Plant and Cell Physiology Advance Access published January 15, 2009

Running title: Arabidopsis carotenoid hydroxylase evolution

Corresponding author:

Dean DellaPenna, Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824-1319

Office: 517-432-9284, Fax: 517-353-9334, E-mail: dellapen@msu.edu

Subject area:

Proteins, enzymes and metabolism

Number of Figures and Tables:

Six black and white figures, two color figures and two tables

The evolution and function of carotenoid hydroxylases in Arabidopsis

Joonyul Kim[1,4], James J. Smith[2], Li Tian[3] and Dean DellaPenna[1,*]

[1]Department of Biochemistry and Molecular Biology,

[2]Lyman Briggs College and Department of Entomology, Michigan State University

[3]Department of Plant Sciences, University of California, Davis

Abbreviations:

α-Carotene-derived xanthophylls, α-xanthophylls; β-carotene-derived xanthophylls, β-xanthophylls; light-harvesting complexes, LHCs; MRCA, most recent common ancestor; Non-photochemical quenching, NPQ; Photosystem II, PS II;

Footnotes:

The nucleotide sequences of *CYP97A5* and *CYP97C3* reported in this paper have been submitted to the GenBank under the accession numbers EF587911 and EF587910, respectively.

[4]Current address of Joonyul Kim: Department of Chemistry and Biochemistry, Auburn University, Auburn, AL 36849 (jzk0015@auburn.edu)

To gain insight into the evolution of xanthophyll synthesis in *Arabidopsis thaliana*, we analyzed two pairs of duplicated carotenoid hydroxylase enzymes in *Arabidopsis thaliana:* the cytochrome P450 enzymes CYP97A3 and CYP97C1, and non-heme di-iron enzymes, BCH1 and BCH2. Hydroxylated carotenes did not accumulate in a quadruple mutant for these four genes, demonstrating that they encode the full complement of carotenoid hydroxylases in *Arabidopsis thaliana*. We were thus able to definitively infer the activity of each enzyme *in vivo* based on the phenotypes of selected double and triple mutant genotypes. The *CYP97* and *BCH* gene pairs are primarily responsible for hydroxylation of α- and β-carotenes, respectively, but exhibit some overlapping activities, most notably in hydroxylation of the β-ring of α-carotene. Surprisingly, triple mutants containing only CYP97C1 or CYP97A3 activity produced 74% and 6% of the WT lutein level indicating CYP97C1 can efficiently hydroxylate both the β- and ε-rings of α-carotene and that CYP97A3 also has low activity toward the ε-ring of α-carotene. The modes of functional divergence for the gene pairs appear distinct with the CYP97 duplicates being strongly co-expressed but encoding enzymes with different *in vivo* substrates while the BCH duplicates encode isozymes that show significant expression divergence in reproductive organs. By integrating the evolutionary history and substrate specificities of each extant enzyme with the phenotypic responses of various mutant genotypes to high light stress we propose two likely scenarios for the evolution of α-xanthophyll biosynthesis in plants from ancestral organisms.

Keywords: *Arabidopsis thaliana*, Carotenoid hydroxylase, Functional divergence, Molecular evolution, Xanthophyll synthesis

Introduction

Oxygenic photosynthesis is an ancient mechanism in cyanobacteria, red algae, green algae, and terrestrial plants that uses light and water to produce ATP and reducing power to drive carbon fixation (Blankenship 1992, Nelson and Ben-Shem 2004). The first step in oxygenic photosynthesis is absorption of light by pigments (e.g. chlorophylls and carotenoids) in the photosynthetic complexes. Carotenoids are divided into two major groups: carotenes, which are enriched in the photosystem reaction centers and xanthophylls, oxygenated carotenes that are most abundant in the light-harvesting complexes (LHCs) where they are key structural and functional components for light harvesting and photoprotection (Dall'Osto et al. 2007, Dall'Osto et al. 2006, Davison et al. 2002, Lokstein et al. 2002, Niyogi et al. 1998, Niyogi et al. 2001, Pogson et al. 1998, Tian et al. 2003). In addition to roles in photosynthesis, xanthophylls scavenge reactive oxygen species (Dall'Osto et al. 2007, Havaux and Niyogi 1999), act as colorants in flowers and fruits, as substrates for volatile production and as biosynthetic precursors for hormone synthesis (Castillo et al. 2005, Galpaz et al. 2006, Gomez et al. 1978, Kishimoto et al. 2005, Schwartz et al. 1997, Van Norman and Sieburth 2007).

From an evolutionary perspective, carotenoid synthesis in photosynthetic organisms has become increasingly more complex and diverse such that specific carotenoid species are used as chemosystematic markers (Liaaen-Jensen 1998). Some photosynthetic bacteria including cyanobacteria contain carotenoids with β-rings (e.g. β-carotene) and many produce various mono and dihydroxy xanthophyll derivatives (e.g. zeaxanthin). Algae and plants introduce additional functional groups to zeaxanthin to generate other β-carotene-derived xanthophylls (β-xanthophylls, antheraxanthin, violaxanthin and neoxanthin), some of which are used in inducible photoprotective mechanisms (e.g. non-photochemical quenching and the xanthophyll cycle). In addition to β-carotene-derived xanthophylls, some red algae and cyanobacteria (i.e. *Acaryochloris* and *Prochlorococcus*) and all green algae and plants also contain a pathway bifurcation at the level of lycopene cyclization that yields α-carotene and from which α-carotene-derived xanthophylls (α-xanthophylls) are synthesized (Adams et al. 1993, Hu et al. 1998, Schubert et al. 2006, Stickforth et al. 2003, Thayer and Björkman 1990, Yoshii 2006). The xanthophyll composition of higher plants is remarkably conserved and generally consists of three major xanthophylls, the most abundant being the α-xanthophyll, lutein, and two β-xanthophylls (violaxanthin and neoxanthin) (Demmig-Adams and Adams 1992, Königer et al. 1995).

Experiments using Arabidopsis carotenoid biosynthetic mutants with altered xanthophyll compositions make clear that the wild type combination of lutein, violaxanthin and neoxanthin is the most functionally adaptive (Dall'Osto et al. 2007, Dall'Osto et al. 2006, Lokstein et al. 2002, Pogson et al. 1998, Tian et al. 2004a, Tian et al. 2003), however, how this composition was initially established and then maintained in higher plants over evolutionary time is unknown.

Prior studies of carotenoid biosynthetic mutants in Arabidopsis coupled with the genome sequence and associated tools available in this organism have advanced understanding of xanthophyll synthesis in higher plants at the molecular level (Fig. 1 and reviewed in DellaPenna and Pogson, 2006). Production of α-xanthophylls from α-carotene requires one β- and one ε-ring hydroxylation while synthesis of β-xanthophylls from β-carotene requires two β-ring hydroxylations. Two classes of structurally unrelated enzymes catalyze these ring hydroxylations; two heme-containing cytochrome P450 hydroxylases (CYP97A3 and CYP97C1) (Kim and DellaPenna 2006, Tian et al. 2004b) and a pair of non-heme di-iron hydroxylases (BCH1 and BCH2) (Sun et al. 1996, Tian and DellaPenna 2001, Tian et al. 2003). Mutant studies have suggested that the CYP97A3 and CYP97C1 enzymes are primarily responsible for catalyzing hydroxylation of the β- and ε-rings of α-carotene, respectively, while the BCH1 and BCH2 enzymes primarily catalyze the two β-ring hydroxylations of β-carotene (Kim and DellaPenna 2006, Sun et al. 1996, Tian and DellaPenna 2001, Tian et al. 2004b). These studies have also suggested a significant degree of functional overlap among these four enzymes and the possibility of additional as yet uncharacterized carotenoid hydroxylase activities (Fiore et al. 2006, Kim and DellaPenna 2006, Tian et al. 2003).

In the current study we assess the evolution and expression of each carotenoid hydroxylase, conclusively define the *in vivo* biochemical activity of each enzyme using selected double, triple and quadruple knockout genotypes and analyze the phenotypic impact at the organismal level of mutating one or more enzyme. Integrating these data allows us to address why both classes of carotenoid hydroxylases are needed in plants and what the driving forces and biological consequences of functional divergence were within each enzyme group. These studies provide significant insight into how each gene pair functionally diverged after duplication and the essential role this has played in the evolution of the xanthophyll pathway in higher plants.

Results

Defining the full complement of carotenoid hydroxylases in Arabidopsis

To understand evolution of the xanthophyll biosynthetic pathway in photosynthetic organisms, we focused on the molecular evolution of carotenoid hydroxylases because the reactions catalyzed by these enzymes are key determinants for xanthophyll synthesis. We selected Arabidopsis as our model system because four carotenoid hydroxylase genes (*CYP97A3*, *CYP97C1*, *BCH1* and *BCH2*) had been previously identified and individually studied in detail (Kim and DellaPenna 2006, Sun et al. 1996, Tian and DellaPenna 2001, Tian et al. 2004b). A prerequisite for our study was to first determine whether these four genes represented the full complement of carotenoid hydroxylases in Arabidopsis or whether additional hydroxylase activities are present as had been previously suggested (Fiore et al. 2006, Kim and DellaPenna 2006, Tian et al. 2003, Tian et al. 2004b). To address this issue, we created and analyzed a mutant genotype that was null for the four known carotenoid hydroxylases. Two different parental genotypes were generated that were homozygous for knockouts in three of the Arabidopsis carotenoid hydroxylase genes and heterozygous for the fourth (i.e., *CYP97A3* or *CYP97C1*). When selfed, the progeny of each line segregated in a 3:1 ratio for green:white seedlings with χ^2 *p*-values of 0.50 and 0.26. White seedlings were confirmed as quadruple mutants by PCR and were lethal in soil but viable in tissue culture when supplied with a carbon source (cf. 1.5 % sucrose) (Supplementary Fig. S1). HPLC analysis showed that tissue culture grown quadruple mutants from both crosses contained approximately 10% of the WT level of α- and β-carotenes and lacked all α and β- xanthophylls (Table 1) within the sensitivity of our HPLC system (0.4% of WT levels).

Functional divergence of Arabidopsis CYP97 and BCH enzyme pairs

Having defined four genes as the full complement of carotenoid hydroxylases in Arabidopsis allowed us to use the leaf and seed xanthophyll compositions of various multiple mutant genotypes to unambiguously deduce the *in vivo* activity of individual carotenoid hydroxylases. The leaf carotenoid compositions of seven such informative mutant genotypes in reference to the respective wild type (Col-0 or Ws) are shown in Table 2. The phenotype of the *bch1 bch2* double mutant demonstrated that in the absence of BCH enzymes, CYP97A3 and/or CYP97C1 was capable of hydroxylating the β-rings of β-carotene, albeit to a lower level than wild type. Similarly, the *cyp97a3 cyp97c1* double mutant showed that in the absence of the CYP97 enzymes, the two BCH isozymes could hydroxylate the β-ring of α-carotene, though again to a lower extent than the full four enzyme complement in wild type. The *in vivo* activities of the CYP97A3 and CYP97C1 enzymes were further clarified by two triple gene knockouts: the *bch1 bch2 cyp97c1* and *bch1 bch2 cyp97a3* mutants (in which only CYP97A3 or CYP97C1 are functional, respectively). β-Xanthophylls are produced at a much higher level in *bch1 bch2 cyp97c1* than in *bch1 bch2* indicating increased *in vivo* activity of CYP97A3 toward the β-rings of β-carotene in the absence of CYP97C1. In contrast, *bch1 bch2 cyp97a3* only contains 2% of the WT (Ws) β-xanthophyll level, indicating that CYP97C1 has only a low level of *in vivo* activity toward the β-rings of β-carotene. However, lutein levels in *bch1 bch2 cyp97a3* were 74% of WT (Col-0), indicating that in addition to the previous ε-ring hydroxylation activity ascribed to CYP97C1 the enzyme also has strong activity toward the β-ring of α-carotene. Similarly, the presence of lutein at 6% of the WT level in *bch1 bch2 cyp97c1* indicates that in addition to strong α-carotene β-ring hydroxylase activity, CYP97A3 also has a low level of activity toward the ε-ring of α-carotene.

The carotenoid composition in Arabidopsis seed differs significantly from that in leaf tissue (Tian et al. 2003) (Supplementary Fig. S2). Xanthophylls account for 99% of the seed carotenoids with lutein being the most abundant at 80% of total followed by zeaxanthin. β-Carotene accounts for less than 1% of total seed carotenoids. Despite the differing carotenoid compositions of WT leaves and seed, the activities inferred from carotenoid hydroxylase mutant xanthophyll compositions in the two tissues are generally in agreement. For example, seed β-xanthophyll levels in *bch1 bch2* (deficient in both BCH activities) are reduced to 50% that of wild type (WT) but lutein is unaffected. Similarly, the *cyp97c1* mutation severely impacts seed lutein levels without affecting β-xanthophyll levels while in the *cyp97a3* mutant lutein is unaffected but β-xanthophyll levels decrease approximately 40%. In the *cyp97a3 cyp97c1* double mutant seed lutein is still severely reduced but β-xanthophylls return to near WT levels.

To assess any divergence in gene expression between the *CYP97* and *BCH* duplicates, we determined the Pearson's correlation coefficient (r) as an index of expression similarity for each gene pair (Makova and Li 2003, Wagner 2000) using data retrieved from publicly available DNA microarray datasets (Supplementary Fig. S3). The value of r is such that $-1 \leq r \leq 1$ and represents how well the corresponding data fit to the regression equation and the degree of coexpression for the duplicated genes. The + and - signs are used for positive and negative linear correlations, respectively. Because carotenoid hydroxylases are primarily involved in producing pigments for LHCs, we compared gene expression in non-stressed photosynthetic tissues (shoot apex and leaf) and in reproductive organs (floral organs and seed) (Fig. 2A). r-Values for both the *CYP97* and *BCH* gene pairs are statistically different from zero in shoot apex and leaf (*p* value <0.01) with expression of the *CYP97* gene pair being more highly correlated than the *BCH* gene pair in both tissues. In floral organs only the *CYP97* gene pair correlation is statistically significant (*p* value <0.05) while in seed correlation of neither gene pair is statistically significant. Under abiotic stress conditions correlation of the *CYP97* gene pair is significant (*p* value <0.05) in five of the seven conditions while the *BCH* gene pair correlation is only significant in two conditions.

The impact of functional divergence on plant adaptation under high light stress

To assess the biological implications of carotenoid hydroxylase diversification we examined the biochemical and physiological consequence of various knockout genotypes to infer the impact of the missing enzyme activities on fitness of the organism. Because xanthophylls play important structural and functional roles in photosynthesis we assessed the response of mutant genotypes to high light stress using non-invasive *in vivo* chlorophyll fluorescence. Non-photochemical quenching (NPQ) and the maximum photosynthetic efficiency of photosystem II (Fv/Fm) are two fluorescence-derived parameters that are routinely used as quantitative measures of photosystem adaptation and response to short-term and long-term high light stress, respectively (Holt et al. 2004, Külheim et al. 2002, Maxwell and Johnson 2000). NPQ is rapidly induced (< 1 min) in response to high light and increases non-radiative energy (e.g. heat) dissipation within the photosystems. The kinetics of NPQ induction and the maximal NPQ level reflect the coordinate induction of complex structural and functional changes in the photosystems. Fv/Fm is a measure of the maximal efficiency of photosystem II (PS II) and changes to Fv/Fm generally occur over a longer time frame than NPQ (hours versus minutes), respectively. The Fv/Fm of healthy non-stressed tissues is 0.8 and approaches zero as PS II is progressively damaged.

Fig. 3 shows changes in NPQ and Fv/Fm in the indicated mutant genotypes and WT as a function of time. The absence of either class of carotenoid hydroxylases (e.g., in the *cyp97a3 cyp97c1* and *bch1 bch2* double mutants) strongly impacted the kinetics and level of NPQ induction relative to WT, indicating the activities of both enzyme classes are required for this aspect of high light adaptation. The impact of the *cyp97a3 cyp97c1* double mutant on NPQ is partially compensated in the *cyp97a3* or *cyp97c1* single mutants suggesting that like the two BCH enzymes (Tian et al. 2003) the two CYP97 enzymes have some functional redundancy in this regard. Following 9 min of actinic light illumination the relaxation of NPQ was assessed in all genotypes for an additional 11 minutes. While some significant differences were observed between mutant genotypes and wild type (e.g., *cyp97a3*, *cyp97c1*, *cyp97a3 cyp97c1* and *npq1* being higher than WT at 800 sec) these differences were small in absolute value suggesting that only minor impacts on qI, the photoinhibitory component of NPQ, occurred between genotypes under the experimental conditions used.

Under longer-term high light stress conditions (1800 $\mu mol \cdot m^{-2} \cdot sec^{-1}$ for up to 360 min) dramatic differences in the levels and kinetics of photoinhibition were observed between genotypes (Fig. 3B). The *bch1*, *bch2* and *cyp97c1* single mutants showed modest but significant differences in photoinhibition relative to WT with the effect in the *bch1bch2* double mutant appearing additive. In contrast the *cyp97a3* and *cyp97a3 cyp97c1* genotypes were extremely rapidly and strongly photoinhibited and photobleached relative to WT (Fig. 3 and 4). These data contradict a recent report where the lack of photoinhibition following exposure of another *cyp97a3* allele to 750 $\mu mol \cdot m^{-2} \cdot sec^{-1}$ (approximately 37% of full sunlight) was used to conclude that *cyp97a3* mutants are not light sensitive (Dall'Osto et al. 2007). These apparently contradictory results can be readily explained as photoinhibition is dependent on light intensity. When 5-week old WT, *lut5-1* and *lut5-2* plants (strong and weak *cyp97a3* alleles, respectively, refer to Kim and DellaPenna, 2006) were exposed to 6 h of 1000 $\mu mol \cdot m^{-2} \cdot sec^{-1}$ (to approximate the conditions of Dall'Osto et al., 2007) *lut5-1* indeed showed only a modest increase in photoinhibition. In contrast, exposure to 1800 $\mu mol \cdot m^{-2} \cdot sec^{-1}$, which closely approximates full sunlight and the selective pressure plants would experience in nature, caused severe and enhanced photoinhibition in *lut5-1* and *lut5-2*, respectively, relative to wild type (Supplementary Fig. S4). These data indicate that had Dall'Osto et al. (2007) used a light treatment that approximates the conditions found in nature (e.g., 1800 $\mu mol \cdot m^{-2} \cdot sec^{-1}$) they too would have observed severe photoinhibition in their *cyp97a3* allele.

Evolution of the *CYP97* and *BCH* genes in plants

A catalogue of *CYP97* and *BCH* genes in photosynthetic eukaryotes that synthesize both α- and β-xanthophylls (Six et al. 2005, Yoshii 2006) was produced based on blast searches against publicly available databases (NCBI, TIGR and JGI) using the three *CYP97* gene family members (*CYP97A3*, *CYP97B3* and *CYP97C1*) and two *BCH* family members (*BCH1* and *BCH2*) of Arabidopsis as queries. tBlastn searches (Altschul et al. 1997) against the full or nearly full genome sequences of *Populus trichocarpa* (poplar), *Oryza sativa* (rice) and two green algae, *Chlamydomonas reinhardtii* (*C. reinhardtii*) and *Ostreococcus tauri* (*O. tauri*) identified the number of *CYP97* and *BCH* homologs in each organism and allowed the pattern of gene duplication in each organism to be deduced.

Phylogenetic analysis of CYP97 protein sequences (Fig. 5) supports the hypothesis that two consecutive duplications of *CYP97* ancestral genes occurred before the higher plant/green algae split with further lineage-specific gene duplications occurring in *C. reinhardtii* and *O. tauri*. All three methods of analysis (NJ, MP and ML) support one to one orthology among the Arabidopsis, poplar and rice sequences within the *CYP97A*, *CYP97B* and *CYP97C* clades. Also present in the NJ tree, but not as well supported by MP and ML are two *CYP97A* and *CYP97B* genes in *C. reinhardtii* and *O. tauri*, respectively (Fig. 5). In contrast, the *BCH* gene tree (Fig. 6) indicates that duplication of *BCH* genes in higher plants occurred in a lineage-specific fashion with three of four monocots and six of thirteen dicots in the analysis having more than one *BCH* gene. Interestingly, these gene duplications appear to have occurred relatively recently, after the monocot/dicot split.

To assess the mechanisms of these recent duplications we examined the exon-intron structure of each *BCH* gene and the homology between pairs of chromosomal segments encoding *BCH* genes in the fully sequenced genomes of Arabidopsis, poplar and rice (Fig. 7). In all cases a significant degree of conservation in *BCH* exon-intron structure and gene colinearity in adjacent chromosomal segments was observed, suggesting that duplication likely resulted from whole or segmental genome duplication. The Arabidopsis *BCH* duplication time inferred from the tree topology (Fig. 6) and Ks value (0.62 in Supplementary Fig. S5) is consistent with *BCH* duplication resulting from a recent polyploidy event (Blanc et al. 2003). Though our Ks value is somewhat lower than the median Ks for paired blocks from this polyploidy event (0.72~0.99), the confidence interval range of our calculations 0.42~0.81 (Supplementary Fig. S5) are consistent with *BCH* gene duplication resulting from this recent polyploidy event. Using the median Ks value for Arabidopsis-*Brassica rapa* orthologs (0.46) for the time calibration (Blanc et al. 2003, Henry et al. 2006) yields a *BCH* gene duplication estimate of 24~40 million years ago.

Discussion

Duplication and subsequent functional divergence of genes are being increasingly recognized as important mechanisms of evolution (Lynch and Conery 2000, Moore and Purugganan 2005, Ohno 1970). Functional divergence can occur in protein coding regions, gene expression patterns or both. In this study, we assessed the evolution and functional divergence of two duplicate gene pairs involved in carotenoid hydroxylation in Arabidopsis, *CYP97A3*/*C1* and *BCH1*/*2*. Our phylogenetic analyses (Fig. 5 and 6) showed that the *CYP97A3*/*C1* genes appear to be the result of an ancient duplication with no further recent duplication whereas the *BCH1*/*2* gene pair resulted from recent gene duplication. In order to define whether individual enzymes have distinct or overlapping activities *in vivo*, we have generated and biochemically characterized a series of mutant genotypes for one or more of the four genes. Also, we used gene phylogeny as a tool to understand the molecular evolution of carotenoid hydroxylase genes and assessed what forces may have driven the evolution, selection and maintenance of carotenoid hydroxylase activities in Arabidopsis by interpreting the phenotypic consequences of mutant genotypes under a predicted selective pressure (e.g. high light).

Prior attempts to define the *in vivo* activities of carotenoid hydroxylases in Arabidopsis based on mutant phenotypes were inconclusive because the possible existence of additional genes encoding carotenoid hydroxylase activities could not be excluded (Fiore et al. 2006, Kim and DellaPenna 2006, Tian et al. 2003, Tian et al. 2004b). In this regard the generation and characterization of the xanthophyll-deficient *bch1 bch2 cyp97c1 cyp97a3* quadruple mutant (Table 1) is critical as it demonstrates that these four genes account for the full complement of carotenoid hydroxylases in Arabidopsis thereby allowing one to accurately define the *in vivo* activities of the encoded proteins based on selected double and triple mutant phenotypes. The two classes of carotenoid hydroxylases share significant overlap in their substrate specificities, i.e. all four enzymes have the ability to hydroxylate the β-ring of α-carotene, at least to some degree *in vivo*. *In vivo* analysis showed the BCH isozymes are most active in the synthesis of β-xanthophylls, have limited activity toward the ε-ring of α-carotene but significant activity toward the β-ring of α-carotene (Tables 2 and Supplementary Fig. S2). In contrast, the CYP97 enzymes have evolved to preferentially function in α-xanthophyll synthesis and show substantial divergence in their preferred *in vivo* substrates (Table 2 and Supplementary Fig. S2). CYP97A3 is highly active toward the β-rings of both β- and α-carotene but has only limited activity toward the ε-ring of α-carotene. CYP97C1 has high activity toward both the ε-rings (Tian et al. 2004b) and β-rings of α-carotene but has only low activity toward the β-rings of β-carotene.

The activities of the individual carotenoid hydroxylases deduced from xanthophyll accumulation data in seed are consistent with that in leaves with one major discrepancy: there is a virtual absence of monohydroxy α-carotene (i.e. zeinoxanthin) and α-carotene in seed of all *cyp97c1* and *cyp97a3* containing genotypes, respectively, while these same genotypes accumulate high levels of both compounds in leaves. Rather than postulating a fundamental difference in carotenoid hydroxylase activities in the two tissues it is more likely that this discrepancy is due to differing stability of the zeinoxanthin and α-carotene produced in leaves and seed. The fact that total carotenoid levels in leaves of the nine genotypes shown in Table 2 do not significantly differ while total seed carotenoid levels of the same genotypes are decreased by as much as 85% is consistent with differential carotenoid turnover in seed and leaves. This is likely due to a combination of increased carotenoid stability in leaves due to their incorporation into protein complexes (e.g. LHCs) and elevated enzyme mediated carotenoid degradation in seed. Consistent with this idea a null mutation in Arabidopsis carotenoid cleavage dioxygenase 1 (CCD1) has been shown to significantly increase seed carotenoid levels while leaf carotenoids are unaffected (Auldridge et al. 2006).

The *in vitro* activities of individual hydroxylases against β- or ε-rings have been extensively studied by their heterologous expression in *E. coli* engineered to accumulate β-, δ- or ε-carotenes (Cunningham and Gantt 2007, Quinlan et al. 2007, Sun et al. 1996, Tian and DellaPenna 2001). While this approach has been highly informative for delineating possible *in vitro* substrate(s) and activities for each enzyme, it has inherent limitations. The number of possible *in vitro* substrates tested may be limited, may not reflect those that occur *in vivo* or, because the enzymes are produced in the absence of other endogenous pathway enzymes, their activities and regulation *in planta* may not be accurately reflected in *E. coli*. The BCH enzymes are highly active toward β-carotene and weakly active toward δ-carotene (which has a single ε-ring) when expressed in *E. coli* (Sun et al. 1996, Tian and DellaPenna 2001), in good agreement with *in vivo* activities deduced from mutant analysis (Table 2 and Supplementary Fig. S2). However, data obtained from *in vivo* and *E. coli* assays of CYP97 activities are less so. For example, when rice CYP97C was expressed in *E. coli* accumulating ε-carotene (not an *in vivo* substrate) or β-carotene as substrates it showed weak ε-ring hydroxylase activity and no β-ring hydroxylase activity, respectively (Quinlan et al. 2007), in sharp contrast to *in vivo* derived activity data (Table 2). Had an *E. coli* strain accumulating the likely preferred CYP97C *in vivo* substrate α-carotene been used, results more similar to those obtained *in vivo* might have been obtained (i.e., that CYP97C can efficiently hydroxylate both the ε-rings and β-rings of α-carotene). Similarly, the major product of rice CYP97A expressed in β-carotene-accumulating *E. coli* was β-cryptoxanthin (a mono hydroxy β-carotene derivative) while in the *bch1 bch2 cyp97c1* triple mutant, which also only contains CYP97A3 activity, dihydroxy β-carotene derivatives (violaxanthin and neoxanthin) accumulate *in vivo* and β-cryptoxanthin was below detection. This may reflect poor activity of the engineered CYP97A enzyme in *E. coli* or a requirement for CYP97A to be associated in a complex for maximal activity.

Evolution of the β-xanthophyll pathway

β-Xanthophylls are widely distributed in nature and found in all photosynthetic eukaryotes and many photosynthetic and non-photosynthetic prokaryotes. β-Carotene hydroxylation in these different organism groups is primarily carried out by three subfamilies of the non-heme di-iron monooxygenase superfamily: the plant and green algal BCH-type enzyme which shares sequence homology to the non-photosynthetic bacterial-type enzyme and the cyanobacterial CrtR-type enzyme that is more closely related to bacterial carotenoid ketolases (Tian and DellaPenna 2004). The hydroxylation of β-carotene in photosynthetic eukaryotes outside the plant lineage (e.g., diatoms and red and brown algae) is less clear as the genomes of the few organisms currently sequenced lack entire hydroxylation families and where present, enzymatic activity of putative carotene hydroxylases have not yet been reported (Armbrust et al. 2004, Barbier et al. 2005, Bowler et al. 2008, Coesel et al. 2008, Cunningham et al. 2007, Glöckner et al. 2000). For example, the red alga *Cyanidioschyzon merolae* contains only a single *CrtR*-type gene and the activity of the encoded enzyme could not be demonstrated when expressed in *E. coli* (Cunningham et al. 2007). In contrast, the sequenced genomes *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (diatoms) lack *CrtR*-type genes and *T. pseudonana* encodes only a single *BCH*-type gene and no P450-type (*CYP97*-like) genes while the *P. tricornutum* genome lacks *BCH*-type sequences and instead contains two P450-type (*CYP97*-like) genes (Coesel et al. 2008). It is difficult to draw conclusions from such a limited data set but it is clear that the *BCH*–type genes in diatoms and red and brown algal lineages have not been under strong selection pressure like in the green plant lineage and have instead evolved different ways to produce β-xanthophylls.

In green plant lineage, Arabidopsis BCH1 and BCH2 are isozymes (Tian and DellaPenna 2001, Tian et al. 2003) and have relatively strong expression correlation in non-stressed photosynthetic tissues (Fig. 2), suggesting their most recent common ancestor (MRCA) had a similar activity and was expressed in photosynthetic tissues. The *bch1 bch2* double mutant has negatively impacted NPQ and Fv/Fm (Fig. 3) suggesting that photoprotection was likely an important function of the BCH containing ancestor and a strong selective pressure for retention of this activity during evolution. The necessity of β-xanthophylls as substrates for abscisic acid (ABA) synthesis may have been an additional selective pressure for retention of BCH activity as *bch1 bch2* has insufficient ABA production to control water loss under drought conditions (Tian et al. 2004a).

Regardless of the underlying selection pressure, the retention of *BCH* paralogs appears to be widespread in higher plants (Fig. 6). In Arabidopsis, duplication of the ancestral *BCH* gene most likely occurred by whole genome or segmental genome duplication (Fig. 7) and the duplicates seem to have functionally diverged primarily at the gene expression level (Fig. 2). Our carotenoid analyses suggest the BCH2 enzyme is more actively involved in β-xanthophyll synthesis in seeds than in leaves, consistent with the rapid and strong induction of *BCH2* gene expression during seed development (Supplementary Fig. S2 and S3). Tissue-specific expression divergence of *BCH* genes is not restricted to Arabidopsis, as one of the two *BCH* members in bell pepper, tomato and saffron (*C. sativus*) also shows preferential expression in flowers or during fruit development (Bouvier et al. 1998, Castillo et al. 2005, Galpaz et al. 2006). In tomato a *bch2* mutant results in a colorless petal phenotype with no impact on β-xanthophyll synthesis in leaves (Galpaz et al. 2006) and the massive accumulation of β-xanthophylls during maturation of the saffron stigma was correlated with high expression of a single *BCH* gene (*C. sativus BCH1* in Fig. 6) (Castillo et al. 2005). Flower development involves the transformation of chloroplasts to chromoplasts (Whatley and Whatley 1987) and it is likely that expression divergence of recently duplicated *BCH* genes in reproductive organs (e.g. chromoplast-specific expression) provides the biochemical flexibility to differentially regulate β-xanthophyll synthesis in these tissues.

Evolution of α-xanthophyll synthesis

Unlike β-xanthophyll synthesis, which is widespread in nature, the synthesis of α-xanthophylls occurs in only a few lineages of photosynthetic eukaryotes: some red algae and all green algae and plants. The prevalence of α-xanthophylls in both green algae and plants suggests that their MRCA synthesized α-xanthophylls and that strong selection pressure has maintained this trait. Four reaction steps are required for the synthesis of a dihydroxy α-xanthophyll (i.e., lutein): β- and ε-ring formation from lycopene by the action of β- and ε-cyclases followed by hydroxylation of each ring by β- and ε-ring hydroxylase activities (Fig. 1). Like β-xanthophylls, β-cyclases are widespread in nature and have been recruited for use in α-xanthophyll synthesis. ε-Cyclase genes have only been identified in green algae, plants and *Prochlorococcus marinus* MED4 (a cyanobacterium) and appear to have arisen from β-cyclases by gene duplication and subsequent functional divergence (Cunningham et al. 2007, Hess et al. 2001, Krubasik and Sandmann 2000, Partensky et al. 1993, Stickforth et al. 2003). Duplication in *P. marinus* MED4 appears to have occurred independently rather than gene duplication from a MRCA of green plants and cyanobacteria (Stickforth et al. 2003). The phylogeny of *CYP97* β- and ε-ring hydroxylases similarly shows that duplication of their MRCA also occurred before the evolutionary divergence of green algae and higher plants and that the *CYP97A* and *CYP97C* genes have been under purifying selection (Fig. 5). Though these phylogenetic inferences provide insight into when the genetic materials for α-xanthophyll synthesis were generated they cannot answer when the function of each gene evolved and how each function has been maintained during evolutionary time. However, the numerous mutant genotypes affecting specific carotenoid biosynthetic enzymes in Arabidopsis do provide important insights into these questions.

The conservation of α-xanthophyll synthesis during the evolution of green algae and plants suggests strong selective pressures were involved in the generation and maintenance of the necessary enzymatic activities, the ε-cyclase and the two carotenoid hydroxylases required sequentially in α-xanthophyll synthesis, CYP97A3 and CYP97C1 (Kim and DellaPenna 2006). To assess these functional constraints, we used several informative carotenoid biosynthetic mutant genotypes to infer the impact that various pathway intermediates would have when accumulated by a hypothetical ancestral organism (i.e., a β-xanthophyll producing organism) as it acquired one or more of the three necessary enzymatic activities. The *lut2* mutant, which is defective in ε-cyclase activity and cannot synthesize α-xanthophylls (Pogson et al. 1996) is a reasonable approximation of a hypothetical β-xanthophyll accumulating ancestral organism. The absence of lutein and all α-xanthophyll intermediates in *lut2* results in elevated levels of leaf β-xanthophylls, partial impairment of NPQ, a smaller photosystem cross sectional area and lower LHC trimer stability (Dall'Osto et al. 2007, Lokstein et al. 2002, Pogson et al. 1996, Pogson et al. 1998). While mature *lut2* plants grow as well as wild type under moderate light conditions they are slightly less resistant to high light stress (Dall'Osto et al. 2006, Niyogi et al. 2001, Pogson et al. 1998).

The consequences of the hypothetical ancestral organism first acquiring ε-cyclase activity is approximated by the *cyp97a3 cyp97c1* genotype, which has a functional ε-cyclase, can synthesize α-carotene, and has BCH enzyme activity but lacks both CYP97A and CYP97C1 activities. In addition to β-carotene and β-xanthophylls, *cyp97a3 cyp97c1* accumulates high levels of α-carotene and zeinoxanthin in approximately equal molar ratios (Table 2), due to inefficient hydroxylation of the β-ring of α-carotene by the extant BCH enzymes present in this mutant background. Like *cyp97a3, cyp97a3 cyp97c1* is highly susceptible to photooxidation in high light (Fig. 3 and 4). The high light susceptibility of *cyp97a3 cyp97c1* is associated with the *cyp97a3* mutation (and presence of α-carotene) rather than *cyp97c1* (and presence of zeinoxanthin) as *cyp97c1*, which accumulates an amount of zeinoxanthin equivalent to *cyp97a3 cyp97c1* but lacks α-carotene, is much less photosensitive (Fig. 3B and 4). Correlation of the presence of α-carotene in leaf tissue and increased susceptibility to photodamage has been previously suggested from several ecophysiological studies of shade-grown plants (Demmig-Adams and Adams 1992, Krause et al. 2004, Thayer and Björkman 1990). Elevated singlet oxygen production by isolated pigment:protein complexes and increased lipid peroxidation in detached leaves subjected to combined high light and low temperature treatment (1000 $\mu mol \cdot m^{-2} \cdot sec^{-1}$ at $10^{o}C$) (Dall'Osto et al. 2007, Havaux and Niyogi 1999, Muller-Moule et al. 2003) have reported for various xanthophyll biosynthetic mutant genotypes and may contribute to the enhanced photosensitivity of *cyp97a3* containing genotypes. Regardless of the precise mechanism(s) responsible, the phenotype of the *cyp97a3 cyp97c1* mutant suggests that acquisition of ε-cyclase activity by an ancestral organism would have been detrimental under full sunlight and would have set in place a situation that would strongly select for the evolution of enzymes that could efficiently hydroxylate α-carotene.

Given the strong selection pressure imposed by acquisition of ε-cyclase activity from an ancestral organism it is surprising that the extant BCH-type enzymes did not evolve to more efficiently hydroxylate the β- and ε-rings of α-carotene. This suggests that there is a structural constraint on the BCH class of enzymes that makes evolving efficient hydroxylation of α-carotene impossible. The fact that none of the BCH-type enzymes assayed in *E. coli* to date can efficiently hydroxylate ε-ring containing carotenoids is consistent with this idea (Sun et al. 1996, Tian and DellaPenna 2001, 2004). In this light it is possible to understand why and how a separate class of carotene hydroxylases, the CYP97 family, evolved and was selected for during the evolution of α-xanthophyll synthesis (Fig. 8). The original CYP97 enzyme acquired by the ancestral organism likely had at least α-carotene β-ring hydroxylation activity (CYP97A-like activity) as both the extant CYP97A3 and CYP97C1 enzymes have this activity. Comparison of the *cyp97c1* and *cyp97a3 cyp97c1* mutant phenotypes suggests the evolution of an efficient CYP97A-like α-carotene β-ring hydroxylation activity would have been strongly selected for as it would have significantly alleviated the photooxidation observed in *cyp97a3* (Fig. 3B and 4).

Subsequent duplication of the ancestral CYP97A-like enzyme and evolution of a CYP97C-like (ε-ring hydroxylase) activity would have allowed efficient synthesis of lutein without accumulation of α-carotene or zeinoxanthin. A likely driving force for selection of CYP97C-like activity is efficient photosystem structure and function which though less intuitive than the photooxidation driven selection of CYP97A-like activity is no less important (Bassi et al. 1999, Lokstein et al. 2002, Niyogi et al. 1997, Pogson et al. 1998). In plants, lutein is essential for the assembly and stability of large light harvesting photosystems and for optimal NPQ kinetics and maximal NPQ (Dall'Osto et al. 2007, Dall'Osto et al. 2006, Lokstein et al. 2002, Pogson et al. 1998). While the impact of xanthophyll alterations on NPQ is most readily quantified under experimental constant high light conditions, such as those used in Fig. 3, the impact of altered NPQ on plant fitness is most apparent under natural light conditions. Külheim *et* al. (2002) showed that NPQ deficiency had no discernable impact on fitness when plants were grown under constant light in growth chamber conditions but when grown in natural lighting, in which large swings in light intensity occur on the order of seconds to minutes, the fitness of NPQ deficient plants was severely impaired and the mutants produced 30-50% fewer seed per generation than wild type. Given such a strong negative impact on fitness it is clear how evolution of a CYP97C-like activity that positively impacts photosystem superstructure and associated processes like NPQ efficiency would be strongly selected for and maintained.

The scenario we have described above (the sequential acquisition of ε-cyclase, CYP97A and CYP97C activities) is not the only possible sequence of events in the evolution of α-xanthophyll synthesis based on the data provided. It is equally probable that a CYP97A-like enzyme was already present in the ancestral organism and along with a BCH type enzyme, was involved in β-xanthophyll synthesis. In this case the acquisition of ε-cyclase activity would have resulted in monohydroxy α-carotene (zeinoxanthin) production and duplication of the CYP97A-like gene and evolution and selection of CYP97C-like activity would then occur as described. Regardless of the exact sequence of events, the phenotypes of mutants in extant carotenoid hydroxylases (Bassi et al. 1999, Dall'Osto et al. 2007, Dall'Osto et al. 2006, Fiore et al. 2006, Kim and DellaPenna 2006, Niyogi et al. 2001, Pogson et al. 1998, Tian and DellaPenna 2004, Tian et al. 2003, Tian et al. 2004b) highlight the strong selective pressures that likely operated during the evolution and selection of particular genes and activities leading from β-xanthophyll accumulating ancestors to the current biosynthetic pathway producing both β- and α-xanthophylls found in all green algae, plants and many red algae. These same selective pressures continue to operate today to maintain the suite of carotenoid biosynthetic enzymes that have been shown to be optimal for light harvesting, photosystem structure, NPQ and adaptation of plants to the ever changing light conditions in nature.

Materials and Methods

Identification of triple and quadruple gene knockout mutants

To isolate an *Arabidopsis thaliana* triple gene knockout in which two *BCH*s and *CYP97A3* are disrupted (*b1b1 b2b2 a3a3 C1C1*), the double *BCH* gene knockout (*b1b1 b2b2*) and *cyp97a3* gene knockout (*a3a3*) were crossed, and F2 progeny screened by HPLC for α-carotene accumulation (the *cyp97a3* mutant phenotype) and a low level of β-xanthophylls (the *bch1 bch2* double mutant phenotype). Putative triple mutant lines were confirmed by genotyping with primer pairs specific for each T-DNA insertion (Kim and DellaPenna 2006, Tian et al. 2003). The two different parental genotypes used for quadruple gene knockout selection (*b1b1 b2b2 a3a3 c1C1* and *b1b1 b2b2 a3A3 c1c1)* were identified by PCR from a segregating F2 population of a *b1b1 b2b2 a3a3 C1C1* x *b1b1 b2b2 A3A3 c1c1* cross. Progeny from each quadruple mutant parent genotype (*b1b1 b2b2 a3a3 c1C1* and *b1b1 b2b2 a3A3 c1c1)* were scored on Petri plates containing 0.5X Murashige and Skoog basal medium (MS salts; Sigma, St. Louis, MO) and 1.0% phytoagar under a 12-h photoperiod (50 µmol· m^{-2}·sec^{-1}) following 5 days imbibition at 6^{o}C to break dormancy as previously described (Sattler et al. 2004). More than 99% of the progeny from both crosses germinated. The genotype of putative quadruple mutants were confirmed by PCR using gene specific primers as described (Tian et al. 2003)

Plant growth and pigment and *in vivo* chlorophyll fluorescence analyses

Leaves of four and five-week old mutant genotypes and corresponding wild type ecotypes for each mutant line (Col-0 and *Ws* backgrounds) were grown under a 12 h photoperiod (100 µmol m^{-2}·sec^{-1}, 22/18 ºC) and used for pigment analyses and high light stress experiments (1600~1800 µmol·m^{-2}·sec^{-1}, 50% humidity and 22 ºC). Seed were harvested and stored in a desiccator for one month prior to pigment analysis (Auldridge et al. 2006, Tian et al. 2003). HPLC separation and quantification of carotenoids by spectra and retention time were performed as previously described (Kim and DellaPenna 2006, Tian and DellaPenna 2001). To obtain the two photosynthetic parameters NPQ and Fv/Fm, *in vivo* chlorophyll fluorescence was measured over a time course of 5 minutes in dark-adapted intact leaves by applying a saturation pulse with and without actinic light (530 µmol·m^{-2}·sec^{-1}) using the Imaging-PAM Chlorophyll Fluorometer (Walz, Germany) (Berger et al. 2004). NPQ was calculated according to the equation, NPQ =(Fm

– F′m) /F′m where Fm is the maximum Chl fluorescence from dark-adapted leaves and F′m is the maximum Chl fluorescence under actinic light. To measure recovery in Fv/Fm, plants exposed to high light for 360 min were placed in the dark for twelve hours followed by a 100 $\mu mol \cdot m^{-2} \cdot sec^{-1}$ light illumination for 360 min.

Expression data analyses

Expression data was obtained from the AtGenExpress database (http://www.arabidopsis.org/info/expression/ATGenExpress.jsp) for the *CYP97* and *BCH* gene pairs except for high-light stressed leaf tissue data which was provided by Dr. Dirk Inzé (Vandenabeele et al. 2004). The four tissue types and seven stress conditions used for correlation analysis were independent and non-redundant. The individual sources for each dataset are listed in Supplementary Fig. S3. For each tissue and stress condition, only datasets in which at least one of the two copies for each family member was expressed (p value <0.05 for present call) were selected for analysis (Makova and Li 2003). Pearson's coefficient was calculated using Microsoft Excel 2003. To test whether the r-values obtained were significantly different from zero we used a student's t-test and the equation $t= r/\mathrm{sqrt}[(1-r^2)/(N-2)]$ where it is assumed that t is normally distributed with df=N-2 and where N is a sample size equal to or greater than 6. (Makova and Li 2003, Spiegel et al. 2000)

Protein sequence alignments and Phylogenetic analyses

CYP97 and *BCH* homologs (E value cutoff $<10^{-30}$) in photosynthetic eukaryotes were identified by tblastn searches using the three Arabidopsis *CYP97* genes (*CYP97A3*, *CYP97B3* and *CYP97C1*) and two *BCH* genes (*BCH1* and *BCH2*) as queries of publicly available databases (NCBI, TIGR and JGI). To assist in genome sequence annotation, we experimentally determined the full-length cDNA sequences for two *C. reinhardtii* homologs (*CYP97A5*, EF587911 and *CYP97C3*, EF587910) amplified from cDNA libraries (Davies et al. 1996). All sequences used in this study are summarized in Supplementary Fig. S6.

Twenty-seven CYP97 and thirty-three BCH amino acid sequences were aligned using ClustalX 1.81 with default parameters (Thompson et al. 1997) and alignments were further refined manually (Supplementary Fig. S7 and S8). Protein sequence rather than DNA sequence was used for alignments as protein sequence is more suitable for long-term evolution studies

(Glazko and Nei 2003, Hashimoto et al. 1994, Nam et al. 2003, Russo et al. 1996). The CYP97 sequence alignment (Supplementary Fig. S7) consisted of 789 aligned positions, and positions 160 to 705 were used for phylogenetic analyses. In the BCH alignment (Supplementary Fig. S8) all 414 aligned amino acid positions were used for phylogenetic analysis.

For phylogenetic analyses, three different methods were applied: Neighbor-Joining (NJ), Maximum Parsimony (MP) and Maximum Likelihood (ML). To construct NJ trees, Mega 3.1 (Kumar et al. 2004) was used employing the Poisson correction (PC) distance method with pairwise deletion of gaps. MP analyses of CYP97 and BCH amino acid sequences were carried out using PAUP* 4.0 beta 10 (Swofford 2003) using the heuristic search option with 10 replicates of random taxon addition and gaps treated as missing data. ML analysis was carried out using PhyML (Guindon and Gascuel 2003) with the JTT model (Jones et al. 1992) of protein sequence evolution and gaps treated as unknown characters. For all three analysis types, branch support was assessed by bootstrapping (500 replicates). CYP86A1 and *C. reinhardtii* BCH were selected as outgroups for each dataset, respectively. CYP86A1 was selected because its substrates, fatty acids with chain lengths from C_{12} to C_{18} (Benveniste et al. 1998), are the most similar to carotenoids and the CYP86 clade is most closely related to CYP97 clade (Werck-Reichhart et al. 2002). All P450 nomenclature follows standard convention (http://drnelson.utmem.edu/CytochromeP450.html).

Non-synonymous (Ka) and synonymous,(Ks) substitution rates in *BCH* gene pairs in Arabidopsis, poplar and rice (Supplementary Fig. S9) were calculated using K-estimator 6.0 with Kimura-2p as the multiple-hits correcting method (Comeron 1999). The 99% confidence intervals (C.I.) were calculated using Monte Carlo simulations with 500 replicates.

Colinearity of chromosomal segments

Protein sequences in Arabidopsis and rice chromosomal segments were obtained from publicly available annotated sequence data (AT4G25580 to AT4G25750 and AT5G52300 to AT5G52860 in Arabidopsis and Os03g0122400 to Os03g0127500, Os04g0576300 to Os04g0581100 and Os10g0531300 to Os10g0536000 in rice). Poplar genome segments including the two poplar *BCH* genes (Scaffold 44 and NZ_AARH01002240.1) were retrieved by tblastn search using poplar BCH homolog 1 and 2 protein sequences as queries against the poplar genome database (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=3694) and a genome segment adjacent to NZ_AARH01002240.1 (NZ_AARH01002243.1) was also retrieved using the end sequence of NZ_AARH01002240.1. The overlapping regions of NZ_AARH01002240.1 and NZ_AARH01002243.1 were aligned and used to produce the 108 kb contig shown in Figure 5. Putative open reading frames in the two poplar segments were deduced by FGENESH algorithm available in Softberry, Inc. (www.softberry.com). Annotated proteins in Arabidopsis, rice or poplar genome segments were compared in a pairwise fashion using bl2seq software available at NCBI. The cutoff for declaring a homologous gene pair between genome segments was an *E*-value of less than 10^{-10}.

Supplementary data

Supplementary data are available at PCP Online.

Funding

This work was supported by grants from the National Science Foundation [IBN 0131253] and from HarvestPlus (www.HarvestPlus.org), a global alliance of agriculture and nutrition research institutions working to increase the micronutrient density of staple food crops through biofortification.

Acknowledgements

We would like to thank Dr. Dirk Inzé for providing their high light stress microarray dataset, Hoo-Sun Chung for cloning of *Chlamydomonas CYP97* cDNAs and Dr. Shin-Han Shiu and Jongmin Nam for critical review of the manuscript.

References

Adams, W.W.III., Demmig-Adams, B. and Lange, O.L. (1993) Carotenoid composition and metabolism in green and blue-green-algal lichens in the field. *Oecologia* 94: 576-584.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., et al. (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306: 79-86.

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

Barbier, G., Oesterhelt, C., Larson, M.D., Halgren, R.G., Wilkerson, C., Garavito, R.M., et al. (2005) Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. *Plant Physiol.* 137: 460-474.

Bassi, R., Croce, R., Cugini, D. and Sandona, D. (1999) Mutational analysis of a higher plant antenna protein provides identification of chromophores bound into multiple sites. *Proc. Natl Acad. Sci. U S A* 96: 10056-10061.

Benveniste, I., Tijet, N., Adas, F., Philipps, G., Salaun, J.P. and Durst, F. (1998) CYP86A1 from *Arabidopsis thaliana* encodes a cytochrome P450-dependent fatty acid omega-hydroxylase. *Biochem. Biophys. Res. Commun.* 243: 688-693.

Berger, S., Papadopoulos, M., Schreiber, U., Kaiser, W. and Roitsch, T. (2004) Complex regulation of gene expression, photosynthesis and sugar levels by pathogen infection in tomato. *Physiologia Plantarum* 122: 419-428.

Blanc, G., Hokamp, K. and Wolfe, K.H. (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Res.* 13: 137-144.

Blankenship, R.E. (1992) Origin and early evolution of photosynthesis. *Photosynth. Res.* 33: 91-111.

Bouvier, F., Keller, Y., d'Harlingue, A. and Camara, B. (1998) Xanthophyll biosynthesis: molecular and functional characterization of carotenoid hydroxylases from pepper fruits (*Capsicum annuum* L.). *Biochim. Biophys. Acta* 1391: 320-328.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456: 239-244.

Castillo, R., Fernandez, J.A. and Gomez-Gomez, L. (2005) Implications of carotenoid biosynthetic genes in apocarotenoid formation during the stigma development of *Crocus sativus* and its closer relatives. *Plant Physiol.* 139: 674-689.

Coesel, S., Obornik, M., Varela, J., Falciatore, A. and Bowler, C. (2008) Evolutionary origins and functions of the carotenoid biosynthetic pathway in marine diatoms. *PLoS ONE* 3: e2896.

Comeron, J.M. (1999) K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics* 15: 763-764.

Cunningham, F.X., Jr. and Gantt, E. (2007) A portfolio of plasmids for identification and analysis of carotenoid pathway enzymes: *Adonis aestivalis* as a case study. *Photosynth. Res.* 92: 245-259.

Cunningham, F.X., Jr., Lee, H. and Gantt, E. (2007) Carotenoid biosynthesis in the primitive red alga *Cyanidioschyzon merolae*. *Eukaryot. Cell* 6: 533-545.

Dall'Osto, L., Fiore, A., Cazzaniga, S., Giuliano, G. and Bassi, R. (2007) Different roles of α- and β-branch xanthophylls in photosystem assembly and photoprotection. *J. Biol. Chem*. 282: 35056-35068.

Dall'Osto, L., Lico, C., Alric, J., Giuliano, G., Havaux, M. and Bassi, R. (2006) Lutein is needed for efficient chlorophyll triplet quenching in the major LHCII antenna complex of higher plants and effective photoprotection *in vivo* under strong light. *BMC Plant Biol.* 6: 32.

Davies, J.P., Yildiz, F.H. and Grossman, A. (1996) Sac1, a putative regulator that is critical for survival of *Chlamydomonas reinhardtii* during sulfur deprivation. *EMBO J.* 15: 2150-2159.

Davison, P.A., Hunter, C.N. and Horton, P. (2002) Overexpression of β-carotene hydroxylase enhances stress tolerance in *Arabidopsis*. *Nature* 418: 203-206.

DellaPenna, D. and Pogson, B. J. (2006) Vitamin synthesis in plants: tocopherols and carotenoids. *Annu. Rev. Plant Biol.* **57: 711-738**

Demmig-Adams, B. and Adams, W.W.III. (1992) Carotenoid composition in sun and shade leaves of plants with different life forms *Plant Cell Environ.* 15: 411-419.

Fiore, A., Dall'osto, L., Fraser, P.D., Bassi, R. and Giuliano, G. (2006) Elucidation of the β-carotene hydroxylation pathway in *Arabidopsis thaliana*. *FEBS Lett.* 580: 4718-4722.

Galpaz, N., Ronen, G., Khalfa, Z., Zamir, D. and Hirschberg, J. (2006) A chromoplast-specific carotenoid biosynthesis pathway is revealed by cloning of the tomato *white-flower* locus. *Plant Cell* 18: 1947-1960.

Glazko, G.V. and Nei, M. (2003) Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* 20: 424-434.

Glöckner, G., Rosenthal, A. and Valentin, K. (2000) The structure and gene repertoire of an ancient red algal plastid genome. *J. Mol. Evol.* 51: 382-390.

Gomez, R., Goni, F.M. and Macarulla, J.M. (1978) Carotenoids from marigold (*Tagetes erecta*) petals and their esterified fatty acids. *Rev. Esp. Fisiol.* 34: 253-256.

Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52: 696-704.

Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., et al. (1994) Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. *Mol. Biol. Evol.* 11: 65-71.

Havaux, M. and Niyogi, K.K. (1999) The violaxanthin cycle protects plants from photooxidative damage by more than one mechanism. *Proc. Natl Acad. Sci. U S A* 96: 8762-8767.

Henry, Y., Bedhomme, M. and Blanc, G. (2006) History, protohistory and prehistory of the *Arabidopsis thaliana* chromosome complement. *Trends Plant Sci.* 11: 267-273.

Hess, W.R., Rocap, G., Ting, C.S., Larimer, F., Stilwagen, S., Lamerdin, J., et al. (2001) The photosynthetic apparatus of *Prochlorococcus*: Insights through comparative genomics. *Photosynth. Res.* 70: 53-71.

Holt, N.E., Fleming, G.R. and Niyogi, K.K. (2004) Toward an understanding of the mechanism of nonphotochemical quenching in green plants. *Biochemistry* 43: 8281-8289.

Hu, Q., Miyashita, H., Iwasaki, I., Kurano, N., Miyachi, S., Iwaki, M., et al. (1998) A photosystem I reaction center driven by chlorophyll *d* in oxygenic photosynthesis. *Proc. Natl Acad. Sci. U S A* 95: 13319-13323.

Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8: 275-282.

Kim, J. and DellaPenna, D. (2006) Defining the primary route for lutein synthesis in plants: the role of *Arabidopsis* carotenoid β-ring hydroxylase CYP97A3. *Proc. Natl Acad. Sci. U S A* 103: 3474-3479.

Kishimoto, S., Maoka, T., Sumitomo, K. and Ohmiya, A. (2005) Analysis of carotenoid composition in petals of calendula (*Calendula officinalis* L.). *Biosci. Biotechnol. Biochem.* 69: 2122-2128.

Königer, M., Harris, G.C., Virgo, A. and Winter, K. (1995) Xanthophyll-cycle pigments and photosynthetic capacity in tropical forest species: a comparative field study on canopy, gap and understory plants. *Oecologia* 104: 280-290.

Krause, G.H., Grube, E., Koroleva, O.Y., Barth, C. and Winter, K. (2004) Do mature shade leaves of tropical tree seedlings acclimate to high sunlight and UV radiation? *Funct. Plant Biol.* 31: 743-756.

Krubasik, P. and Sandmann, G. (2000) Molecular evolution of lycopene cyclases involved in the formation of carotenoids with ionone end groups. *Biochem. Soc. Trans.* 28: 806-810.

Külheim, C., Ågren, J. and Jansson, S. (2002) Rapid regulation of light harvesting and plant fitness in the field. *Science* 297: 91-93.

Kumar, S., Tamura, K. and Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5: 150-163.

Liaaen-Jensen, S. (1998) Carotenoids in Chemosystematics. In *Carotenoids: Biosynthesis and Metabolism*. Edited by Britton, G., Liaaen-Jensen, S. and Pfander, H. pp. 217-247. Birkhäuser, Basel.

Lokstein, H., Tian, L., Polle, J.E. and DellaPenna, D. (2002) Xanthophyll biosynthetic mutants of *Arabidopsis thaliana*: altered nonphotochemical quenching of chlorophyll fluorescence is due to changes in Photosystem II antenna size and stability. *Biochim. Biophys. Acta* 1553: 309-319.

Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151-1155.

Makova, K.D. and Li, W.H. (2003) Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* 13: 1638-1645.

Maxwell, K. and Johnson, G.N. (2000) Chlorophyll fluorescence-a practical guide. *J. Exp. Bot.* 51: 659-668.

Moore, R.C. and Purugganan, M.D. (2005) The evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.* 8: 122-128.

Muller-Moule, P., Havaux, M. and Niyogi, K.K. (2003) Zeaxanthin deficiency enhances the high light sensitivity of an ascorbate-deficient mutant of *Arabidopsis*. *Plant Physiol.* 133: 748-760.

Nam, J., dePamphilis, C.W., Ma, H. and Nei, M. (2003) Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol. Biol. Evol.* 20: 1435-1447.

Nelson, N. and Ben-Shem, A. (2004) The complex architecture of oxygenic photosynthesis. *Nat. Rev. Mol. Cell Biol.* 5: 971-982.

Niyogi, K.K., Björkman, O. and Grossman, A.R. (1997) The roles of specific xanthophylls in photoprotection. *Proc. Natl Acad. Sci. U S A* 94: 14162-14167.

Niyogi, K.K., Grossman, A.R. and Björkman, O. (1998) *Arabidopsis* mutants define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion. *Plant Cell* 10: 1121-1134.

Niyogi, K.K., Shih, C., Soon Chow, W., Pogson, B.J., Dellapenna, D. and Björkman, O. (2001) Photoprotection in a zeaxanthin- and lutein-deficient double mutant of *Arabidopsis*. *Photosynth. Res.* 67: 139-145.

Ohno, S. (1970) Evolution by gene duplication. p. 160. Springer-Verlag.

Partensky, F., Hoepffner, N., Li, W., Ulloa, O. and Vaulot, D. (1993) Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) Strains Isolated from the North Atlantic and the Mediterranean Sea. *Plant Physiol.* 101: 285-296.

Pogson, B., McDonald, K.A., Truong, M., Britton, G. and DellaPenna, D. (1996) *Arabidopsis* carotenoid mutants demonstrate that lutein is not essential for photosynthesis in higher plants. *Plant Cell* 8: 1627-1639.

Pogson, B.J., Niyogi, K.K., Björkman, O. and DellaPenna, D. (1998) Altered xanthophyll compositions adversely affect chlorophyll accumulation and nonphotochemical quenching in *Arabidopsis* mutants. *Proc. Natl Acad. Sci. U S A* 95: 13324-13329.

Quinlan, R.F., Jaradat, T.T. and Wurtzel, E.T. (2007) *Escherichia coli* as a platform for functional expression of plant P450 carotene hydroxylases. *Arch. Biochem. Biophys.* 458: 146-157.

Russo, C.A., Takezaki, N. and Nei, M. (1996) Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol.* 13: 525-536.

Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M. and DellaPenna, D. (2004) Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *Plant Cell* 16: 1419-1432.

Schubert, N., Garcı´a-Mendoza, E. and Pacheco-Ruiz, I. (2006) Carotenoid composition of marine red algae. *J Phycol.* 42: 1208-1216.

Schwartz, S.H., Tan, B.C., Gage, D.A., Zeevaart, J.A. and McCarty, D.R. (1997) Specific oxidative cleavage of carotenoids by VP14 of maize. *Science* 276: 1872-1874.

Six, C., Worden, A.Z., Rodriguez, F., Moreau, H. and Partensky, F. (2005) New insights into the nature and phylogeny of prasinophyte antenna proteins: *Ostreococcus tauri*, a case study. *Mol. Biol. Evol.* 22: 2217-2230.

Spiegel, M., Schiller, J. and Srinivasan, A. (2000) Schaum's outline of Probability and Statistics. p. 407. McGraw-Hill.

Stickforth, P., Steiger, S., Hess, W.R. and Sandmann, G. (2003) A novel type of lycopene ε-cyclase in the marine cyanobacterium *Prochlorococcus marinus* MED4. *Arch. Microbiol.* 179: 409-415.

Sun, Z., Gantt, E. and Cunningham, F.X., Jr. (1996) Cloning and functional analysis of the β-carotene hydroxylase of *Arabidopsis thaliana*. *J. Biol. Chem.* 271: 24349-24352.

Swofford, D.L. (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). version 4. Sinauer Associates, Sunderland, Massachusetts.

Thayer, S. and Björkman, O. (1990) Leaf xanthophyll content and composition in sun and shade determined by HPLC. *Photosynth. Res.* 23: 331-343.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24: 4876-4882.

Tian, L. and DellaPenna, D. (2001) Characterization of a second carotenoid β-hydroxylase gene from *Arabidopsis* and its relationship to the *LUT1* locus. *Plant Mol. Biol.* 47: 379-388.

Tian, L. and DellaPenna, D. (2004) Progress in understanding the origin and functions of carotenoid hydroxylases in plants. *Arch. Biochem. Biophys.* 430: 22-29.

Tian, L., DellaPenna, D. and Zeevaart, J.A.D. (2004a) Effect of hydroxylated carotenoid deficiency on ABA accumulation in *Arabidopsis*. *Physiologia Plantarum* 122: 314-320.

Tian, L., Magallanes-Lundback, M., Musetti, V. and DellaPenna, D. (2003) Functional analysis of β- and ε-ring carotenoid hydroxylases in *Arabidopsis*. *Plant Cell* 15: 1320-1332.

Tian, L., Musetti, V., Kim, J., Magallanes-Lundback, M. and DellaPenna, D. (2004b) The *Arabidopsis LUT1* locus encodes a member of the cytochrome p450 family that is required for carotenoid ε-ring hydroxylation activity. *Proc. Natl Acad. Sci. U S A* 101: 402-407.

Van Norman, J.M. and Sieburth, L.E. (2007) Dissecting the biosynthetic pathway for the *bypass1* root-derived signal. *Plant J.* 49: 619-628.

Vandenabeele, S., Vanderauwera, S., Vuylsteke, M., Rombauts, S., Langebartels, C., Seidlitz, H.K., et al. (2004) Catalase deficiency drastically affects gene expression induced by high light in *Arabidopsis thaliana*. *Plant J.* 39: 45-58.

Wagner, A. (2000) Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. *Proc. Natl Acad. Sci. U S A* 97: 6579-6584.

Werck-Reichhart, D., Bak, S. and Paquette, S.M. (2002) Cytochromes P450. In *Arabidopsis book*. Editted by Somerville, C. and Meyerowitz, E. pp. 1-28. American Society of Plant Biologists, Rockville, MD.

Whatley, J.M. and Whatley, F.R. (1987) When is a chromoplast? *New Phytologist* 106: 667-678.

Yoshii, Y. (2006) Diversity and evolution of photosynthetic antenna systems in green plants. *Phycol. Res.* 54: 220-229.

Table 1 Carotenoid composition (nmol·g^{-1} fresh weight) of green and albino progeny from the indicated genotypes grown on sterile sucrose containing medium.

		Green plants[a]	Albino plants[a]	Green plants[b]	Albino plants[b]
Xanthophylls	β-Xanthophylls[c]	1.00±0.10	n.d.	0.10±0.01	n.d.
	Lutein	0.61±0.02	n.d.	3.01±0.17	n.d.
	α-Car-OH	7.18±0.34	n.d.	0.43±0.00	n.d.
Carotenes	α-Carotene	0.23±0.02	0.72±0.19	3.84±0.67	0.53±0.21
	β-Carotene	6.84±0.28	0.17±0.03	1.69±0.24	0.17±0.03

[a] progeny from *b1b1b2b2c1c1A3a3* parent, albino plants are homozygous quadruple mutants, green plants contain at least one *CYP97A3* allele.

[b] progeny from *b1b1b2b2C1c1a3a3* parent, albino plants are quadruple mutants, green plants contain at least one *CYP97C1* allele.

[c] Sum of zeaxanthin, antheraxanthin, violaxanthin and neoxanthin.

NOTE.- n.d., below the HPLC detection limit (0.5 ng); α-Car-OH, monohydroxy α-carotene.

The values shown are the mean of at least two biological replicates analyzed in triplicate ± standard deviation.

Table 2. Leaf carotenoid composition (mmol pigment mol^{-1} chlorophyll) from soil grown plants of the indicated genotypes

	Col-0	Ws	*cyp97c1*	*cyp97a3*	*bch1 bch2*	*cyp97a3 cyp97c1*	*bch1 bch2 cyp97c1*	*bch1 cyp97a3*	*bch1 bch2 cyp97a3*
Violaxanthin	31±4.6 (12)	28±3.1 (12)	55±4.0 (23)	21±2.0 (9)	11±1.0 (4)	41±2.7 (17)	10±3.0 (4)	14±0.8 (7)	0.2±0.1 (0)
Antheraxanthin	2.5±0.6 (1)	1.8±0.4 (0.7)	24±3.0 (10)	2.3±1.0 (1)	1.1±0.3 (0)	22±2.5 (9)	8.6±0.6 (4)	4.3±1.6 (2)	0.6±0.3 (0)
Zeaxanthin	n.d.	n.d.	8.1±1.4 (3)	n.d.	n.d.	7.0±1.3 (3)	4.3±0.4 (2)	n.d.	n.d.
Neoxanthin	32±3.0 (12)	30±2.5 (13)	26±1.8 (11)	18±2.6 (8)	3.8±1.0 (2)	20±2.4 (8)	2.0±0.9 (1)	4.1±0.7 (2)	0.5±0.9 (0)
β-Xanthophylls[a]	66±8.2 (25)	60±6.0 (26)	113±10 (47)	41±5.6 (18)	16±2.3 (6)	90±8.9 (37)	25±4.9 (11)	22±3.1 (11)	1.3±1.3 (0)
Lutein	133±10 (52)	122±8.9 (51)	4.0±0.4 (2)	104±6.7 (46)	156±8.0 (65)	3.2±0.4 (1)	7.2±0.7 (3)	94±6.2 (43)	99±14 (48)
α-Car-OH	0.8±0.3 (0)	0.6±0.1 (0.2)	61±5.8 (26)	5.4±0.6 (2)	1.0±0.3 (0)	53±5.2 (22)	101±3.8 (43)	8.8±0.7 (4)	10±5.9 (5)
α-Carotene	1.6±0.6 (1)	1.5±0.2 (0.6)	1.4±0.3 (1)	49±0.8 (22)	1.5±0.2 (1)	54±3.1 (23)	1.6±0.3 (1)	59±6.3 (27)	56±9.2 (27)
β-Carotene	56±3.6 (22)	54±3.7 (23)	57±4.7 (24)	26±3.1 (12)	67±2.5 (28)	38±3.5 (16)	99±4.7 (42)	33±5.4 (15)	39±6.9 (19)
Total carotenoids	257±21	238±16	236±17	227±10	241±11	237±18	234±7.6	217±16	205±20

[a] Sum of violaxanthin, antheraxanthin, zeaxanthin and neoxanthin.

NOTE- The values shown are the mean of at least three biological replicates analyzed in triplicate ± standard deviation. The numbers in parenthesis indicate percent of total carotenoids for each carotenoid in the indicated genotype. n.d., below the detection limit of HPLC (0.5 ng).

Fig. 1 Xanthophyll biosynthetic pathway in wild-type *Arabidopsis thaliana*. The genes encoding lycopene cyclases and carotenoid hydroxylases are italicized with the gene(s) encoding the major activity at each step bolded. The predominant carotenoids in unstressed WT leaf tissue are shown in black while those present at less than 1% of total carotenoids are shown in gray.

Fig. 2 Expression correlation coefficient (r) of *CYP97* (white) and *BCH* gene pairs (black) in the indicated tissues (A) and abiotic stress conditions in leaf (B). The significance of each correlation coefficient is indicated by ** and * where r is significantly different from zero at a *p* value of less than 0.01 or 0.05, respectively (Spiegel et al. 2000). The number of independent microarray datasets used in each condition or tissue is indicated in parenthesis.

Fig. 3 Non-photochemical quenching (NPQ) (A) and maximum photosynthetic efficiency of PSII (Fv/Fm) (B) in the indicated mutant genotypes and WT. The values shown are the averages and standard deviations from nine datapoints in three biological replicates. * $p < 0.05$, ** $p < 0.01$ by Student's t test of mutant leaves relative to corresponding WT leaves.

Fig. 4 Visual phenotypes of plants of the indicated genotypes grown under standard lab conditions (100 µmol m^{-2} sec^{-1}, 12h/12h light-dark cycle, 22 °C / 18 °C) and after 2 days of high light treatment (1800 µmol m^{-2} sec^{-1}, 16h/ 8h light-dark cycle, 22 °C).

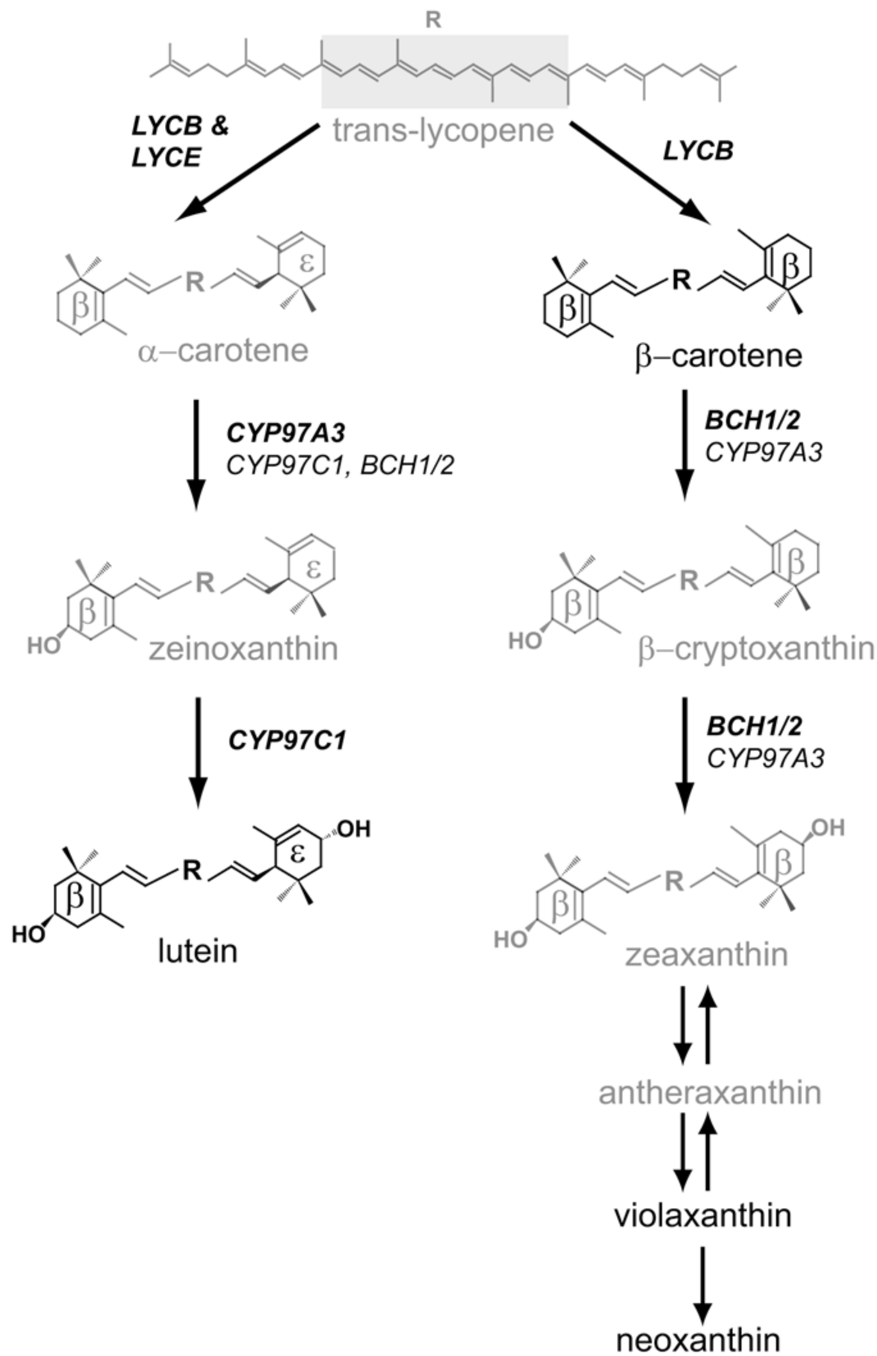
Fig. 5 Neighbor-joining tree of CYP97 amino acid sequences. Numbers on branches represent neighbor-joining, maximum parsimony and maximum likelihood bootstrap support, respectively (NJ/MP/ML); a hyphen indicates support of less than 50%. Bootstrap values are omitted from branches where support was less than 50% in all three analyses.

Fig. 6 Neighbor-joining tree of BCH amino acid sequences. BCH1, BCH2 indicates paralogs in an organism that have been demonstrated to have β-ring hydroxylase activity by assay in carotenoid containing *E. coli* lines. BCH homolog 1, BCH homolog 2, BCH homolog 3 indicate BCH paralogs in an organism whose functions have not yet been experimentally defined.

Fig. 7 Exon-intron structures of *BCH* genes (A) and colinearity of chromosomal segments containing parologous *BCH* genes (B) in Arabidopsis, poplar and rice. In (A), exons and introns are represented as boxes and lines in scale, respectively. In (B), syntenic gene pairs (*e*-value < 10^{-10} for amino acid identities) are represented as thick solid bars on the side of chromosomal regions with syntenic partners between regions matched with double-headed arrows. Split arrows indicate instances where apparent tandem duplications have occurred for a syntenic partner. Numbers on the sides of the chromosomes refer to *e*-value exponents between two syntenic proteins. Note that for clarity annotated genes without syntenic partners are omitted from each chromosomal region. The *BCH* genes in each chromosomal segment are marked with asterisks.

Fig. 8 Two different scenarios for evolution of the α-xanthophyll biosynthetic pathway in Arabidopsis from a hypothetical β-xanthophyll accumulating ancestral organism. The acquisition of newly evolved enzymatic activities in each scenario is indicated in bold.

Fig. 1



R
trans-lycopene
LYCB &
LYCE
LYCB
α–carotene
β–carotene
CYP97A3
CYP97C1, BCH1/2
BCH1/2
CYP97A3
zeinoxanthin
β–cryptoxanthin
CYP97C1
BCH1/2
CYP97A3
lutein
zeaxanthin
antheraxanthin
violaxanthin
neoxanthin

Fig. 2

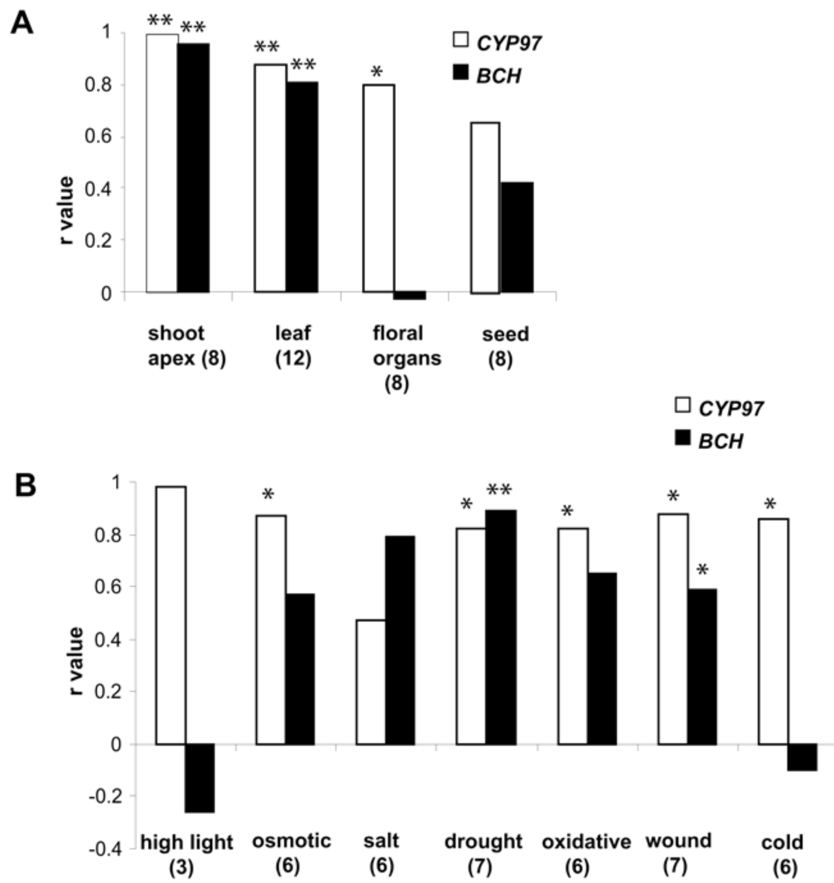
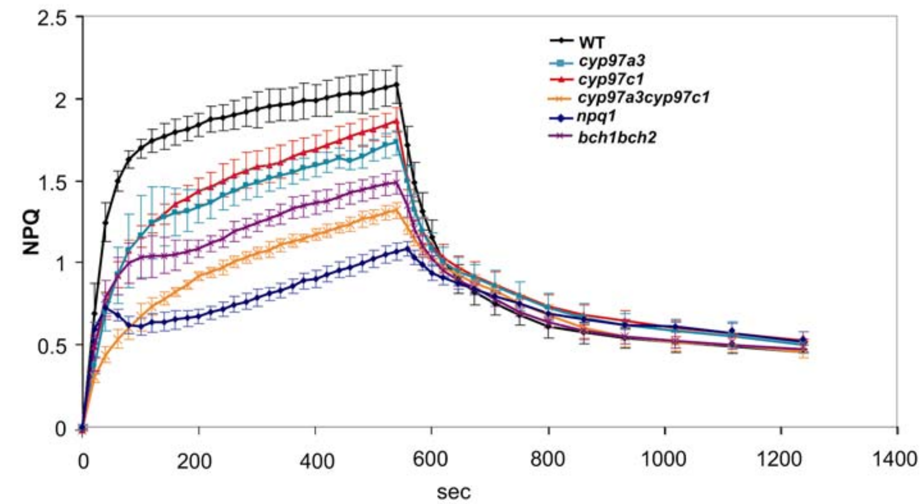

A
B
r value
CYP97
BCH
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
**
*
shoot apex (8)
leaf (12)
floral organs (8)
seed (8)
high light (3)
osmotic (6)
salt (6)
drought (7)
oxidative (6)
wound (7)
cold (6)

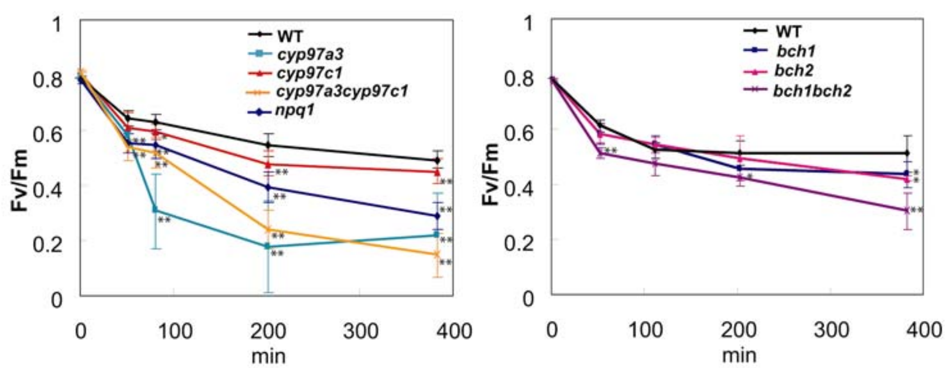
Fig. 3

A



WT
cyp97a3
cyp97c1
cyp97a3cyp97c1
npq1
bch1bch2
NPQ
sec

B



WT
cyp97a3
cyp97c1
cyp97a3cyp97c1
npq1
Fv/Fm
min
WT
bch1
bch2
bch1bch2
Fv/Fm
min

Fig. 4

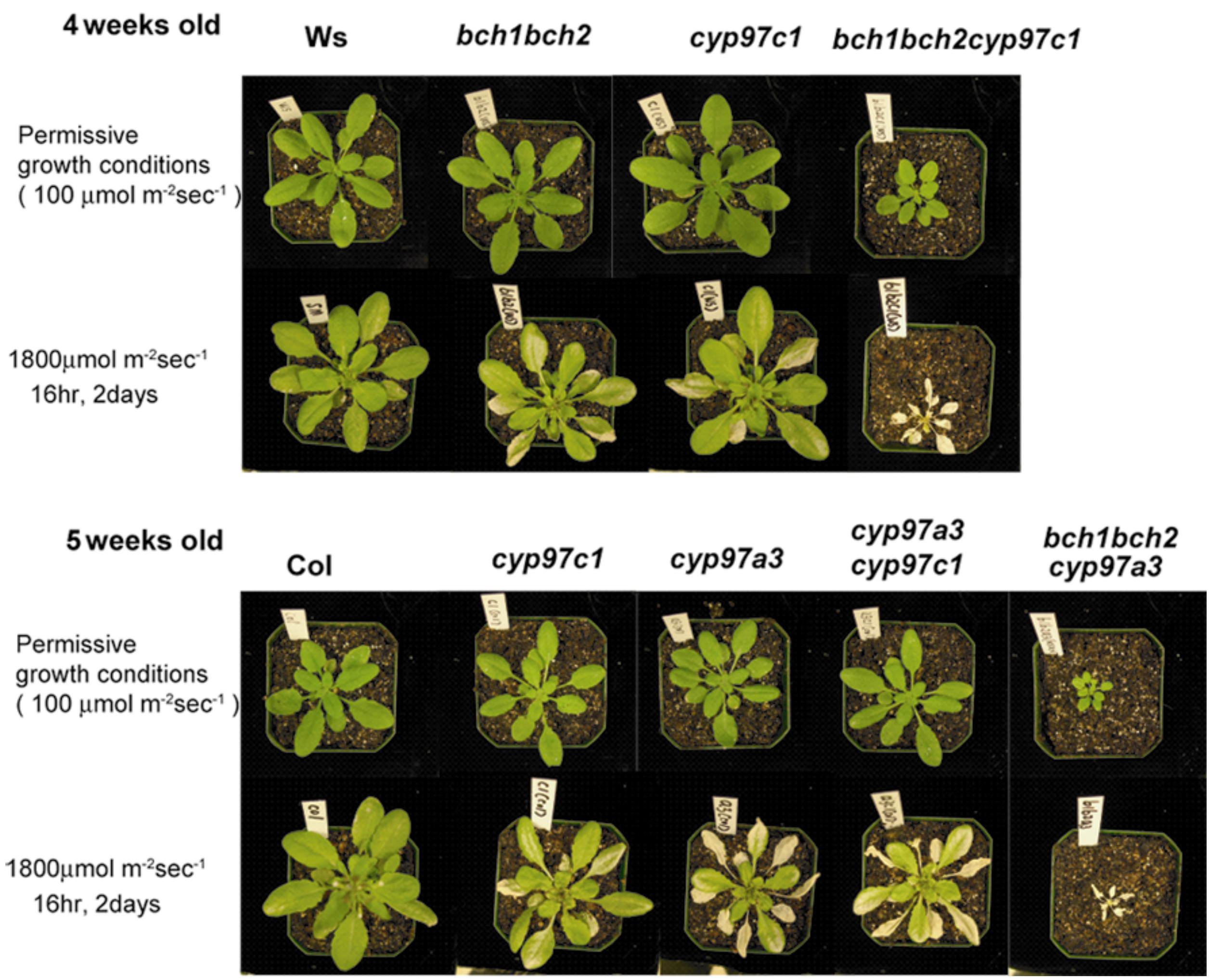

4 weeks old
Ws
bch1bch2
cyp97c1
bch1bch2cyp97c1
Permissive
growth conditions
(100 µmol m-2sec-1)
1800µmol m-2sec-1
16hr, 2days
5 weeks old
Col
cyp97c1
cyp97a3
cyp97a3
cyp97c1
bch1bch2
cyp97a3
Permissive
growth conditions
(100 µmol m-2sec-1)
1800µmol m-2sec-1
16hr, 2days

Fig. 5

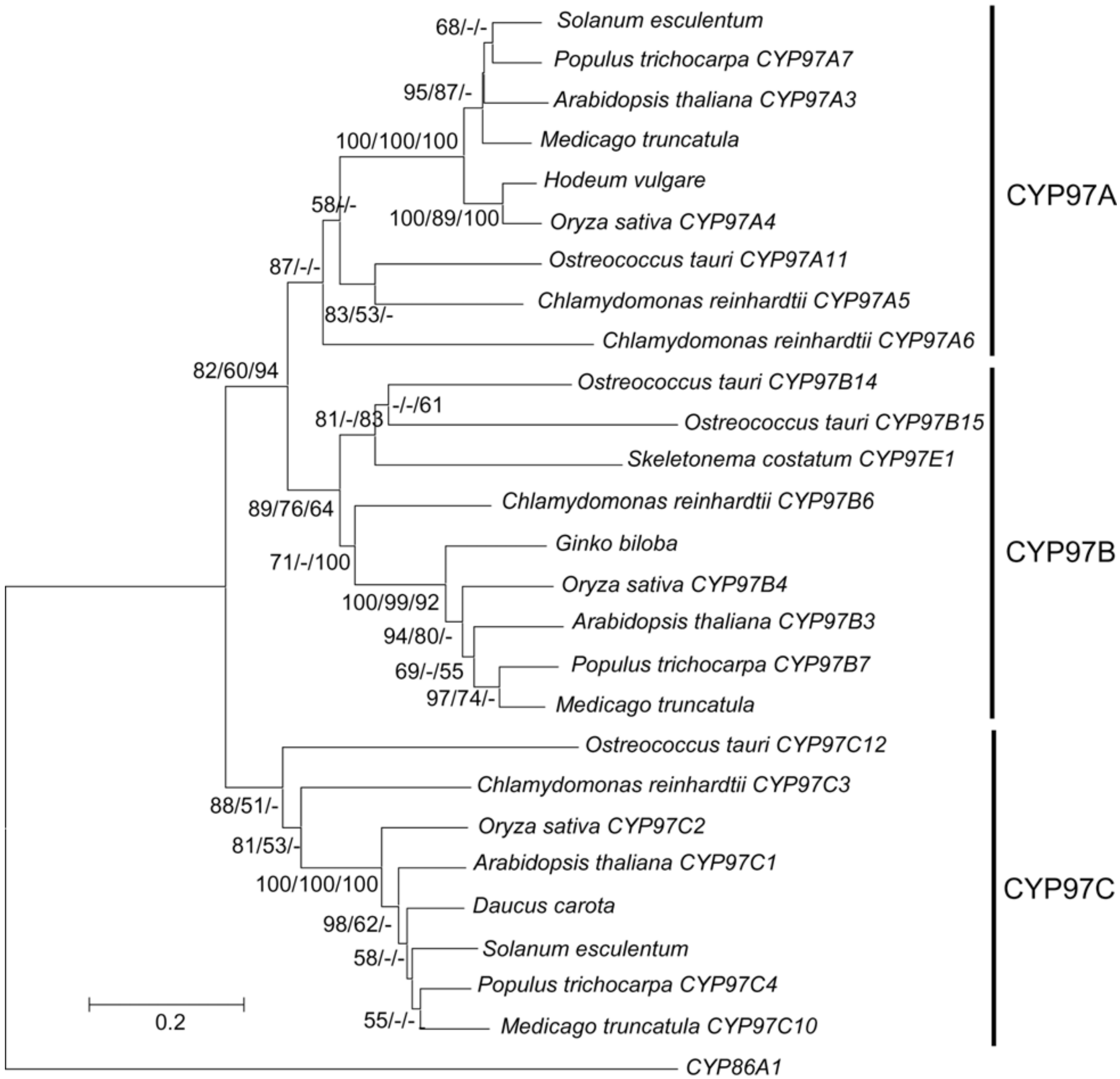
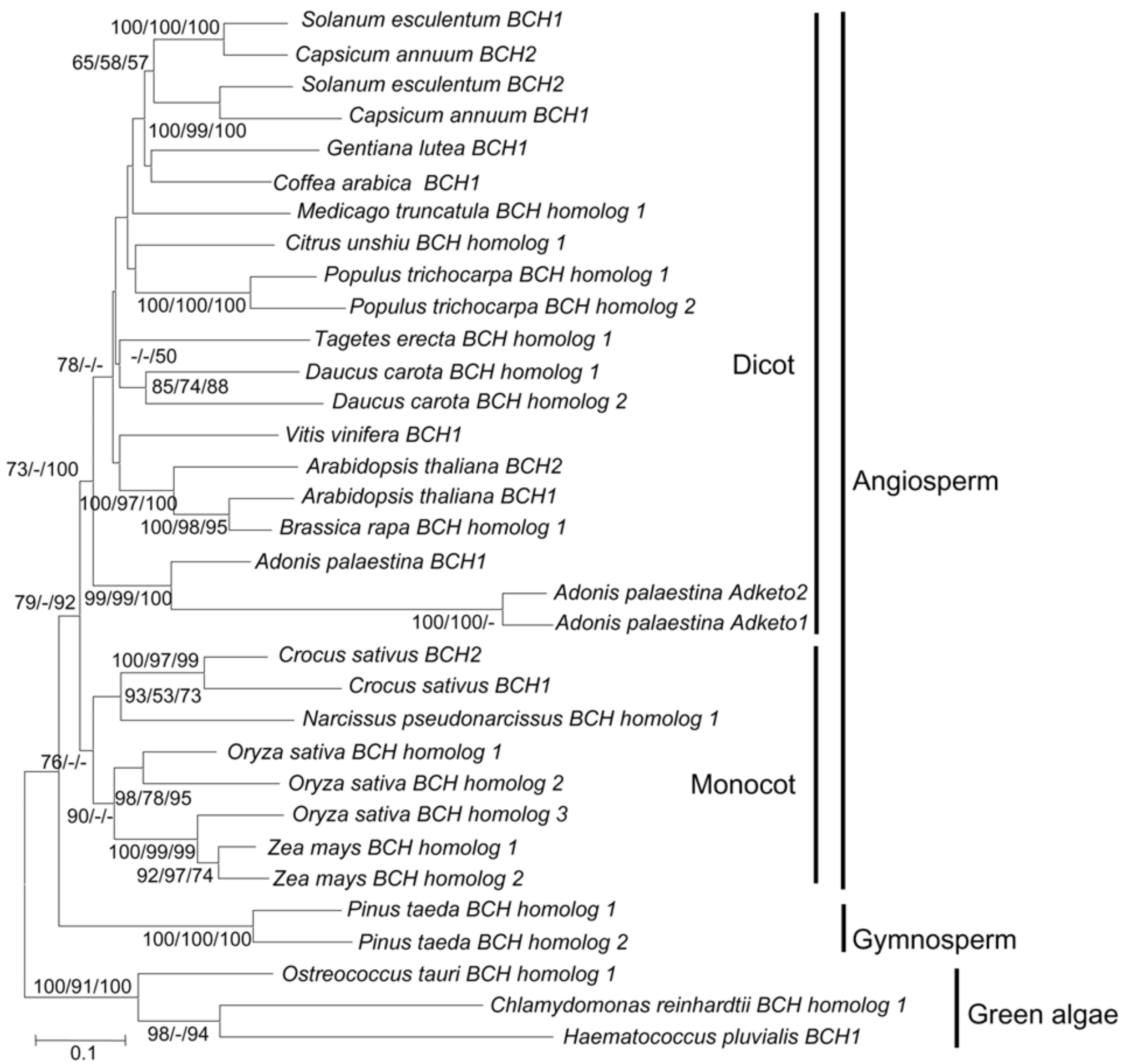
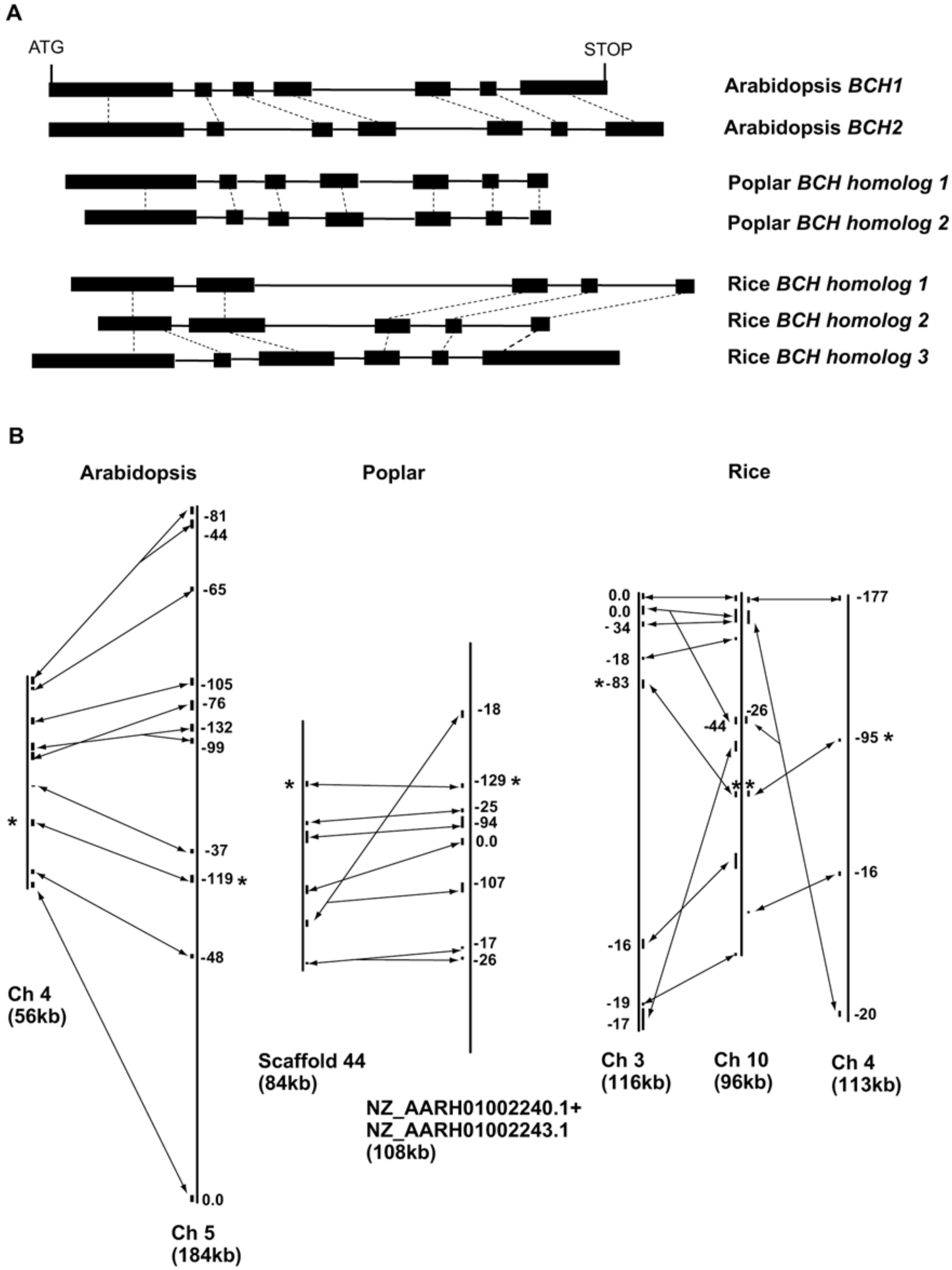

Solanum esculentum
Populus trichocarpa CYP97A7
Arabidopsis thaliana CYP97A3
Medicago truncatula
Hodeum vulgare
Oryza sativa CYP97A4
Ostreococcus tauri CYP97A11
Chlamydomonas reinhardtii CYP97A5
Chlamydomonas reinhardtii CYP97A6
Ostreococcus tauri CYP97B14
Ostreococcus tauri CYP97B15
Skeletonema costatum CYP97E1
Chlamydomonas reinhardtii CYP97B6
Ginko biloba
Oryza sativa CYP97B4
Arabidopsis thaliana CYP97B3
Populus trichocarpa CYP97B7
Medicago truncatula
Ostreococcus tauri CYP97C12
Chlamydomonas reinhardtii CYP97C3
Oryza sativa CYP97C2
Arabidopsis thaliana CYP97C1
Daucus carota
Solanum esculentum
Populus trichocarpa CYP97C4
Medicago truncatula CYP97C10
CYP86A1
CYP97A
CYP97B
CYP97C
68/-/-
95/87/-
100/100/100
58/-/-
100/89/100
87/-/-
83/53/-
82/60/94
-/-/61
81/-/83
89/76/64
71/-/100
100/99/92
94/80/-
69/-/55
97/74/-
88/51/-
81/53/-
100/100/100
98/62/-
58/-/-
55/-/-
0.2

Fig. 6



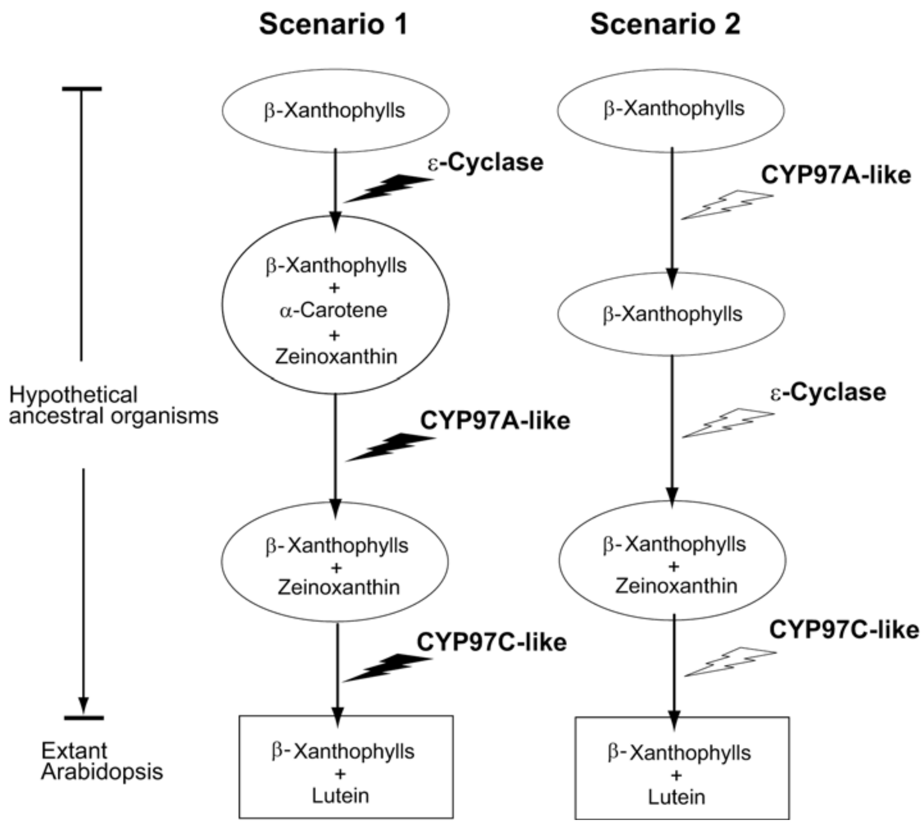
100/100/100
Solanum esculentum BCH1
65/58/57
Capsicum annuum BCH2
Solanum esculentum BCH2
Capsicum annuum BCH1
100/99/100
Gentiana lutea BCH1
Coffea arabica BCH1
Medicago truncatula BCH homolog 1
Citrus unshiu BCH homolog 1
Populus trichocarpa BCH homolog 1
100/100/100
Populus trichocarpa BCH homolog 2
Tagetes erecta BCH homolog 1
-/-/50
78/-/-
Daucus carota BCH homolog 1
85/74/88
Daucus carota BCH homolog 2
Vitis vinifera BCH1
73/-/100
Arabidopsis thaliana BCH2
100/97/100
Arabidopsis thaliana BCH1
100/98/95
Brassica rapa BCH homolog 1
Adonis palaestina BCH1
79/-/92
99/99/100
Adonis palaestina Adketo2
100/100/-
Adonis palaestina Adketo1
100/97/99
Crocus sativus BCH2
93/53/73
Crocus sativus BCH1
Narcissus pseudonarcissus BCH homolog 1
Oryza sativa BCH homolog 1
76/-/-
Oryza sativa BCH homolog 2
98/78/95
90/-/-
Oryza sativa BCH homolog 3
100/99/99
Zea mays BCH homolog 1
92/97/74
Zea mays BCH homolog 2
Pinus taeda BCH homolog 1
100/100/100
Pinus taeda BCH homolog 2
Ostreococcus tauri BCH homolog 1
100/91/100
Chlamydomonas reinhardtii BCH homolog 1
98/-/94
Haematococcus pluvialis BCH1
0.1
Dicot
Monocot
Angiosperm
Gymnosperm
Green algae

Fig. 7



A
ATG
STOP
Arabidopsis BCH1
Arabidopsis BCH2
Poplar BCH homolog 1
Poplar BCH homolog 2
Rice BCH homolog 1
Rice BCH homolog 2
Rice BCH homolog 3
B
Arabidopsis
Poplar
Rice
-81
-44
-65
-105
-76
-132
-99
-37
-119 *
-48
0.0
Ch 4 (56kb)
Ch 5 (184kb)
-18
-129 *
-25
-94
0.0
-107
-17
-26
Scaffold 44 (84kb)
NZ_AARH01002240.1+ NZ_AARH01002243.1 (108kb)
0.0
0.0
- 34
-18
*-83
-44
-26
-16
-19
-17
-177
-95 *
-16
-20
Ch 3 (116kb)
Ch 10 (96kb)
Ch 4 (113kb)

Fig. 8



Scenario 1
Scenario 2
Hypothetical
ancestral organisms
Extant
Arabidopsis
β-Xanthophylls
ε-Cyclase
β-Xanthophylls
+
α-Carotene
+
Zeinoxanthin
CYP97A-like
β-Xanthophylls
+
Zeinoxanthin
CYP97C-like
β-Xanthophylls
+
Lutein
β-Xanthophylls
CYP97A-like
β-Xanthophylls
ε-Cyclase
β-Xanthophylls
+
Zeinoxanthin
CYP97C-like
β-Xanthophylls
+
Lutein