel.

ACKNOWLEDGMENTS

We thank Payam Mehrshahi for valuable comments and discussion on the article. This work was supported by the National Science Foundation (grant DBI-0822493 to D.D.P. and C.R.B.).

AUTHOR CONTRIBUTIONS

S.G.-J. designed and performed research and wrote the manuscript. S.-H.H., M.M.-L., L.U.G., A.Z., Holly L.L., and Y.-N.N. performed research. A.E.L. and R.A. performed data analysis. Haining L., J.C., and C.R.B. contributed new computational data analysis tools. M.A.G. designed and performed data analysis. D.D.P. designed research and wrote the manuscript.

Received October 16, 2013; revised December 3, 2013; accepted December 10, 2013; published December 24, 2013.

REFERENCES


CAROTENOID CLEAVAGE DIOXYGENASE4 Is a Negative Regulator of β-Carotene Content in Arabidopsis Seeds
Sabrina Gonzalez-Jorge, Sun-Hwa Ha, Maria Magallanes-Lundback, Laura Ullrich Gilliland, Ailing Zhou, Alexander E. Lipka, Yen-Nhu Nguyen, Ruthie Angelovici, Haining Lin, Jason Cepela, Holly Little, C. Robin Buell, Michael A. Gore and Dean DellaPenna

Plant Cell; originally published online December 24, 2013;
DOI 10.1105/tpc.113.119677

This information is current as of June 23, 2017

Supplemental Data /content/suppl/2013/12/23/tpc.113.119677.DC1.html /content/suppl/2014/01/15/tpc.113.119677.DC2.html

Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain

CiteTrack Alerts Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain

Subscription Information Subscription Information for The Plant Cell and Plant Physiology is available at: http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY