This is to certify that the
dissertation entitled

FUNCTIONAL AND EVOLUTIONARY CHARACTERIZATION OF ARABIDOPSIS CAROTENOID HYDROXYLASES

presented by

JOONYUL KIM

has been accepted towards fulfillment
of the requirements for the

Ph. D degree in Biochemistry and Molecular Biology

Major Professor's Signature

March, 14, 2008

Date

FUNCTIONAL AND EVOLUTIONARY CHARACTERIZATION OF ARABIDOPSIS CAROTENOID HYDROXYLASES

By

Joonyul Kim

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Biochemistry and
Molecular Biology

2008

ABSTRACT

FUNCTIONAL AND EVOLUTIONARY CHARACTERIZATION OF ARABIDOPSIS CAROTENOID HYDROXYLASES

By

Joonyul Kim

Xanthophylls are a group of more than 500 different oxygenated carotenes that serve a variety of functions in procaryotes and eucaryotes. Xanthophyll composition is highly conserved in photosynthetic tissues of higher plants but how changes in xanthophyll composition occurred in ancestral photosynthetic organisms and why specific changes have been retained in lineages leading to higher plants remain open questions. To study on evolution of the xanthophyll biosynthetic pathway, I focused on the molecular evolution of four Arabidopsis carotenoid hydroxylases (CYP97A3, CYP97C1, CRTR-B1 and CRTR-B2) that catalyze key reactions in xanthophyll synthesis.

The protein encoded by *CYP97A3* was identified as the primary α-carotene β-ring hydroxylase (Chapter 2). Mutation of *CYP97A3* (*lut5* locus) caused accumulation of α-carotene but not α-cryptoxanthin and the major route for lutein synthesis to be determined. Lutein synthesis occurred by two successive β- and ε-ring hydroxylation from α-carotene give actual sequence of reaction. The susceptibility of a *lut5* null mutant

to photooxidation under high light stress is likely due to the massive accumulation of α-carotene in this mutant.

In Chapter 3, functional divergence of the four carotenoid hydroxylases (CYP97A3, CYP97C1, CRTR-B1 and CRTR-B2) were assessed based on their different *in planta* substrate specificities and gene expression profile. Generation of a quadruple mutant in which all four genes are inactive showed these carotenoid hydroxylases represent the full enzymatic complement in Arabidopsis. Phylogenetic analyses suggested that the *CYP97A3* and *CYP97C1* genes were duplicated before the speciation of Arabidopsis and green algae (cf. *Chlamydomonas reinhardtii* and *Ostreococcus tauri*) while duplication of the *CRTR-B* genes was more recent, after the Arabidopsis/*A. palaestina* split. Although the four enzymes exhibit some overlap in activities, most notably in hydroxylation of the β-ring of α-carotene, the mode of functional divergence in the two gene pairs appears to be distinct. *CYP97* duplicates are strongly coexpressed but the encoded proteins have distinct *in planta* substrates, likely due to divergence in their putative substrate recognition/binding regions. In contrast, the *CRTR-B* duplicates are isozymes that show significant expression divergence in reproductive organs.

ACKNOWLEDGMENTS

I am happy to have an opportunity to call over the names of the people one after another who helped me to continue my Ph. D. Although, only my name is shown on the cover of the dissertation, I would say that it is impossible to complete my Ph. D. without their supports.

I deeply thank to my Ph. D. advisor Dr. Dean DellaPenna for his supports and encouragements during my graduate study. I was lucky to have him who taught me the way of thinking which was the most critical to continue my career as a researcher. I also appreciate him to give me a lifelong word, "Hope for the best, prepare for the worst!" I would also thank to my committee members, Drs. James J. Smith, Kenneth Keegstra, Shelagh Ferguson-Miller and Gregory Zeikus for their advices both humanly and scientifically.

Some former and current members of the DellaPenna lab should be acknowledged. They are Tian Li, Laura Ullrich, Maria Magallanes, Hiroshi Maeda, Dana Chris, Naoko Kobayashi, and Wan Song. I also thank to Dr. Jongmin Nam in CALTECH, a lifelong friend and a collaborator to let me jump into a study on molecular evolution. And also many thanks go to several faculty members in the Biochemistry and Molecular

Biology (BMB) and Plant Biology departments for their advices and supports. They are Drs. Daniel Jones, Shin-Han Shiu, Kaillathe Padmanabhan, Michael Feig, and Robert Last. I also appreciate to my graduate fellows in the BMB department and Plant Research Labs for their supports and friendship, especially Hoo-Sun Chung and Colleen Doherty. Many Korean friends helped me to overcome homesickness. They are Namjoon Kim, Seunghoon Song, Sangheuk Park, Seunghwan Yang, Abe Koo, and Laee Choi.

Most importantly, I have to give a special mention for the support given by my family. Especially, I would like to express my gratitude to my wife, Sangeun for her love and patience.

멀리 고국에서 항상 응원하여 주신 부모님과 가족 친지분, 그리고 포항에 계신 안진흥 교수님께 감사의 마음을 전합니다.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABA	Abscisic acid
α-carotene	β,ε-carotene
α-xanthophylls	α-carotene derived xanthophylls
β-xanthophylls	β-carotene derived xanthophylls
β,β-carotenoids	β-carotene derived carotenoids
β,ε-carotenoids	α-carotene derived carotenoids
β-carotene	β,β-carotene
β-cyclase	lycopene β-ring cyclase
ε-cyclase	lycopene ε-ring cyclase
CCD	Carotenoid-cleavage dioxygenase
Chl *a*	Chlorophyll *a*
Chl *b*	Chlorophyll *b*
GGPP	Geranylgeranyl pyrophosphates
LHC	Light-harvesting complex
MRCA	Most recent common ancestor
Non-heme	Non-heme di-iron
P450	Cytochrome P450
PC	Photosystem core complex
Polyene	Poly-unsaturated organic compound
Xanthophyll	Oxidized carotene
WT	Wild type

CHAPTER 1

LITERATURE REVIEW

1.1. General property and biological function of carotenoids

Carotenoids are a group of more than 700 red, yellow and orange colored pigments and are some of the most widespread of all natural products (Straub, 1987; Kull and Pfander, 1995). These features are achieved by Nature's adoptation of two simple and economical strategies, making use of a universal pathway, the isoprenoid pathway, for carotenoic synthesis and elaborating various downstream modification steps in a lineage-specific fashion (Umeno et al., 2005). The carotenoid backbone is a hydrocarbon chain with conjugated double bonds, resulting from a series of condensations of five carbon isoprene building blocks. Most of carotenoids have a C_{40} backbone resulting from the condensation of two geranylgeranyl pyrophosphates (GGPP, C_{20}) and only a few eubacteria produce C_{30}-based carotenoids (Armstrong, 1997; Umeno et al., 2005). Chemical modifications leading to the diversity of carotenoids in nature include different types of ring cyclizations, oxygenations, glycosylations, prenylations and oxidative cleavages. This tremendous structural diversity that is widely distributed in nature has presumably evolved in relation to a number of independent and interdependent carotenoid biological functions (Vershinin, 1999). In this chapter I summarize the biological functions of carotenoids derived from their general mechanical and chemical properties.

1.1.1. Mechanic rigidity

1.1.1.1. Membrane fluidity The polyene structure of carotenoids provides an extremely rigid backbone in a lipid-soluble environment. This mechanic rigidity may be the basis of carotenoid functions in controlling membrane fluidity (Rohmer et al., 1979). The effects

of carotenoids on the physical properties of saturated phosphatidylcholine membranes were also studied with an electron paramagnetic resonance spin-labeling method (Wisniewska et al., 2006). These effects were monitored at the membrane center as a function of the amount of the carotenoid added to the sample and as a function of the critical temperature for fluid-phase transition. This study showd that carotenoids, especially the dipolar, terminally dihydroxylated carotenoid lutein decreased membrane fluidity and increased the hydrophobicity of the membrane interior.

1.1.2. The unique electron cloud of the carotenoid polyene chain

1.1.2.1. Light absorption Oxygenic photosynthesis is an ancient mechanism in cyanobacteria, red algae, green algae, and terrestrial plants that uses light and molecular oxygen 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. In the light-harvesting complexes (LHCs) and photosystem core complex (PCs) of photosynthetic organisms, carotenoids have slightly different absorption spectra from chlorophylls, the major light-harvesting pigments. This allows the photosynthetic apparatus to maximize light absorption in a gap of the chlorophyll absorption spectra (420~500 nm) thus increasing the active spectral range of photosynthesis. In nonphotosynthetic tissues of plants and animals, this feature provides coloration for sexual displays, pollinator attraction and seed dispersal. Many yellow and orange flowers and fruits of plants (Howitt and Pogson, 2006) and various red and orange coloration of male fish and birds are prominent examples (Olson and Owen, 1998).

1.1.2.2. Energy transfer between carotenoids and chlorophylls The unique

arrangement of π-electrons in the carotenoid polyene chain allows bidirectional excitation energy transfer between carotenoids and chlorophylls. This property is particularly important for photosynthesis and allows the photosynthetic apparatus to be optimized for light usage: (1) the captured light energy in LHCs is channeled to PCs by virtue of the slightly higher excited state (S1) of carotenoids compared to chlorophylls in the PCs (Cogdell and Gardiner, 1993; Liu et al., 2004) and, (2) carotenoids protect chlorophylls from excessive light energy by quenching the excited triplet of chlrophylls directly (Ma et al., 2003; Holt et al., 2005). Many photosynthetic eukaryotes developed these photoprotection mechanisms using carotenoids via xanthophyll cycles (Young and Frank, 1996).

1.1.3. Versatility in carotenoid oxidative cleavage reactions The long polyene chain of carotenoids can be cleaved enzymatically to produce a diverse range of oxidative cleavage products.These carotenoid-derived molecules, so called apocarotenoids, have been identified as bioactive molecules in plants, fungi and animals (Auldridge et al., 2006a).

1.1.3.1. Chromophores In animals, the chromophore of rhodopsins, retinal, is a cleavage product of β-carotene and provide the basis for visual reception. The light-dependent *cis-trans* isomerization of retinal induces a conformational changes in rhodopsin which is transmited to the nerve cell (Palczewski, 2006). Bixin in bixa seed and crocin in saffron stigma are the cleavage products of lycopene and zeaxanthin, respectively, which are extensively used as food and cosmetic additives for coloration (Bouvier et al., 2003; Castillo et al., 2005).

1.1.3.2. Aroma and flavors β-ionone, β-cyclocitral, damascenones, and theaspirone

contribute to the flavor and aroma of flowers and foods. For example, both β-ionone and damascenone are the key fragrant-contributing compounds in flowers (Demole et al., 1970). In fact, the sweet floral smells present in black tea, aged tobacco, grape, and many fruits are due to in large part to apocarotenoids (Auldridge et al., 2006a).

1.1.3.3. Hormones Oxidaized derivatives of retinal are involved in vertebrate morphogenesis, growth, cellular differentiation, and tissue homeostasis by receptor-mediated signal transduction pathways (Mark et al., 2006). In plants, abscisic acid (ABA) is the best studied carotenoid-derived phytohormone, having vital functions in plant adaptation to stressful environments by regulating stomatal aperture and the expression of stress responsive genes, and in plant development such as seed maturation, germination and seedling growth (Leung and Giraudat, 1998). In addition, lateral branching phenotypes shown in two Arabidopsis mutants (*max3* and *max4*) which are dirsupted in specific carotenoid-cleavage dioxygenases (CCDs), suggest a new class of carotenoid-derived mobile signaling molecules involving in controlling branching. Orthologous mutants in pea, petunia and rice exhibit increased branching and some can be restored to wild-type branching if the mutant is grafted to wild-type tissues (Auldridge et al., 2006a), indicating that this carotenoid derived signaling molecule is conserved through angiosperm.

1.1.3.4. Fungal growth regulators Arbuscular mycorrhizae represent the most widespread symbiosis on the earth and form in association with the roots of more than 80% of land plants. Arbuscular mycorrhizae (AM) fungi facilitate the uptake of phosphate, by plants, and in return obtain carbohydrates from their hosts. Massive accumulation of apocarotenoids including mycorradicin and cyclohexenone derivatives is

initiated during root colonization by AM fungi. For instance, strigolactone, previously isolated as a seed-germination stimulant for root parasitic weeds, acts as a chemical signal for arbuscular mycorrhizal fungi during presymbiotic stages. The significance of apocarotenoids in AM symbiosis has been shown by experiments with maize mutants deficient in carotenoid biosynthesis. The maize *y9* mutant in which a carotenoid isomerase is disrupted showed remarkable decrease in the mycorrhizal colonization rate and in exudation of strigolactones (Matusova et al., 2005; Akiyama, 2007).

1.2. Xanthophylls, structural and functional components of light-harvesting complexes (LHC) in photosynthetic eukaryotes

The more than 700 structurally distinct carotenoid compounds are classified into two subgroups; the carotenes, acyclic or cyclic hydrocarbons and the xanthophylls, oxygenated derivatives of carotenes. The introduction of oxygen functional groups into carotenes seems to be a very ancient event that has been under strong selective pressure during evolutionary time because photosynthetic eukaryotes produce xanthophylls but no organism has been identified that only synthesizes carotenes (Table 1). One important role of xanthophylls is their association with LHC apoproteins to form functional light-harvesting complexes. *In vitro* LHC reconstitution studies with various carotenes and xanthophylls revealed that xanthophyll oxygen as a hydroxyl or epoxi is critical for LHC structure and function (Ruban et al., 1999; Bassi and Caffarri, 2000; Phillip et al., 2002). For example, lutein, the predominant xanthophyll in higher plants has been shown not only to induce correct folding of the LHC II apoprotein (Paulsen et al., 1993) but also to be involved in stability of LHC II trimer (Lokstein et al., 2002). Interestingly, the

function of xanthophylls in LHC assembly and function is conserved in photosynthetic eucaryotes but significant structural diversity still exists and fulfills their function in different organisms. For example, red and green algae have a variety of xanthophylls in which the backbone is α-carotene or β-carotene (Table 1), but most brown algae such as chromophytes and dinophytes only produce β-carotene derived xanthophylls (Alberte and Andersen, 1986; Britton, 1998; Pascal et al., 1998). Although LHC apoproteins are one of the highly structurally conserved proteins between taxa with having a flexibility to functionally bind xanthophylls (Bassi and Caffarri, 2000; Grabowski et al., 2001), different binding affinities for each xanthophyll in each LHC binding site suggests that the diversity of xanthophyll is somehow related and perhaps correlated to the evolution of LHC apoproteins.

LHC II associated with three different xanthophylls (lutein, violaxanthin and neoxanthin) is the most abundant integral membrane protein in higher plant chloroplasts and exists as a trimer which binds half of the chlorophyll molecules in the plastids. Beside light-harvesting, LHC II has also been shown to function in the non-radiative dissipation of excess excitation energy under high light stress. Recently, the crystal structure of spinach LHC II was determined with high enough resolution (2.72 Å) to assign all pigments including chlorophylls and xanthophylls at the atomic level (Liu et al., 2004). In the three dimensional structure, the position of pigments provide a rationale for how the rate of singlet excitation energy transfer between xanthophylls and chlorophylls is correlated: two luteins are found to be in favorable orientations and distances to Chl *a* for efficient singlet energy transfer from lutein to Chl *a*. One neoxanthin is found to transfer its energy mostly towards Chl *b*. Therefore, lutein and neoxanthin may function

as effective accessory light-harvesting antennae, absorbing light in the blue-green spectral region as a complement to Chl *a*/*b* absorbing in the red region. The photoprotective role by the xanthophyll cycle could be also deduced by proposing an efficient non-photochemical energy –transfer pathway.

1.3. The xanthophyll biosynthetic pathway in higher plants

The C_5 building blocks for carotenoid synthesis are isopentenyl diphosphate (IPP) and its isomer, dimethylallyl diphosphate (DMAPP). In higher plants, two independent pathways to synthesize these molecules are located in separate intracellular compartments, the cytosol and the plastid. In the cytosol, IPP is derived from the mevalonic acid pathway that starts from the condensation of acetyl-CoA and is used for the biosynthesis of sterols, sesquiterpenes and triterpenoids (Qureshi and Porter, 1981) with few exceptions (Dudareva et al., 2005). In plastids, IPP is formed from pyruvate and glyceraldehyde 3-phosphate via the 1-deoxy-d-xylulose-5-phosphate (DOXP) pathway and utilized for the synthesis of carotenoids (Lichtenthaler, 1999; Rohmer, 1999). In nature, more than two third of carotenoids are xanthophylls (>500) but the initial steps of xanthophyll biosynthesis leading to β-carotene are believed to have been established in very ancient organisms based on the widespread distribution of these compounds in both eubacteria and eukaryotes.

In higher plants, xanthophyll composition is highly conserved, generally consisting of three predominant xanthophylls, lutein, violaxanthin and neoxanthin. How changes in xanthophyll composition occurred in ancestral photosynthetic organisms and why specific changes have been retained in lineages leading to higher plants remain open

questions. In practical terms, understanding synthesis of the major xanthophylls in plants is an essential component for increasing xanthophyll production or creating new xanthophyll molecules by modifying the pathway (Britton, 1998). The aim of the following section of this literature review is to provide an overview of the current state of knowledge about xanthophyll biosynthesis in higher plants. Fig. 1 is a schematic diagram to illustrate the synthesis of major xanthophylls in higher plants.

1.3.1. From colorless to colorful carotenoids All C_{40} backbone carotenoids are derived from phytoene, a colorless carotenoid. The formation of a phytoene is achieved from the condensation of two GGPP molecules catalyzed by phytoene synthase, a 40 kDa protein encoded by *PSY* in plants and *crtB* in bacteria. There is significant amino acid identity between PSY and crtB, suggesting the functional conservation of phytoene synthase was established in a very early ancestor. Phytoene synthase in pepper could be partially purified from a multiprotein complex containing other biosynthetic enzymes required for the preparation of GGPP, the substrate for phytoene synthase (Dogbo et al., 1988), indicating substrate channeling for phytoene synthesis in the multiprotein complex. Phytoene is oxidized to the red colored compound lycopene by the introduction of four symmetric double bonds in its polyene chain. Each double bond is introduced in the *cis* configuration and isomerized to the *trans* configuration by a pair of carotenoid isomerase (Isaacson et al., 2002; Park et al., 2002) to yield all *trans* lycopene, which is the prefered substrate for the next reaction step, lycopene cyclization. In higher plants, the two distinct reaction steps from phytoene leading to lycopene, desaturation and isomerazation are catalyzed by two desaturases encoded by *PDS* and *ZDS*, and two isomerases encoded by *Z-ISO* and *CRTISO*. Interestingly, PDS and ZDS are unrelated to the bacterial desaturase,

but CRTISO appear to arise from a progenitor bacterial desaturase (Giuliano et al., 2002; Isaacson et al., 2002; Park et al., 2002; Li et al., 2007).

1.3.2. From acyclic to cyclic carotenoids Cyclization of lycopene is a key step in generating carotenoid diversity as it marks a branch point to two major cyclic carotenoid groups: the β,β- and β,ε-carotenoids. β,β-carotenoids (β-carotene derived carotenoids) contain two identical β-rings formed by the symmetrical action of the β-ring cyclase (β-cyclase) whereas β,ε-carotenoids (α-carotene derived carotenoids) contain two different ring structures (β and ε) formed by the action of the β-cyclase and ε-cyclase. β-rings contain a double bond in conjugation with the polyene chain which results in a rigid ring structure with only one conformation. In contrast, the ε-ring double bond is not in conjugation and thus has relatively free rotation around the C6'–C7' carbon. Unlike the ubiquitous β-carotene derived carotenoids which occur in archaea, eubacteria and plants, α-carotene derived carotenoids are found exclusively in the green plant lineage (Adams et al., 1993; Schagerl and Pichler, 2000; Yoshii et al., 2004), red algae (Marquardt and Hanelt, 2004b), and one extant prochlorophyte (a cyanobacterium) (Partensky et al., 1993; Stickforth et al., 2003). Therefore, from an evolutionary perspective, ε-ring formation and modifications in the α-carotene derived branch should postdate those of β-rings in the β-carotene derived branch in evolutionary time and would be expected to have evolved only in this subgroup of ε-ring carotenoid containing organisms. Plant β- and ε-cyclases are the archetypal monomeric lycopene cyclases encoded by *LYCB* and *LYCE*, respectively. Both enzymes show high similarities in their amino acid sequence and it is very likely they originated from duplication of a common ancestral gene (Krubasik and Sandmann, 2000).

1.3.3. From nonoxygenated to oxygenated carotenoids The most common and varied carotenoids among the hundreds of carotenoids are those that have cyclic end groups containing at least one oxygen function. The most commonly encountered oxygen function is a hydroxyl group at C3 (Britton, 1998) because hydroxylation at this position is the initial step in converting carotenes to xanthophylls in eubacteria, fungi and plants. Hydroxylation of the rings of α-carotene and β-carotene is catalyzed by a class of carotenoid hydroxylases. Production of β-carotene derived xanthophylls (β-xanthophylls) require two β-ring hydroxylations while α-carotene derived xanthophylls (α-xanthophylls) requires one β- and one ε-ring hydroxylation.

Two successive ring hydroxylations of β-carotene and α-carotene give rise to the dihydroxyl carotenoids lutein and zeaxanthin, respectively. Lutein is the most abundant carotenoid in plants, accounting for over 50% of the carotenoids in Arabidopsis leaves while zeaxanthin is only accumulated transiently under stress conditions. The major β-xanthophylls are violaxanthin and neoxanthin resulting from further ring modifications of zeaxanthin such as epoxidation and formation of allene group. These account for approximately 30% of total carotenoids in Arabidopsis. Experiments with Arabidopsis carotenoid biosynthetic mutants that have altered xanthophyll compositions make clear this conserved composition of lutein, violaxanthin and neoxanthin is the most functionally adaptive xanthophyll combination (Lokstein et al., 2002; Tian et al., 2003; Tian et al., 2004a; Dall'Osto et al., 2006), however, how this specific composition was established and maintained over evolutionary time remains an open question.

1.3.4. Molecular characterization of carotenoid hydroxylases Many carotenoid hydroxylases have been identified across phyla and all of them showed strong β-ring

hydroxylase activity with the exception of two ε-ring hydroxylases identified in Arabidopsis and rice. With very few exceptions (Blasco et al., 2004; Alvarez et al., 2006), these β-ring hydroxylases have been invariably characterized as members of the non-heme di-iron (non-heme) type carotenoid hydroxylases. Non-heme β-ring hydroxylases are further categorized into at least three subgroups (plant and green algal, non-photosynthetic bacterial, cyanobacterial groups) based on their primary structures. The catalytic mechanism using iron coordinated by histidine residues is the same in all three groups (Tian and DellaPenna, 2004), suggesting the existence of the common ancestral enzyme before the eubacteria and eucaryote split.

Molecular characterizations of two ε-ring hydroxylases proved that cytochrome P450 (P450) type enzymes are another class of carotenoid hydroxylase in higher plants. Tian, *et.* al (2004) determined the LUT1 locus responsible for ε-ring hydroxylation by positional cloning. The encoded protein was a P450 type carotenoid hydroxylase (CYP97C1) and suggested that the mechanism of ε-ring hydroxylation evolved independently from that of non-heme type enzymes (Tian et al., 2004b). In rice, CYP97C2, an ortholog of Arabidopsis CYP97C1, has also shown to be an ε-ring hydroxylase (Quinlan et al., 2007).

Although three carotenoid hydroxylases had been identified in Arabidopsis (two non heme oxygenases and CYP97C1) it could not be concluded that these three genes represent the full complement of carotenoid hydroxylases in Arabidopsis. When a null LUT1 mutant allele (*lut1-3*) was introduced into a Arabidopsis mutant background also disrupted for the two non-heme enzymes, carotenoid hydroxylation activity was still present, suggesting that at least one additional unknown carotenoid hydroxylase activity

existed *in vivo* (Tian et al., 2004b). The isolation and characterization of this fourth carotenoid hydroxylase activity became the major focus of my thesis research.

The fact that two P450 types β-ring hydroxylases had been identified in eubacteria and fungi (Blasco et al., 2004; Alvarez et al., 2006) suggested that cytochrome P450 (P450) type β-ring hydroxylase might also exist in higher plants. Such a β-ring hydroxylase activity structurally unrelated to non-heme type enzyme had been suggested based on molecular genetic studies (Tian et al., 2003; Kim and DellaPenna, 2006): an Arabidopsis plant in which all non-heme type β-ring hydroxylases are disrupted (*crtr-b1crtr-b2* hearafter *b1b2*) still produced β-xanthophylls up to 20% of the wild type (WT) level.

1.4. Carotenoid hydroxylases as tools to understand evolution of the xanthophyll biosynthetic pathway

The functional diversity of xanthophylls produced by plants has presumably evolved in relation to evolution of the xanthophyll biosynthetic pathway. Xanthophylls accumulate in nearly all types of plastids, and are thus found in most plant organs and tissues where they perform several independent and interdependent functions. Xanthophylls are predominant in photosynthetic tissues, accounting for about 80% of total carotenoids in an Arabidopsis leaf. Studying the evolutionary history of carotenoid hydroxylases is one of the most appropriate approaches to understand how xanthophyll biosynthesis and diversity was established because the reaction catalyzed by this enzyme is the first and key step in xanthophyll biosynthesis (Fig. 1). Work to date, including my Ph. D work demonstrated that four carotenoid hydroxylase genes are present in

Arabidopsis: two P450-type (e.g. *CYP97A3* and *CYP97C1*) and two non-heme type (e.g. *CRTR-B1* and *CRTR-B2*) genes (Tian and DellaPenna, 2004; Kim and DellaPenna, 2006). The two carotenoid hydroxylases in each class have significant amino acid identity to each other, implying each gene family member occurred by gene duplication and subsequent functional divergence. In addition to Arabidopsis, the rice orthologs CYP97A4 and CYP97C2, have 49% amino acid identity (Quinlan et al., 2007) and two non-heme type enzymes in tomato (75%), pepper (71%) and saffron (77%) were also functionally demonstrated to be carotenoid hydroxylases (Bouvier et al., 1998; Castillo et al., 2005; Galpaz et al., 2006). With these experimental data, further molecular evolutionary analyses such as detailed gene phylogeny and inference of an ancestral enzyme may provide insight into how the function of each gene evolved. Specifically, comparison of substrate specificity and gene expression between homologous genes make it possible to delineate the mode of functional divergence from a common ancestral gene.

1.5. General characteristics of P450 and non-heme type carotenoid hydroxylase gene families

Gene families arise through a process of duplication of an ancestral gene followed by functional divergence (e.g. subfunctionalization or neofunctionalization) or pseudogenization (e.g. nonfunctionalization). Gene families are a basis for classifying proteins based on amino acid similarity, but their broader biological significance is less clear. Our growing knowledge of complete genome sequences and molecular genetics in several organisms now allow us to understand the evolutionary history of gene families

such as gene birth-and-death and functional divergence (Zhang, 2003; Taylor and Raes, 2004; Nei and Rooney, 2005).

Carotenoid hydroxylase genes can be used as a test case to attempt to reach some general conclusions about the evolution of multigene families because the four known carotenoid hydroxylases show functional conservation and divergence of their activities *in vivo* (Tian et al., 2003). Here, I describe the general molecular characteristics of P450 and non-heme type enzymes as a preface for understanding the molecular evolution of carotenoid hydroxylase genes which will be discussed in more detail in Chapter 3.

1.5.1. Heme-containing cytochrome P450 monooxygenase genes (P450) The cytochrome P450 monooxygenase (P450) gene family is the one of the largest superfamilies of divergent genes encoding ~1% of the gene complement in Arabidopsis and rice. (Werck-Reichhart et al., 2002; Nelson et al., 2004). Proteins encoded by P450 genes are heme-thiolate enzymes that catalyze monooxygenation of a variety of hydrophobic substrates. Catalysis is based on the activation of molecular oxygen with insertion of one of its atoms into the substrate and reduction of the other to form water.

$$\mathbf{RH + O_2 + NADPH{+}H^{+} \leftrightarrow ROH{+}H_2O{+}\ NADP^{+}}$$

(R= substrate, ROH= product)

The substrates of plant P450s are diverse, including the precursors of membrane sterols, structural polymers and bioactive secondary metabolites such as pigments, antioxidants and defense compounds. P450s are also involved in biosynthesis and catabolism of hormones and signaling molecules, thus contribute to the control of

hormone homeostasis. In addition to their physiological substrates, exogenous molecules such as pesticides and pollutants are usually detoxified in an organism by P450s (Werck-Reichhart et al., 2002).

Interestingly, known substrates are generally classified along with gene phylogeny of P450s, suggesting the possibility to use P450s as markers. The *CYP97* subfamily, which includes carotenoid hydroxylase genes, is a deep branch of the P450 gene phylogeny in Arabidopsis (Werck-Reichhart et al., 2002), indicating divergence of the *CYP97* clade is relatively ancient compared to other P450 clades. Although *CYP97* orthologs in other organisms such as nonvascular plants and green algae have not yet been functionally characterized, the function of *CYP97* genes are believed to be conserved through plants based on the ubiquitous distribution of their product, lutein. At minimum, the function of *CYP97* is very likely to be conserved between monocotyledon and dicotyledon because the rice orthologs, CYP97A4 and CYP97C2 showed the same enzymatic activities as Arabidopsis CYP97A3 and CYP97C1 (Quinlan et al., 2007).

1.5.2. Non-heme di-iron monooxygenases (non-heme) With few exceptional cases (Blasco et al., 2004; Alvarez et al., 2006), all β-carotene β-ring hydroxylases in eubacteria and plants were characterized as a non-heme di-iron monooxygenase (non-heme type), having conserved histidine signature histidine motifs originally identified in membrane fatty acid desaturases. These enzymes require iron, ferredoxin, and ferredoxin oxidoreductase for activity and all ten of the conserved iron-coordinating histidines are required for activity (Tian and DellaPenna, 2004). Arabidopsis, pepper, tomato and saffron contain more than one member of the non-heme CRTR-B family that generally have identical substrates (β-carotene) but differ in their gene expression patterns. In

Adonis aestivalis, two of three CRTR-B homologs have diverged in enzymatic activity and are ketolases that are preferentially expressed in flowers (Cunningham and Gantt, 2005).

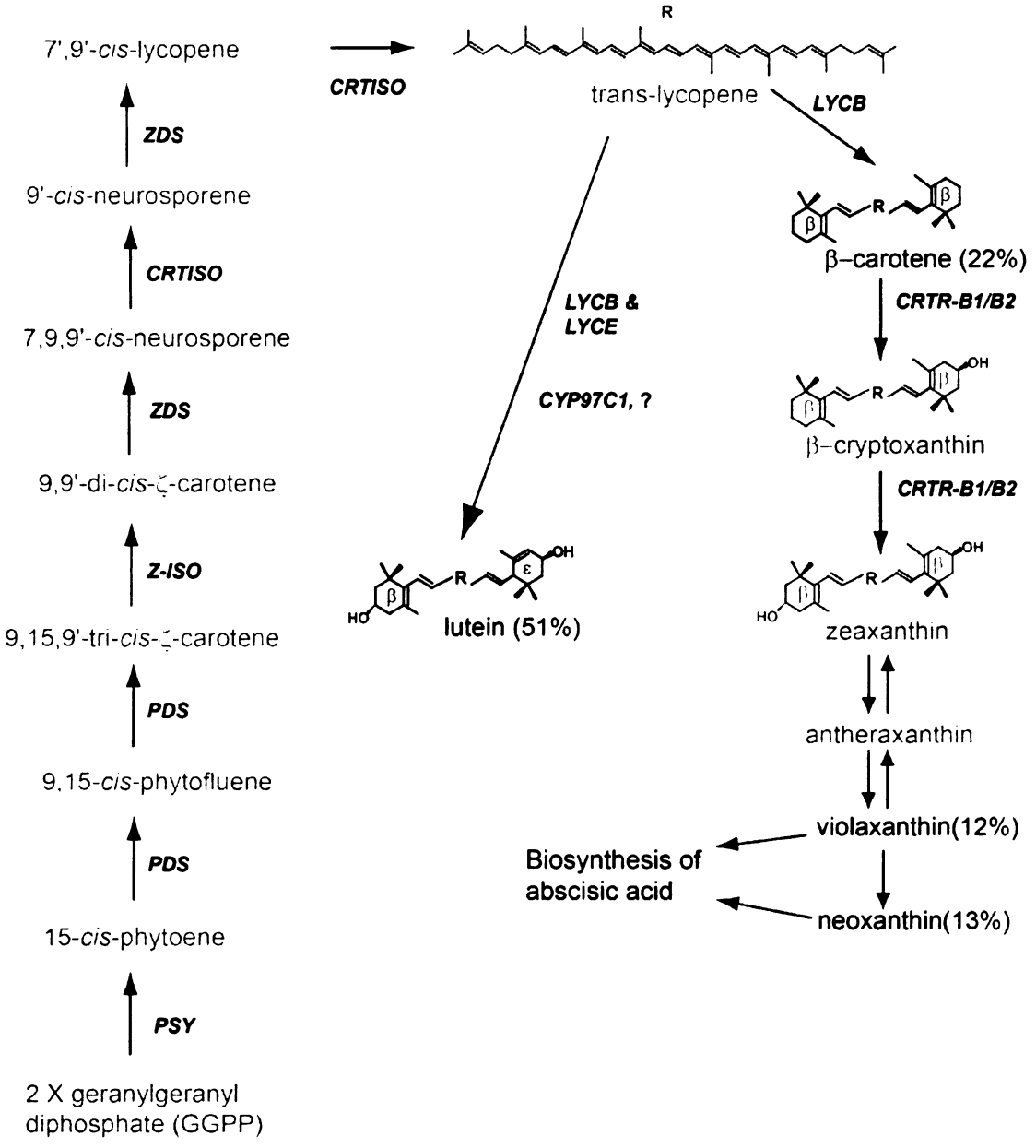
1.6. Goals of my research

Quantitative measurement to determine how molecular evolution of each carotenoid hydroxylase gene family member contributed to the evolution of xanthophyll biosynthetic pathway in plants is still a difficult task. My research aim was to develop a comprehensive understanding of the mode of functional divergence in each carotenoid hydroxylase gene pair and its biological impact on Arabidopsis. To reach this goal, determining the full complement of carotenoid hydroxylases in Arabidopsis was a prerequisite. Chapter 2 describes the identification of the primary α-carotene β-ring hydroxylase, which represents the last uncharacterized carotenoid hydroxylase in Arabidopsis. Chapter 2 also provided data for the primary route and metabolic channel for lutein biosynthesis to be proposed. In Chapter 3, the function and molecular evolution of each carotenoid hydroxylase gene in Arabidopsis is investigated. In Chapter 4, I summarize future work that could provide additional insight into the evolution of xanthophyll synthesis in photosynthetic eucaryotes.

Table 1 Pigment composition in the indicated lineages of photosynthetic eucaryotes. Lutein derivatives include prasinoxanthin, loroxanthin, siphonaxanthin derivatives and lutein 5,6-epoxide etc. The presence and absence of a pigment in the indicated lineage is indicated as Y and hyphen, respectively. Superscript numbers indicate to the reference describing pigment compositions of the corresponding lineages. 1, (Johnson and Schroeder, 1996; Miller et al., 2005); 2, (Partensky et al., 1993; Tomitani et al., 1999; Hess et al., 2001; Chen et al., 2005); 3, (Marquardt and Hanelt, 2004a; Schubert et al., 2006); 4, (Yoshii, 2006); 5, (Thayer and Björkman, 1990; Königer et al., 1995).

Pigment type			Cyanobacteria[1]	Prochlorophyte[2]	Red algae[3]	Green algae[4]	Land plants[5]
Chlorophylls	Chlorophyll *a*		Y	Y	Y	Y	Y
	Chlorophyll *b*		-	Y	-	Y	Y
	Chlorophyll *c*		-	Y	-	-	-
	Chlorophyll *d*		Y	Y	-	-	-
Carotenoids	α-carotene-derived	α-carotene	-	Y	Y	Y	Y
		lutein	-	-	Y	Y	Y
		lutein-derivatives	-	-	-	Y	Y
	β-carotene-derived	β-carotene	Y	Y	Y	Y	Y
		zeaxanthin	-	Y	Y	Y	Y
		violaxanthin	-	-	Y	Y	Y
		neoxanthin	-	-	-	Y	Y

Fig. 1 Xanthophyll synthesis in *Arabidopsis thaliana*. The genes are bold italicized. Carotenoids present at less than 1% of total carotenoids in unstressed WT leaf tissue are shown in gray.



CHAPTER 2

DEFINING THE PRIMARY ROUTE FOR LUTEIN SYNTHESIS IN PLANTS: THE ROLE OF ARABIDOPSIS CAROTENOID β-RING HYDROXYLASE CYP97A3[1]

[1] This chapter was published in " Kim, J. & DellaPenna, D. (2006) Proceedings of the National Academy of Sciences of the United States of America 103, 3474-3479".

2.1. Summary

Lutein, a dihydroxy derivative of α-carotene (β,ε-carotene), is the most abundant carotenoid in photosynthetic plant tissues where it plays important roles in LHCII structure and function. The synthesis of lutein from lycopene requires at least four distinct enzymatic reactions: β- and ε-ring cyclizations and hydroxylation of each ring at the C-3 position. Three carotenoid hydroxylases have already been identified in Arabidopsis, two non-heme di-iron β-ring monooxygenases (the *CRTR-B1* and *CRTR-B2* loci) that primarily catalyze hydroxylation of the β-ring of β,β-carotenoids and one heme-containing cytochrome P450 monooxygenase (CYP97C1, the *LUT1* locus) that catalyzes hydroxylation of the ε-ring of β,ε-carotenoids. In this study I demonstrate that Arabidopsis CYP97A3 (the *LUT5* locus) encodes a fourth carotenoid hydroxylase with a major *in vivo* activity toward the β-ring of α-carotene and a minor activity on β-rings of β-carotene. A *cyp97a3* null allele, *lut5-1*, causes an accumulation of α-carotene at a level equivalent to β-carotene in wild type, which is stably incorporated into photosystems and a 35% reduction in β-xanthophylls. That *lut5-1* still produces 80% of wild type lutein levels indicating at least one of the other carotene hydroxylases can partially compensate for the loss of CYP97A3 activity. From these data I propose a model for the preferred pathway for lutein synthesis in plants: ring cyclizations to form α-carotene, β-ring hydroxylation of α-carotene by CYP97A3 to produce zeinoxanthin, followed by ε-ring hydroxylation of zeinoxanthin by CYP97C1 to produce lutein.

2.2. Introduction

Numerous studies have shown that a class of non-heme type carotenoid hydroxylases that are present in most carotenoid containing organisms (Tian and DellaPenna, 2004) can efficiently hydroxylate the β-rings of several carotenoid substrates (Sun et al., 1996; Tian and DellaPenna, 2001). Arabidopsis contains two genes encoding non-heme type β-ring hydroxylases (the *CRTR-B1* and *CRTR-B2* loci) (Fig. 1). Their primary *in planta* substrates are β-rings of β-carotene, because a *crtr-b1crtr-b2* (*b1b2*) double null mutant reduced β-xanthophyll levels 80% but the level of lutein was slightly increased relative to wild type (Tian et al., 2003).

Recently, several P450 type enzymes have been identified as new classes of carotenoid hydroxylases. The archetypal members are the Arabidopsis ε-ring hydroxylase (CYP97C1) encoded by the *LUT1* locus (Tian et al., 2004b) and the β-carotene hydroxylase (CYP175A1) of the eubacteria, *Thermus thermophilus* (Blasco et al., 2004). Expression of *T. thermophilus* CYP175A1 in *E. coli* engineered to produce β-carotene resulted in the production of zeaxanthin, a dihydroxy β-carotene, and is a clear example of convergent evolution in β-ring hydroxylases (the non-heme and P450 type). In Arabidopsis, a T-DNA insertion mutant in the *CYP97C1* (*lut1-3*) gene accumulates zeinoxanthin (α-carotene with a hydroxylated β-ring) in place of lutein, consistent with CYP97C1 being the primary enzyme responsible for ε-ring hydroxylation in Arabidopsis (Tian et al., 2004b).

In the triple carotenoid hydroxylase mutant, *b1b2lut1-3*, β-xanthophylls are reduced 80% relative to WT but zeinoxanthin is still accumulated to high levels suggesting at least one additional carotene hydroxylase must exist in Arabidopsis with activity toward the β-rings of β- and α-carotene (Tian et al., 2003). Here, I report that a second member

of the Arabidopsis CYP97 family, CYP97A3, encodes a β-ring hydroxylase with a major activity toward the β-ring of α-carotene.

2.3. Results

2.3.1 The *lut5* mutation is in Arabidopsis CYP97A3 and causes accumulation of high levels of α-carotene My initial search for a fourth Arabidopsis carotenoid hydroxylase focused on members of the CYP97 clade because it already contains one carotenoid hydroxylase (CYP97C1, the *LUT1* locus). Although all three Arabidopsis CYP97 clade members are predicted to be targeted to the chloroplast by ChloroP 1.1 (Emanuelsson et al., 1999), CYP97A3 (At1g31800) was selected as it has the highest (52%) amino acid identity with CYP97C1. To determine whether loss of CYP97A3 activity affected the carotenoid composition in Arabidopsis leaf tissue, five-week old leaves of two independent *cyp97a3* mutant alleles were analyzed. *lut5-1* contains a T-DNA insertion in the third exon of *CYP97A3* while *lut5-2* contains a single amino acid change (E283K). Fig. 3 shows the HPLC profiles of leaf extracts from WT, *b1b2*, *lut1-4* (a *cyp97c1* null mutant), and *lut5-1*. All lines are in the Col-0 background except *b1b2lut1-3*, which are Wassilewskija (Ws). Individual and total carotenoid levels in Col-0 and Ws were not significantly different (data not shown) and only Col-0 is shown for WT. WT accumulates four major carotenoids: three xanthophylls (neoxanthin, violaxanthin, and lutein) and one carotene (β-carotene). Though the total carotenoid level is not significantly different between mutants and their respective WT, the carotenoid composition of each mutant genotype dramatically differs from WT (Table 2). As previously reported (Tian et al., 2003), the *b1b2* mutant has lower levels of β-

xanthophylls and increased β-carotene and lutein. *lut1-4* (a null *lut1* allele) virtually lacks lutein and accumulates a high level of zeinoxanthin, α-carotene with a hydroxylated β-ring, and elevated levels of β-xanthophylls (zeaxanthin, antheraxanthin, violaxanthin and neoxanthin) (Pogson et al., 1996; Tian et al., 2004b).

lut5-1 has a 18% reduction in lutein and contains two novel peaks (minor and major) relative to WT with retention times (24.3 and 32.7 min), mass and absorption spectra that are consistent with those of monohydroxy α-carotene derivatives (zeinoxanthin and/or α-cryptoxanthin, see Fig. 2) and α-carotene, respectively (Fig. 3). Zeinoxanthin and α-cryptoxanthin cannot be distinguished based on HPLC retention time or spectra. However, carotenoids that have an allylic hydroxyl group, for example at C-3′ of an ε-ring, readily eliminate water when positively ionized, while a non allylic hydroxyl group, such as at the C-3 of a β-ring, does not (Pogson et al., 1996). The major ion of the 24.3 min unknown was $[MH^+-H_2O]$ (see Supplementary data Fig. 1) and based on this comparatively easy loss of water I can conclude its single hydroxyl group is on the ε-ring of α-carotene and hence it is α-cryptoxanthin. Accumulation of high and low levels of α-carotene and α-cryptoxanthin, respectively, occurs in both *lut5* alleles, is accompanied by a reduction in total β,β-carotenoids and an increase in total α-carotenoids (Table 2). The *lut5-1* phenotype is generally more severe than *lut5-2*, consistent with *lut5-2* being a weaker allele. The *lut5* and *lut1* mutations are additive: the *lut5-1lut1-4* double mutant accumulates α-carotene and zeinoxanthin at levels nearly identical to *lut5-1* and *lut1-4*, respectively.

2.3.2 α-carotene is incorporated into *lut5-1* photosystems To determine the location of α-carotene in *lut5* I purified thylakoid membranes from WT, *lut1-4* and *lut5-1*, separated

the pigment:protein complexes by non-denaturing gel electrophoresis (Fig. 4) and analyzed their pigment compositions (Table 3). Fig. 4 shows that the photosystem I holocomplex and photosystem II core complex co-migrated and are well separated from LHC monomers and trimers in all genotypes (Lee and Thornber, 1995; Lokstein et al., 2002). The photosystems are the most likely place for α-carotene incorporation, as they contain β-carotene, a structural isomer of α-carotene (Alfonso et al., 1994; Lee and Thornber, 1995; Croce et al., 2002; Klimmek et al., 2005). Like β-carotene in WT, α-carotene in *lut5-1* was localized almost exclusively in photosystems, accounting for 39% of photosystem carotenoids (Table 3). In terms of carotenoid stoichiometry in *lut5-1* photosystems, the increase in α-carotene was mirrored by a corresponding decrease in β-carotene. Fig. 4 also shows *lut1-4* had a higher level of LHC monomers and lower level of LHC trimers, consistent with a previous report for this mutant showing a key role for lutein in LHC trimerization (Lokstein et al., 2002). *lut5-1* also shows an increase in LHC monomers, suggesting the 18% reduction in lutein in the mutant also impacts LHC trimer stability. The faint bands migrating between the photosystem and LHC trimer bands in *lut1* and *lut5* were variable in occurrence and levels between experiments and presumable reflect reduced photosystem stability in the mutants.

2.3.3 The impact of mutating CYP97-type carotenoid hydroxylase on the expression of other carotenoid biosynthetic enzymes As shown in Table 2, the ratio of total β,ε-carotenoids to β,β-carotenoids in *lut5-1* was over twice that of WT, suggesting that β- and/or ε-cyclase activities might be affected in the mutant. Because it is not possible to directly assay carotenoid cyclase or hydroxylase activities in Arabidopsis leaf extracts, I determined the steady-state transcript levels of these genes to indirectly assess the impact

of the *lut5* and *lut1* mutations on the pathway. As shown in Fig. 5, both cyclase mRNAs are modestly increased in *lut1-4* and *lut5-1* relative to WT, the β-cyclase more so than the ε-cyclase, but this increase appears unrelated to changes in the β,ε- to β,β-carotenoid ratios in the mutants. I also quantified mRNAs for the four known carotenoid hydroxylase genes (*CRTR-B1, CRTR-B2, CYP97C1*, and *CYP97A3*) to determine whether altered gene expression plays a role in compensating for the absence of P450-type carotenoid hydroxylases in the mutants. The detection of each CYP97 gene transcript in the corresponding gene knockout mutant was possible because the real-time probe is positioned upstream of each T-DNA insertion site. With the exception of a slight increase in *CYP97A3* mRNA levels, the expression of other carotene hydroxylases is not impacted in the *lut1-4* mutant. In contrast, *B1*, *B2* and *CYP97C1* mRNAs were all up-regulated in *lut5-1* and *lut5-1lut1-4*, most notably *CYP97C1* mRNA which was more than 4-fold higher than WT.

2.4. Discussion

2.4.1. At least four enzymes are involved in carotenoid hydroxylation The current study and prior work (Pogson et al., 1996; Sun et al., 1996; Tian and DellaPenna, 2001) have defined a minimum of four carotenoid hydroxylase genes involved in xanthophyll biosynthesis in Arabidopsis: two non-heme type β-ring hydroxylases encoded by the *CRTR-B1* and *CRTR-B2* genes, a P450-type ε-ring hydroxylase encoded by the *CYP97C1* (a *LUT1* locus) and now a P450-type β-ring hydroxylase encoded by the *CYP97A3* (a *LUT5* locus). Analyses of mutants defective in one or more carotenoid hydroxylase activities (Table 2) (Tian et al., 2003) provide insight into the specific and overlapping

activities of each enzyme *in vivo*. All mutant genotypes exhibit specific alterations in xanthophyll and carotene compositions (Table 2) (Tian et al., 2003), indicating the loss of any single activity cannot be fully compensated by the remaining activities. The degree of compensation observed in a given mutant genotype reflects the preferential activity of the missing enzyme(s) and the regulation and substrate specificities of the remaining active enzymes in the mutant. For example, CYP97C1 (LUT1) is the primary ε-ring hydroxylase activity in Arabidopsis as lutein synthesis is nearly completely blocked in the null *lut1-3* (Tian et al., 2004b) and *lut1-4* (Table 2) alleles and the presumed CYP97C1 monohydroxy substrate, zeinoxanthin, accumulates. Similarly, disruption of both non-heme type β-ring hydroxylases in the *b1b2* double null mutant reduces β-xanthophylls 76% without affecting β-ring hydroxylation in lutein synthesis, suggesting that the β-rings of β,β-carotenoids are the preferred *in planta* substrates for CRTR-B1 and CRTR-B2.

2.4.2. CYP97A3, the new type of β-ring hydroxylase in Arabidopsis CYP97A3 (LUT5) is a P450 type β-ring hydroxylase with activity toward the β-rings of both α-carotene and β-carotene *in vivo*. The major impact of the null *lut5-1* mutation is an accumulation of α-carotene at a level equivalent to β-carotene in WT, consistent with the β-ring of α-carotene being a preferred CYP97A3 substrate *in planta.* Because lutein is only reduced 18% in *lut5-1* relative to WT, at least one of the other carotene hydroxylases must also be able to catalyze hydroxylation of the β-ring of α-carotene, though less efficiently than CYP97A3. CRTR-B1 and CRTR-B2 may be able to hydroxylate the β-ring of α-carotene as in the *lut5-1lut1-4* double mutant zeinoxanthin is still present at levels similar to *lut1-4*. CYP97C1 (LUT1) may also have some level of β-ring

hydroxylase activity *in vivo* as *CYP97C1* expression is strongly up-regulated in the *lut5-1* background, as one might expect for a compensating enzyme (Fig. 5), and the level of total hydroxylated β-rings is further reduced in the *b1b2lut1-3* mutant relative to the *b1b2* mutant (Table 2) (Tian et al., 2004b).

CYP97A3 also appears to have a minor *in vivo* activity toward the β-rings of β-carotene as β-xanthophylls (primarily violaxanthin and neoxanthin) are reduced 35% in *lut5-1*, versus a 76% reduction in the *b1b2* genotype (Table 2). While CYP97A3 is likely to be responsible for synthesis of at least a portion of the β-xanthophylls present in *b1b2*, contributions from CYP97C1, an additional uncharacterized hydroxylase activity or indirect impacts resulting from the elevated levels of α-carotene produced in *lut5-1* can also not be excluded. Unfortunately, attempts to assay CYP97C1 and CYP97A3 *in vitro* by heterologous expression in *Saccharomyces cerevisiae* and *Escherichia coli* (*E. coli*) and *in vivo* in *E. coli* engineered to accumulate carotenoid substrates have not been successful (data not shown) and I therefore cannot directly assay potential substrates for the two enzymes. The generation of triple and quadruple mutant genotypes remains would be informative in this regard.

2.4.3. A model for lutein biosynthesis and regulation in plants As shown in Fig. 2, there are several potential biosynthetic routes leading from lycopene to lutein that have only partially been delineated by prior genetic studies in Arabidopsis (Cunningham et al., 1996; Pogson et al., 1996; Pogson and Rissler, 2000). The reactions leading to lutein must be highly efficient or tightly associated as many of the potential pathway intermediates in WT Arabidopsis leaf tissue are near or below the one ng carotenoid HPLC detection limit (Table 2). The activities of Arabidopsis β- and ε-cyclases expressed in lycopene

producing *E. coli* suggested the preferential substrate for the ε-cyclase is lycopene rather than γ-carotene (Cunningham et al., 1996). Consistent with this, mutation of the Arabidopsis LYCE, ε-cyclase encoded by *LUT2* locus eliminates lutein synthesis without causing accumulation of the pathway intermediates rubixanthin or γ-carotene (Pogson et al., 1996). These data indicate that ε-ring cyclization of lycopene to produce δ-carotene is the first step in lutein synthesis. δ-Carotene is undetectable in WT Arabidopsis leaf tissue but α-carotene is detectable suggesting the δ-carotene produced by LUT2 is efficiently utilized by the β-cyclase. The isolation and characterization of mutations in the *CYP97A3* and *CYP97C1* genes now allow us to infer the primary reaction sequence for lutein synthesis in plant tissues from among the remaining possible steps in Fig. 2.

The pathway shown in Fig. 2 has two possible routes leading to lutein from α-carotene: β-ring hydroxylation to zeinoxanthin followed by ε-ring hydroxylation to lutein or ε-ring hydroxylation to α-cryptoxanthin followed by β-ring hydroxylation to lutein. Whether one pathway is favored or both occur depends on the substrate specificities and regulation of the hydroxylases involved, CYP97A3 and CYP97C1. Accumulation of zeinoxanthin in *lut1* while the α-carotene level remains identical to that in WT is consistent with zeinoxanthin, rather than α-carotene, being a preferred substrate for ε-ring hydroxylation by LUT1. If α-carotene were a preferred CYP97C1 substrate the *lut5* mutants (which still has CYP97C1 activity) would be expected to accumulate significant amounts of α-cryptoxanthin, which it does not. *lut5* instead accumulates an α-carotene level ten times that of α-cryptoxanthin, which is consistent with α-carotene being a preferred substrate for β-ring hydroxylation by CYP97A3 and CYP97C1 having at best a

minor activity toward the ε-ring of α-carotene. Biochemical regulation (e.g. feedback inhibition by α-cryptoxanthin) may also contribute to the *lut5-1* phenotype but this seems less likely as the *lut5-1lut1-4* double mutant, that cannot synthesize α-cryptoxanthin, has a phenotype that is essentially the combination of the *lut1* and *lut5* single mutants. If α-cryptoxanthin played a regulatory role one would expect a different phenotype when the compound were removed in the double mutant. Together, the data presented are consistent with the preferred pathway for lutein synthesis being two ring cyclizations to yield α-carotene and then proceeding to zeinoxanthin and lutein by the sequential action of CYP97A3 and CYP97C1.

The general reaction sequence for lutein synthesis, ring cyclizations followed by hydroxylations, is the same as that of its closest structural isomer, zeaxanthin, a dihydroxy β-carotene derivative. However, zeaxanthin is produced primarily by the action of the non-heme type β-ring hydroxylases (CRTR-B1 and CRTR-B2) while lutein is produced primarily by the action of P450-mediated α-carotene ring hydroxylases and the regulatory mechanisms of these two types of enzymes appear to differ significantly. In lutein synthesis the bicyclic intermediate α-carotene is barely detectable in WT whereas the corresponding intermediate in zeaxanthin synthesis, β-carotene, is 22% of total WT leaf carotenoids. Accumulation of high levels of α-carotene and zeinoxanthin in the *lut5* and *lut1* mutants excludes the possibility these intermediates do not accumulate in WT simply because they are unstable *in vivo*. The high level of α-carotene in *lut5* is especially intriguing as it suggests the presence of CYP97A3 is specifically required for efficient lutein synthesis and that the two non-heme β-ring hydroxylases, *CRTR-B1* and *CRTR-B2*, cannot substitute for CYP97A3 in this regard.

Why would two members of the P450 type carotenoid hydroxylases be specifically required for efficient lutein biosynthesis *in vivo*? I propose that in WT the reaction sequence from lycopene to lutein is catalyzed by a protein complex composed of two lycopene cyclases (the β and ε-cyclases) and two P450 type hydroxylases (CYP97A3 and CYP97C1) that allows channeling of substrates between reactions. There is precedent for this as cytochrome P450 enzymes are known to form dimers (Scott et al., 2003; Schoch et al., 2004), functionally interact with other cytochrome P450 enzymes (Kaminsky and Guengerich, 1985; Dutton et al., 1987; Kelley et al., 2005) or act as anchors for soluble and membrane biosynthetic complexes (Burbulis and Winkel-Shirley, 1999) in other systems. Given the clear importance of lutein in LHC structure and photosystem function (Lokstein et al., 2002; Liu et al., 2004; Wentworth et al., 2004), it is relatively easy to rationalize a strong selective pressure for the evolution of ε-ring cyclization and hydroxylation activities for lutein synthesis. However, the forces driving the evolution of different biochemical machineries (P450 and non-heme carotenoid hydroxylases) for hydroxylation of α-carotene and β-carotene are less obvious. Perhaps the answer lies in the need to efficiently synthesize lutein for LHC structure and function while simultaneously tightly controlling α-carotene production, due to the potential negative consequences of producing α-carotene containing photosystems.

Plants produce β-carotene containing photosystems almost exclusively under most environmental conditions (e.g., high light). α-Carotene containing photosystems are also produced in many plant genera but generally only in shade-grown or low light adapted plants, where α-carotene can be present in excess of β-carotene (Thayer and Bjorkman, 1990; Demming-Adams and Adams III, 1992; Koniger et al., 1995).

Presumably α-carotene containing photosystems provide a competitive advantage under low light conditions but at higher light levels such photosystems show increased photooxidation, α-carotene levels decrease and β-carotene containing photosystems predominate (Barth et al., 2001). These data suggest tight control of the α/β-carotene ratio is an important adaptive response to low and high light (Krause et al., 2001; Krause et al., 2004). Consistent with this hypothesis, the constitutive production of α-carotene in *lut5* renders the mutant much more sensitive than WT to high light exposure (see Supplementary data Fig. 2). CYP97A3 clearly plays a key role in carotenoid synthesis by allowing efficient lutein production while limiting α-carotene accumulation and provides a straightforward biochemical mechanism for producing α-carotene and lutein when both are needed under low light conditions: by regulation of CYP97A3.

When taken together, the current genetic data in Arabidopsis are consistent with *in vivo* synthesis of the two major groups of xanthophylls in plants being preferentially catalyzed by two different classes of carotene hydroxylases: P450-type enzymes for synthesis of lutein and non-heme type enzymes for synthesis of β-xanthophylls. The evolution of what at first appearance seems to be an unnecessarily complex system can be understood in considering the contrasting needs of plants in response to changing light conditions. In high (normal) light it is competitively advantageous to produce lutein without α-carotene while under low light conditions it is advantageous to produce both lutein and α-carotene. β-Carotene and β-xanthophylls are needed at different levels and ratios under these conditions. Thus the P450 type carotenoid hydroxylases may be best suited to fulfilling these contrasting demands by allowing formation of a separate biosynthetic complex for efficient metabolic channeling to lutein while providing a

regulatory mechanism for α-carotene production that is independent of the synthesis and regulation of β-carotene and β-xanthophylls. Such independent regulation of the two branches of the carotenoid pathway allows plants to respond efficiently, effectively and adaptively to ever changing light conditions.

2.5. Material and Methods

2.5.1. Plant materials Plants were grown under a 12 h photoperiod (100-120 μmol m^{-2} s^{-1}, 22 °C) and 18 °C at night. *lut1-4* and *lut5-1* are T-DNA knockout mutant alleles of *CYP97C1* (At3g53130) and *CYP97A3* (At1g31800), respectively. *lut5-2* was identified by HPLC screening of nine missense mutant alleles isolated by TILLING screening (Till et al., 2003). The *lut5-1lut1-4* double mutant was selected by PCR screening of F_2 progeny from a cross of *lut5-1* and *lut1-4*. HPLC separation, identification and quantification by spectra and retention time were performed as previously described (Tian and DellaPenna, 2001) except quantification of monohydroxy α-carotenes was performed at 475 nm.

2.5.2. TaqMan Real-Time PCR Assays The transcript levels of six different carotenoid biosynthetic genes (*β-cyclase*, *ε-cyclase*, *CRTR-B1*, *CRTR-B2*, *CYP97C1* and *CYP97A3*) were quantified by TaqMan real-time PCR using elongation factor 1α mRNA levels for normalization. The *CYP97A3* primers and TaqMan probe are: 5'-GTTTGATTGGACTGGTTCTGACC-3' (forward primer), 5'-TTCCGGACCGCCTGAAT-3' (reverse primer), 5'-ACCCCAAGGTTCCTGAGGCTAAAGGCT-3' (TaqMan probe). Primers and probes for other genes are as described (Tian et al., 2003). The relative quantity of transcripts was

calculated using the comparative threshold cycle (C_T) method (Livak, 1997).

2.5.3. Isolation of thylakoid membranes and nondenaturing polyacrylamide gel electrophoresis (PAGE) Thylakoid membranes from Columbia-0 (Col-0), *lut1-4* and *lut5-1* were isolated, and the photosystems and peripheral LHCs were separated by non-denaturing PAGE as described (Lokstein et al., 2002). Individual pigment containing bands were excised and homogenized in gel running buffer (Lee and Thornber, 1995) and pigments extracted and analyzed as described (Tian and DellaPenna, 2001).

2.5.4. Structural determination of unknown monohydroxy α-carotene TLC separation on Si250F•PA silica plates (Mallinckrodt Baker, Phillipsburg, NJ) with hexane:isopropanol solvent (9:1) was used to enrich the unknown monohydroxy α-carotene from *lut5-1* saponified leaf extracts (Pogson et al., 1996). A band showing the same R_f value as that of zeinoxanthin was isolated and subjected to further mass analysis. The elutant from HPLC was chemically ionized by atmospheric pressure chemical ionization (APCI) and subsequently analyzed by MS. Lutein was used as a control to show water loss from a hydroxylated ε-ring when ionized, while zeaxanthin and zeinoxanthin were used to show no water loss from hydroxylated β-rings when ionized (Pogson et al., 1996).

Table 2 Leaf tissue carotenoid composition of the indicated genotypes. Carotenoids are expressed as mmol pigment mol^{-1} chlorophyll *a* + *b*, with the relative molar percentage of each carotenoid given in parentheses. Each value is the mean result of four experiments ± SD. Student's *t* test for two samples; *, $P<0.05$. α-crypto, α-cryptoxanthin; VAZ, the sum of violaxanthin, antheraxanthin, and zeaxanthin; neo, neoxanthin; Hβ, the moles of hydroxylated β-rings; β,ε/β,β, the molar ratio of total β,ε-carotenoids to total β,β-carotenoids. [a] indicates a monohydroxy α-carotene derivative that could not be identified because of the low levels present. n.d., not detectable.

	lutein	α-crypto	zeino-xanthin	α-carotene	β-carotene	VAZ	neo	tot.car	Hβ	β,ε/β,β
WT	127.0±9.1 (51)	0.7±0.3[a] (0.3)		1.4±0.5 (0.6)	55.4±4.0 (22)	32.1±5.1 (13)	32.5±3.8 (13)	249±21	257±26	1.1
b1b2	157.2±0.8* (65)	1.0±0.3[a] (0.4)		1.4±0.2 (0.6)	68.4±0.8* (28)	11.9±1.2* (5)	3.9±1.1* (1.6)	244±2	190±3*	1.9
lut1-4	0.6±0.1* (0.3)	n.d.	60.5±5.8* (26)	1.4±0.3 (0.6)	57.1±4.7 (25)	87.1±7.4* (37)	26.3±1.8* (11)	233±17	288±22	0.4
b1b2lut1-3	0.7±0.3* (0.3)	n.d.	100.6±4.4* (45)	1.5±0.2 (0.7)	97.4±3.3* (43)	22.4±3.6* (10)	1.9±0.9* (0.8)	225±4	150±5*	0.8
lut5-1	103.9±7.4* (46)	5.3±0.7* (2.3)	n.d.	49.6±0.7* (22)	27.2±2.8* (12)	23.2±3.2* (10)	18.7±2.7* (8)	228±10	193±8*	2.3
lut5-2	110.7±1.8* (49)	3.1±0.7* (1.3)	n.d.	34.0±4.8* (15)	32.5±4.6* (14)	22.4±1.8* (10)	24.7±1.4* (11)	228±4	208±2*	1.9
lut5-1lut1-4	0.4±0.1* (0.2)	n.d.	52.8±5.2* (23)	53.9±3.1* (23)	37.9±3.5* (16)	69.6±5.6* (30)	19.9±2.4* (9)	235±17	232±20	0.8

Table 3 Carotenoid composition in photosystems (PS I holocomplex and PS II core complex) of the indicated genotypes. The amount of carotenoid is expressed as mmol pigment mol^{-1} chlorophyll *a*. Each value is the mean result from three experiments ± SD, with the relative molar percentage of each carotenoid given in parentheses. Student's *t* test for two samples; *, $P<0.05$. n.d., not detectable.

	lutein	α-crypto	zeino-xanthin	α-carotene	β-carotene	VAZ	neo
WT	54.0±8.1 (28)	n.d.	n.d.	1.9±0.3 (1)	112.6±7.7 (59)	26.1±3.1 (12)	0.8±0.4 (0.6)
lut1-4	1.3±1.6* (1.7)	n.d.	14.6±0.9* (12)	1.3±0.5 (1)	86.1±2.9* (68)	21.6±1.2 (17)	0.5±1.1 (<0.5)
lut5-1	51.4±3.3 (26)	1.0±1.1 (0.5)	n.d.	77.3±5.8* (39)	47.7±1.7* (24)	17.9±1.2 (9.1)	0.6±1.1 (<0.5)

Fig. 2 Pathway showing all possible routes to xanthophyll synthesis in Arabidopsis. Enzymatic reactions are indicated by numbers: 1, ε-cyclization; 2, β-cyclization; 3, β-ring hydroxylation of β,ε- and β,ψ- carotenoids; 4, ε-ring hydroxylation; 5, β-ring hydroxylation of β,β-carotenoids. Reactions blocked by mutation of the indicated loci are shown: *cyp97c1* (*lut1,* ε-ring hydroxylase), *lyce* (*lut2*, ε-cyclase), *cyp97a3* (*lut5*, β-ring hydroxylase), *crtr-b1*, and *crtr-b2* (two β-ring hydroxylases). Solid arrows indicate a reaction sequence that is supported by mutant phenotypes and/or enzyme activity assays in E. *coli* while dashed arrows are not. Black arrows, compounds and mutant genes indicate major biosynthetic routes.

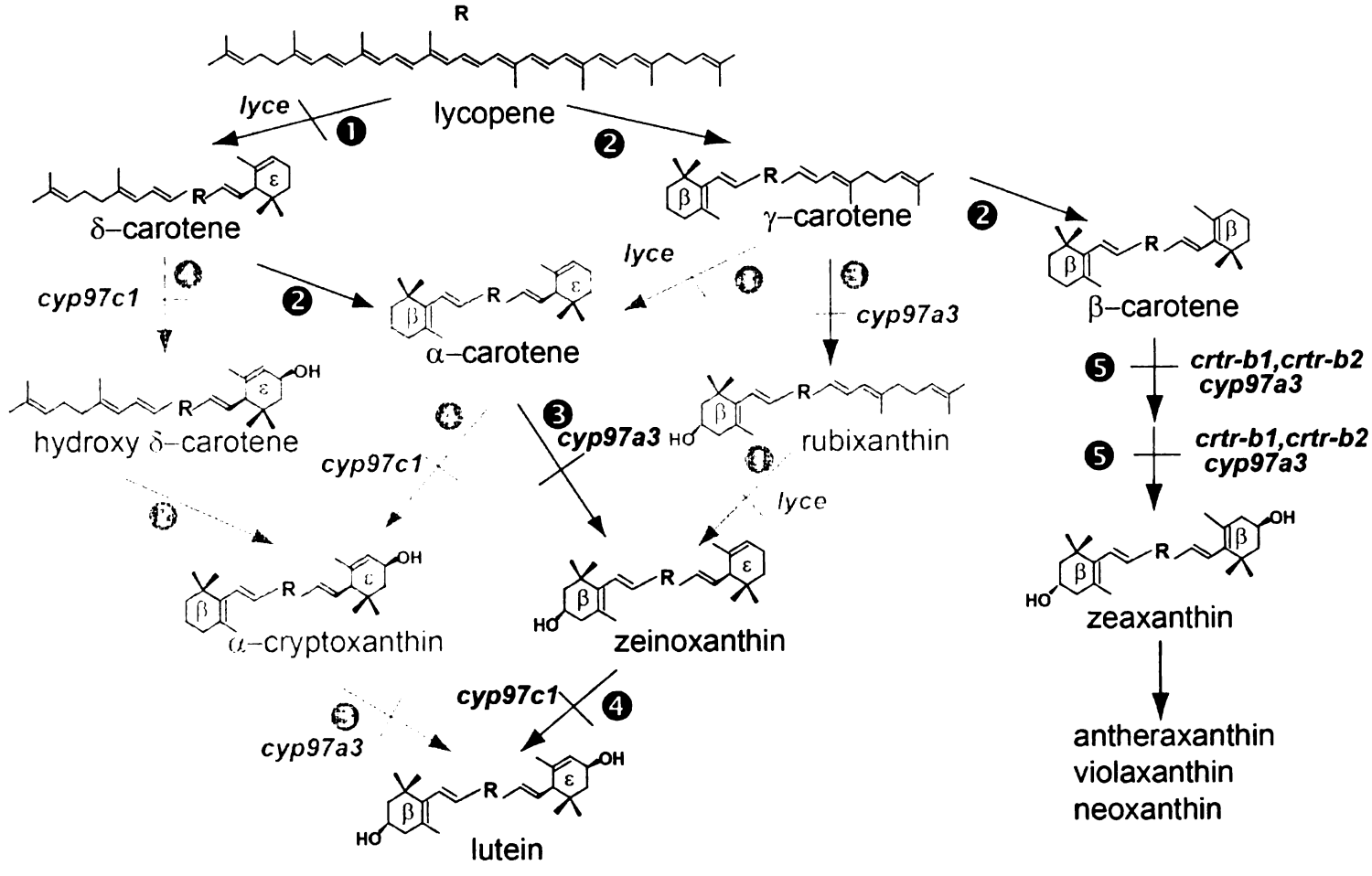
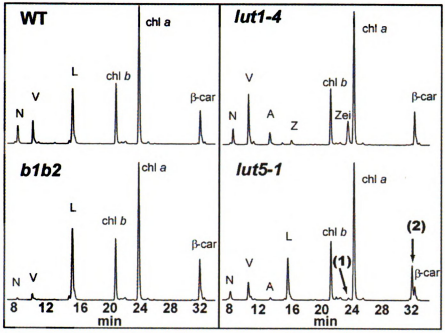

R
lycopene
lyce
δ–carotene
γ–carotene
β–carotene
cyp97c1
α–carotene
lyce
cyp97a3
crtr-b1,crtr-b2
cyp97a3
crtr-b1,crtr-b2
cyp97a3
hydroxy δ–carotene
cyp97a3
rubixanthin
cyp97c1
lyce
α–cryptoxanthin
zeinoxanthin
zeaxanthin
cyp97c1
cyp97a3
lutein
antheraxanthin
violaxanthin
neoxanthin

Fig. 3 Left panel: HPLC analyses (440 nm absorption) of leaf extracts of the indicated genotypes. N, neoxanthin; V, violaxanthin; A, antheraxanthin; L, lutein; Z, zeaxanthin; chl *a*, chlorophyll *a*; chl *b*, chlorophyll *b*; zei, zeinoxanthin; β-car, β-carotene. Right panel: an overlay of sections of the *lut5-1* (black) and *lut1-4* (gray) HPLC chromatograms containing unknown peaks 1 and 2 and UV-visible absorption spectra of zeinoxanthin, α-carotene and unknown peaks 1 and 2.


WT
chl a
L
chl b
N
V
β-car
lut1-4
chl a
V
chl b
N
A
Z
Zei
β-car
b1b2
chl a
L
chl b
β-car
N
V
lut5-1
chl a
L
chl b
(2)
V
(1)
N
A
β-car
8 12 16 20 24 28 32
min
8 12 16 20 24 28 32
min

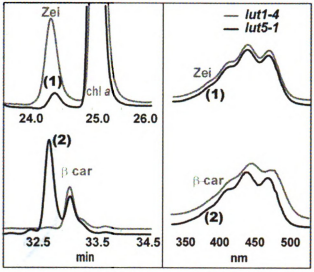

Zei
(1)
chl a
24.0 25.0 26.0
(2)
β-car
32.5 33.5 34.5
min
lut1-4
lut5-1
Zei
(1)
β-car
(2)
350 400 450 500
nm

Fig. 4 Non-denaturing gel electrophoretic separation of pigment:protein complexes from thylakoid membranes of the indicated genotypes. PS, photosystem I holocomplex and photosystem II core complex; LHCT, trimeric form of LHC; LHCM, monomeric form of LHC; FP, free pigment zone.

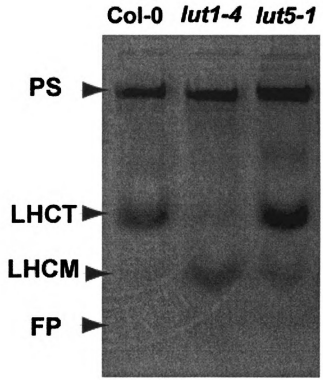
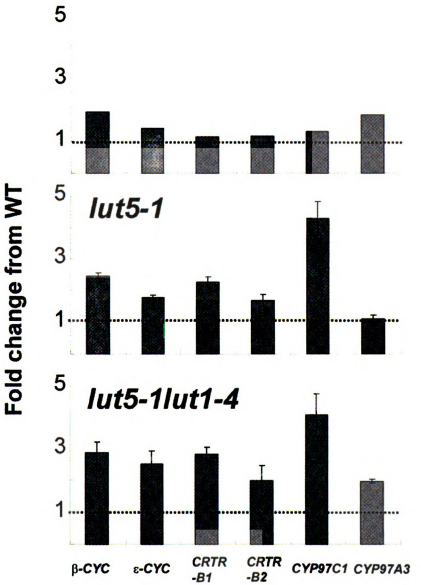


Fig. 5 Expression of carotenoid biosynthetic genes in *lut1-4*, *lut5-1* and *lut5-1lut1-4* relative to WT. The dotted line refers to the expression level of each gene in WT. Transcripts were quantified by real time PCR using elongation factor 1α as a reference control.



CHAPTER 3

MOLECULAR EVOLUTION OF ARABIDOPSIS CAROTENOID HYDROXYLASES[2]

[2] The part of this chapter was submitted to Plant Physiology.

3.1. Summary

Xanthophylls are a group of more than 500 different oxygenated carotenes that serve a variety of functions in prokaryotes and eucaryotes. Xanthophyll composition is highly conserved in photosynthetic tissues of higher plants but how changes in xanthophyll composition occurred in ancestral photosynthetic organisms and why specific changes have been retained in lineages leading to higher plants remain open questions. To gain insight into the evolution of xanthophyll synthesis, we analyzed two pairs of duplicated enzymes catalyzing key carotenoid hydroxylation steps in *Arabidopsis thaliana*. Recent work has suggested that α-carotene hydroxylation is catalyzed primarily by a pair of cytochrome P450 enzymes, CYP97A3 and CYP97C1, while β-carotene hydroxylation is catalyzed primarily by two non-heme di-iron enzymes, CRTR-B1 and CRTR-B2. We have used a series mutant genotypes null for one to four of these enzymes to demonstrate they represent the full complement of carotenoid hydroxylases in Arabidopsis and to infer the activity of each enzyme *in vivo*. Phylogenetic analyses suggested that the *CYP97A3* and *CYP97C1* genes were duplicated before the speciation of Arabidopsis and green algae (cf. *Chlamydomonas reinhardtii* and *Ostreococcus tauri*) while duplication of the *CRTR-B* genes was more recent, after the Arabidopsis/poplar split. Although the four enzymes exhibit some overlap in activities, most notably in hydroxylation of the β-ring of α-carotene, the mode of functional divergence in the *CYP97* and *CRTR-B* gene pairs appears to be distinct. *CYP97* duplicates are strongly coexpressed but the encoded enzymes have distinct *in vivo* substrates likely due to divergence in their putative substrate recognition/binding regions. In contrast, the CRTR-B duplicates are 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 mutant genotypes to high light stress we propose likely scenarios for the evolution of xanthophylls biosynthesis in Arabidopsis.

3.2. Introduction

Prior studies of xanthophyll biosynthetic mutants in Arabidopsis coupled with the genome sequence and associated wealth of gene expression data in this organism have advanced understanding of xanthophyll synthesis at the molecular level. Two classes of structurally unrelated enzymes catalyze ring hydroxylations of α- and β-carotene; P450 type carotenoid hydroxylase (CYP97A3 and CYP97C1) (Tian et al., 2004b; Kim and DellaPenna, 2006) and non-heme type carotenoid hydroxylases (CRTR-B1 and CRTR-B2) (Tian and DellaPenna, 2001; Tian et al., 2003). Mutation of one or more of these enzymes has shown that CYP97A3 and CYP97C1 are primarily responsible for catalyzing hydroxylation of the β-ring and ε-ring of α-carotene, respectively, for α-xanthophyll synthesis while CRTR-B1 and CRTR-B2 primarily catalyze the two β-ring hydroxylations of β-carotene in β-xanthophyll synthesis.

The fact that ring hydroxylations in each branch of xanthophyll synthesis are primarily catalyzed by one of two structurally unrelated enzyme groups (CYP97 and CRTR-B) *in vivo* raises interesting questions about the evolution of the two pathway branches and their associated enzymes. Additionally, that the two successive ring hydroxylations are catalyzed by homologous enzymes in each pathway branch suggests that duplication and subsequent functional diversification of each gene pair has been important for xanthophyll pathway evolution. Therefore, comparing and contrasting the

evolution of the *CYP97* and *CRTR-B* genes may provide important insight into the evolution of xanthophyll biosynthesis in higher plants.

3.3. Results

3.3.1. Defining the full complement of caroteonid hydroxylases in Arabidopsis To understand evolution of the xanthophyll synthetic pathway in photosynthetic organisms, I focused on the molecular evolution of carotenoid hydroxylases because the reactions catalyzed by these enzymes are key determinants for xanthophyll synthesis. I selected Arabidopsis enzymes as our model system because four carotenoid hydroxylase genes (*CYP97A3*, *CYP97C1*, *CRTR-B1* and *CRTR-B2*) had been previously identified and individually studied in detail (Sun et al., 1996; Tian and DellaPenna, 2001; Tian et al., 2004b; Kim and DellaPenna, 2006). A prerequisite for my study was to first determine whether these four genes represent the full complement of carotenoid hydroxylases or whether additional hydroxylase activities are present as had been previously suggested (Tian et al., 2003; Tian et al., 2004b; Kim and DellaPenna, 2006). To address this issue, I 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 lethal in soil but viable in tissue culture if supplied with a carbon source (cf. 1.5 % sucrose). HPLC analysis showed that white seedlings from both crosses contained trace amounts of α- and β-carotenes and lacked all xanthophylls (Table 4).

3.3.2 Gene duplication of the *CYP97* and *CRTR-B* genes Having established that CYP97A3, CYP97C1, CRTR-B1 and CRTR-B2 are the full complement of carotenoid hydroxylases in Arabidopsis, the molecular evolution of each gene was assessed. A catalogue of *CYP97* and *CRTR-B* 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* genes (*CYP97A3*, *CYP97B3* and *CYP97C1*) and two *CRTR-B* genes (*CRTR-B1* and *CRTR-B2*) of Arabidopsis as queries (Table 5 and 6). tBlastn searches (Altschul et al., 1997) against the full or nearly full genome sequences of *Populus trichocarpa* (poplar), *Oryza sativa* (rice) and two green algaes, *Chlamydomonas reinhardtii* (*C. reinhardtii*) and *Ostreococcus tauri* (*O. tauri*) identified the number of *CYP97* and *CRTR-B* homologs in each organism which allowed the pattern of gene duplication in each organism to be deduced.

All three *CYP97* gene trees (NJ, MP and ML) support 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 *CYP97* gene trees show one to one orthology among Arabidopsis, poplar and rice in each *CYP97A*, *CYP97B* and *CYP97C* clade, and two *CYP97A* and *CYP97B* genes in *C. reinhardtii* and *O. tauri*, respectively (Fig. 6A). In contrast, the *CRTR-B* gene tree (Fig. 6B) indicates that duplication of *CRTR-B* genes in higher plants occured in a lineage-specific fashion with three of four monocots and six of thirteen dicots in the analysis having more than one *CRTR-B* gene. Interestingly, these gene

duplications appear to have occurred relatively recently, after the monocot/dicot and stem eudicot/core eudicot splits, respectively. To assess the mechanisms of these recent duplications I examined the exon-intron structure of each *CRTR-B* gene and homology between pairs of chromosomal segments encoding *CRTR-B* genes in the fully sequenced genomes of Arabidopsis, poplar and rice (Fig. 7). In all cases, a significant degree of conservation in *CRTR-B* exon-intron structure and colinearity in adjacent chromosomal segments was observed, suggesting that duplication likely resulted from whole or segmental genome duplication. The *CRTR-B* duplication time infered from tree topology (Fig. 6B) and the Ks value (0.45) for Arabidopsis (See Supplementary data Table 1) suggest duplication likely occur 24~40 million years ago (Blanc et al., 2003; Wang et al., 2006).

3.3.3 Functional divergence of Arabidopsis CYP97 and CRTR-B enzyme pairs

3.3.3.1 Substrate divergence 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 combinations to unambiguously deduce the *in vivo* activity of the individual carotenoid hydroxylases. The leaf carotenoid compositions of seven such informative mutant genotypes in reference to wild type (Col-0 or *Ws*) are shown in Table 7. The phenotype of the *b1b2* mutant demonstrated that CYP97A3 and/or CYP97C1 could hydroxylate the β-rings of β-carotene, though at lower efficiency than the CRTR-B enzymes. Likewise, the *a3c1* double mutants (homozygous for the *cyp97a3cyp97c1* mutations) showed that the two CRTR-B isozymes could hydroxylate the β-ring of α-carotene, though again with lower efficiency than the CYP97 enzymes. The activities of the CYP97A3 and CYP97C1 enzymes were further clarified by two

triple gene knockouts: the *b1b2c1* (in which only CYP97A3 is functional) and *b1b2a3* mutants (in which only CYP97C1 is functional). β-xanthophylls are produced at a much higher level in *b1b2c1* than in *b1b2*, indicating the increased activity of CYP97A3 toward the β-rings of β-carotene in the absence of CYP97C1. In contrast, *b1b2a3* only contains 2% of the WT β-xanthophyll level, indicating that CYP97C1 has almost no *in vivo* activity toward the β-rings of β-carotene. However, lutein levels in *b1b2a3* are 74% of WT, indicating that in addition to ε-ring hydroxylation activity CYP97C1 also has strong activity toward β-ring of α-carotene.

The carotenoid composition in WT mature seeds differs significantly from that in leaf tissue (Tian et al., 2003) (Table 8). 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 the total seed carotenoids. Despite the differing carotenoid compositions of leaves and seeds, the activities inferred from carotenoid hydroxylase mutant xanthophyll composition in the two tissues are generally in agreement. For example, the β-xanthophyll levels in the *b1b2* double mutant (deficient in both *CRTR-B* genes) are reduced to 50% that of wild type (WT) but lutein is unaffected. Similarly, the *cyp97c1* mutation (*c1*) severely impacts lutein levels without affecting β-xanthophyll levels while in the *cyp97a3* mutant (*a3*) lutein is unaffected but β-xanthophyll levels decrease approximately 40%. In the *a3c1* double mutant lutein is still severely reduced but β-xanthophylls return to WT levels.

In an attempt to delineate any domains/residues in the CYP97A and CYP97C enzymes that may have contributed to their functional divergence at the protein level, I performed two different types of molecular evolutionary analyses. First, I scanned the

aligned amino acid sequences of each paralogous CYP97A/C pair in Arabidopsis, rice, *C. reinhardtii* and *O. tauri* using a statistical method that calculates the Z-score from the null hypothesis that the evolutionary rates are the same between two sequences in a sliding window (window size = 30 a.a.) (Nam et al., 2005). In this analysis, it is assumed that diversification of protein function is reflected by the relaxation or intensification of functional constraints at the protein level (Li, 1983). Because the activities of *CYP97A* and *CYP97C* are conserved between Arabidopsis and rice (Tian et al., 2004b; Kim and DellaPenna, 2006; Quinlan et al., 2007), one would expect that protein regions important for differentiating CYP97A and CYP97C substrates would occur consistently in Arabidopsis and rice. These regions would likely be shared as well in *C. reinhardtii* and *O. tauri* CYP97A/CYP97C pairs if functional divergence were initiated before the speciation of higher plants and green algae. I found only one region (from column 160 to 199) that showed a consistent evolutionary rate difference between CYP97A and CYP97C in all four organisms ($Z > 1.0$, $p < 0.16$). Fig. 8A shows the average of the four Z-scores of this region is 1.57 ($p = 0.06$). When the same analysis was applied to the three *CRTR-B* gene pairs from Arabidopsis, rice and pine (Fig. 8B), no consistent pattern was observed and the maximum Z-score was 0.94 ($p=0.17$) for any gene pair.

In a second approach, I estimated the posterior probability for cluster-specific functional divergence in a given residue (Gu, 2006) in order to identify those that might be critical for functional divergence of CYP97A and CYP97C. Fig. 8C illustrates the site-specific profile calculated from comparison between the *CYP97A* clade including Arabidopsis, *Medicago truncatula*, tomato, rice and barley genes, and the *CYP97C* clade including Arabidopsis, *Medicago truncatula*, tomato, carrot and rice genes. The estimated

coefficient of cluster-specific functional divergence (θ_{II}) is 0.29±0.04, indicating the analysis is statistically meaningful. Twenty-nine residues were identified with a 7.1 % prediction error (false-positive rate) and six of these were also conserved in *C. reinhardtii* and *O. tauri*.

Based on the structural conservation of the cytochrome P450 superfamily across phyla (Graham and Peterson, 2002; Schoch et al., 2003; Mestres, 2005), I developed two CYP97A3 structural models (i.e. closed and open conformation) and mapped the single domain and six residues identified from these two approaches onto the three-dimensional structures (Fig. 9). As templates for homology modeling, I used the crystal structures of human microsomal P450 3A4 (1tqnA) and mammalian cytochrome P450 2B4 (1po5A) which had the best 3D-jury scores (333.71 and 311.71) in the protein data bank (PDB) for closed and open conformations (Ginalski et al., 2003). The domain identified by sliding window analysis includes the β-strand that sits on the putative substrate access channel (blue in Fig. 9) (Graham and Peterson, 2002). Similarly, one of the six conserved amino acids identified from the second approach was mapped onto the middle of the I helix, (red in Fig. 9, alanine in CYP97A and serine in CYP97C) which has been shown to be important for positioning of substrate aromatic rings in the cytochrome P450 active site (Rupasinghe et al., 2003; Schoch et al., 2003). Identified domains and residues are also shown in amino acid alignment which is used for NJ tree construction (Supplementary data 3A).

3.3.3.2 Expression divergence To assess any divergence in gene expression between the *CYP97* and *CRTR-B* duplicates, I determined Pearson's correlation coefficient (r) as an index of expression similarity for each gene pair retrieved from the publicly available

DNA microarray datasets (Fig. 10) (Wagner, 2000). Only datasets in which at least one of the two copies for each family was expressed were selected for analysis (Makova and Li, 2003). Because carotenoid hydroxylases are primarily involved in producing pigments for LHCs, I compared gene expression in photosynthetic tissues (shoot apex, leaf and green seedling) and in reproductive organs (carpel, sepal, stamen, petal and seed). r-Values for the *CYP97* gene pair are high in nearly all tissues surveyed, indicating that expression of the genes are relatively highly correlated. r-Values for the *CRTR-B* gene pair are also high in unstressed and stressed photosynthetic tissues, with the exception of cold stress and high light stress, but much lower in all reproductive organs.

3.3.4 The impact of functional divergence under high light 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(s) on fitness of the organism. Because xanthophylls play many important structural and functional roles in photosynthesis we measured 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 (Maxwell and Johnson, 2000; Holt et al., 2004). 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 nonstressed tissues is 0.8 and approaches zero as PS II is progressively damaged.

Fig. 11A and 11B show changes in NPQ and Fv/Fm in the indicated genotypes as a function of time. The absence of either class of carotenoid hydroxylases (*a3c1* and *b1b2*) significantly reduced NPQ and Fv/Fm relative to WT level after high light stress, indicating both the CYP97 and CRTR-B type enzymes are required for high light adaptation. However, within an enzyme class, differing degrees of functional complementation were observed for individual class members. Impairment of NPQ and Fv/Fm in *b1b2* was partially complemented in the *b1* or *b2* single mutants, consistent with a prior study suggesting functional redundancy (Tian et al., 2003). In contrast, impairment of NPQ and Fv/Fm in the *a3c1* mutant is due largely to the *cyp97c1* (*c1*) and the *cyp97a3* (*a3*) mutations, respectively. Fv/Fm in the *c1* mutant was indistinguishable from WT during a 400 min high light treatment (Fig. 11B) but the kinetics of the NPQ rise and maximal NPQ level was slower and lower, respectively, than WT (Fig. 11A). The impact of the *a3* mutation on NPQ and Fv/Fm was opposite to that of the *c1* mutation. The kinetics of the rise in NPQ in *a3* was somewhat slower and more variable than WT but the maximal NPQ achieved after 260 sec was not significantly different (Fig. 11A). However, Fv/Fm in the *a3* mutant was strongly negatively impacted relative to WT after only 80 min of high light treatment (Fig. 11B). The *a3c1* double mutant had significantly slower NPQ induction kinetics, lower maximal NPQ and lower Fv/Fm and is essentially an additive phenotype of the two single mutations. After 10 hours illumination at 1600~1800 $\mu mol \cdot m^{-2} \cdot s^{-1}$, the *a3* mutant had irreversibly reduced CO_2 fixation rate under

100 $\mu mol \cdot m^{-2} \cdot s^{-1}$ (Fig. 12).

3.4. Discussion

3.4.1. Gene duplication and *in vivo* activity of carotenoid hydroxylase genes

Duplication and subsequent functional divergence of genes are being increasingly recognized as important mechanisms of evolution (Ohno, 1970; Lynch and Conery, 2000; Moore and Purugganan, 2005). Functional divergence can occur in protein coding regions, gene expression patterns or both. Here, I compared the evolutionary histories and *in vivo* function of two duplicate gene pairs involved in carotenoid hydroxylation in Arabidopsis, *CYP97A3/C1* and *CRTR-B1/B2* which together account for the full complement of carotenoid hydroxylases in this organism (Table 4). Our phylogenetic analyses (Fig. 6) showed that the *CYP97A3/C1* genes appear to be the result of an ancient duplication with no further recent duplication (Fig. 6A) whereas the *CRTR-B1/B2* gene pair results from a relatively recent gene dupication. The *in vitro* activities of individual hydroxylases against β- or ε-rings have been studied by heterologous expression in *E. coli* engineered to accumulate β-, δ- or ε-carotenes (Sun et al., 1996; Tian and DellaPenna, 2001; Quinlan et al., 2007). While this approach has been quite informative in delineating the possible *in vitro* substrate(s) for each enzyme, the approach has inherent limitations. The substrates tested may not occur *in vivo* and the enzymes are produced and assayed in isolation from the other pathway enzymes. Hence, the *in vivo* activity, regulation and role of each hydroxylase in this complex biochemical pathway may not accurately be reflected in the *E. coli* assay system. In order to better understand molecular evolution of the four extant carotenoid hydroxylases, whether they have distinct or overlapping enzymatic activities

in vivo and what the forces may have driven their evolution, selection and maintenance in plants we have generated a series of mutant genotypes for one or more of the four genes and assessed their consequences *in vivo*.

The two classes of carotenoid hydroxylases share significant overlap in substrate specificities, i.e. all four carotenoid hydroxylases have the ability to hydroxylate the β-ring of α-carotene, at least to some degree. The CRTR-B enzymes have been previously shown to be isozymes with indistinguishable activities toward β-carotene when expressed in *E. coli*. *In vivo* analysis indicates the CRTR-B enzymes are most active in the synthesis of β-xanthophylls, but that they also have significant activity toward the β-ring of α-carotene. 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 7) likely due to changes in substrate recognition and binding (Fig. 8 and 9). CYP97A3 has high activity toward the β-rings of β-carotene and α-carotene but no activity toward the ε-ring of α-carotene. CYP97C1 has high activity toward ε-rings (Tian et al., 2004b) and the β-ring of α-carotene but almost no 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 zeinoxanthin and α-carotene in seed of all *c1* and *a3* containing genotypes, respectively, while the same genotypes accumulate high levels of these carotenoids 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 zeinoxanthin and α-carotene in leaves and seed. The fact

that total carotenoid levels in the leaves of nine genotypes shown in Table 7 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 but not leaves. This is likely due to a combination of carotenoid associations with proteins in leaves but not seed (e.g. LHCs) and enzyme mediated carotenoid degradation. Indeed, a null mutant for carotenoid cleavage dioxygenase 1 has been shown to have significantly increase seed carotenoid levels while leaf carotenoids are unaffected (Auldridge et al., 2006b).

3.4.2. The evolution of β-xanthophyll synthesis β-xanthophylls and CRTR-B-type enzymes are widely distributed in nature and found in all photosynthetic eukaryotes and many photosynthetic and non-photosynthetic prokaryotes. The Arabidopsis CRTR-B1 and CRTR-B2 are isozymes (Tian and DellaPenna, 2001; Tian et al., 2003) and relatively strong expression correlation in photosynthetic tissues (Fig. 10), suggesting their most recent common ancestor (MRCA) had a similar biochemical activities and was expressed in photosynthetic tissues. The *b1b2* mutant has negatively impacted NPQ and Fv/Fm (Fig. 11) suggesting that photoprotection was likely an important function of the CRTR-B 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 be an additional selective pressure for retention of CRTR-B activity as the *b1b2* mutant also has insufficient ABA production under drought conditions (Tian et al., 2004a).

The retention of *CRTR-B* paralogs appears to be widespread in higher plants whether underlying selective pressures are common. In Arabidopsis, after duplication of the ancestral *CRTR-B* gene, most likely by whole genome or segmental genome

duplication 24~48 million years ago (Fig. 7 and Supplementary data Table 1), the duplicates seem to have diverged primarily at the gene expression level (Fig. 10). Our carotenoid analyses suggest CRTR-B2 is more actively involved in β-xanthophyll synthesis in seeds than in leaves, consistent with the rapidly induced *CRTR-B2* gene during seed development (Supplementary material 5). Expression divergence of *CRTR-B* genes is not restricted to Arabidopsis, as one of the two *CRTR-B* members in bell pepper, tomato and saffron (C. *sativus*) also shows preferential expression in flower or during fruit development (Bouvier et al., 1998; Castillo et al., 2005; Galpaz et al., 2006). Unlike Arabidopsis, *CRTR-B* expression divergence is strongly associated with tissue specific functional divergence in these other organisms. In tomato *crtr-b2* mutant results in a colorless petal phenotype with no impact on β-xanthophyll synthesis in leaves (Galpaz et al., 2006). The massive accumulation of β-xanthophylls during maturation of the saffron stigma was correlated with high expression of a single *CRTR-B* gene (C. *sativus C1* in Fig. 6B) (Castillo et al., 2005). Flower development involves the transformation of chloroplasts to chromoplasts (Whatley and Whatley, 1987) and it is likely that the expression divergence of the *CRTR-B* genes in reproductive organs (e.g. chromoplast-specific expression) provides the biochemical flexibility to differentially regulate β-xanthophyll synthesis in these tissues.

3.4.3. The 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 eucaryotes, 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 selective 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 formations from lycopene by β- and ε- cyclase followed by hydroxylation of each ring by β- and ε-ring hydroxylases (Fig. 2). Like β-xanthophylls, β-cyclases are widespread in nature and have been recruited for use in α-xanthophyll synthesis. ε-Cyclases have only been identified in green algae, plants and *Prochlorococcus* (a cyanobacterium) (Partensky et al., 1993; Krubasik and Sandmann, 2000; Hess et al., 2001; Cunningham et al., 2007) and appear to have arisen from β-cyclases by gene duplication and subsequent functional divergence before the green algae and plant split (Krubasik and Sandmann, 2000; Cunningham et al., 2007). They phylogeny of CYP97 β- and ε-ring hydroxylases similarly shows that duplication of the MRCA also occurred before he speciation of green algae and higher plants and that the *CYP97A* and *CYP97C* genes have been under purifying selection (Fig. 6A). 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 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 activites, the ε-ring cyclase, CYP97A3 and CYP97C1. To assess these functional constraints, we used several informative carotenoid biosynthetic mutant genotypes to infer the impact that various pathway intermediates have when accumulated by a hypothetical ancestral organism (i.e., β-xanthophyll

producing organism) as it acquired one or more of the three necessary enzymatic activities. The *lut2* mutant, which is defective in ε-ring cyclase activity, cannot synthesize α-xanthophylls (Pogson et al., 1996) and is a reasonable approximation of a hypothetical β-xanthophyll accumulating ancestral organism. The absence of lutein in *lut2* results in elevated levels of β-xanthophylls, partial impairment of NPQ, a smaller photosystem cross sectional area and lower LHC trimer stability . While mature *lut2* plants grow as well as wild type under moderate light conditions they are slightly less resistant to high light stress (Pogson et al., 1998; Niyogi et al., 2001; Dall'Osto et al., 2006).

The consequences of the hypothetical ancestral organism first acquiring ε-ring cyclase activityis approximated by the *a3c1* genotype, which has a functional ε-ring cyclase, can synthesize α-carotene and has CRTR-B enzyme activity but lacks both CYP97A and CYP97C1 activities. In addition to β-carotene and β-xanthophylls the *a3c1* mutant produces α-carotene and zeinoxanthin in approximately equal molar ratios (Table 7), due to inefficient hydroxylation of the β-ring of the α-carotene by the extant CRTR-B enzymes present in the mutant background. The *a3c1* mutant is highly susceptible to photooxidation in high light, more so than any other single xanthophyll biosynthetic mutant. The high light susceptibility of *a3c1* is due to the presence of α-carotene rather than zeinoxanthin as the *c1* single mutant, which accumulates an equivalent amount of zeinoxanthin as *a3c1* but lacks α-carotene, does not exhibit such photosensitivity (Table 7 and Fig. 11). The phentoype of the *a3c1* mutant suggests that acquisition of ε-cyclase activity by the 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 α-carotene accumulation and the presence of CRTR-B-type enzymes in the ancestral organism, it is surprising that the extant CRTR-B-type enzymes have not evolved to more efficiently hydroxylate the β-ring of α-carotene. This suggests that there is a structural constraint on the CRTR-B class of enzymes that makes the efficient hydroxylation of the β-rings of both α-carotene and β-carotene impossible. 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 in α-xanthophyll synthesis (Fig. 13). The original CYP97 enzyme acquired by the ancestral organism likely had α-carotene β-ring hydroxylation activity (CYP97A-like activity) as both the extant CYP97A3 and CYP97C1 have this activity. The *a3* mutant phenotype suggests the evolution of an efficient CYP97A-like α-carotene β-ring hydroxylation activity would have been strongly selected for as it would have alleviated α-carotene-dependent photooxidation.

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. The driving force for selection of a CYP97C-like activity is efficient photosystem structure and function, which is less intuitive than the α-carotene-dependent photooxidation driving selection of CYP97A-like activity, but no less important. In plants, lutein is essential for the assembly and stability of large light harvesting photosystems and for optimal NPQ kinetics and maximal NPQ. While the impact of xanthophyll alterations on NPQ is most readily quantified under experimental constant high light conditions, such as those used in Fig. 11A, the impact of altering NPQ on plant fitness is most apparent under natural light conditions. Kulheim *et.*

al (2002) showed that NPQ deficiency had no discernable impact on plants grown under constant light in growth chamber conditions but when grown in natural lighting, which has large swings in light intensity occurring 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 (Kulheim et al., 2002). Given this fitness impact it is clear how evolution of a CYP97C-like activity would be selected for and maintained.

The scenario we have described above (the sequential acquisition of ε-cyclase, CYP97A and CYP97C activities in Fig. 13) 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 CRTR-B type enzyme was involved in β-xanthophyll synthesis. In this case, and assuming the CYP97A-like enzyme had at least some endogenous α-carotene β-ring hydroxylation activity, the acquisition of ε-cyclase activity would have resulted in monohydroxy α-carotene (zeinoxanthin) production. If the enzyme was inefficient and some α-carotene was produced, this would still provide strong selection for improvement of α-carotene β-ring hydroxylation activity and against loss of the CYP97A-like gene (which would lead to α-carotene-dependent photooxxidation). Duplication of the CYP97A-like gene and evolution of CYP97C-like activity would then occur as described. Regardless of the exact sequence of events, the phenotypes of mutants defective in extant carotenoid hydroxylases highlight the strong selective pressures that likely operated in the evolution and selection of particular genes and activities that have led to the current biosynthetic pathway found in Arabidopsis. These same strong 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.

3.5. Materials and methods

3.5.1 The *CYP97* and *CRTR-B* genes used for phylogenetic analyses To search for *CYP97* and *CRTR-B* homologs in photosynthetic eucaryotes, I performed tblastn search using three Arabidopsis *CYP97* genes (*CYP97A3*, *CYP97B3* and *CYP97C1*) and two *CRTR-B* genes (*CRTR-B1* and *CRTR-B2*) as queries in publicly available databases (NCBI, TIGR and JGI) (*E* value cutoff = 10^{-30}). To assist in genome sequence annotation, I experimentally determined the full length sequences for two *C. reinhardtii* homologs (*CYP97A5*, EF587911 and *CYP97C3*, EF587910) amplified from cDNA pool as for the CYP97 family members. All sequences used were summarized in Supplementary data 1.

3.5.2. Sequence alignments and computational analyses 27 CYP97 and 33 CRTR-B sequences were aligned by the computer program ClustalX 1.81 with default parameters (Thompson et al., 1997) and alignments further refined manually (Supplementary data Fig. 3A and 3B). Protein sequence rather than DNA sequence was used for alignments, as protein sequence is more suitable for long-term evolution studies (Hashimoto et al., 1994; Russo et al., 1996; Glazko and Nei, 2003; Nam et al., 2003). In the CYP97 sequence alignment (Supplementary data 3A), an alignment block corresponding to columns 160 to 705 was selected for phylogenetic analyses, sliding window analyses and the estimation of cluster-specific functional divergence. In the CRTR-B alignment (Supplementary data 3B), a block corresponding to column 1 to 414 was used for phylogenetic analyses and column 134 to 403 for sliding window analyses. For phylogenetic analyses, three different methods were applied: Neighbor-Joining (NJ), Maximum Parsimony (MP) and Maximum Likelihood (ML). To construct NJ tree, I used Poisson correction (PC) distance method with pairwise deletion of gaps was used with the MEGA 3.1 program

(Kumar et al., 2004). Each MP tree of *CYP97* and *CRTR-B* gene was generated with the JTT model (Jones et al., 1992), after removing uninformative residues in the computer program PAUP* 4.0 beta 10 (Swofford, 2003). The ML tree was constructed using the PhyML algorithm with gaps treated as unknown characters (Guindon and Gascuel, 2003). For all trees, branch support was assessed by bootstrapping (500 replicates). The *CYP86A1* and *C. reinhardtii CRTR-B* genes were selected as outgroups for each tree, respectively. *CYP86A1* was selected because its substrates, fatty acids with chain lengths from C12 to C18 (Benveniste et al., 1998), are the most similar to carotenoids and the *CYP86* clade is the most closely related to *CYP97* clade (http://www.p450.kvl.dk/cyp_allsubfam_NJ_102103.pdf). All P450 nomenclature follows convention (http://drnelson.utmem.edu/CytochromeP450.html).

For sliding window analysis, seven taxa in each alignment block were selected and gaps were eliminated. Rates of amino acid substitutions (*p*-distance) in each sliding window of two aligned protein sequences with one outgroup (i.e. CYP86A1 in CYP97A/C and *C. reinhardtii* C1 in CRTR-B comparisons) since their most recent common ancestor (MRCA) were estimated using least-square method and compared using a program developed by Nam *et.* al (Nam et al., 2005). Because *C. reinhardtii* and rice have multiple *CYP97A* and *CRTR-B* paralogs, respectively, *CYP97A5* and *CYP97C3* were used for the *C. reinhardtii* CYP97A/C comparison, and *O. sativa C2* and *C3* were selected for the rice *CRTR-B* comparison. The posterior probability to lead cluster-specific functional divergence was estimated for each residue of CYP97 alignment by the DIVERGE 2 program (Gu, 2006) with ten taxa in the angiosperm lineage. Ka, Ks and confidence interval (C.I.) were calculated by K-estimator 6.0 software (Comeron, 1999)

from the CRTR-B alignment block used for the sliding window analysis. C.I was retrieved from 500 replicates.The templates for CYP97A3 homology modeling were retrieved from Metaserver (http://meta.bioinfo.pl/submit_wizard.pl) and the configuration of side chains were determined by the computer program maxsprout (Holm and Sander, 1991) (http://www.ebi.ac.uk/maxsprout/) and scrawl 3.0 (Canutescu et al., 2003). The CYP97A3 homology model was visualized with VMD software available in (http://www.ks.uiuc.edu/Research/vmd/).

3.5.3 Plant materials and pigment analyses The leaves of four-week old *Ws* background and five-week old Col-0 background genotypes grown under 12 h photoperiod (100 μmol m^{-2}·sec^{-1}, 22/18 °C) were used for pigment analyses and high light stress experiment. High light was subjected to plants in the chamber which was set to 1600~1800 μmol·m^{-2}·sec^{-1}, 50% humidity and 22 °C. The seeds were harvested and stored in the box containing a desiccator for one month before pigment analysis (Tian et al., 2003; Auldridge et al., 2006b). HPLC separation and quantification by spectra and retention time were performed as described in (Tian and DellaPenna, 2001; Kim and DellaPenna, 2006).

3.5.4. Expression data analyses AtGenExpress database was the source of expression data for the *CYP97* and *CRTR-B* gene pairs (http://www.arabidopsis.org/info/expression/ATGenExpress.jsp) with the exception that the data obtained from high-light stressed leaf tissue was provided by Dr. Dirk Inzé (Vanderauwera et al., 2005). Used eight tissue types and stress conditions were independent and nonredundant.

3.5.5. Colinearity of chromosomal segments Protein sequence in Arabidopsis and rice

chromosomal segments was obtained from publicly available annotated sequence data. The putative open reading frames in poplar genome sequence were deduced by FGENESH software available in SoftBerry (www.softberry.com). The proteins in chromosomal segments were compared in a pairwise fashion using bl2seq software available in NCBI. Cutoff for homologous pair was an *e*-value less than 10^{-10}.

3.5.6 Measurement of *in vivo* chlorophyll fluorescence and CO_2 fixation rates To obtain two photosynthetic parameters (NPQ and Fv/Fm), *in vivo* chlorophyll fluorescence was measured from the dark-adapted intact leaves for five minutes by applying a saturation pulse with and without actinic light (530 $\mu mol \cdot m^{-2} \cdot sec^{-1}$) using the Imaging-PAM Chlorophyll Fluorometer (Walz, Germany) (Berger et al., 2004). CO_2 fixation rate was measured after five minute in Arabidopsis chamber of LI-COR 6400 (Li-COR Biosciences), a gas exchanging system for CO_2 and H_2O analyses. Reference CO_2 concentration was set to 400 $\mu mol \cdot sec^{-1}$ (slightly above ambient pressure) by the CO_2 mixer and the temperature in the block was 22 °C. Air flow was set to 100 $\mu mol \cdot sec^{-1}$. Before CO_2 fixation rate measurements, 10 h- treated plants were put in dark for twelve-hours, followed by 100 $\mu mol \cdot m^{-2} \cdot sec^{-1}$ illumination for seven hours.

Table 4 Carotenoid composition (nmol·g^{-1} fresh weight) of green and albino progeny from the indicated genotypes.

		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.

[b] progeny from *b1b1b2b2C1c1a3a3* parent.

[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 5 List of *CYP97* homologs used for constructing a neighbor-joining tree.

Organism	Gene name	Accession #/locus name/ TC number/JGI annotation #
Solanum esculentum	*S. esculentum CYP97A*	TC126862
Populus trichocarpa	*CYP97A7*	gw1.87.99.1
Arabidopsis thaliana	*CYP97A3*	At1g31800
Medicago truncatula	*M. truncatula CYP97A*	ABD28565
Hodeum vulgare	*H. vulgare CYP97A*	TC76166
Oryza sativa	*CYP97A4*	AK068163
Ostreococcus tauri	*CYP97A11*	Ot13g02550
Chlamydomonas reinhardtii	*CYP97A5*	EF587911
Chlamydomonas reinhardtii	*CYP97A6*	DNE_DNE_e_gwW.42.59.1
Ostreococcus tauri	*CYP97B14*	Ot01g05440
Ostreococcus tauri	*CYP97B15*	Chr15.00010014
Skeletonema costatum	*CYP97E1*	AF459441
Chlamydomonas reinhardtii	*CYP97B6*	DNE_e_gwW.1.53.1
Ginko biloba	*G.biloba CYP97B*	AY601887
Oryza sativa	*CYP97B4*	TC299269
Arabidopsis thaliana	*CYP97B3*	AT4G15110
Populus trichocarpa	*CYP97B7*	fgenesh1_pg.C_LG_VI000069
Medicago truncatula	*M. truncatula CYP97B*	ABE94036
Ostreococcus tauri	*CYP97C12*	Ot09g02560
Chlamydomonas reinhardtii	*CYP97C3*	EF587910
Oryza sativa	*CYP97C2*	AK065689
Arabidopsis thaliana	*CYP97C1*	At3g53130
Daucus carota	*D.carota CYP97C*	ABB52076
Solanum esculentum	*S. esculentum CYP97C*	SGN E542349 and E346934

Table 5 con't

Populus trichocarpa	*CYP97C4*	eugene3.00280258
Medicago truncatula	*CYP97C10*	DQ335801

Table 6 List of *CRTR-B* homologs used for constructing a neighbor-joining tree.

Organism	Gene name	Accession #/locus name/ TC number/JGI annotation #
Solanum esculentum	*S. esculentum CRTR-B1*	CAB55625
Capsicum annuum	*C. annuum CRTR-B2*	CAA70888
Solanum esculentum	*S. esculentum CRTR-B2*	CAB55626
Capsicum annuum	*C. annuum CRTR-B1*	CAA70427
Gentiana lutea	*G. lutea CRTR-B homolog 1*	AB027187
Medicago truncatula	*M. truncatula CRTR-B homolog 1*	ABE85312
Citrus unshiu	*C. unshiu CRTR-B homolog 1*	AAG33636
Coffea arabica	*C. arabica CRTR-B1*	DQ157169
Tagetes erecta	*T. erecta CRTR-B homolog 1*	AAG10430
Arabidopsis thaliana	*A. thaliana CRTR-B2*	AT5G52570
Arabidopsis thaliana	*A. thaliana CRTR-B1*	AT4G25700
Brassica rapa	*B. rapa CRTR-B homolog 1*	DQ156907
Daucus carota	*D. carota CRTR-B homolog 1*	ABB52074
Daucus carota	*D. carota CRTR-B homolog 2*	ABB52075
Vitis vinifera	*V. vinifera CRTR-B1*	AAM77007
Poplus trichocarpa	*P. trichocarpa CRTR-B homolog 1*	estExt_fgenesh1_pg_v1.C_440227
Poplus trichocarpa	*P. trichocarpa CRTR-B homolog 2*	estExt_fgenesh1_pg_v1.C_LG_IV0070
Adonis palaestina	*A. palaestina CRTR-B1*	ABI93208
Adonis palaestina	*A.palaestina Adketo2*	AY644758
Adonis palaestina	*A.palaestina AdKeto1*	AY644757
Crocus sativus	*C. sativus CRTR-B2*	CAC95130

Table 6 con't

Crocus sativus	*C. sativus CRTR-B1*	AAT84408
Narcissus pseudonarcissus	*N. pseudonarcissus CRTR-B homolog 1*	CAC06712.1
Oryza sativa	*O. sativa CRTR-B homolog 1*	Os03g0125100
Oryza sativa	*O. sativa CRTR-B homolog 2*	Os10g0533500
Oryza sativa	*O. sativa C CRTR-B homolog 3*	Os04g0578400
Zea mays	*Z. mays CRTR-B homolog 1*	AY844956
Zea mays	*Z. mays CRTR-B homolog 2*	AY844958
Zea mays	*Z. mays CRTR-B homolog 3*	BQ619575
Pinus taeda	*P. taeda CRTR-B homolog 1*	TC67291
Pinus taeda	*P. taeda CRTR-B homolog 2*	TC67290
Ostreococcus tauri	*O. tauri CRTR-B homolog 1*	gw1.10.00.289.1
Chlamydomonas reinhardtii	*C. reinhardtii CRTR-B homolog 1*	AAX54907
Haematococcus pluvialis	*H. pluvialis CRTR-B1*	AAD54243

Table 7 Carotenoid composition in the leaves of the two wild types (Col-0 and *Ws*) and seven informative genotypes. Carotenoid levels are expressed as mmol pigment mol^{-1} of chlorophylls. Each value is the mean ± SD of at least three biological replicates. The numbers in parenthesis indicate the percentage of total carotenoids

	violaxanthin	antheraxanthin	zeaxanthin	neoxanthin	lutein	α-car-OH	α-carotene	β-carotene	β-xanthophylls	tot.car
Col-0	31±4.6 (12)	2.5±0.6 (1)	n.d (0)	32±3.0 (12)	133±10 (52)	0.8±0.3 (0)	1.6±0.6 (1)	56±3.6 (22)	66±8.2	257±21
Ws	28±3.1 (12)	1.8±0.4(0.7)	n.d (0)	30±2.5 (13)	122±8.9 (51)	0.6±0.1(0.2)	1.5±0.2(0.6)	54±3.7(23)	60±6.0	238±16
c1	55±4.0 (23)	24±3.0 (10)	8.1±1.4 (3)	26±1.8 (11)	4.0±0.4 (2)	61±5.8 (26)	1.4±0.3 (1)	57±4.7 (24)	113±10	236±17
a3	21±2.0 (9)	2.3±1.0 (1)	n.d (0)	18±2.6 (8)	104±6.7 (46)	5.4±0.6 (2)	49±0.8 (22)	26±3.1 (12)	41±5.6	227±10
b1b2	11±1.0 (4)	1.1±0.3 (0)	n.d (0)	3.8±1.0 (2)	156±8.0 (65)	1.0±0.3 (0)	1.5±0.2 (1)	67±2.5 (28)	16±2.3	241±11
a3c1	41±2.7 (17)	22±2.5 (9)	7.0±1.3 (3)	20±2.4 (8)	3.2±0.4 (1)	53±5.2 (22)	54±3.1 (23)	38±3.5 (16)	90±8.9	237±18
b1b2c1	10±3.0 (4)	8.6±0.6 (4)	4.3±0.4 (2)	2.0±0.9 (1)	7.2±0.7 (3)	101±3.8 (43)	1.6±0.3 (1)	99±4.7 (42)	25±4.9	234±7.6
b1a3	14±0.8 (7)	4.3±1.6 (2)	n.d (0)	4.1±0.7 (2)	94±6.2 (43)	8.8±0.7 (4)	59±6.3 (27)	33±5.4 (15)	22±3.1	217±16
b1b2a3	0.2±0.1 (0)	0.6±0.3 (0)	n.d (0)	0.5±0.9 (0)	99±14 (48)	10±5.9 (5)	56±9.2 (27)	39±6.9 (19)	1.3±1.3	205±20

Table 8 Carotenoid composition in the seeds of the two wild types (Col-0 and *Ws*) and seven informative genotypes. Carotenoid levels are expressed as nmol pigment·g^{-1} of dry weight. Each value is the mean ± SD of at least three biological replicates. The numbers in parenthesis indicate the percentage of total carotenoids.

	violaxanthin	antheraxanthin	zeaxanthin	neoxanthin	lutein	α-car-OH	α-carotene	β-carotene	β-xanthophylls	tot.car
Col-0	2.4±0.3 (6)	1.9±0.2 (5)	3.5±0.3 (9)	1.0±0.1 (3)	31±2.1 (78)	n.d.	n.d.	0.1±0.0 (0)	8.8±0.9 (22)	40±2.9
Ws	1.9±0.2 (6)	1.5±0.1 (4)	2.5±0.2 (8)	0.8±0.1 (3)	25±1.2 (81)	n.d.	n.d.	0.1±0.0 (0)	6.7±0.6 (22)	31±1.6
c1	3.4±0.7 (34)	2.0±0.2 (20)	1.9±0.3 (19)	1.2±1.6 (12)	1.0±0.1 (10)	0.5±0.2 (5)	n.d.	0.1±0.0 (1)	8.5±2.8 (86)	9.9±2.8
a3	1.4±0.2 (3)	1.4±0.1 (3)	1.8±0.2 (4)	0.8±0.1 (2)	38±2.8 (86)	n.d.	0.2±0.0 (0)	0.1±0.1 (0)	5.4±0.6 (12)	44±3.1
b1b2	1.0±0.0 (3)	0.9±0.1 (3)	1.1±0.0 (4)	0.6±0.0 (2)	26±1.5 (87)	n.d.	n.d.	0.3±0.0 (1)	3.6±0.1 (12)	30±1.6
a3c1	3.2±0.7 (27)	1.9±0.3 (16)	1.9±0.2 (16)	0.5±0.2 (4)	1.5±0.6 (13)	2.2±0.1 (18)	0.4±0.1 (3)	n.d.	7.5±1.4 (63)	12±1.6
b1b2c1	1.9±0.3 (33)	0.6±0.1 (10)	0.5±0.0 (9)	0.5±0.1 (9)	1.1±0.2 (19)	1.3±0.4 (22)	n.d.	0.4±0.2 (7)	3.5±0.5 (60)	5.8±1.0
b1a3	2.4±0.4 (5)	2.4±0.1 (5)	1.6±0.2 (3)	1.1±0.1 (2)	45±2.3 (85)	n.d.	0.4±0.2 (1)	0.2±0.1 (0)	7.5±0.8 (14)	53±2.8
b1b2a3	1.0±0.1 (2)	1.0±0.1 (2)	0.7±0.1 (2)	0.6±0.1 (1)	40±4.7 (91)	n.d.	0.4±0.1 (1)	0.7±0.1 (2)	3.3±0.4 (8)	44±4.9

Fig. 6A Neighbor-joining tree of CYP97 sequences. Scale bar = 0.2 (Poisson distance). 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.

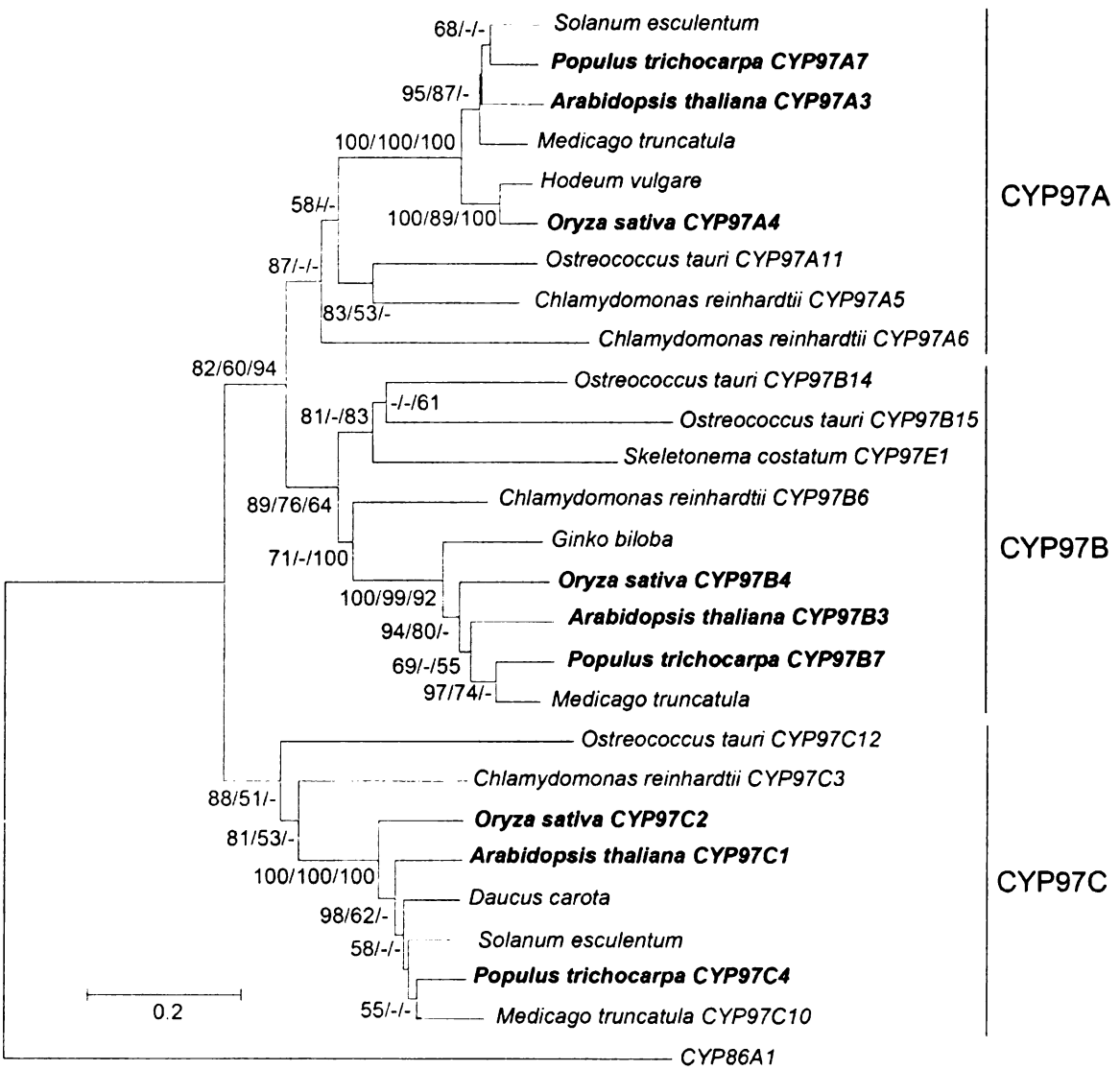


Fig. 6B Neighbor-joining tree of CRTR-B sequences. Scale bar = 0.1 (Poisson distance). CRTR-B1, CRTR-B2 indicates paralogs in an organism that have been demonstrated to have β-ring hydroxylase activity by assay in carotenoid containing *E. coli* lines. *CRTR-B homolog 1*, *CRTR-B homolog 2*, *CRTR-B homolog 3* indicate CRTR-B paralogs in an organism whose functions have not yet been experimentally defined. 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.

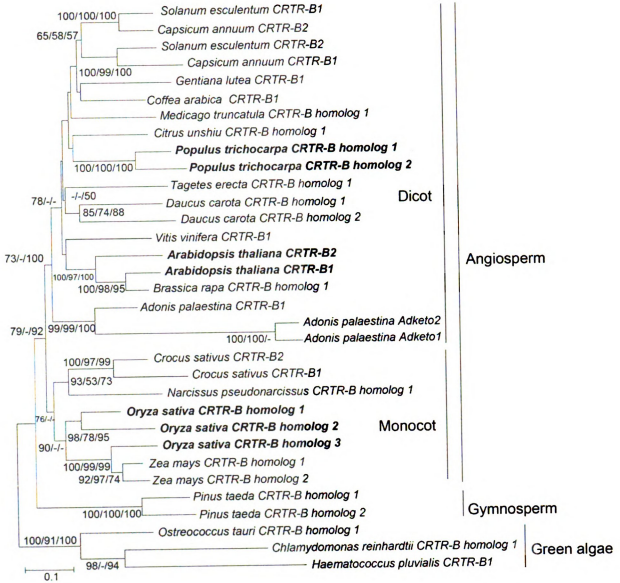

Solanum esculentum CRTR-B1
Capsicum annuum CRTR-B2
Solanum esculentum CRTR-B2
Capsicum annuum CRTR-B1
Gentiana lutea CRTR-B1
Coffea arabica CRTR-B1
Medicago truncatula CRTR-B homolog 1
Citrus unshiu CRTR-B homolog 1
Populus trichocarpa CRTR-B homolog 1
Populus trichocarpa CRTR-B homolog 2
Tagetes erecta CRTR-B homolog 1
Daucus carota CRTR-B homolog 1
Daucus carota CRTR-B homolog 2
Vitis vinifera CRTR-B1
Arabidopsis thaliana CRTR-B2
Arabidopsis thaliana CRTR-B1
Brassica rapa CRTR-B homolog 1
Adonis palaestina CRTR-B1
Adonis palaestina Adketo2
Adonis palaestina Adketo1
Crocus sativus CRTR-B2
Crocus sativus CRTR-B1
Narcissus pseudonarcissus CRTR-B homolog 1
Oryza sativa CRTR-B homolog 1
Oryza sativa CRTR-B homolog 2
Oryza sativa CRTR-B homolog 3
Zea mays CRTR-B homolog 1
Zea mays CRTR-B homolog 2
Pinus taeda CRTR-B homolog 1
Pinus taeda CRTR-B homolog 2
Ostreococcus tauri CRTR-B homolog 1
Chlamydomonas reinhardtii CRTR-B homolog 1
Haematococcus pluvialis CRTR-B1
100/100/100
65/58/57
100/99/100
100/100/100
78/-/-
-/-/50
85/74/88
73/-/100
100/97/100
100/98/95
79/-/92
99/99/100
100/100/-
100/97/99
93/53/73
76/-/-
98/78/95
90/-/-
100/99/99
92/97/74
100/100/100
100/91/100
98/-/94
0.1
Dicot
Angiosperm
Monocot
Gymnosperm
Green algae

Fig. 7A Exon-intron structures of *CRTR-B* genes in Arabidopsis, poplar and rice. Exons and introns are represented as boxes and lines in scale, respectively.

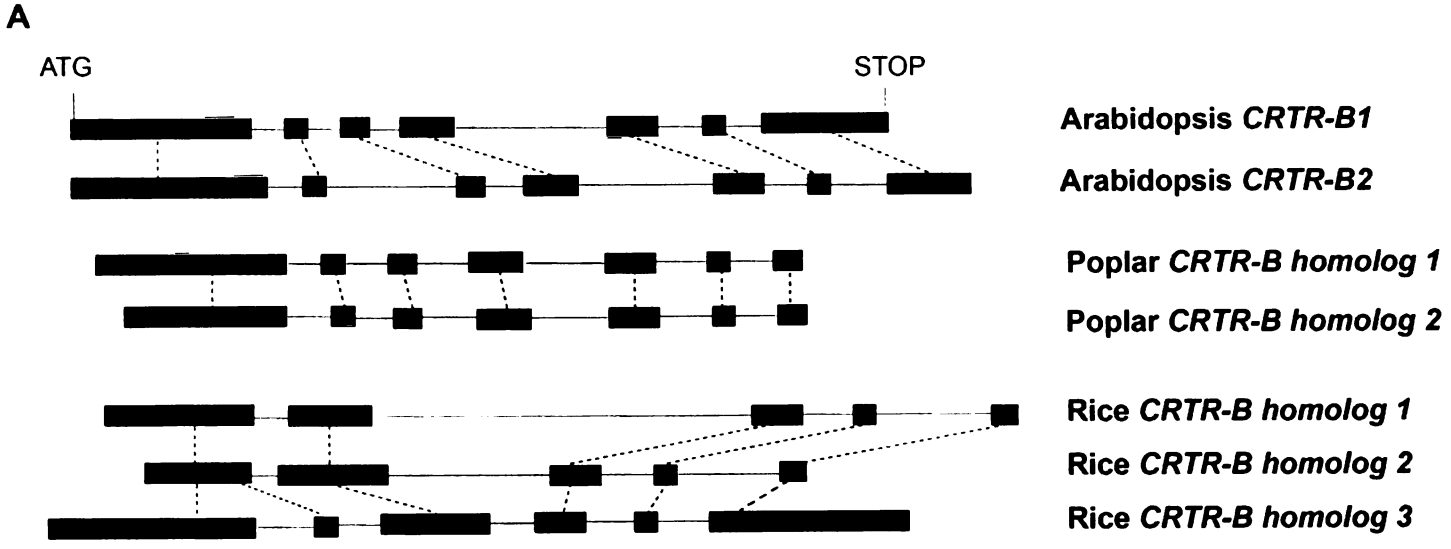

A
ATG
STOP
Arabidopsis CRTR-B1
Arabidopsis CRTR-B2
Poplar CRTR-B homolog 1
Poplar CRTR-B homolog 2
Rice CRTR-B homolog 1
Rice CRTR-B homolog 2
Rice CRTR-B homolog 3

Fig. 7B Colinearity of chromosomal segments containing parologous *CRTR-B* genes in Arabidopsis, poplar and rice. 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 *CRTR-B* genes in each chromosomal segment are marked with asterisks.

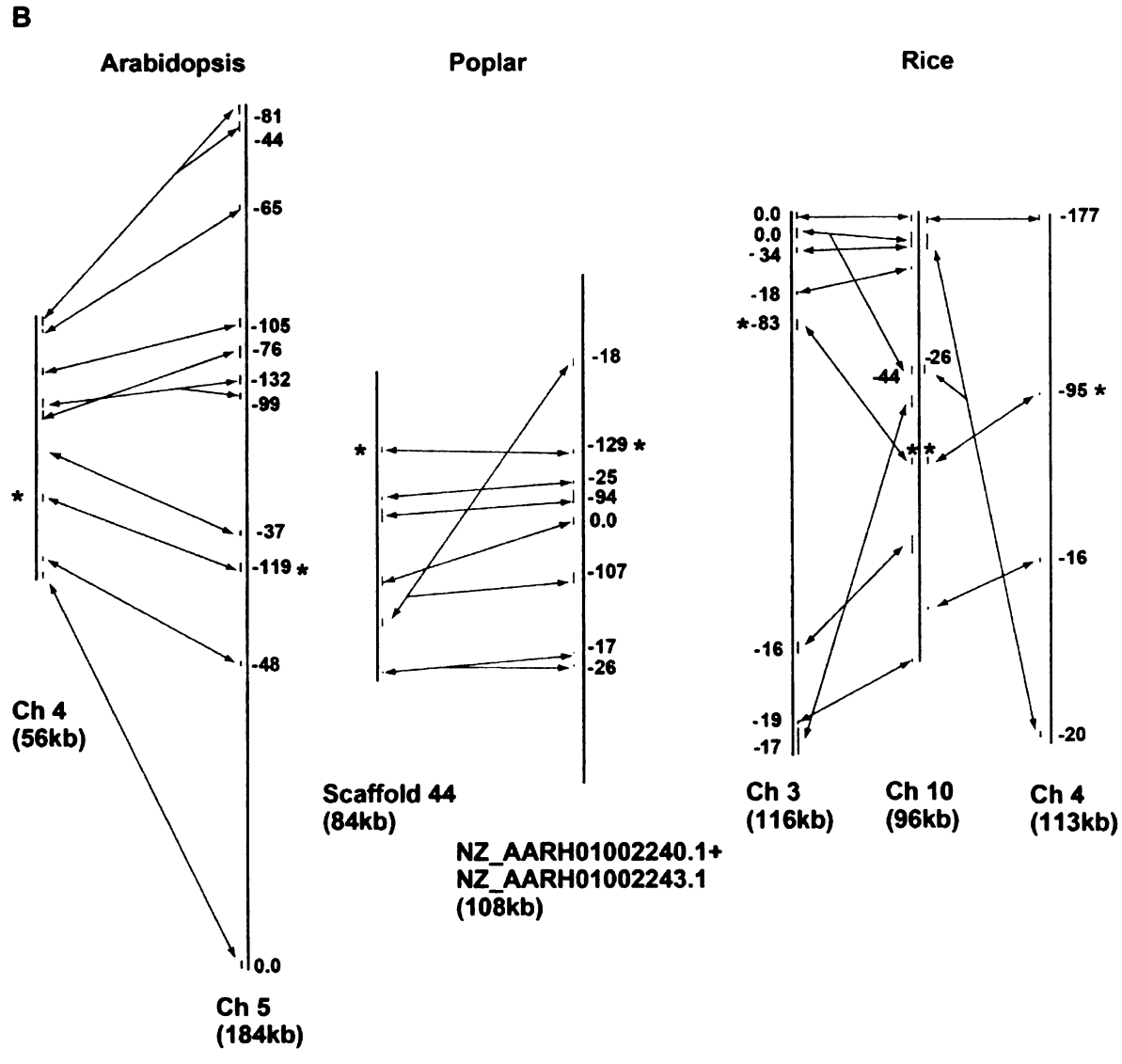

B
Arabidopsis
Poplar
Rice
-81
-44
-65
-105
-76
-132
-99
-37
-119 *
-48
0.0
Ch 4
(56kb)
Ch 5
(184kb)
-18
-129 *
-25
-94
0.0
-107
-17
-26
Scaffold 44
(84kb)
NZ_AARH01002240.1+
NZ_AARH01002243.1
(108kb)
0.0
0.0
-34
-18
*-83
-44
-26
-16
-19
-17
-177
-95 *
-16
-20
Ch 3
(116kb)
Ch 10
(96kb)
Ch 4
(113kb)

Fig. 8A Averaged Z-scores from each paralogous CYP97A/C pair from Arabidopsis, rice, *C. reinhardtii* and *O. tauri*. The identified region is indicated with an arrow. The numbers on the X-axis represents column positions in the CYP97 alignment. Note that the length of the CYP97 protein used for the alignment is 436 amino acid residues.

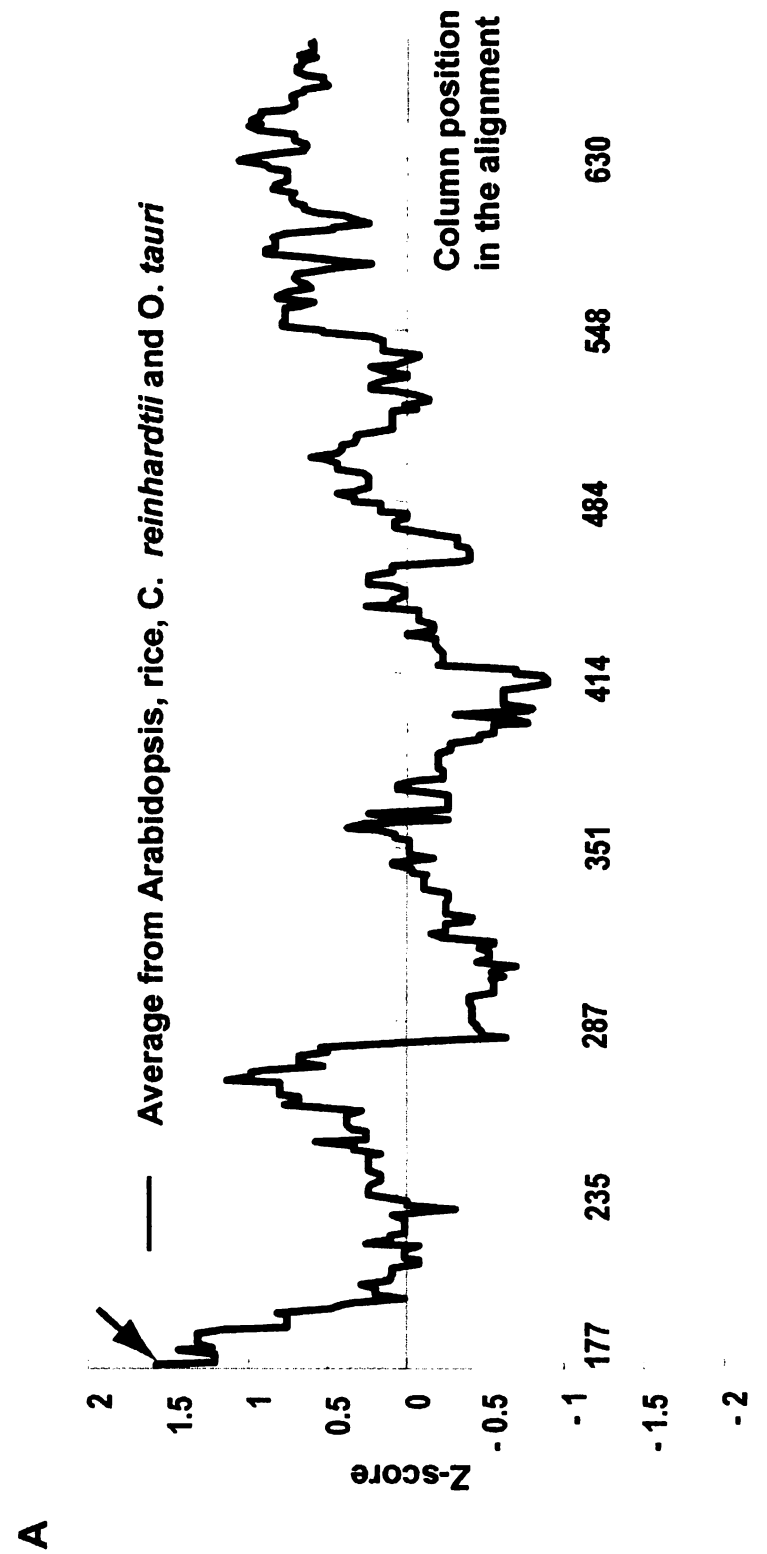

A
Average from Arabidopsis, rice, C. reinhardtii and O. tauri
Z-score
2
1.5
1
0.5
0
- 0.5
- 1
- 1.5
- 2
177
235
287
351
414
484
548
630
Column position
in the alignment

Fig. 8B Z-scores from the two CRTR-B pairs of each Arabidopsis, rice and pine. The numbers on the X-axis represents column positions in the CRTR-B alignment. Note that the length of the CRTR-B protein used for the alignment is 214 amino acid residues.

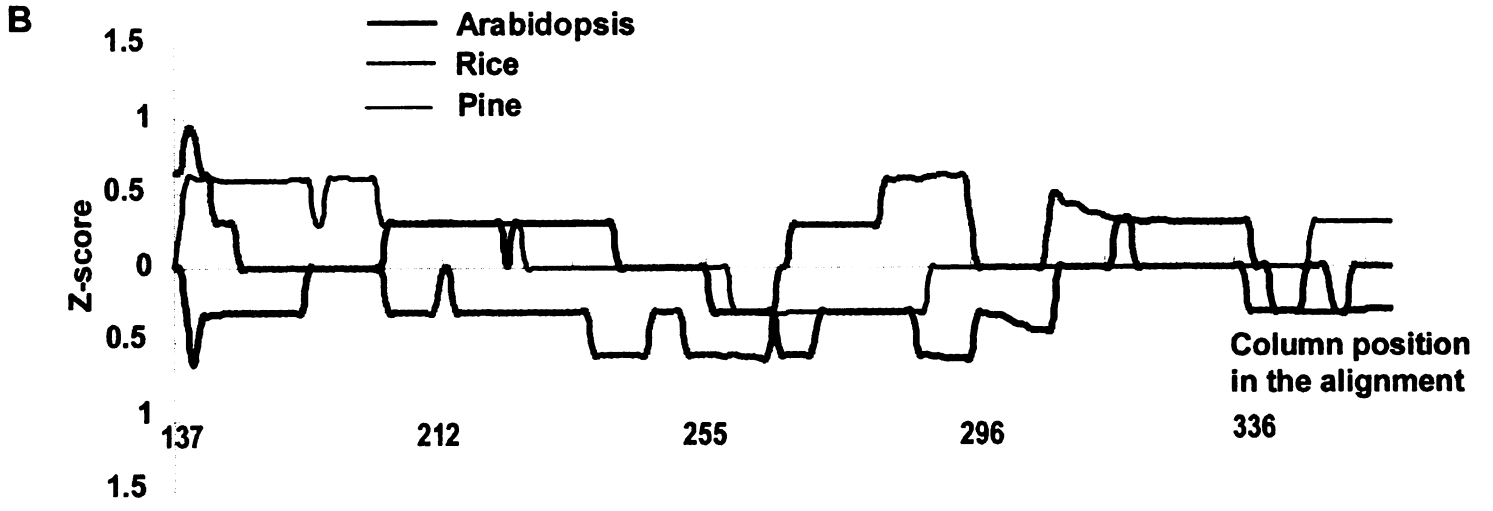

B
Arabidopsis
Rice
Pine
1.5
1
0.5
0
0.5
1
1.5
Z-score
137
212
255
296
336
Column position
in the alignment

Fig. 8C Posterior probability of cluster-specific functional divergence in each amino acid residue. Among the twenty-nine residues having the highest score the six indicated by arrows are also conserved when both *C. reinhardtii* and *O. tauri* are included in the clades. The number on X-axis represents amino acid position in CYP97 alignment.

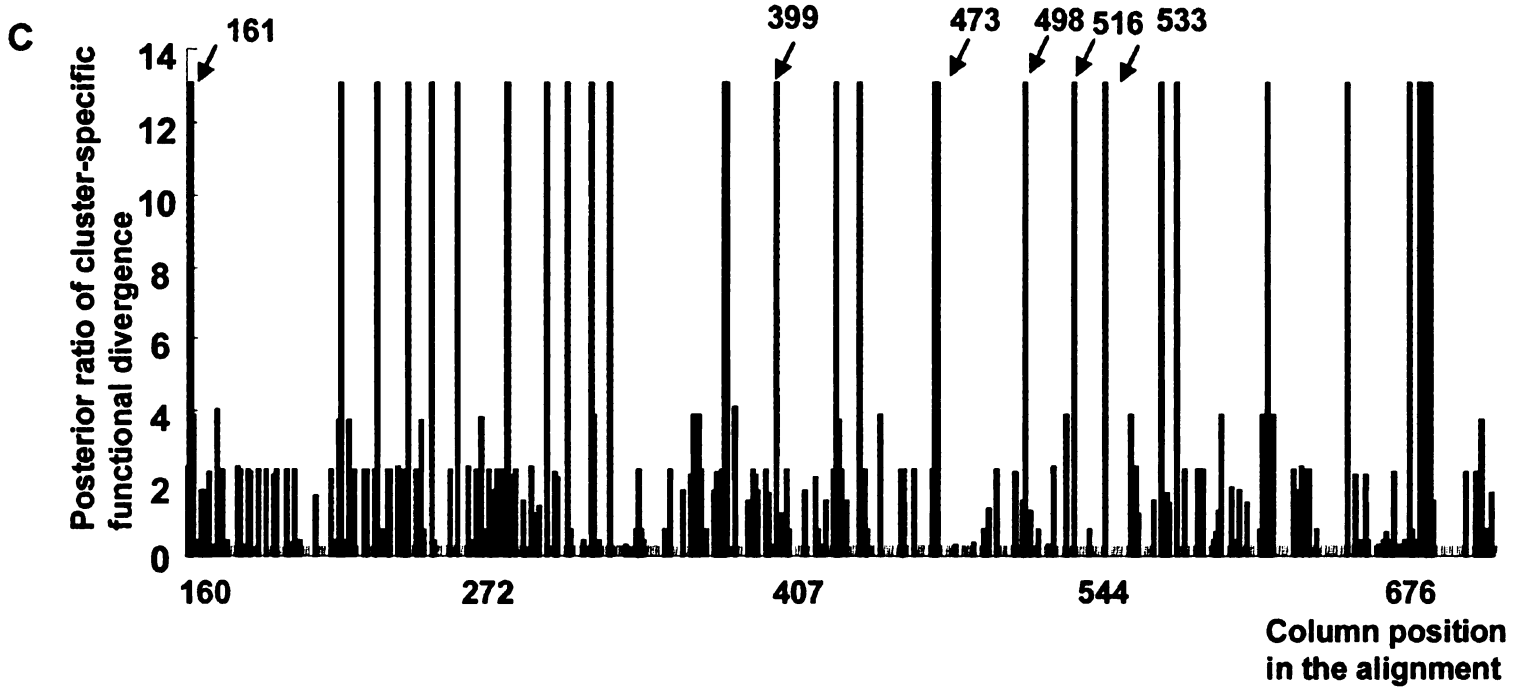

C
161
399
473
498
516
533
Posterior ratio of cluster-specific functional divergence
14
12
10
8
6
4
2
0
160
272
407
544
676
Column position in the alignment

Fig. 9 Two homology models (open and closed conformations) of CYP97A3. The single domain (blue) and six amino acid residues (red and pink) identified in Fig. 4 are mapped onto the model. The red indicates the amino acid residue in the middle of I helix which may bind to the substrate ring in the active site. The image in this dissertation is presented in color.

Closed conformation **Open conformation**

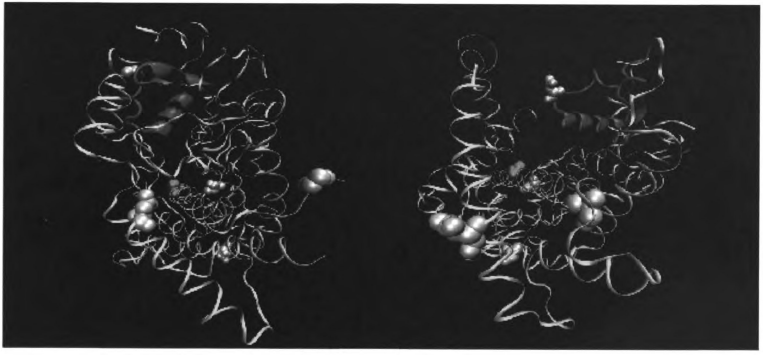


Fig. 10 Expression divergence of *CYP97* and *CRTR-B* gene pairs in the indicated tissues (A) and stress conditions in leaf (B). The expression correlation coefficient (r) of the *CYP97* and *CRTR-B* gene pairs is represented as white and black, respectively. The number of independent microarray datasets used to test for correlation in each condition is indicated in parenthesis.

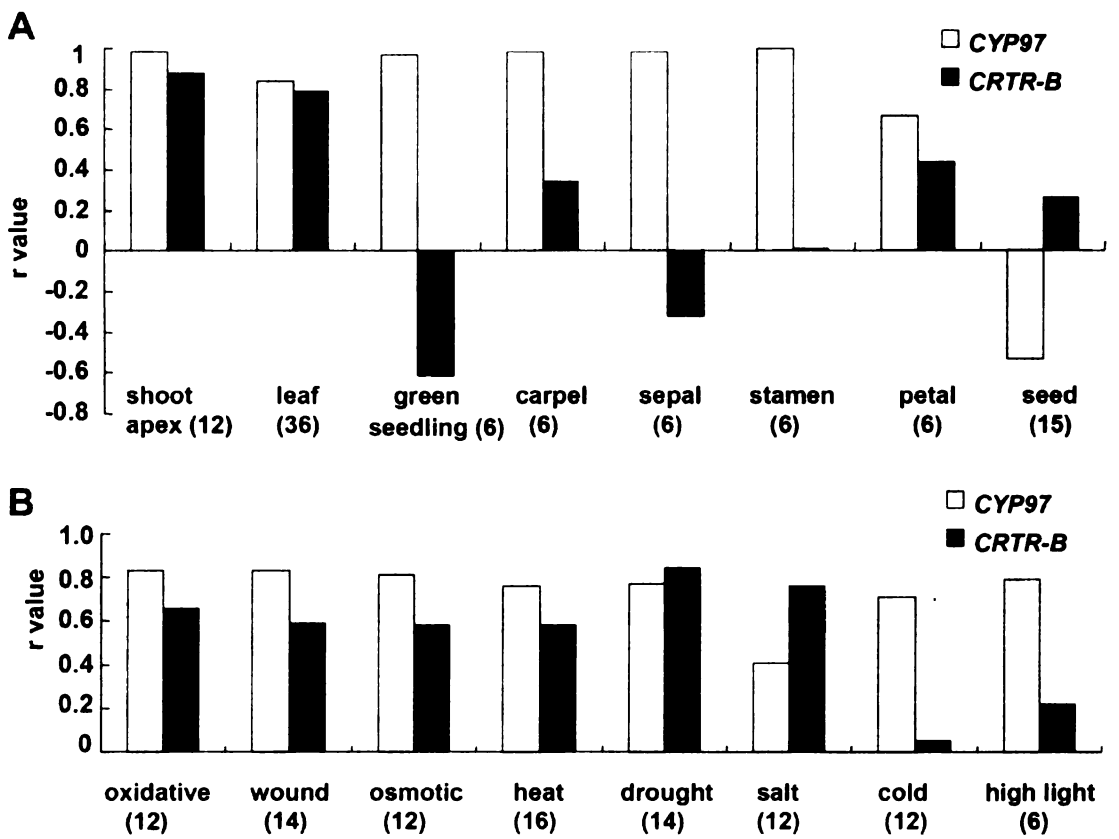


Fig. 11A Non-photochemical quenching (NPQ) kinetics under 530 μmol·m^{-2}·s^{-1} . The values are from three biological replicates. For clarity, error bars are omitted having less than 10% error.

A

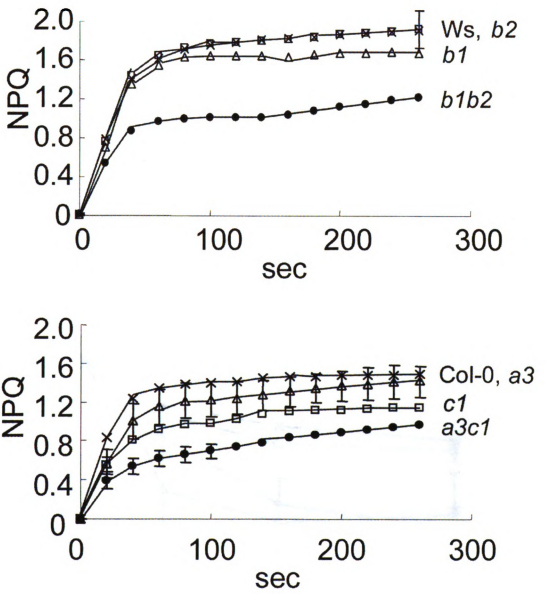


Fig. 11B Kinetics of maximum photosynthetic efficiency of PSII (Fv/Fm) under 1600~1800 $\mu mol \cdot m^{-2} \cdot s^{-1}$. The values are from three biological replicates. For clarity, error bars are omitted having less than 10% error.

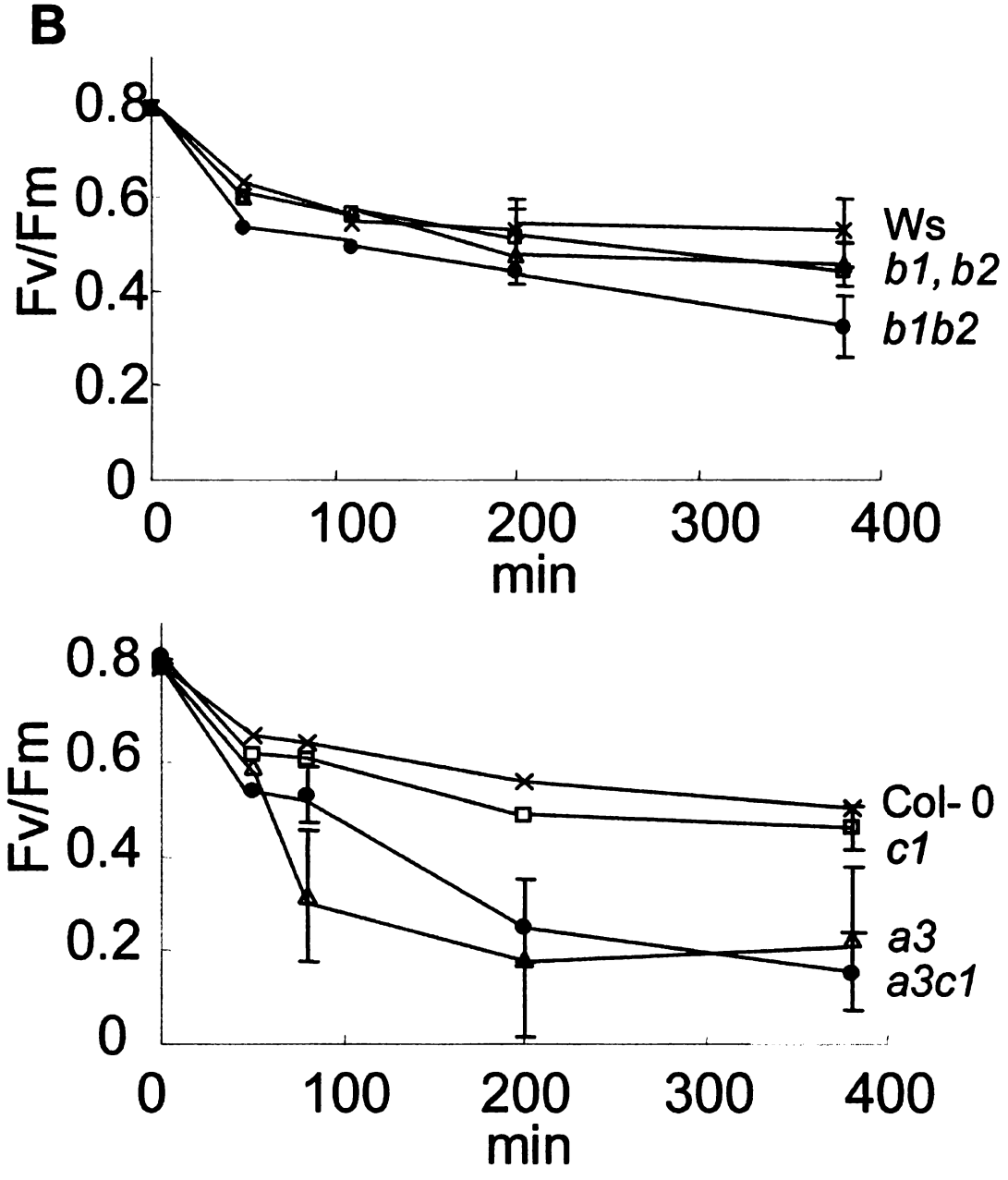


Fig. 12 CO_2 fixation rates of WT and *a3* before and after ten hour high light exposure. Data are means ± SD (n = 6; one leaf of six independents for CO_2 fixation measurement,). * $p < 0.05$, by Student's t test of mutant leaves relative to corresponding WT leaves. BHL, efore high light; AHL, after high light.

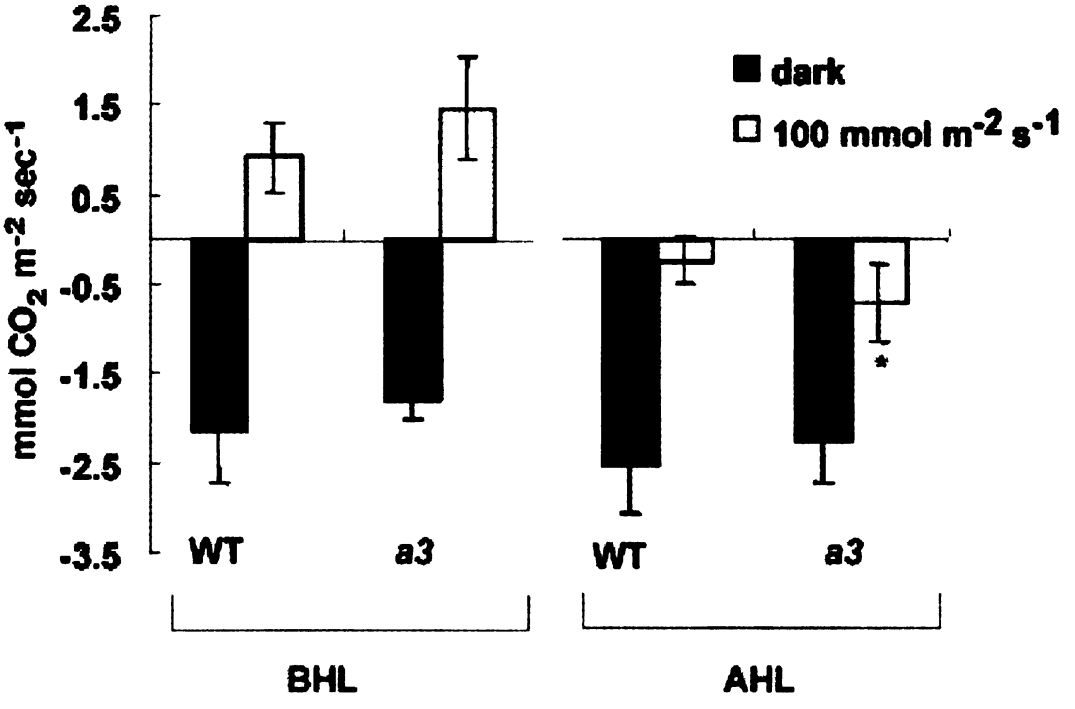
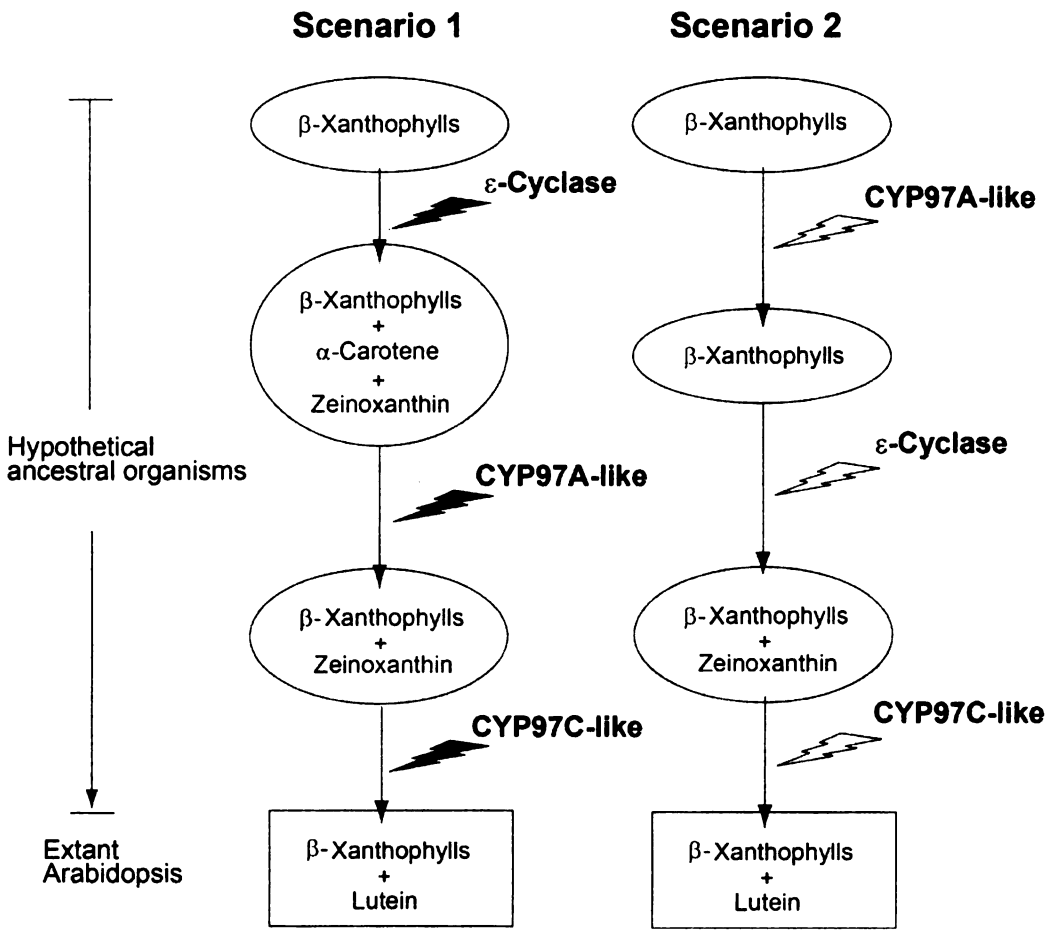


Fig. 13 Two different scenarios for evolution of the α-xanthophyll biosynthetic pathway in Arabidopsis from a hypothetical β-xanthophyll accumulating ancestral organism. The acquisition of newly evolved enzymatic activities in each scenario is indicated in bold.



CHAPTER 4

FUTURE WORK

4.1. Fitness test of *lut5-1* (*a3*) mutant under low light condition

In contrast to the undetectable level of α-carotene in canopy plants, the ratio of α-/β-carotene in understory and gap plants is elevated. One hypothesis is that these differences are dependent on light-intensity and that α-carotene accumulation has an advantage to adapt under low light, although it clearly has a negative impact on fitness under high light. The null mutant allele of CYP97A3, *lut5-1*, in which α-carotene constitutively accumulates can be used to test this hypothesis. My preliminary data (Supplementary data Fig. 4) showed that CO_2 fixation rate was not significantly different between WT and *lut5-1* mature leaves under light intensities ranging from 0 to 1300 $\mu mol \cdot m^{-2} sec^{-1}$. This indicates that at least the short term carbon fixation efficiency of plants with α-carotene containing photosystem is indistinguishable to that of β-carotene containing (wild type) photosystems. However, it is still open question whether α-carotene accumulation is beneficial to long term adaption under low light condition or at specific developmental stages such as young seedlings.

4.2. Identification of P450-type carotenoid hydroxylase in *C. reinhardtii*

The computational analyses with CYP97 amino acid sequences of green algae and higher plants only allowed prediction of domains and residues in which functional divergence may have occurred before the speciation of higher plants and green algae. These predictions are made with the assumption that CYP97 enzymatic activities are conserved at some extents in both green algae and higher plants. However, experimental

data to support this assumption has not yet been reported because the enzymatic activities of CYP97 enzymes have only been determined in Arabidopsis and rice. In this regard, the functional characterization of *C. reinhardtii* CYP97 orthologs could be quite insightful. Examining the activities of CYP97A5 and CYP97A6, which are orthologous to Arabidopsis CYP97A3, to determine whether these two enzymatic activities have "true" CYP97A3 activity and/or whether one or both have evolved new enzymatic activities (i.e. hydroxylation of lutein at C-19 to produce loroxanthin in *C. reinhardtii*). The amount of loroxanthin in *C. reinhardtii* is approximately 60 $mmol \cdot mol^{-1}$ chlorophyll *a*, accounting for the half of the amount of lutein (Niyogi et al., 1997).

4.3. In-depth phenotypic analyses of *crtr-b1* and *crtr-b2* mutants

NPQ and Fv/Fm measurement under high light stress with a series of mutants showed that the three enzymatic activities (CYP97A3, CYP97C1 and either CRTR-B1 or CRTR-B2) have the minimal complement to adapt under high light although all four enzymes have the full complement of carotenoid hydroxylases. Although the distinct phenotype in both *crtr-b1* and *crtr-b2* was not observed in the study, expression divergence observed in reproductive organs and stressed leaves (e.g. cold and high light stress) still remains as interesting subject to be answered in relation to their functional divergence. Therefore, future study should be focused on the phenotypic difference between *crtr-b1* and *crtr-b2* mutants. One possible role of CRTR-B2 is to produce β-xanthophylls for ABA synthesis. Highly induced its mRNA during seed development (Supplementary Fig. 5) and almost fully recovered β-xanthophyll level by sole activity of CRTR-B2 in dry seeds (Table 8) suggest that highly expressed *CRTR-B2* is related to β-

xanthophyll level. In accordance with highly accumulation of ABA, the increasing mRNA level of *CRTR-B2* gene during seed development is somewhat related with ABA biosynthesis because β-xanthophylls (neoxanthin and violaxanthin) are the precursor for ABA. *In silico* analyses to find *cis*-acting elements also supports this idea, detecting the ABA-responsive element (CACGTGGC) which is located on -291 upstream of *CRTR-B2* gene but not on *CRTR-B1* gene.

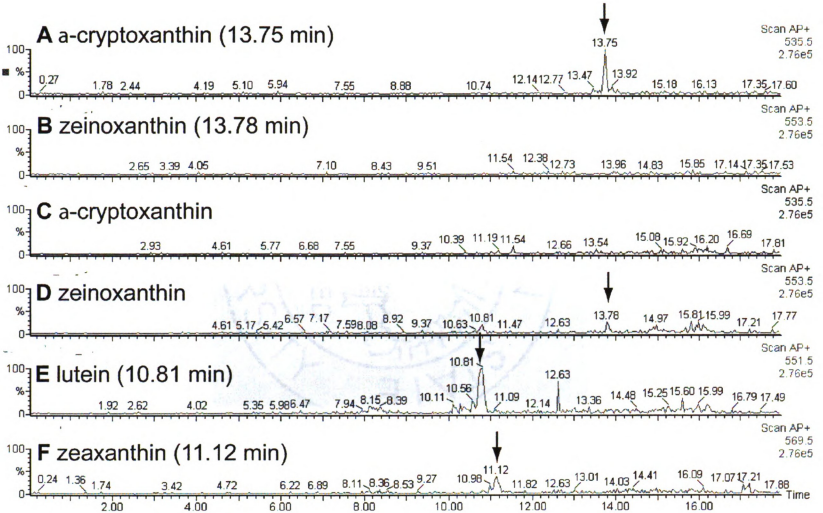
4.4. Identification of carotenoid cleavage enzymes recognizing the ε-ring

The major difference in carotenoid composition in dried seeds is the elevated mol % of total xanthophylls compared to leaves. During de-greening of Arabidopsis silique, proteolysis of photosynthetic machineries allows their pigments (chlorophylls and carotenoids) to be exposed without an association with LHC. Seeds contain trace levels of α- and β-carotene regardless of mutant genotypes and the level of zeinoxanthin in *c1*-containing mutant seeds was also decreased. These results illustrate that nonhydroxylated β- and/or ε-rings are unstable in seeds. One of possible reason may be that nonhydroxylated ring (β- and ε-rings) containing carotenoids are targets for carotenoid cleavage enzymes (CCDs) in seeds. Some Arabidopsis CCDs had already been shown to have activity toward β-carotene by functional complementation in β-carotene producing E. *coli* system (Schwartz et al., 2004). Considering the elevated mol % of total xanthophylls in seeds than in leaf tissues, the low level of zeinoxanthin in *cyp97c1* (*c1*) mutant seed is noteworthy, suggesting the existence of CCDs in seed that are highly reactive toward nonhydroxylated ε- or β-rings.

APPENDIX

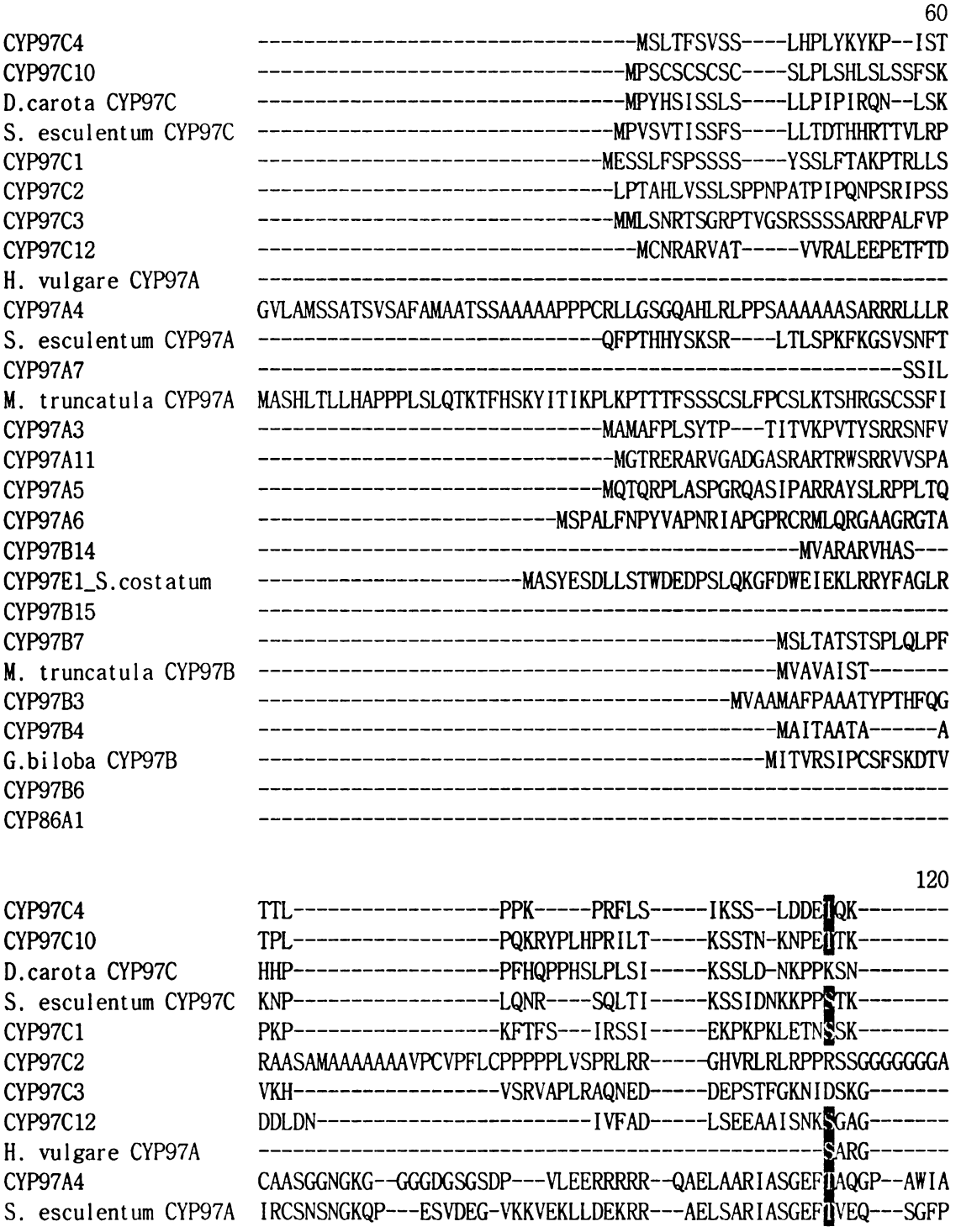
Supplementary data

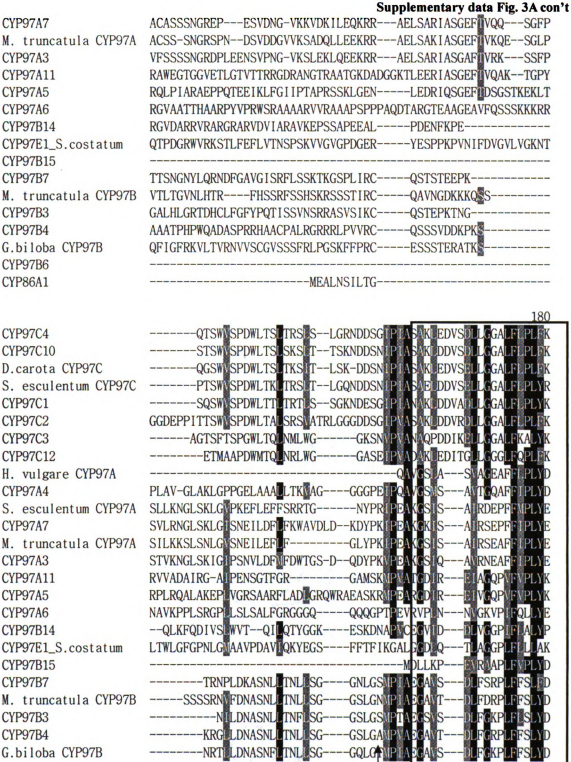
Supplementary data Fig. 1 HPLC chromatograms of the indicated mass ions. A and B, Mass traces corresponding to major quasimolecular ions for α-cryptoxanthin (A) and zeinoxanthin (B) in the TLC- purified unknown monohydroxy α-carotene derivative from *lut5-1* saponified leaf extracts. C-F, Mass traces corresponding to major quasimolecular ions for α-cryptoxanthin (C), zeinoxanthin (D), lutein (E) and zeaxanthin (F) in a mixture of *lut1-4* and *aba1-5* saponified leaf extracts. Arrows indicate the position of the respective carotenoid in each mass trace. Note that the *lut5-1* trace in (B) lacks a zeinoxanthin peak and that the *lut1-4* + *aba1-5* trace in (D) lacks a α-cryptoxanthin peak. Zeinoxanthin would be readily detectable if present at a level > 1% of the *alpha*-cryptoxanthin peak in *lut5-1*. The mass of quasimolecular ions ($[MH^+]$ and $[MH^+-H_2O]$) of the indicated carotenoids are below. α-cryptoxanthin (535.5), zeinoxanthin (553.5), lutein (551.5), zeaxanthin (569.5). Both A and B were separated as described (Tian and DellaPenna, 2001) except that the following gradient was initiated after sample injection: 0–13.0 min, 0% to 66.6%; 13.0–13.2 min, 66.6% to 100%.


A a-cryptoxanthin (13.75 min)
Scan AP+
535.5
2.76e5
B zeinoxanthin (13.78 min)
Scan AP+
553.5
2.76e5
C a-cryptoxanthin
Scan AP+
535.5
2.76e5
D zeinoxanthin
Scan AP+
553.5
2.76e5
E lutein (10.81 min)
Scan AP+
551.5
2.76e5
F zeaxanthin (11.12 min)
Scan AP+
569.5
2.76e5
%
Time

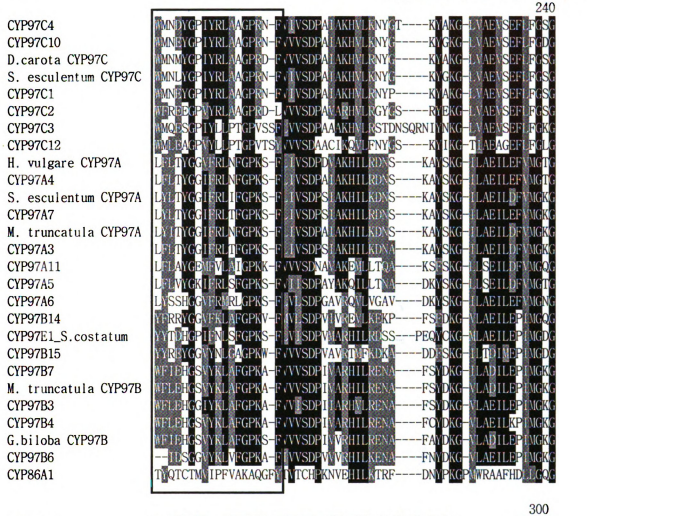
Supplementary data Fig. 2 Whole-plant phenotype of WT and *lut5-1* under normal light level (upper panel, 100 μmol $m^{-2}s^{-1}$) and two days of high light exposure (lower panel, 1700 μmol $m^{-2}s^{-1}$ for 8 hrs, followed by a 12hr dark period and an additional 12 hr of high light). The image in this dissertation is presented in color.

Supplementary data Fig. 3A Amino acid alignment of *CYP97* homologs which is used for NJ-tree construction. A domain detected by sliding window analysis is boxed and six residues identified from estimation of type II functional divergence are pointed with arrows.

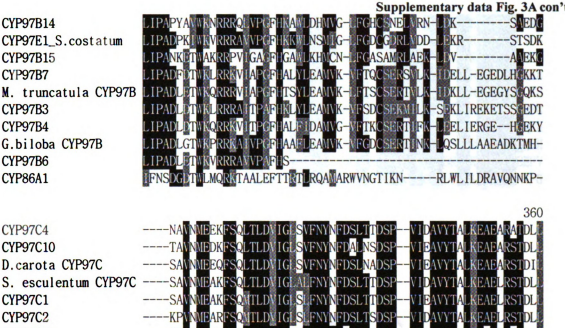


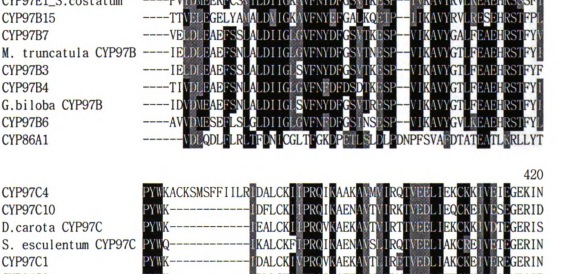

CYP97A7
M. truncatula CYP97A
CYP97A3
CYP97A11
CYP97A5
CYP97A6
CYP97B14
CYP97E1_S.costatum
CYP97B15
CYP97B7
M. truncatula CYP97B
CYP97B3
CYP97B4
G.biloba CYP97B
CYP97B6
CYP86A1
180
CYP97C4
CYP97C10
D.carota CYP97C
S. esculentum CYP97C
CYP97C1
CYP97C2
CYP97C3
CYP97C12
H. vulgare CYP97A
CYP97A4
S. esculentum CYP97A
CYP97A7
M. truncatula CYP97A
CYP97A3
CYP97A11
CYP97A5
CYP97A6
CYP97B14
CYP97E1_S.costatum
CYP97B15
CYP97B7
M. truncatula CYP97B
CYP97B3
CYP97B4
G.biloba CYP97B
CYP97B6
CYP86A1

Supplementary data Fig. 3A con't

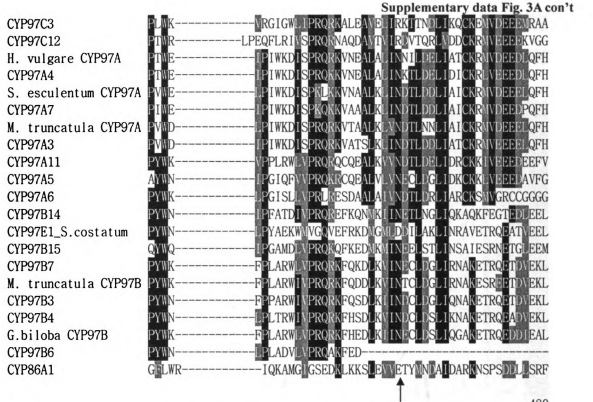


Supplementary data Fig. 3A con't

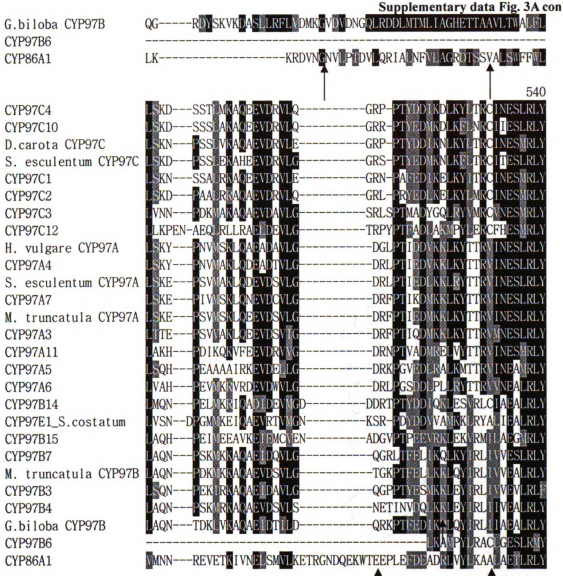


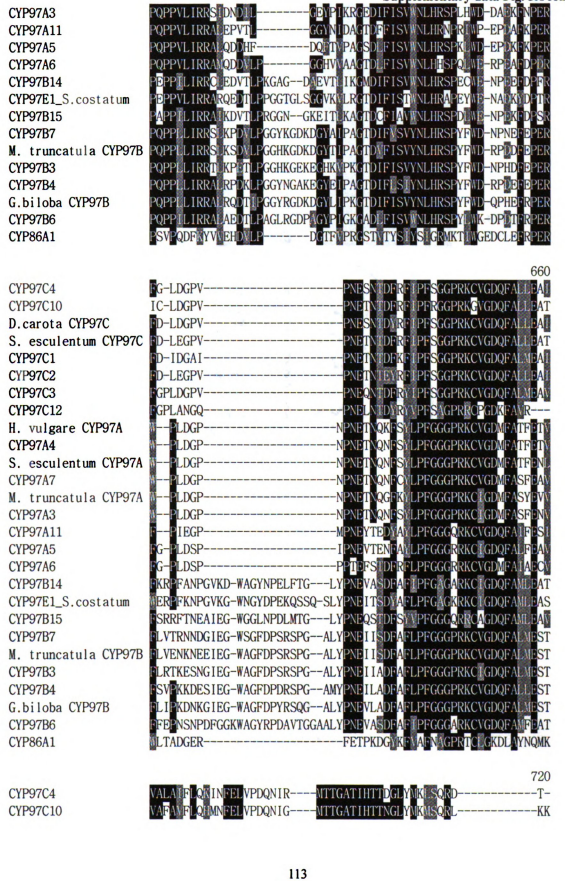


Supplementary data Fig. 3A con't

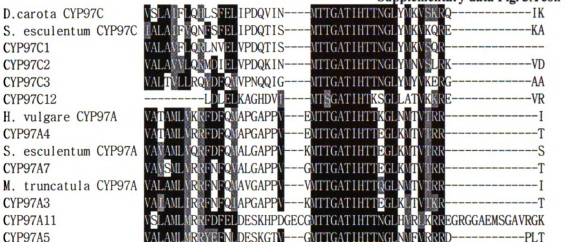


Supplementary data Fig. 3A con't




CYP97A3
CYP97A11
CYP97A5
CYP97A6
CYP97B14
CYP97E1_S.costatum
CYP97B15
CYP97B7
M. truncatula CYP97B
CYP97B3
CYP97B4
G.biloba CYP97B
CYP97B6
CYP86A1
660
CYP97C4
CYP97C10
D.carota CYP97C
S. esculentum CYP97C
CYP97C1
CYP97C2
CYP97C3
CYP97C12
H. vulgare CYP97A
CYP97A4
S. esculentum CYP97A
CYP97A7
M. truncatula CYP97A
CYP97A3
CYP97A11
CYP97A5
CYP97A6
CYP97B14
CYP97E1_S.costatum
CYP97B15
CYP97B7
M. truncatula CYP97B
CYP97B3
CYP97B4
G.biloba CYP97B
CYP97B6
CYP86A1
720
CYP97C4
CYP97C10

Supplementary data Fig. 3A con't

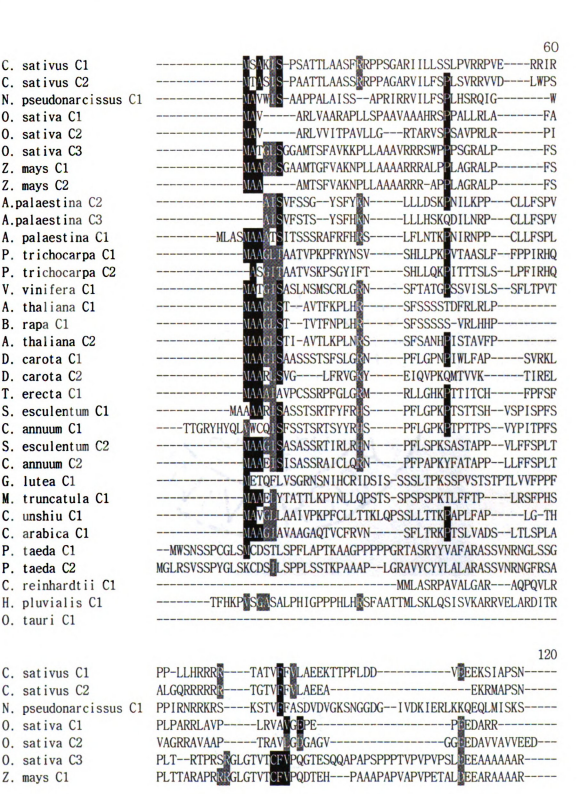


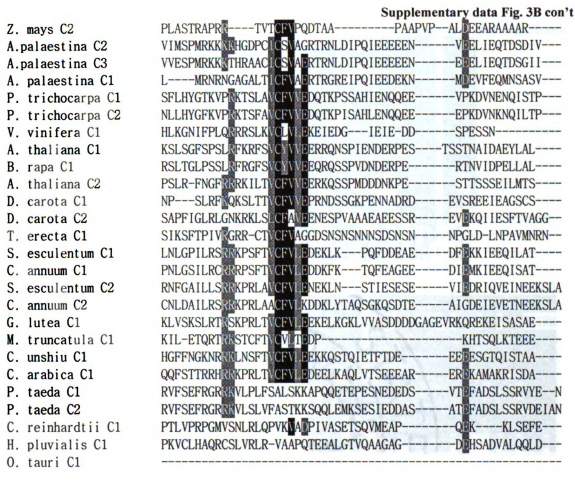
780

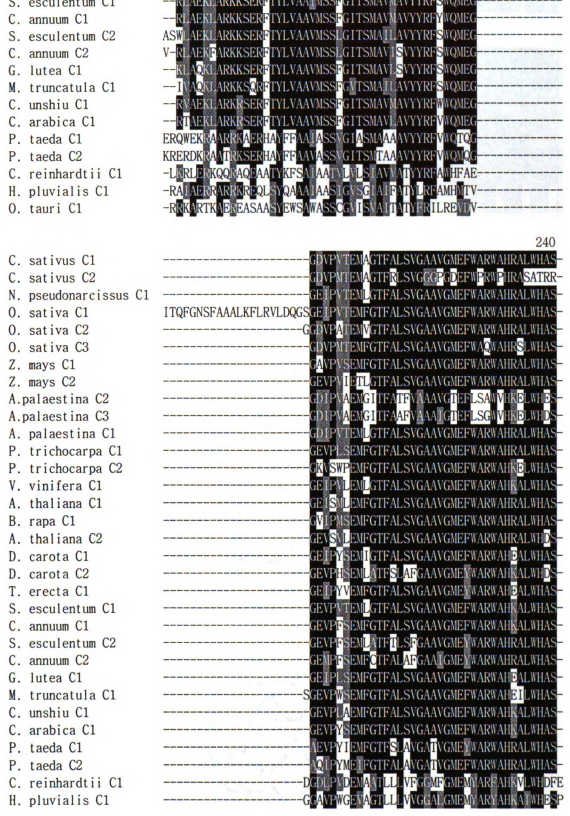
Supplementary data Fig. 3A con't

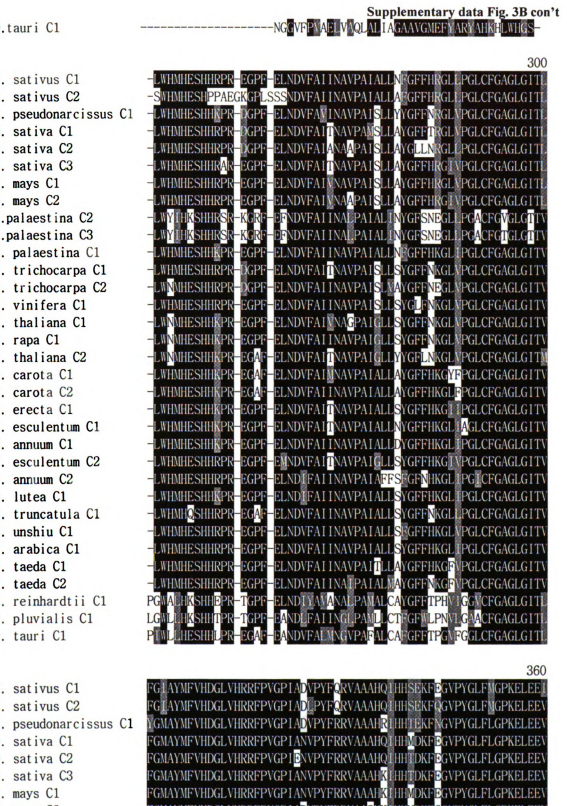
```
CYP97B7                VH----------
M. truncatula CYP97B   AH----------
CYP97B3                ----------
CYP97B4                ----------
G.biloba CYP97B        YSESLQ----------
CYP97B6                SGGSGSGAPGAAAKVPATV----------
CYP86A1                VLA----------


CYP97C4                ---------
CYP97C10               ---------
D.carota CYP97C        ---------
S. esculentum CYP97C   ---------
CYP97C1                ---------
CYP97C2                ---------
CYP97C3                ---------
CYP97C12               ---------
H. vulgare CYP97A      ---------
CYP97A4                ---------
S. esculentum CYP97A   ---------
CYP97A7                ---------
M. truncatula CYP97A   ---------
CYP97A3                ---------
CYP97A11               R--------
CYP97A5                LRPTGPPSA
CYP97A6                Q--------
CYP97B14               ---------
CYP97E1_S.costatum     ---------
CYP97B15               ---------
CYP97B7                ---------
M. truncatula CYP97B   ---------
CYP97B3                ---------
CYP97B4                ---------
G.biloba CYP97B        ---------
CYP97B6                ---------
CYP86A1                ---------
```

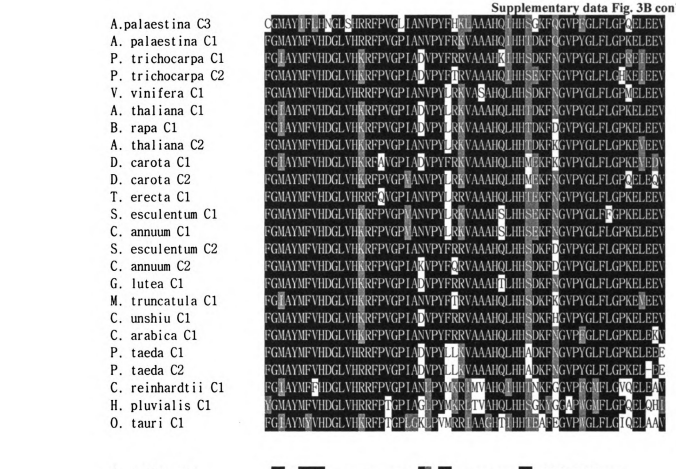
Supplementary data Fig. 3B Amino acid alignment of *CRTR-B* homolog genes used for NJ tree construction.

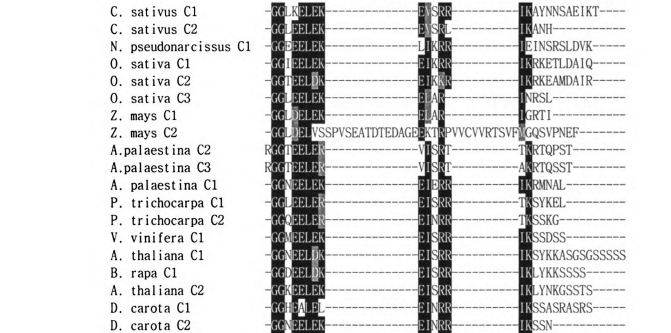


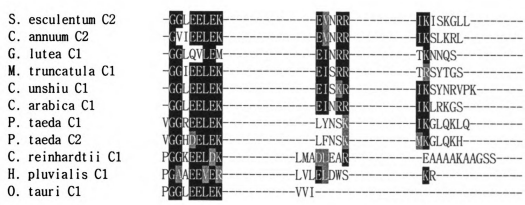

Z. mays C2
A.palaestina C2
A.palaestina C3
A. palaestina C1
P. trichocarpa C1
P. trichocarpa C2
V. vinifera C1
A. thaliana C1
B. rapa C1
A. thaliana C2
D. carota C1
D. carota C2
T. erecta C1
S. esculentum C1
C. annuum C1
S. esculentum C2
C. annuum C2
G. lutea C1
M. truncatula C1
C. unshiu C1
C. arabica C1
P. taeda C1
P. taeda C2
C. reinhardtii C1
H. pluvialis C1
O. tauri C1
180
C. sativus C1
C. sativus C2
N. pseudonarcissus C1
O. sativa C1
O. sativa C2
O. sativa C3
Z. mays C1
Z. mays C2
A.palaestina C2
A.palaestina C3
A. palaestina C1
P. trichocarpa C1
P. trichocarpa C2
V. vinifera C1
A. thaliana C1
B. rapa C1
A. thaliana C2
D. carota C1
D. carota C2
T. erecta C1


S. esculentum C1
C. annuum C1
S. esculentum C2
C. annuum C2
G. lutea C1
M. truncatula C1
C. unshiu C1
C. arabica C1
P. taeda C1
P. taeda C2
C. reinhardtii C1
H. pluvialis C1
O. tauri C1
240
C. sativus C1
C. sativus C2
N. pseudonarcissus C1
O. sativa C1
O. sativa C2
O. sativa C3
Z. mays C1
Z. mays C2
A.palaestina C2
A.palaestina C3
A. palaestina C1
P. trichocarpa C1
P. trichocarpa C2
V. vinifera C1
A. thaliana C1
B. rapa C1
A. thaliana C2
D. carota C1
D. carota C2
T. erecta C1
S. esculentum C1
C. annuum C1
S. esculentum C2
C. annuum C2
G. lutea C1
M. truncatula C1
C. unshiu C1
C. arabica C1
P. taeda C1
P. taeda C2
C. reinhardtii C1
H. pluvialis C1


O.tauri C1
300
C. sativus C1
C. sativus C2
N. pseudonarcissus C1
O. sativa C1
O. sativa C2
O. sativa C3
Z. mays C1
Z. mays C2
A.palaestina C2
A.palaestina C3
A. palaestina C1
P. trichocarpa C1
P. trichocarpa C2
V. vinifera C1
A. thaliana C1
B. rapa C1
A. thaliana C2
D. carota C1
D. carota C2
T. erecta C1
S. esculentum C1
C. annuum C1
S. esculentum C2
C. annuum C2
G. lutea C1
M. truncatula C1
C. unshiu C1
C. arabica C1
P. taeda C1
P. taeda C2
C. reinhardtii C1
H. pluvialis C1
O. tauri C1
360
C. sativus C1
C. sativus C2
N. pseudonarcissus C1
O. sativa C1
O. sativa C2
O. sativa C3
Z. mays C1
Z. mays C2
A.palaestina C2


A.palaestina C3
A. palaestina C1
P. trichocarpa C1
P. trichocarpa C2
V. vinifera C1
A. thaliana C1
B. rapa C1
A. thaliana C2
D. carota C1
D. carota C2
T. erecta C1
S. esculentum C1
C. annuum C1
S. esculentum C2
C. annuum C2
G. lutea C1
M. truncatula C1
C. unshiu C1
C. arabica C1
P. taeda C1
P. taeda C2
C. reinhardtii C1
H. pluvialis C1
O. tauri C1

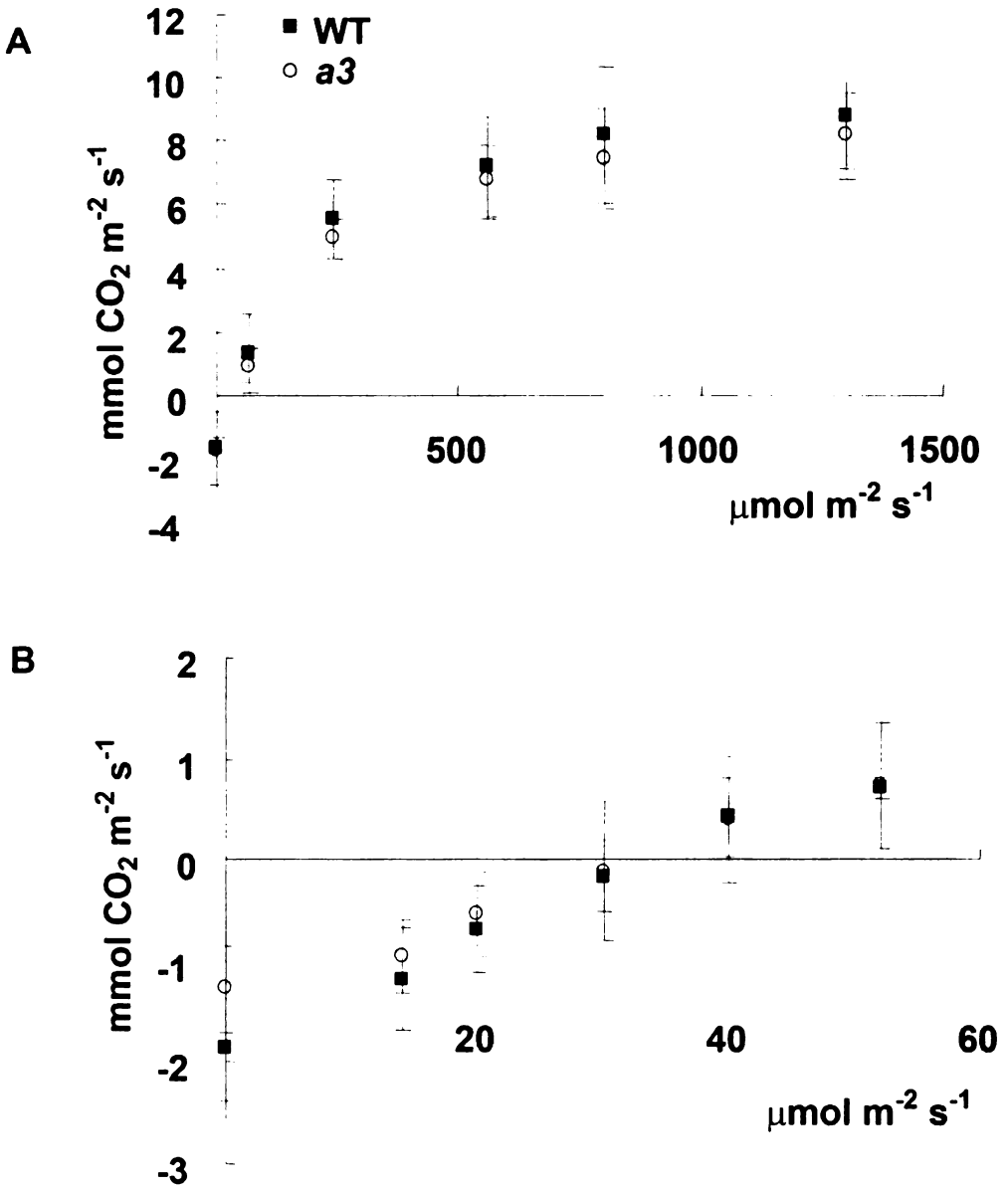

C. sativus C1
C. sativus C2
N. pseudonarcissus C1
O. sativa C1
O. sativa C2
O. sativa C3
Z. mays C1
Z. mays C2
A.palaestina C2
A.palaestina C3
A. palaestina C1
P. trichocarpa C1
P. trichocarpa C2
V. vinifera C1
A. thaliana C1
B. rapa C1
A. thaliana C2
D. carota C1
D. carota C2
T. erecta C1
S. esculentum C1
C. annuum C1


S. esculentum C2
C. annuum C2
G. lutea C1
M. truncatula C1
C. unshiu C1
C. arabica C1
P. taeda C1
P. taeda C2
C. reinhardtii C1
H. pluvialis C1
O. tauri C1

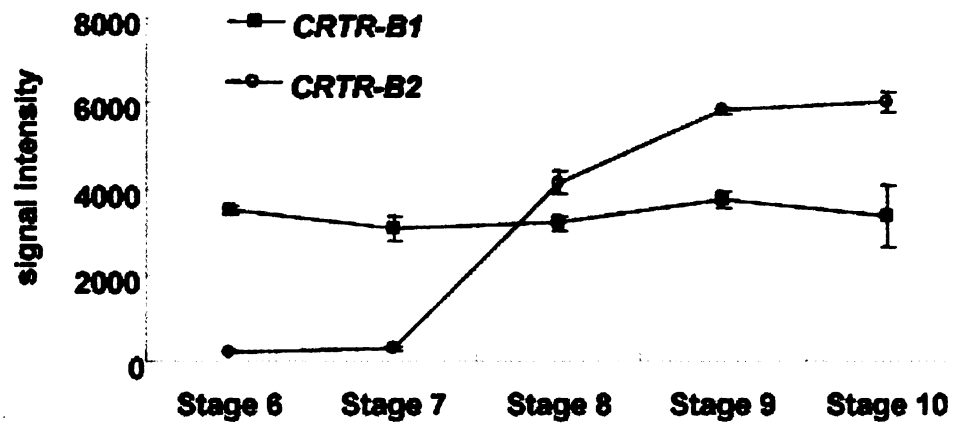
Supplementary data Table 1. Non-synonymous substitutions per site (Ka) and synonymous substitutions per site (Ks), and corresponding confidence intervals (C.I.) in *CRTR-B* genes of the indicated organisms. [a] Number of nucleotide base pairs used. [b] Comparison between *CRTR-B* homolog 1 and 2 [c] Comparison between *CRTR-B* homolog 1 and 3.

	No. bp[a]	Ka	C.I p=0.01		Ks	C.I p=0.01	
			Min	Max		Min	Max
Arabidopsis	903	0.165	0.124	0.218	0.617	0.420	0.809
Poplar	927	0.093	0.068	0.123	0.262	0.174	0.396
Rice[b]	849	0.146	0.100	0.212	0.339	0.227	0.489
Rice[c]	825	0.197	0.143	0.255	0.657	0.435	1.010

Supplementary data Fig. 4 CO_2 fixation rates of WT and *a3* mature leaves in a light dependent manner. Data are means ± SD from three biological replicates.



Supplementary data Fig. 5 Signal intensity of *CRTR-B1* and *CRTR-B2* obtained from microarray data obtained from developing seeds. Stage 6, mid torpedo to late torpedo; Stage 7, late torpedo to early walking-stick; Stage 8, walking-stick to early curled cotyledon; Stage 9, curled cotyledon to early green cotyledon; Stage 10, green cotyledons.



BIBLIOGRAPHY

Adams, W.W., 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.

Akiyama, K. (2007). Chemical identification and functional analysis of apocarotenoids involved in the development of arbuscular mycorrhizal symbiosis. Bioscience, biotechnology, and biochemistry **71,** 1405-1414.

Alberte, R.S., and Andersen, R.A. (1986). Antheraxanthin, a Light Harvesting Carotenoid Found in a Chromophyte Alga. Plant Physiol **80,** 583-587.

Alfonso, M., Montoya, G., Cases, R., Rodriguez, R., and Picorel, R. (1994). Core antenna complexes, CP43 and CP47, of higher plant photosystem II. Spectral properties, pigment stoichiometry, and amino acid composition. Biochemistry **33,** 10494-10500.

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

Alvarez, V., Rodriguez-Saiz, M., de la Fuente, J.L., Gudina, E.J., Godio, R.P., Martin, J.F., and Barredo, J.L. (2006). The crtS gene of Xanthophyllomyces dendrorhous encodes a novel cytochrome-P450 hydroxylase involved in the conversion of beta-carotene into astaxanthin and other xanthophylls. Fungal Genet Biol **43,** 261-272.

Armstrong, G.A. (1997). Genetics of eubacterial carotenoid biosynthesis: A colorful tale. Annu. Rev. Microbiol. **51,** 629-659.

Auldridge, M.E., McCarty, D.R., and Klee, H.J. (2006a). Plant carotenoid cleavage oxygenases and their apocarotenoid products. Current opinion in plant biology **9,** 315-321.

Auldridge, M.E., Block, A., Vogel, J.T., Dabney-Smith, C., Mila, I., Bouzayen, M., Magallanes-Lundback, M., DellaPenna, D., McCarty, D.R., and Klee, H.J. (2006b). 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.

Barth, C., Krause, G.H., and Winter, K. (2001). Responses of photosystem I compared with photosystem II to high-light stress in tropical shade and sun leaves. Plant Cell and Environment **24,** 163-176.

Bassi, R., and Caffarri, S. (2000). Lhc proteins and the regulation of photosynthetic light harvesting function by xanthophylls. Photosynthesis research **64,** 243-256.

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. Biochemical and biophysical research communications **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 research **13,** 137-144.

Blankenship, R.E. (1992). Origin and early evolution of photosynthesis. Photosynthesis research **33,** 91-111.

Blasco, F., Kauffmann, I., and Schmid, R.D. (2004). CYP175A1 from Thermus thermophilus HB27, the first beta-carotene hydroxylase of the P450 superfamily. Applied microbiology and biotechnology **64,** 671-674.

Bouvier, F., Dogbo, O., and Camara, B. (2003). Biosynthesis of the food and cosmetic plant pigment bixin (annatto). Science (New York, N.Y **300,** 2089-2091.

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.). Biochimica et biophysica acta **1391,** 320-328.

Britton, G. (1998). Overview of Carotenoid Biosynthesis. In Carotenoids, G. Britton, S. Liaaen-Jensen, and H. Pfander, eds (Basel: Birkhäuser Verlag), pp. 30-83.

Burbulis, I.E., and Winkel-Shirley, B. (1999). Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. Proc Natl Acad Sci U S A **96,** 12929-12934.

Canutescu, A.A., Shelenkov, A.A., and Dunbrack, R.L., Jr. (2003). A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci **12,** 2001-2014.

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.

Chen, M., Bibby, T.S., Nield, J., Larkum, A.W., and Barber, J. (2005). Structure of a large photosystem II supercomplex from Acaryochloris marina. FEBS letters **579,** 1306-1310.

Cogdell, R.J., and Gardiner, A.T. (1993). Functions of carotenoids in photosynthesis.

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

Croce, R., Morosinotto, T., Castelletti, S., Breton, J., and Bassi, R. (2002). The Lhca antenna complexes of higher plants photosystem I. Biochim Biophys Acta **1556,** 29-40.

Cunningham, F.X., Jr., and Gantt, E. (2005). A study in scarlet: enzymes of ketocarotenoid biosynthesis in the flowers of Adonis aestivalis. Plant J **41,** 478-492.

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

Cunningham, F.X., Jr., Pogson, B., Sun, Z., McDonald, K.A., DellaPenna, D., and Gantt, E. (1996). Functional analysis of the beta and epsilon lycopene cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation. Plant Cell **8,** 1613-1626.

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 biology **6,** 32.

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

Demole, E., Enggist, P., Sauberli, U., and Stoll, M. (1970). Structure and synthesis of damascenone [2,6,6-trimethyl-1-*trans*-crotonyl-1,3-cyclohexadiene], odorous constituent in rose oil (*Rosa damascena* Mill.). Helv. Chim. Acta **53,** 541-551.

Dogbo, O., Laferriere, A., D'Harlingue, A., and Camara, B. (1988). Carotenoid biosynthesis: Isolation and characterization of a bifunctional enzyme catalyzing the synthesis of phytoene. Proceedings of the National Academy of Sciences of the United States of America **85,** 7054-7058.

Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., Boland, W., and Gershenzon, J. (2005). The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. Proceedings of the National Academy of Sciences of the United States of America **102,** 933-938.

Dutton, D.R., McMillen, S.K., Sonderfan, A.J., Thomas, P.E., and Parkinson, A.

(1987). Studies on the rate-determining factor in testosterone hydroxylation by rat liver microsomal cytochrome P-450: evidence against cytochrome P-450 isozyme:isozyme interactions. Arch Biochem Biophys **255,** 316-328.

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci **8,** 978-984.

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. The Plant cell **18,** 1947-1960.

Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003). 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics (Oxford, England) **19,** 1015-1018.

Giuliano, G., Giliberto, L., and Rosati, C. (2002). Carotenoid isomerase: a tale of light and isomers. Trends in plant science **7,** 427-429.

Glazko, G.V., and Nei, M. (2003). Estimation of divergence times for major lineages of primate species. Molecular biology and evolution **20,** 424-434.

Grabowski, B., Cunningham, F.X., Jr., and Gantt, E. (2001). Chlorophyll and carotenoid binding in a simple red algal light-harvesting complex crosses phylogenetic lines. Proceedings of the National Academy of Sciences of the United States of America **98,** 2911-2916.

Graham, S.E., and Peterson, J.A. (2002). Cytochrome P450. In Methods in Enzymology, E.F. Johnson and M.R. Waterman, eds (San Diego: Academic Press), pp. 15-28.

Gu, X. (2006). A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. Molecular biology and evolution **23,** 1937-1945.

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

Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K., and Hasegawa, M. (1994). Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, Giardia lamblia. Molecular biology and evolution **11,** 65-71.

Hess, W.R., Rocap, G., Ting, C.S., Larimer, F., Stilwagen, S., Lamerdin, J., and Chisholm, S.W. (2001). The photosynthetic apparatus of Prochlorococcus:

Insights through comparative genomics. Photosynthesis research **70,** 53-71.

Holm, L., and Sander, C. (1991). Database algorithm for generating protein backbone and side-chain co-ordinates from a C alpha trace application to model building and detection of co-ordinate errors. Journal of molecular biology **218,** 183-194.

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.

Holt, N.E., Zigmantas, D., Valkunas, L., Li, X.P., Niyogi, K.K., and Fleming, G.R. (2005). Carotenoid cation formation and the regulation of photosynthetic light harvesting. Science (New York, N.Y **307,** 433-436.

Howitt, C.A., and Pogson, B.J. (2006). Carotenoid accumulation and function in seeds and non-green tissues. Plant Cell Environ **29,** 435-445.

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

Johnson, E.A., and Schroeder, W.A. (1996). Microbial carotenoids. Advances in biochemical engineering/biotechnology **53,** 119-178.

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.

Kaminsky, L.S., and Guengerich, F.P. (1985). Cytochrome P-450 isozyme/isozyme functional interactions and NADPH-cytochrome P-450 reductase concentrations as factors in microsomal metabolism of warfarin. Eur J Biochem **149,** 479-489.

Kelley, R.W., Reed, J.R., and Backes, W.L. (2005). Effects of ionic strength on the functional interactions between CYP2B4 and CYP1A2. Biochemistry **44,** 2632-2641.

Kim, J., and DellaPenna, D. (2006). Defining the primary route for lutein synthesis in plants: the role of Arabidopsis carotenoid beta-ring hydroxylase CYP97A3. Proceedings of the National Academy of Sciences of the United States of America **103,** 3474-3479.

Klimmek, F., Ganeteg, U., Ihalainen, J.A., van Roon, H., Jensen, P.E., Scheller, H.V., Dekker, J.P., and Jansson, S. (2005). Structure of the higher plant light harvesting complex I: in vivo characterization and structural interdependence of the Lhca proteins. Biochemistry **44,** 3065-3073.

Koniger, M., Harris, G., 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.

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., Koroleva, O.Y., Dalling, J.W., and Winter, K. (2001). Acclimation of tropical tree seedlings to excessive light in simulated tree-fall gaps. Plant Cell and Environment **24,** 1345-1352.

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? Functional Plant Biology **31,** 743-756.

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

Kulheim, C., Agren, J., and Jansson, S. (2002). Rapid regulation of light harvesting and plant fitness in the field. Science (New York, N.Y **297,** 91-93.

Kull, D., and Pfander, H. (1995). Carotenoids. (Birkhauser, Basel).

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

Lee, A.I., and Thornber, J.P. (1995). Analysis of the pigment stoichiometry of pigment-protein complexes from barley (Hordeum vulgare). The xanthophyll cycle intermediates occur mainly in the light-harvesting complexes of photosystem I and photosystem II. Plant Physiol **107,** 565-574.

Leung, J., and Giraudat, J. (1998). Abscisic Acid Signal Transduction. Annu Rev Plant Physiol Plant Mol Biol **49,** 199-222.

Li, F., Murillo, C., and Wurtzel, E.T. (2007). Maize Y9 encodes a product essential for 15-cis-zeta-carotene isomerization. Plant Physiol **144,** 1181-1189.

Li, W.H. (1983). Evolution of duplicated genes. In Evolution of genes and proteins, M. Nei and R.K. Koehn, eds (Sunderland, MA: Sinauer associates), pp. 14-37.

Lichtenthaler, H.K. (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **50,** 47-65.

Liu, Z., Yan, H., Wang, K., Kuang, T., Zhang, J., Gui, L., An, X., and Chang, W.

(2004). Crystal structure of spinach major light-harvesting complex at 2.72 A resolution. Nature **428,** 287-292.

Livak, K.J. (1997). User Bulletin No. 2: ABI PRISM 7700 Sequence Detection System (Foster City, CA: PE Applied Biosystems), pp. 11-15.

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 (New York, N.Y **290,** 1151-1155.

Ma, Y.Z., Holt, N.E., Li, X.P., Niyogi, K.K., and Fleming, G.R. (2003). Evidence for direct carotenoid involvement in the regulation of photosynthetic light harvesting. Proceedings of the National Academy of Sciences of the United States of America **100,** 4377-4382.

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

Mark, M., Ghyselinck, N.B., and Chambon, P. (2006). Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. Annual review of pharmacology and toxicology **46,** 451-480.

Marquardt, J., and Hanelt, D. (2004a). Carotenoid composition of Delesseria lancifolia and other marine red algae from polar and temperate habitats. EUROPEAN JOURNAL OF PHYCOLOGY **39,** 285-292.

Marquardt, J., and Hanelt, D. (2004b). Carotenoid composition of Delesseria lancifolia and other marine red algae from polar and temperate habitats. European Journal ofPhycology
39, 285-292.

Matusova, R., Rani, K., Verstappen, F.W., Franssen, M.C., Beale, M.H., and Bouwmeester, H.J. (2005). The strigolactone germination stimulants of the plant-parasitic Striga and Orobanche spp. are derived from the carotenoid pathway. Plant Physiol **139,** 920-934.

Maxwell, K., and Johnson, G.N. (2000). Chlorophyll fluorescence--a practical guide. Journal of experimental botany **51,** 659-668.

Mestres, J. (2005). Structure conservation in cytochromes P450. Proteins **58,** 596-609.

Miller, S.R., Augustine, S., Olson, T.L., Blankenship, R.E., Selker, J., and Wood, A.M. (2005). Discovery of a free-living chlorophyll d-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene. Proceedings of the National Academy of Sciences of the United States of America **102,** 850-855.

Moore, R.C., and Purugganan, M.D. (2005). The evolutionary dynamics of plant duplicate genes. Current opinion in plant biology **8,** 122-128.

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. Molecular biology and evolution **20,** 1435-1447.

Nam, J., Kaufmann, K., Theissen, G., and Nei, M. (2005). A simple method for predicting the functional differentiation of duplicate genes and its application to MIKC-type MADS-box genes. Nucleic acids research **33,** e12.

Nei, M., and Rooney, A.P. (2005). Concerted and birth-and-death evolution of multigene families. Annual review of genetics **39,** 121-152.

Nelson, D.R., Zeldin, D.C., Hoffman, S.M., Maltais, L.J., Wain, H.M., and Nebert, D.W. (2004). Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. Pharmacogenetics **14,** 1-18.

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

Niyogi, K.K., Bjorkman, O., and Grossman, A.R. (1997). The roles of specific xanthophylls in photoprotection. Proceedings of the National Academy of Sciences of the United States of America **94,** 14162-14167.

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

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

Olson, V.A., and Owen, I.P.F. (1998). Costly sexual signals: are carotenoids rare, risky or required? Trends in Ecology and Evolution **13,** 510-514.

Palczewski, K. (2006). G protein-coupled receptor rhodopsin. Annual review of biochemistry **75,** 743-767.

Park, H., Kreunen, S.S., Cuttriss, A.J., DellaPenna, D., and Pogson, B.J. (2002). Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. The Plant cell **14,** 321-332.

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.

Pascal, A.A., Caron, L., Rousseau, B., Lapouge, K., Duval, J., and Robert, B. (1998). Resonance Raman spectroscopy of a light-harvesting protein from the brown alga Laminaria saccharina. Biochemistry **37,** 2450-2457.

Paulsen, H., Finkenzeller, B., and Kuhlein, N. (1993). Pigments induce folding of light-harvesting chlorophyll a/b-binding protein. European journal of biochemistry / FEBS **215,** 809-816.

Phillip, D., Hobe, S., Paulsen, H., Molnar, P., Hashimoto, H., and Young, A.J. (2002). The binding of Xanthophylls to the bulk light-harvesting complex of photosystem II of higher plants. A specific requirement for carotenoids with a 3-hydroxy-beta-end group. The Journal of biological chemistry **277,** 25160-25169.

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., and Rissler, H.M. (2000). Genetic manipulation of carotenoid biosynthesis and photoprotection. Philos Trans R Soc Lond B Biol Sci **355,** 1395-1403.

Pogson, B.J., Niyogi, K.K., Bjorkman, O., and DellaPenna, D. (1998). Altered xanthophyll compositions adversely affect chlorophyll accumulation and nonphotochemical quenching in Arabidopsis mutants. Proceedings of the National Academy of Sciences of the United States of America **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. Archives of biochemistry and biophysics **458,** 146-157.

Qureshi, N., and Porter, W. (1981). Biosynthesis of isoprenoid compounds. (New York: Wiley).

Rohmer, M. (1999). Comprehensive natural product chemistry. (Pergamon: Oxford).

Rohmer, M., Bouvier, P., and Ourisson, G. (1979). Molecular evolution of biomembranes: structural equivalents and phylogenetic precursors of sterols. Proceedings of the National Academy of Sciences of the United States of America **76,** 847-851.

Ruban, A.V., Lee, P.J., Wentworth, M., Young, A.J., and Horton, P. (1999). Determination of the stoichiometry and strength of binding of xanthophylls to the

photosystem II light harvesting complexes. The Journal of biological chemistry **274,** 10458-10465.

Rupasinghe, S., Baudry, J., and Schuler, M.A. (2003). Common active site architecture and binding strategy of four phenylpropanoid P450s from Arabidopsis thaliana as revealed by molecular modeling. Protein engineering **16,** 721-731.

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

Schagerl, M., and Pichler, C. (2000). Pigment composition of freshwater charophyceae. Aquatic Botany **67,** 117-129.

Schoch, G.A., Attias, R., Le Ret, M., and Werck-Reichhart, D. (2003). Key substrate recognition residues in the active site of a plant cytochrome P450, CYP73A1. Homology guided site-directed mutagenesis. European journal of biochemistry / FEBS **270,** 3684-3695.

Schoch, G.A., Yano, J.K., Wester, M.R., Griffin, K.J., Stout, C.D., and Johnson, E.F. (2004). Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. J Biol Chem **279,** 9497-9503.

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., Qin, X., and Loewen, M.C. (2004). The biochemical characterization of two carotenoid cleavage enzymes from Arabidopsis indicates that a carotenoid-derived compound inhibits lateral branching. The Journal of biological chemistry **279,** 46940-46945.

Scott, E.E., He, Y.A., Wester, M.R., White, M.A., Chin, C.C., Halpert, J.R., Johnson, E.F., and Stout, C.D. (2003). An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. Proc Natl Acad Sci U S A **100,** 13196-13201.

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. Molecular biology and evolution **22,** 2217-2230.

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

Straub, O. (1987). Key to Carotenoids. (Birkhauser, Basel).

Sun, Z., Gantt, E., and Cunningham, F.X., Jr. (1996). Cloning and functional analysis of the beta-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. (Sunderland, Massachusetts.: Sinauer Associates).

Taylor, J.S., and Raes, J. (2004). Duplication and divergence: the evolution of new genes and old ideas. Annual review of genetics **38,** 615-643.

Thayer, S., and Bjorkman, O. (1990). Leaf xanthophyll content and composition in sun and shade determined by HPLC. Photosynthesis research **23,** 331-343.

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

Tian, L., and DellaPenna, D. (2001). Characterization of a second carotenoid beta-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. Archives of biochemistry and biophysics **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 beta- and epsilon-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 epsilon-ring hydroxylation activity. Proceedings of the National Academy of Sciences of the United States of America **101,** 402-407.

Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R., Young, K., Taylor, N.E., Henikoff, J.G., Comai, L., and Henikoff, S. (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res **13,** 524-530.

Tomitani, A., Okada, K., Miyashita, H., Matthijs, H.C., Ohno, T., and Tanaka, A. (1999). Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. Nature **400,** 159-162.

Umeno, D., Tobias, A.V., and Arnold, F.H. (2005). Diversifying carotenoid biosynthetic pathways by directed evolution. Microbiol Mol Biol Rev **69,** 51-78.

Vanderauwera, S., Zimmermann, P., Rombauts, S., Vandenabeele, S., Langebartels, C., Gruissem, W., Inze, D., and Van Breusegem, F. (2005). Genome-wide analysis of hydrogen peroxide-regulated gene expression in Arabidopsis reveals a high light-induced transcriptional cluster involved in anthocyanin biosynthesis. Plant Physiol **139,** 806-821.

Vershinin, A. (1999). Biological functions of carotenoids--diversity and evolution. BioFactors (Oxford, England) **10,** 99-104.

Wagner, A. (2000). Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. Proceedings of the National Academy of Sciences of the United States of America **97,** 6579-6584.

Wang, X., Shi, X., Li, Z., Zhu, Q., Kong, L., Tang, W., Ge, S., and Luo, J. (2006). Statistical inference of chromosomal homology based on gene colinearity and applications to Arabidopsis and rice. BMC bioinformatics **7,** 447.

Wentworth, M., Ruban, A.V., and Horton, P. (2004). The functional significance of the monomeric and trimeric states of the photosystem II light harvesting complexes. Biochemistry **43,** 501-509.

Werck-Reichhart, D., Bak, S., and Paquette, S.M. (2002). Cytochromes P450. (Rockville, MD: American Society of Plant Biologists).

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

Wisniewska, A., Widomska, J., and Subczynski, W.K. (2006). Carotenoid-membrane interactions in liposomes: effect of dipolar, monopolar, and nonpolar carotenoids. Acta biochimica Polonica **53,** 475-484.

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

Yoshii, Y., Hanyuda, T., Wakana, I., Miyaji, K., Arai, S., Ueda, K., and Inouye, I. (2004). Carotenoid compositions of Cladophora balls (Aegagropila linnaei) and some members of the Cladophorales (Ulvophyceae, Chlorophyta): Their taxonomic and evolutionary implication. Journal of Phycology **40,** 1170-1177.

Young, A.J., and Frank, H.A. (1996). Energy transfer reactions involving carotenoids: quenching of chlorophyll fluorescence. Journal of photochemistry and photobiology **36,** 3-15.

Zhang, J. (2003). Evolution by gene duplication: an update. Trends in Ecology and Evolution **18,** 292-297.