Received Date : 03-Oct-2014
Revised Date : 27-Feb-2015
Accepted Date : 06-Mar-2015
Article type : Original Article

A lycopene β-cyclase/lycopene ε-cyclase/light-harvesting complex-fusion protein from the green alga *Ostreococcus lucimarinus* can be modified to produce α-carotene and β-carotene at different ratios

Andreas Blatt*, Matthias E. Bauch*, Yvonne Pörschke*, Martin Lohr
Institut für Allgemeine Botanik, Johannes Gutenberg-Universität, 55099 Mainz, Germany

* These authors contributed equally to the work

Corresponding author:
Dr. Martin Lohr
Institut für Allgemeine Botanik, Johannes Gutenberg-Universität, 55099 Mainz, Germany
Tel.: +49-6131-3924201; Fax: +49-6131-3923075; E-Mail: Lohr@uni-mainz.de

E-mail addresses of coauthors:
Andreas Blatt: aphidina@web.de
Dr. Matthias Bauch: matthias@drbauch.com
Yvonne Poerschke: leinchen@gmx.de

Running Title: Lycopene cyclase-fusion protein from Ostreococcus

Keywords: *Ostreococcus lucimarinus*, Mamiellales, carotenoid biosynthesis, lycopene cyclase, *Chlamydomonas reinhardtii*, LHC, *Bigelowiella natans*

Summary

Biosynthesis of asymmetric carotenoids such as α-carotene and lutein in plants and green algae involves the two enzymes lycopene β-cyclase (LCYB) and lycopene ε-cyclase (LCYE). The two cyclases are closely related and probably resulted from an ancient gene

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/tpj.12826

duplication. While in most plants investigated so far the two cyclases are encoded by separate genes, prasinophyte algae of the order Mamiellales contain a single gene encoding a fusion protein comprised of LCYB, LCYE and a C-terminal light-harvesting complex (LHC) domain. Here we show that the lycopene cyclase-fusion protein from *Ostreococcus lucimarinus* catalyzed the simultaneous formation of α-carotene and β-carotene when heterologously expressed in *Escherichia coli*. The stoichiometry of the two products in *E. coli* could be altered by gradual truncation of the C-terminus, suggesting that the LHC domain may be involved in modulating the relative activities of the two cyclase domains in the algae. Partial deletions of the linker region between the cyclase domains or replacement of one or both cyclase domains with the corresponding cyclases from the green alga *Chlamydomonas reinhardtii* resulted in pronounced shifts of the α-carotene-to-β-carotene ratio, indicating that both the relative activities of the cyclase domains and the overall structure of the fusion protein have a strong impact on the product stoichiometry. The possibility to tune the product ratio of the lycopene cyclase-fusion protein from Mamiellales renders it useful for the biotechnological production of the asymmetric carotenoids α-carotene or lutein in bacteria or fungi.

Introduction

Carotenoids are essential components in the photosynthetic machineries of all oxygenic phototrophs, fulfilling crucial functions in photoprotection by quenching of triplet chlorophylls and scavenging of various reactive oxygen species. In addition, they serve as accessory light-harvesting pigments in the photosynthetic antenna complexes (Goodwin, 1980; Siefermann-Harms, 1987; Telfer et al., 2008; Ballottari et al., 2012; Jahns and Holzwarth, 2012). Carotenoids are isoprenoids consisting of a C_{40}-backbone generated by condensation of eight isoprene units (Cunningham and Gantt, 1998). Most carotenoids have a cyclic structure at both ends, and such dicyclic carotenoids are ubiquitous in oxygenic phototrophs (Goodwin, 1980;

Young, 1993; Jeffrey et al., 2011). Several types of rings can be formed from the linear ends, namely β-, ε-, and γ-ionone rings, and aromatic rings derived from the β-ring (Britton, 1998). The most prominent types are the β-ring, present e.g. at both ends of β-carotene, and the ε- ring, differing from the β-ring only in the position of the double bond (Figure 1).

β-carotene and xanthophylls derived thereof are ubiquitous and occur in the photosynthetic apparatus of all oxygenic phototrophs (Goodwin, 1980; Young, 1993; Telfer et al., 2008). Carotenoids with ε-ionone rings are common to all Viridiplantae, certain groups of cyanobacteria and various red algae. In addition, they have been found in species of several algal groups with secondary plastids such as chlorarachniophytes, haptophytes and cryptophytes (Jeffrey et al., 2011; Lohr, 2011; Takaichi, 2011).

In land plants and green algae (Viridiplantae), the biosynthetic pathway of carotenoids splits at the level of lycopene (Figure 1) into the α- and the β-branch (Cunningham and Gantt, 1998; Lohr, 2011). The introdution of a β-ionone ring at one end and an ε-ionone ring at the other end of lycopene yields α-carotene, whereas the formation of β-ionone rings at both ends yields β-carotene. Consequently, all α-branch carotenoids contain a β-ionone and an ε-ionone ring, while the β-branch carotenoids contain two β-ionone rings.

Formation of the β-rings is catalyzed by lycopene β-cyclase (LCYB) and that of the ε-ring by lycopene ε-cyclase (LCYE). Both enzymes are closely related and likely resulted from gene duplication during early evolution of Viridiplantae (Krubasik and Sandmann, 2000). As the lipophilic nature of their substrate lycopene suggests, lycopene cyclases are membrane-associated enzymes (Camara and Dogbo, 1986, Beyer et al., 1991) and contain two hydrophobic regions that may form transmembrane helical domains (Cunningham et al., 1996). LCYB is able to introduce β-rings at both ends of lycopene and at the linear end of the monocyclic carotenoids

δ-carotene and γ-carotene. LCYE, with rare exceptions (Cunningham and Gantt, 2001), generally does not accept monocyclic carotenoids as substrates, but catalyzes the formation of a single ε-ring at one end of lycopene (Cunningham et al., 1996; Yu and Beyer, 2012). Therefore, LCYB alone is sufficient for the formation of β-carotene via the intermediate γ-carotene, whereas both LCYE and LCYB are necessary for the formation of α-carotene via δ-carotene (Figure 1).

α-carotenoids and β-carotenoids are unevenly distributed in the various pigment-protein-complexes comprising the photosynthetic machinery of plants and algae. The α-carotenoid lutein is only present in the peripheral light-harvesting antennae of the two photosystems, with a higher concentration in LHCII than in LHCI (Schmid, 2008). The β-carotenoids neoxanthin and violaxanthin also are present only in the peripheral light-harvesting complexes, with neoxanthin limited to LHCII (Schmid, 2008). β-Carotene, on the other hand, is present only in the inner antennae of both photosystems (Green and Durnford, 1996). Under light limitation, plants increase the number of LHC units per photosystem. Consequently, the relative demand of α- and β-carotenoids varies depending on the photoacclimation state (Ballottari et al., 2007). As the two cyclases are encoded by separate genes in most Viridiplantae, the flow of substrate between the two branches can be adjusted by varying the expression levels of the cyclases (Cunningham et al., 1996). This has been experimentally confirmed for various plants (Pogson et al., 1996; Ronen et al., 1999; Harjes et al., 2008; Bai et al., 2009; Giorio et al., 2013).

In the genomes of two prasinophytes of the order Mamiellales, however, both cyclases were predicted to be encoded by a single gene and expressed as a fusion protein that in addition contains a C-terminal domain with high similarity to light-harvesting-proteins of the chlorophyll *a*/*b*-binding protein family (Derelle et al., 2006; Palenik et al., 2007; Cunningham et al., 2007) (Figure 2a). As a consequence of the gene fusion, both cyclases have a fixed stoichiometry of

1:1 and their relative activities cannot be adjusted by differential expression. Therefore, these prasinophyte algae likely evolved other mechanisms to adjust the flux through the two branches to changing demands. In fact, light acclimation experiments with the two Mamiellales *Ostreococcus tauri* and *Micromonas pusilla* demonstrated a decrease of the ratio of α-carotenoids to β-carotenoids from about 2-2.5 to almost 1 with increasing growth irradiances (Laviale and Neveux, 2011) suggesting a corresponding adjustment of the lycopene cyclase activities in the algae. The mechanism by which the activities of the fused cyclases are modulated is unknown. It may be speculated that the quite unusual fusion of an LHC sequence to the two carotenoid biosynthesis enzymes is involved in this modulation.

A better understanding of the presumptive mechanisms modulating the activities of the two cyclase domains in the fusion protein from Mamiellales may also be beneficial for biotechnological applications. The α-carotenoid lutein, for example, has been implicated in preventing age-related macular degeneration in humans, and therefore gained increasing interest as food supplement (Kijlstra et al., 2012). Chemical synthesis of the highly asymmetric lutein is difficult and results in low yields (Mayer and Rüttimann, 1980; Khachik and Chang, 2009), and currently a major source of the lutein used as food supplement are the petals of marygold (*Tagetes erecta*) (Ferndandez-Sevilla et al., 2010). The efficient biosynthesis of lutein in a heterologous host would require a well-balanced activity of the two cyclases to maximize the yield of α-carotene over β-carotene.

In order to learn more about the unusual lycopene cyclase-LHC fusion protein, we have cloned the corresponding gene of the prasinophyte *O. lucimarinus* and performed a detailed investigation of the catalytic activity of the encoded fusion protein by heterologous expression in carotenogenic strains of *E. coli*. Key questions were (i) if the cyclase fusion protein is capable of catalyzing the parallel formation of α- and β-carotene, (ii) if the LHC domain or the fusion of

the two cyclase domains have become essential for catalytic activity, (iii) given that the fusion enzyme generates α- and β-carotene concurrently, whether there is a way to alter the stoichiometry of the two products and which mechanisms may be involved, (iv) how replacement of either of the two cyclase domains in the fusion protein with the corresponding enzyme from an alga with separate cyclase genes may alter its performance, and (v) if introduction of the cyclase fusion protein into already available carotenogenic plasmids may enable the constitutive synthesis of α-carotene in *E. coli*.

Results

The lycopene cyclase fusion is conserved in Mamiellales and a Chlorarachniophyte

We first investigated the occurrence of the cyclase fusion among the complete genome sequences publicly available from green algae, with a particular focus on the Mamiellales. BLAST searches of the genomes from the core chlorophytes *C. reinhardtii* and *Volvox carteri* and from the Trebouxiophytes *Chlorella subellipsiodea* and *Coccomyxa variabilis* confirmed that LCYB and LCYE are encoded by separate genes in all four genomes. On the other hand, the fusion of LCYB, LCYE and the LHC domain was conserved in all six available genomes from Mamiellales, namely in three species of the genus *Ostreococcus*, two species of *Micromonas* and in *Bathycoccus prasinos*. Based on the genome sequences and protein alignments, a careful analysis and manual editing of the *LCY* fusion gene models from all six algae suggested that the gene from *Micromonas pusilla* CCMP1545 contains a single intron whereas the genes from the other five Mamiellales lack introns.

We also included the genome of the chlorarachniophyte alga *Bigelowiella natans* in our searches. Chlorarachniophytes acquired their plastid by the endosymbiotic uptake of a green alga-like organism and contain α-carotenoids, suggesting the presence of genes for both LCYB and LCYE. Indeed, we found the two cyclase genes next to each other, and one of the gene

models in the JGI genome browser predicts them to be fused in a single mRNA. The predicted gene model contains 16 introns, but does not encode a C-terminal LHC-domain. In the genomic region downstream of the fused *LCY* genes we could not find any indication of a nucleotide sequence encoding such a domain, further supporting the absence of a LHC domain. Nevertheless, phylogenetic analyses (Figure 2b) supported a close affiliation of the cyclase fusion protein from *B. natans* with the fusion proteins from Mamiellales, because both the β-cyclase and the ε-cyclase domains from these algae clustered to the exclusion of the separate cyclases from other green algae.

In silico *analyses of the lycopene cyclase fusion proteins from Mamiellales*

To predict length and borders of the functional domains, we compared the sequences of the fusion proteins from the six Mamiellales with the corresponding separate proteins from *C. reinhardtii* and *Arabidopsis thaliana*. The first (N-proximal) cyclase domain in the fusion protein is highly similar to the separate LCYB, while the second cyclase domain is closely related to the LCYE from other Viridiplantae. The LHC domain is most similar to the minor LHC proteins CP24 (Lhcb6) and CP29 (Lhcb4; Figure 2c). All amino acids known to be involved in binding of chlorophylls and carotenoids in CP24, CP29 and related LHC proteins (Ballotari et al., 2012) are conserved in the LHC domain (Figure S1), although at position 1359 in helix C a glutamate residue conserved in most LHC proteins has been replaced by a histidine residue.

We used the LCY sequence from *O. lucimarinus* (OluLCY) as reference and, based on a detailed comparison of the aligned sequences, assigned the LCYB domain from amino acid (aa) 123 to 560 (438 aa), the LCYE domain from aa 642 to 1107 (466 aa) and the LHC domain from aa 1241 to 1442 (202 aa) (Figure 2a). At the N-terminus, a putative transit peptide of 73 aa length for import of the nucleus-encoded protein into the plastid was predicted by ChloroP

(Emanuelsson et al., 2007). Based on our domain assignment, the LCYB and LCYE domains are fused by a linker of about 80 aa (S1 in Figure 2a) and the LCYE and LHC by a linker of about 135 aa (S2 in Figure 2a). Although significant parts of the linker regions were well conserved between the fusion proteins from the different Mamiellales (Figures S1 and S2), they had no similarity to other proteins in the NCBI protein database and did not contain any functionally conserved motifs (no significant Pfam matches with E-values set at 10 (http://pfam.sanger.ac.uk/search)). A conspicuous feature found in linker region S1 of all six proteins, however, was the enrichment of proline residues (7-9 proline residues per 80 aa), while linker region S2 contained one or two stretches with lengths of 5-15 amino acids composed of glycine, serine and alanine residues (Figure S2).

In dinoflagellates and euglenophytes, two phylogenetically unrelated algal groups with secondary plastids, some nucleus-encoded plastid proteins such as light-harvesting apoproteins are expressed as polyprotein precursors that are cleaved into the individual units upon import into the plastids (Houlné and Schantz, 1988; Rikin and Schwartzbach, 1988; Hiller et al., 1995). In these cases, the subunits are linked by a conserved motif of 5-10 amino acids length containing one (dinoflagellates) or two (euglenophytes) specific cleavage sites (Muchhal and Schwartzbach, 1992; Hiller et al., 1995; Koziol et al., 2008; Boldt et al., 2012). The linker regions S1 and S2 in the cyclase fusion proteins from the five Mamiellales, however, do not share any obvious motif indicative of a conserved cleavage site. Moreover, *in silico* analyses of the cyclase fusion proteins from Mamiellales with TargetP (Emanuelsson et al., 2007) consistently predicted the presence of a transit peptide with a corresponding cleavage site at the N-terminus, while analyses of the linker regions provided no evidence for the presence of transit peptide-like sequences. An analysis of the secondary structure of the cyclase fusion proteins from the five Mamiellales using HMMTOP, TMpred and TopPred predicted the presence of the three canonical transmembrane α-helices in the LHC domain and of two potential

transmembrane regions in each of the cyclase domains which were also detected in the separate cyclases from *A. thaliana* and *C. reinhardtii* (Figure S1 and Table S3). The S2 linker regions in the LCY proteins from *O. lucimarinus* and the two *Micromonas* species contain a hydrophobic stretch predicted to represent another potentially membrane spanning α-helix, albeit with lower support. Analyses of the fusion protein from from *O. lucimarinus* with HMM-TM, MEMSAT-SVM and DAS-TMfilter, however, predicted only the transmembrane helices in the LHC domain with confidence, and TMHMM found no transmembrane domain at all (Table S3).

Parallel formation of α- and β-carotene by the lycopene cyclase fusion protein from **Ostreococcus**

For functional analyses, we cloned the *OluLCY* gene encoding the cyclase fusion protein from genomic DNA of *O. lucimarinus*. All functional analyses were performed by heterologous expression of the gene or parts thereof in lycopene-accumulating strains of *E. coli*. We used the pBAD-TOPO® plasmid for arabinose-inducible expression of the *OluLCY* genes and co-transformed *E. coli* TOP10 cells with each pBAD-construct and plasmid pLYCO2 leading to constitutive accumulation of lycopene in the bacteria (see Experimental Procedures for further details). Liquid cultures of the transformants were induced over night at 28 °C, harvested and the pigment content analyzed by HPLC. Reproducibility was checked by assaying three different clones from each transformation.

Expression of the nearly full-length *OluLCY* gene (shortened only by the first 49 amino acids of the predicted transit peptide) in the lycopene-accumulating bacteria resulted in parallel formation of α-carotene and β-carotene with a molar ratio of about 1.4 (Figure 3a). Separation of the two cyclase domains did not noticeably compromise their catalytic activity. Expression of only the β-cyclase domain (OluLCYB) yielded the dicyclic all-*trans*-β-carotene as major product and some *cis*-isomers of β-carotene as minor products (Figure 3b), whereas expression

of the ε-cyclase domain (OluLCYE) yielded the monocyclic δ-carotene (Figure 3c). When comparing the product specificities of the two isolated domains with those of the separate cyclases CrLCYB and CrLCYE from *C. reinhardtii* (Figure 3b,c) we found no difference.

The length of the C-terminus of the lycopene cyclase fusion protein affects the ratio between the products α- and β-carotene

Next, we removed the LHC domain of the fusion protein (196 C-terminal aa) by truncating the *OluLCY* gene at the 3'end. Expression of the resulting partial gene *OluLCYΔ196* in lycopene-accumulating bacteria led to a reproducible shift of the product stoichiometry towards α-carotene (Figure 4). Obviously, the LHC domain is not essential for catalytic activity, but its presence affects the relative activities of the cyclase domains at least in *E. coli*.

To analyse in more detail a potential influence of the C-terminus on the stoichiometry of α- and β-carotene, we generated five additional mutant genes (*OluLCYΔ228* - *OluLCYΔ312*, Figure 4) in which the linker region S2 was gradually shortened by stepwise removal of 20-30 amino acids. The functional assays of these constructs in *E. coli* revealed a pronounced effect of the C-terminus on the relative amounts of α-carotene and β-carotene accumulating in the bacteria (Figure 4). The maximum percentage of α-carotene was synthesized by proteins with a C-terminal truncation of 248 or 272 amino acids corresponding to the LHC domain and about half of the linker S2. Further truncation resulted in a reversal of the shift.

Deletions in the linker region between the lycopene cyclase domains also affect the ratio between α- and β-carotene

Based on the observation that the relative activity of the two cyclase domains is sensitive to the nature of the amino acid sequence behind the second domain, we speculated that the sequence of linker S1 between the cyclase domains may have a similar influence on the product

ratio. Using the truncated protein OluLCYΔ312 (already lacking the LHC domain and the linker S2) as reference, we generated three mutant proteins with varying deletions in linker region S1 and mutant in which we flipped the order of the two cyclase domains.

The corresponding genes were assembled from smaller DNA modules by a restriction enzyme-based strategy (see Experimental Procedures and Figure S3 for details). Introduction of the restriction sites resulted in some amino acid exchanges affecting, however, only amino acids not conserved in the six sequences from Mamiellales. To assess the effect of these exchanges, we assayed the construct *B-S-E* assembled from separate gene fragments encoding the LCYB domain (B), the linker S1 (S) and the LCYE domain (E). The protein encoded by *B-S-E* differed from that encoded by *OluLCYΔ312* in four amino acids, namely F562S (S is present at this position in the protein from *B. prasinos*, T in *O. tauri* and both *Micromonas* species), E630S (S present in proteins from *Ostreococcus* spec. RCC899 and both *Micromonas* species), and the two additional amino acids P and R at the C-terminus. As shown in Figure 5, the product ratios of α-carotene and β-carotene in *E. coli* were similar for both constructs.

The mutant protein B-Sa-Sb-E lacked only a short motif containing two conserved proline residues (IPXPX) located in the central part of the linker (aa 593-597). Deletion of these five amino acids resulted in halving of the α-carotene/β-carotene ratio from about 3 to below 1.5 (Figure 5). Deletion of the C-proximal half of linker S1 (aa 593-630) in mutant B-Sa-E changed the ratio in the opposite direction by doubling it from 3 to 6.

Deletion of the N-proximal half of linker S1 (aa 562-597) in mutant B-Sb-E resulted in a complete loss of β-carotene formation and the accumulation of significant amounts of δ-carotene besides α-carotene. The exchange of the two cyclase domains in mutant E-S-B had a very similar effect on the product composition with the percentage of δ-carotene being even

higher. Notably, expression of module B alone also resulted in the conversion of only half of the lycopene to β-carotene, while expression of the bipartite module B-Sa resulted in almost complete conversion of lycopene (Figure S4). These observations indicate that the N-proximal half (Sa) of the region we had defined as linker S1 based on sequence comparisons contains some amino acids important for full catalytic activity of the β-cyclase domain. Therefore, the results for the latter two mutants are inconclusive. The results for the mutants B-Sa-Sb-E and B-Sa-E, however, support the idea that the sequence in the linker region between the two cyclase domains also affects their relative activities.

Chimeric lycopene cyclase fusion proteins show a low efficiency of α-carotene formation

The observed susceptibility of the activities of the two cyclase domains to mutations in the linker regions already suggested that the OluLCY protein is not a result of just putting together two cyclases, but that the fusion protein as a whole went through an evolutionary optimization process. We explored this issue further by replacing either one or both cyclase domains in the fusion protein with the corresponding cyclase(s) from *C. reinhardtii*. The hybrid genes were generated by the modular assembly strategy used before (see Experimental Procedures and Figure S3). We combined five gene modules encoding the LCYB domain (B), LCYE domain (E) and the linker S1 (S) of the fusion protein from *O. lucimarinus*, and the LCYB (cb) and LCYE (ce) from *C. reinhardtii*. Before fusing the modules, we verified catalytic activity for each of the cyclase modules by separate expression in lycopene-accumulating *E. coli* (Figure S4). We assayed the activities of three combinations maintaining the original order of the two cyclases (cb-S-E, B-S-ce, cb-S-ce), and three more combinations with the order of the cyclases reverted (E-S-cb, ce-S-B, ce-S-cb). With the exception of ce-S-B, all hybrid proteins produced mainly β-carotene and only very little or no α-carotene (Figure 6). This observation indicates that the β-cyclase domain in these hybrid proteins has a much higher catalytic activity than the

respective ε-cyclase domain. The hybrid ce-S-B was the only protein with a higher relative activity of the ε-cyclase domain. Expression of ce-S-B in the lycopene-accumulating *E. coli* resulted in a product ratio of about 40% α-carotene, 40% δ-carotene and 20% β-carotene. Thus, 80% of the products contained an ε-ionone ring. However, only about 50 % of the substrate lycopene were converted to cyclic carotenoids, indicating that either both cyclase domains in this hybrid protein have a low catalytic activity or that the protein accumulates in *E. coli* at a significantly lower level than the other hybrids (e.g. due to misfolding and subsequent degradation). Taken together, the results further supported our assumption that the two cyclase domains need to have finely coordinated activities to convert lycopene efficiently into α-carotene.

Expanding the plasmid portfolio for color complementation assays in **E. coli** *by constitutive expression of the cyclase fusion protein*

E. coli strains genetically engineered to produce carotenoids have become a valuable tool for the functional characterization of enzymes involved in the bacterial and eukaryotic carotenoid metabolism (Cunningham and Gantt, 2007). By this method, also known as color complementation assay, a number of key enzymes of carotenoid biosynthesis from plants and algae have been identified and further characterized (Cunningham and Gantt, 2007). Moreover, there is growing interest in the biotechnological production of carotenoids in *E. coli* (Misawa, 1998; Immethun et al., 2013). For these applications, a plethora of plasmids have been designed containing different combinations of constitutively expressed genes from bacteria and plants that enable the accumulation of various carotenoids in *E. coli*. (Cunningham and Gantt, 2007). However, to the best of our knowledge, no single plasmid containing the genes for all enzymes necessary for the formation of α-carotene has been reported.

Therefore, we used the cyclase fusion gene from *O. lucimarinus* for engineering two constructs that enable the constitutive accumulation of α-carotene in *E. coli*. As starting point, we used plasmid pBETA2 that we derived from plasmid pACCAR25ΔcrtX (Kajiwara et al., 1995) by inactivating the *crtZ* gene (see Experimental Procedures for full details). We excised a fragment containing *crtY* (encoding a bacterial lycopene β-cyclase) and most of the defective *crtX* gene and replaced it with the truncated gene *OluLCYΔ312*, resulting in plasmid pALPHA1 (Figure 7a). The Shine-Dalgarno sequence of *crtX* was used to drive translation of the alga-derived gene from the polycistronic mRNA. Due to our cloning strategy, the resulting OluLCY protein differed slightly at the N-terminus from the corresponding protein generated by the pBAD construct. The N-terminal extension of 15 amino acids introduced by the pBAD vector (MGSGSGDDDDKLAL) was replaced by the first 8 amino acids encoded by *crtX* and two more amino acids introduced by the restriction site for AvrII (MSHFAAIALG). Nevertheless, *E. coli* cells containing pALPHA1 accumulated α-carotene and β-carotene with a ratio of about 3:1 (Figure 7b), thus essentially matching the product ratio obtained by inducible expression of the corresponding gene from the pBAD vector (Figure 4). Moreover, although the algal *OluLCY* gene was introduced without prior codon optimization for expression in *E. coli*, the bacteria accumulated only minor amounts of the intermediates lycopene and δ-carotene.

For generation of pALPHA2 (Figure 7c), we replaced the 5'-end of the OluLCYD312 gene with the 5'-end of OluLCYD272 by double digest of pALPHA1 with KpnI and SnaBI and insertion of a KpnI-MmeI fragment from pBAD-OluLCYΔ272. *E. coli* cells transformed with pALPHA2 accumulated almost exclusively α-carotene and only small amounts of δ-carotene and β-carotene (Figure 7d).

Discussion

Is the fusion ancestral to LCYE evolution or were LCYB and LCYE fused secondarily?

As has been noted before (Cunningham et al., 1996, Krubasik and Sandmann, 2000, Cunningham et al., 2007, Chan et al., 2012), the high sequence similarity of the LCYB and LCYE proteins from Viridiplantae and their phylogenetic relation to lycopene cyclases from other phototrophs argues for their origin by duplication of a lycopene cyclase gene in the common ancestor of all Viridiplantae. The tandem arrangement of the two cyclases in the fusion protein from Mamiellales may have resulted from unequal crossing over which is regarded as the main mechanism responsible for gene duplication (Graur and Li, 2000). In this case, the fusion would represent the ancestral state of the gene duplication that led to LCYB and LCYE in Viridiplantae (Cunningham et al., 2007), and the gene fragments encoding the two cyclase domains would have been separated only after they had evolved their differing cyclase activities. In line with this hypothesis, the Mamiellales are one of the most basal clades in the green lineage (Turmel et al., 2009, Marin and Melkonian, 2010). Alternatively, the two cyclases may have evolved as separate genes and later became fused only in the Mamiellales. To decide between the two alternative scenarios it will be necessary to increase taxon sampling among extant prasinophytes.

Based on the close phylogenetic relation of the LCYB and LCYE domains in the cyclase fusion protein from the chlorarachniophyte *B. natans* to the corresponding cyclase domains from prasinophytes (Figure 2b), *B. natans* likely recruited the fusion gene from a prasinophyte-like alga. Phylogenetic analyses of plastid-localized proteins from green algae and chlorarachniophytes indicate, however, that the latter acquired their plastids by endosymbiotic uptake of an alga closely related to core chlorophytes such as the genus *Chlorella* (Rogers et al., 2007, Turmel et al., 2009). This suggests that *B. natans* replaced its original *LCY* genes with a

cyclase fusion gene acquired by lateral transfer from a prasinophyte alga. A significant contribution of lateral gene transfer to the gene inventory of *B. natans* has already been well documented by phylogenetic analyses of transcriptome and genome data from this alga (Archibald et al., 2003; Curtis et al., 2012). As the sequences of the LCYB and LCYE domain from *B. natans* are positioned basal to the clusters containing the corresponding cyclases from Mamiellales, the cyclase fusion may have been acquired early in the evolution of chlorarachniophytes. Again, investigation of this hypothesis demands an increased taxon sampling among prasinophytes and chlorarachniophytes.

As already noted, dinophyte and euglenophyte algae synthesize some nucleus-encoded plastid proteins as polyprotein precursors containing conserved cleavage sites for separation of the individual units upon import into the plastids (Muchhal and Schwartzbach, 1992; Hiller et al., 1995; Koziol et al., 2008; Boldt et al., 2012). Here, we found no evidence for any conserved cleavage motifs in the linker regions of the fusion proteins from the five Mamiellales. Moreover, the genomes of *O. lucimarinus* and other Mamiellales have been found to contain a conspicuously high number of gene fusions encoding proteins involved in the same metabolic processes (Derelle et al., 2006; Palenik et al., 2007; Worden et al., 2009). In light of these observations, we have no reason to doubt that the fusion of the two lycopene cyclases and the LHC domain is maintained in the mature protein. For final proof, however, a native fusion protein will have to be isolated from *O. lucimarinus* or one of its relatives.

How does the cyclase fusion protein catalyze the formation of both α- and β-carotene and which mechanisms may modulate the product ratio?

The fusion of genes resulting in the expression of two or more proteins as a single polypeptide is a recurring theme in the molecular evolution of organisms (Marcotte et al., 1999, Enright and Ouzounis, 2001, Kummerfeld and Teichmann, 2005). Often the fusion involves

proteins participating in the same metabolic pathway and even catalyzing consecutive reaction steps (Marcotte et al., 1999; Enright et al., 1999), thereby supporting efficient substrate channeling (Miles et al., 1999, Jørgensen et al., 2005, Zhang, 2011). Different from other known fusion proteins, however, the two cyclase domains in the LCY fusion compete for the same substrate and constitute a gateway for channeling carotenoids into the two different branches of the carotenoid biosynthetic pathway in green algae and land plants.

In principle, the cyclase domains could act on the two linear ends of lycopene either consecutively or simultaneously. The latter option would need the enzymes to be in close proximity or even to form dimers. Indeed, the separate cyclases present in land plants and most other green algae have been suggested to form dimers, with homodimers of LCYB catalyzing the formation of β-carotene, and heterodimers of LCYB and LCYE synthesizing α-carotene (Cunningham and Gantt, 1998, 2001). Experimental evidence from *in vitro* assays, however, supports a successive mode of lycopene cyclization (Yu and Beyer, 2012). LCYB from rice (Oryza sativa) was shown to accept not only lycopene but also the monocyclic δ-carotene as substrate, with the latter giving rise to α-carotene. LCYE, on the other hand, was not able to introduce an ε-ionone ring in the monocyclic γ-carotene (Yu and Beyer, 2012), suggesting the enzyme to sense the configuration of the complete molecule including the end opposite to the end that is cyclized. The necessary physical contact between LYCE and both ends of the substrate probably would prevent a parallel action of LCYB and LCYE on the same lycopene molecule. Therefore, we favor the hypothesis that the two cyclase domains catalyze consecutive independent reactions, with the monocyclic product first dissociating from the enzyme and then either turning around and binding to the same domain or entering the catalytic site of a neighboring domain.

In this scenario, the ratio of the products α- and β-carotene in *E. coli* as well as in the algae would essentially be determined by the relative accessibilities of the competing catalytic centers in the fusion protein for lycopene (Figure 1), and the linker regions would play a major role in positioning the domains to each other, thereby controlling substrate access to the catalytic centers. In agreement with this conclusion, the putative linker regions between the conserved cyclase and LHC domains are longer than reported for linker regions between catalytic domains in other fusion proteins (Chen et al., 2013, and references therein). Moreover, they contain some conspicuous amino acid motifs – the proline rich motif in linker S1 and the stretch of glycine, serine and alanine residues in linker S2 (Figure S2). As proline residues tend to increase the rigidity of a peptide (Chen et al., 2013), the proline-rich motif in linker S1 may serve as a rigid spacer between the cyclase domains thereby controlling access of the substrate to the LCYB domain. Shortening of the linker by 32 amino acids resulted in a strongly reduced formation of β-carotene (Figure 5) indicating a lower activity of the LCYB domain which may result from restricted substrate access to this domain due to steric hindrance by the adjacent LCYE domain. Partial deletion of the proline-rich motif, on the other hand, favored the formation of β-carotene over α-carotene (Figure 5). The deletion should result in increased conformational flexibility of the linker which may alleviate the proposed constrained access of the substrate to the LCYB domain.

In linker S2, the stretches composed of glycine, serine and alanine residues (Figure S2) may serve as flexible hinges between the second cyclase domain and the LHC domain. Poly-G-S stretches interspersed with alanine residues have been shown to introduce a higher conformational mobility in linker peptides (Xiang et al., 2004, Wang et al., 2010, Haga et al., 2013). The increased formation of α-carotene by truncated fusion proteins with a length between 1170 aa (mutant Δ272 in Figure 4) and 1214 aa (Δ228) may be related to an enrichment of charged residues in the regions from aa 1164 to 1181 and aa 1196 to 1211 and the loss of a

hydrophobic stretch between aa 1212 and 1230 in these mutants (see Figure S1). Assuming that the cyclase domains are membrane-localized, the hydrophobic region that is still present in mutant Δ196 may restrict access of lycopene to the catalytic center of the LCYE domain by extending into the membrane.

A potential role of the LHC domain in modulating the activity of the two cyclase domains in Ostreococcus

Although the *E. coli* system is limited by the lack of chlorophylls that would be necessary for proper folding of the LHC domain in the fusion protein, the pronounced effect of the various C-terminal truncations on the product ratio in *E. coli* indicates a potential mechanism of how in *Ostreococcus* the ratio of α-carotene and β-carotene may be adjusted to varying physiological demands. According to this model, the LHC domain may serve as a gate controlling substrate access to the catalytic center of the LCYE domain. As one possibility, the spatial arrangement of the LCYE and LHC domain could be modified by a conformational change of the latter due to an altered pigment composition. In this way, the LHC domain may be involved in sensing the availability of α-carotenoids directly, or it may probe the availability of Chl b for the assembly of peripheral antennae which in turn would necessitate the synthesis of more α-carotenoids. In line with the latter hypothesis, both CP29 and CP24 have been shown to possess a higher plasticity in binding of Chl a versus Chl b than the major LHCII (Giuffra et al., 1996; Pagano et al., 1998).

Alternatively, substrate access to the LCYE domain may be altered by binding of the LHC domain to other pigment-protein-complexes in the thylakoid membrane. Our phylogenetic analyses (Figure 2c) indicate that the LHC domain in the cyclase fusion protein is most closely related to the minor LHC proteins CP24 (Lhcb6) and CP29 (Lhcb4) which in the photosystem II supercomplexes of land plants connect the peripheral LHCII trimer M with CP47 of the core

complex (Ballotari et al., 2012). CP24 is absent from green algae and probably evolved during the transition of plants from aquatic habitats to land (Ballotari et al., 2012). If the LHC domain of the cyclase fusion protein would participate in the presumable photosystem supercomplexes of Mamiellales in a similar fashion to CP24 in land plants, then it may monitor the ratio of its binding partners (i.e. the ratio of core complexes to peripheral LHCII trimers) and an unoccupied binding site for LHCII may favor formation of α-carotenoids.

Although our in-silico analyses indicated that the LHC domain is able to bind both chlorophylls and carotenoids, this needs to be confirmed by isolation of the native cyclase fusion protein. Further studies will be necessary to clarify if the pigment composition of the LHC domain is variable and if the LCY-LHC-fusion protein is part of a photosystem supercomplex in the thylakoid membranes of Mamiellales.

What kind of selective pressure may be involved in maintaining the cyclase fusion in Mamiellales?

The most straightforward explanation for conservation of the cyclase fusion in Mamiellales would be that the fusion has become crucial for catalytic activity of the cyclase domains. The successful functional expression of the separate cyclase domains in *E. coli* (Figure 3), however, argues against this hypothesis. Also, the LHC domain is dispensable for the formation of α-carotene and β-carotene at least in *E. coli*. Still, if the cyclase fusion protein is part of a larger complex of carotenogenic enzymes, parts of the linker regions may be involved in protein-protein interactions necessary for complex stabilization.

The prasinophyte genomes analyzed so far are rather compact, and a striking frequency of gene fusions has been reported for *O. tauri*, *O. lucimarinus* and species of the genus *Micromonas* (Derelle et al., 2006, Palenik et al., 2007, Worden et al., 2009). This observation

suggests that gene fusions in general are under positive selection in the small genomes of prasinophytes. Moreover, the need to maintain the delicate balance in the activities of the two cyclase domains may exert a selective pressure that so far has prevented their physical separation into independent polypeptides in Mamiellales.

Finally, the Mamiellales have been shown to contain several α-carotenoids not present in other green algae (Egeland et al., 1997; Wilhelm et al., 1997; Latasa et al., 2004) (Figure 1). These carotenoids are part of the LHC antennae and help increasing the absorbance efficiency of green light (Böhme et al., 2002; Six et al., 2005). The enzymes involved in the biosynthesis of these carotenoids from α-carotene or lutein are unknown. A potential mechanistic or physiological link between the cyclase fusion and the formation of these carotenoids, however, is not obvious. Before speculating further on such a link, the general co-occurrence of these two traits should be validated by sampling of additional prasinophyte taxons.

Potential biotechnological applications of the lycopene cyclase fusion protein from **Ostreococcus**

Plasmids pALPHA1 and pALPHA2 add to the growing number of carotenogenic plasmids that are useful not only in basic research on the characterization of carotenoid biosynthetic genes by heterologous functional expression, but also for the biotechnological production of carotenoids. Plasmid pALPHA1 induced the parallel accumulation of α- and β-carotene in the bacteria. Such a plasmid may be useful for a direct assessment in *E. coli* of the specificity of carotenoid biosynthetic enzymes accepting both carotenes as potential substrates. Examples for such enzymes are carotene hydroxylases and carotene ketolases (Lohr, 2011).

Transformation of *E. coli* with pALPHA2, on the other hand, resulted in almost exclusive accumulation of α-carotene. The chemical synthesis of asymmetric carotenoids with a β-ionone and an ε-ionone ring such as α-carotene and lutein is rather laborious with low yields (Karrer and Eugster, 1950; Mayer and Rüttimann, 1980; Khachik and Chang, 2009). The biosynthesis of α-carotene in *E. coli* has already been reported in several publications (Cunningham et al., 1996, 2007; Stickforth et al., 2003; Li et al., 2009). In all instances, however, the bacteria were co-transformed with two plasmids and still accumulated considerable amounts of β-carotene as byproducts.

The fixed ratio of the two cyclases in the fusion protein from Mamiellales offers the advantage of a well-defined relative activity of the two enzymes, resulting in a more reproducible product stoichiometry than expression of the two cyclases from separate genes. In principle, the same could be achieved by fusing an *LCYB* and an *LCYE* gene from any other plant or green algal source. Our attempts to achieve this by using the *LCYB* and *LCYE* genes from *C. reinhardtii*, however, were not successful. Either the yield of α-carotene compared to β-carotene was very low or α-carotene accumulated together with δ-carotene, illustrating that the activities of the two cyclases in the fusion protein need to be well-balanced to result in specific formation of α-carotene. Another major advantage of the fusion protein from *O. lucimarinus* is the possibility of tailoring the ratio of its products α-carotene and β-carotene by changing the length of the C-terminus (Figure 4). Therefore, the cyclase fusion protein may become a valuable tool for the biotechnological production of α-carotenoids in *E. coli* or other hosts more suitable for the accumulation of high amounts of carotenoids.

Experimental Procedures

Bioinformatic analyses

The sequence of the cyclase fusion protein from *O. lucimarinus* (JGI Genome Portal, http://genome.jgi.doe.gov/; *O. lucimarinus* genome browser, gene model fgenesh1_pg.C_Chr_20000053; Palenik et al., 2007) was used to search the following databases by BLAST (Altschul et al., 1997), (i) JGI Genome Portal: *B. natans* CCMP2755 (Curtis et al., 2012), *Chlorella variabilis* NC64A (Blanc et al., 2010),*Coccomyxa subellipsoidea* C-169 (Blanc et al., 2012), *M. pusilla* CCMP1545 and *Micromonas sp*. RCC299 (Worden et al., 2009), *O. tauri* (Derelle et al., 2006), *Ostreococcus sp.* RCC809, and *V. carteri* (Prochnik et al., 2010); (ii) Genomes at Bioinformatics Gent (http://bioinformatics.psb.ugent.be/genomes/): *B. prasinos* (Moreau et al., 2012); and (iii) the NCBI protein database (http://www.ncbi.nlm.nih.gov/).

The phylogenetic analysis of lycopene cyclases was based on the alignment in Chan et al. (2012). Database accession codes of additional sequences and sequences used for the phylogenetic analysis of the LHC domain are compiled in Table S1. As the LHC domain had not been included in the phylogenetic analyses of the LHC proteins from *O. tauri* by Six et al. (2005), we first constructed a phylogenetic tree from a larger data set including all LHC sequences from *A. thaliana*, *C. reinhardtii* and *O. tauri*. For the final tree, we used only the sequences of a major branch containing the LHC domains of the cyclase fusion proteins from Mamiellales, thereby enabling the inclusion of additional amino acid positions conserved specifically in this subset of sequences.

Protein alignments were constructed using ClustalW 1.83 (Chenna et al., 2003). Manual refinement and selection of unambiguously aligned amino acid positions were made using BioEdit 5.0.9 (Hall, 1999). Maximum likelihood (ML) trees were inferred from the protein alignments using PThreads in RAxML 8.0.14 (Stamatakis, 2014) and the WAG substitution

model (Whelan and Goldman, 2001) with gamma rate distribution ('PROTGAMMAWAG', four discrete rate categories. Branch support was estimated with 100 bootstrap replicates using the rapid bootstrap algorithm (Stamatakis et al., 2008) implemented in PThreads. Posterior probabilities for nodes in the RAxML trees were calculated with MrBayes 3.1.1 (Huelsenbeck and Ronquist, 2001) as described in Chan et al. (2012).

Generation of expression constructs for OluLCY, CrLCYB and CrLCYE

The *OluLCY* gene devoid of the first 49 predicted codons was cloned from genomic DNA of *O. lucimarinus* into pBAD-TOPO® (Life Technologies, https://www.lifetechnologies.com) by PCR with the GC RICH PCR system (Roche, http://www.roche.com) using the primers specified in Table S2. Correct orientation and in-frame localization of the gene was confirmed by sequencing (Starseq, www.starseq.com). This yielded plasmid pBAD_OluLCY#2rep containing a *OluLCY* gene with a single point mutation resulting in the amino acid exchange A74V. A74 is the first amino acid after the predicted cleavage site of the putative transit peptide and is distant to the first conserved protein domain starting at position 123, making negative effects on enzyme activity unlikely.

Using the PCR primers specified in Table S2, the *LCYB* and *LCYE* genes of *C. reinhardtii* were amplified from EST clones AV641959 and AV643150, respectively (Asamizu et al., 2004; Lohr et al., 2005), and the gradually truncated versions of the *OluLCY* gene were generated using pBAD_OluLCY#2rep as template. The PCR products were cloned into pBAD-TOPO® and their correct orientation and in-frame localization confirmed by sequencing.

The chimeric gene fusions were generated by a modular assembly strategy. Seven DNA modules were amplified by PCR introducing restriction sites at the 5'- and 3'-end of each module for either one of the enzymes AvrII, NheI or SpeI all of which generate compatible 5'

overhangs (see Table S2 for primer sequences and added restrictions sites). PCR products were cloned into pBAD-TOPO® and their correct orientation and in-frame localization confirmed by sequencing. The resulting modules were sequentially assembled in pBAD according to **Figure S3** into various combinations of a β-cyclase module and an ε-cyclase module connected by a linker module.

Functional expression of pBAD cyclase constructs in carotenogenic E. coli

For construction of plasmid pLYCO2, the *crtZ* gene in plasmid pACCARΔ25crtX (Kajiwara et al., 1995) was excised by double digest with AatII and NsiI, the overhangs filled using Phusion® High-Fidelity DNA Polymerase (NEB; https://www.neb.com/) and religated yielding plasmid pBETA2. Double digest of pBETA2 with AsiSI and SnaBI removed crtX and crtY, and filling of overhangs and religation yielded plasmid pLYCO2. For functional analyses, *E. coli* TOP10 were co-transformed with pLYCO2 and either of the various pBAD-LCY constructs. Double transformants were grown in LB medium under induction with 0.2 % arabinose and selection with chloramphenicol and ampicillin at 28 °C and 250 rpm to an OD600 between 0.5 and 1.0. Then, 10 ml culture per sample were harvested by centrifugation (5,000xg, 4 min, 4 °C), the supernatant completely removed and the cell pellet stored at -20 °C until further analysis.

Generation of plasmid pALPHA1

For construction of pALPHA1, we replaced *crtX* and *crtY* in pBETA2 by *OluLCYΔ312*. Digestion of pBETA2 with AsiSI and SnaBI would allow excision of both genes with only the ribosome binding site and the first ten codons of *crtX* remaining, but the *OluLCY* gene also contained an AsiSI site. Therefore, pBETA2 was first digested with AsiSI and a short multiple cloning site introduced by ligation with a duplex of the self-complementary (underlined) oligonucleotide CGC-CCTAGG-TACGTA-CCTAGG-GCGAT composed of the restriction site

sequence AsiSI(5'cut)-AvrII-SnaBI-AvrII-AsiSI(3'cut). The resulting plasmid was digested with SnaBI to excise *crtX* and *crtY* and then religated yielding pLYCO3-MCS. Next, the *OluLCY* gene shortened by the last 312 codons (*OluLCYΔ312*) was amplified by PCR with the GC RICH PCR system using pBAD_OluLCY#2rep as template and the primers specified in Table S2, cloned into pGEM® T-Easy (Promega, http://www.promega.com) and its correctness confirmed by sequencing. After plasmid amplification, the *OluLCYΔ312* gene was excised by digestion with AvrII, inserted between the AvrII sites of pLYCO3-MCS, and the resulting *E. coli* transformants screened for clones with *OluLCYΔ312* inserted in the same orientation as the flanking genes *crtE* and *crtI* (Figure 7). The resulting plasmid was termed pALPHA1.

Extraction and analysis of carotenoids from E. coli

Pigments were extracted from bacterial cell pellets by successive addition of 200 µl methanol, 200 µl acetone and 200 µl dichloromethane with intermittent vortexing and sonication in a Transsonic T460 water bath (Elma, http://www.elma-ultrasonic.com). Samples were centrifuged (11,000xg, 4 min, 4 °C) and the supernatant used for HPLC analysis. HPLC analyses were performed on a Waters Alliance 2795 Separation Module equipped with a Waters 2996 photodiode-array detector and the data analyzed using Waters Empower software (Waters, http://www.waters.com). Pigments were separated on a ProntoSIL 200-5 C30, 5.0 µm, 250 x 4.6 mm column equipped with a ProntoSIL 200-5-C30, 5.0 µm, 20 x 4.0 mm guard column (Bischoff Analysentechnik, http://www.bischoff-chrom.de/) at 28 °C applying a ternary gradient (flow 1.3 ml/min) with linear changes from 10 % eluent A (85 % methanol, 0.075 M ammonium acetate) and 90 % eluent B (90 % acetonitrile) at 0 min to 100 % B at 0.75 min, 60 % B and 40 % C (ethylacetate) at 2.5 min, 50 % B and 50 % C at 8 min, 10 % B and 90 % C at 23 min, 10 % B and 90 % C at 26 min, 100 % C at 26.5 min, 100 % C at 30 min, 100 % B at 32 min, and 10 % A and 90 % B at 33 min.

For pigment identification, we used as reference pigments β-carotene from Calbiochem (Merck, http://www.merckmillipore.com) and α-carotene prepared from a purchased mixture of carotene isomers from carrots (Sigma-Aldrich, http://www.sigmaaldrich.com). δ-carotene, ε-carotene and γ-carotene were identified by HPLC-*online*-absorbance spectra using data in Britton et al. (2004) as reference. For pigment quantification, the following molar extinction coefficients [units of L mol^{-1} cm^{-1}] were used (Britton, 1995): α-carotene, $\varepsilon_{445\ nm}$ = 145300; β-carotene, $\varepsilon_{452\ nm}$ = 138900; δ-carotene, $\varepsilon_{456\ nm}$ = 176300; ε-carotene, $\varepsilon_{440\ nm}$ = 155400; γ-carotene, $\varepsilon_{462\ nm}$ = 147900; and lycopene, $\varepsilon_{470\ nm}$ = 184900.

Acknowledgements

We thank Evelyne Derelle (Observatoire Océanologique de Banyuls, France) for providing DNA from *O. lucimarinus*; the Kazusa DNA Research Institute (Chiba, Japan) for providing *C. reinhardtii* EST clones; Johannes von Lintig (Case Western Reserve University, Cleveland, USA) for plasmid pACCAR25ΔcrtX; Jeannette Plew, Jennifer Elbers, Cecilia Perera, Ute Lehmann and André Busch (Universität Mainz) for valuable help in some of the experiments; and Harald Paulsen, Stephan Hobe, Oliver Dautermann (all Universität Mainz) and two anonymous reviewers for helpful comments and constructive criticism of the manuscript. This work was supported by a research grant from the Forschungsfonds Universität Mainz (to M.L.) and PhD scholarships from the Stipendienstiftung Rheinland-Pfalz (to M.B.) and the Carl-Zeiss-Stiftung (to A.B.).

Supporting Information

Figure S1. Sequence conservation between LCY fusion proteins from Mamiellales and corresponding proteins from *B. natans*, *C. reinhardtii* and *A. thaliana*

Figure S2. Sequence comparison of linker regions S1 and S2 in LCY fusion proteins from Mamiellales

Figure S3. Modular assembly strategy used for generation of mutant cyclase fusion proteins

Figure S4. Catalytic activity of modules used for assembly of mutant cyclase fusion proteins

Table S1. Database accessions of sequences used for the phylogenetic analyses shown in Fig. 2b and 2c.

Table S2. PCR primers, templates and DNA fragments amplified and cloned into pBAD-TOPO

Table S3. Prediction of potential transmembrane helices in the cyclase fusion proteins from Mamiellales and the lycopene cyclases from *A. thaliana* and *C. reinhardtii*.

References

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

Archibald, J.M., Rogers, M.B., Toop, M., Ishida, K. and Keeling, P.J. (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7678-7683.

Asamizu, E., Nakamura, Y., Miura, K., Fukuzawa, H., Fujiwara, S., Hirono, M., Iwamoto, K., Matsuda, Y., Minagawa, J., Shimogawara, K., Takahashi, Y. and Tabata, S. (2004) Establishment of publicly available cDNA material and information resource of *Chlamydomonas reinhardtii* (Chlorophyta) to facilitate gene function analysis. *Phycologia* **43**, 722-726.

Bai, L., Kim, E.H., DellaPenna, D. and Brutnell, T.P. (2009) Novel lycopene epsilon cyclase activities in maize revealed through perturbation of carotenoid biosynthesis. *Plant J.* **59**, 588-599.

Ballottari, M., Dall'Osto, L., Morosinotto, T. and Bassi, R. (2007) Contrasting behavior of higher plant photosystem I and II antenna systems during acclimation. *J. Biol. Chem.* **282**, 8947-8958.

Ballottari, M., Girardon, J., Dall'Osto, L. and Bassi, R. (2012) Evolution and functional properties of photosystem II light harvesting complexes in eukaryotes. *Biochim. Biophys. Acta* **1817**, 143-157.

Beyer, P., Kroncke, U. and Nievelstein, V. (1991) On the mechanism of the lycopene

isomerase/cyclase reaction in *Narcissus pseudonarcissus* L. chromoplasts. *J. Biol. Chem.* **266**, 17072-17078.

Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D.D., Gurnon, J., Ladunga, I., Lindquist, E., Lucas, S., Pangilinan, J., Proschold, T., Salamov, A., Schmutz, J., Weeks, D., Yamada, T., Lomsadze, A., Borodovsky, M., Claverie, J.M., Grigoriev, I.V. and Van Etten, J.L. (2012) The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol.* **13**, R39.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., Salamov, A., Terry, A., Yamada, T., Dunigan, D.D., Grigoriev, I.V., Claverie, J.M. and Van Etten, J.L. (2010) The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* **22**, 2943-2955.

Böhme, K., Wilhelm, C. and Goss, R. (2002) Light regulation of carotenoid biosynthesis in the prasinophycean alga *Mantoniella squamata*. *Photochem. Photobiol. Sci.* **1**, 619-628.

Boldt, L., Yellowlees, D., and Leggat, W. (2012) Hyperdiversity of genes encoding integral light-harvesting proteins in the dinoflagellate *Symbiodinium* sp. *PLoS ONE* **7**, e47456.

Britton, G. (1995) UV/Visible Spectroscopy. In *Carotenoids, Vol. 1B, Spectroscopy* (Britton, G., Liaaen-Jensen, S. and Pfander, H. eds). Basel, Boston, Berlin: Birkhäuser, pp. 13-62.

Britton, G. (1998) Overview of Carotenoid Biosynthesis. In *Carotenoids, Vol. 3, Biosynthesis and Metabolism* (Britton, G., Liaaen-Jensen, S. and Pfander, H. eds). Basel, Boston, Berlin: Birkhäuser, pp. 13-147.

Britton, G., Liaaen-Jensen, S., Pfander, H., Mercadante, A.Z. and Egeland, E.S. (2004) *Carotenoids. Handbook* Basel, Boston, Berlin: Birkhäuser.

Bülow, L. and Mosbach, K. (1991) Multienzyme systems obtained by gene fusion. *Trends Biotechnol.* **9**, 226-231.

Camara, B. and Dogbo, O. (1986) Demonstration and solubilization of lycopene cyclase from *Capsicum* chromoplast cembranes. *Plant Physiol.* **80**, 172-174.

Chan, C.X., Blouin, N.A., Zhuang, Y.Y., Zauner, S., Prochnik, S.E., Lindquist, E., Lin, S.J., Benning, C., Lohr, M., Yarish, C., Gantt, E., Grossman, A.R., Lu, S., Muller, K., Stiller, J.W., Brawley, S.H. and Bhattacharya, D. (2012) Porphyra (Bangiophyceae) transcriptomes provide insights into red algal development and metabolism. *J. Phycol.* **48**, 1328-1342.

Chen, X.Y., Zaro, J.L. and Shen, W.C. (2013) Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliver. Rev.* **64**, 1357-1369.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. (2003) Multiple sequence alignment with the Clustal series of programs.

Nucl. Acids Res. **31**, 3497-3500.

Conrado, R.J., Varner, J.D. and DeLisa, M.P. (2008) Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr. Opin. Biotechnol.* **19**, 492-499.

Cunningham, F.X. and Gantt, E. (1998) Genes and enzymes of carotenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 557-583.

Cunningham, F.X. and Gantt, E. (2001) One ring or two? Determination of ring number in carotenoids by lycopene ε-cyclases. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2905-2910.

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

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

Curtis, B.A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M.C., Ball, S.G., Gile, G.H., Hirakawa, Y., Hopkins, J.F., Kuo, A., Rensing, S.A., Schmutz, J., Symeonidi, A., Elias, M., Eveleigh, R.J.M., Herman, E.K., Klute, M.J., Nakayama, T., Obornik, M., Reyes-Prieto, A., Armbrust, E.V., Aves, S.J., Beiko, R.G., Coutinho, P., Dacks, J.B., Durnford, D.G., Fast, N.M., Green, B.R., Grisdale, C.J., Hempel, F., Henrissat, B., Hoppner, M.P., Ishida, K.I., Kim, E., Koreny, L.K., Kroth, P.G., Liu, Y., Malik, S.B., Maier, U.G., McRose, D., Mock, T., Neilson, J.A.D., Onodera, N.T., Poole, A.M., Pritham, E.J., Richards, T.A., Rocap, G., Roy, S.W., Sarai, C., Schaack, S., Shirato, S., Slamovits, C.H., Spencer, D.F., Suzuki, S., Worden, A.Z., Zauner, S., Barry, K., Bell, C., Bharti, A.K., Crow, J.A., Grimwood, J., Kramer, R., Lindquist, E., Lucas, S., Salamov, A., McFadden, G.I., Lane, C.E., Keeling, P.J., Gray, M.W., Grigoriev, I.V. and Archibald, J.M. (2012) Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**, 59-65.

Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A.Z., Robbens, S., Partensky, F., Degroeve, S., Echeynie, S., Cooke, R., Saeys, Y., Wuyts, J., Jabbari, K., Bowler, C., Panaud, O., Piegu, B., Ball, S.G., Ral, J.P., Bouget, F.Y., Piganeau, G., De Baets, B., Picard, A., Delseny, M., Demaille, J., Van de Peer, Y. and Moreau, H. (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11647-11652.

Egeland, E.S., Guillard, R.R.L. and Liaaen-Jensen, S. (1997) Additional carotenoid prototype representatives and a general chemosystematic evaluation of carotenoids in Prasinophyceae (Chlorophyta). *Phytochemistry* **44**, 1087-1097.

Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**, 953-971.

Enright, A.J., Iliopoulos, I., Kyrpides, N.C. and Ouzounis, C.A. (1999) Protein interaction

maps for complete genomes based on gene fusion events. *Nature* **402**, 86-90.

Enright, A.J. and Ouzounis, C.A. (2001) Functional associations of proteins in entire genomes by means of exhaustive detection of gene fusions. *Genome Biol.* **2**, research00341-0034.7.

Fernandez-Sevilla, J.M., Fernandez, F.G.A. and Grima, E.M. (2010) Biotechnological production of lutein and its applications. *Appl. Microbiol. Biotechnol.* **86**, 27-40.

Giorio, G., Yildirim, A., Stigliani, A.L. and D'Ambrosio, C. (2013) Elevation of lutein content in tomato: A biochemical tug-of-war between lycopene cyclases. *Metab. Eng.* **20**, 167-176.

Giuffra, E., Cugini, D., Croce, R. and Bassi, R. (1996) Reconstitution and pigment-binding properties of recombinant CP29. *Eur. J. Biochem.* **238**, 112-120.

Goodwin, T.W. (1980) *The Biochemistry of the Carotenoids, Vol. I: Plants; 2nd ed.* London, New York: Chapman & Hall.

Graur, D. and Li, W.-H. (2000) *Fundamentals of Molecular Evolution* 2nd. ed.: Sinauer Associates.

Green, B.R. and Durnford, D.G. (1996) The chlorophyll-carotenoid proteins of oxygenic photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 685-714.

Haga, T., Hirakawa, H. and Nagamune, T. (2013) Fine tuning of spatial arrangement of enzymes in a PCNA-mediated multienzyme complex using a rigid poly-L-proline linker. *PloS ONE* **8**, e75114.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Harjes, C.E., Rocheford, T.R., Bai, L., Brutnell, T.P., Kandianis, C.B., Sowinski, S.G., Stapleton, A.E., Vallabhaneni, R., Williams, M., Wurtzel, E.T., Yan, J.B. and Buckler, E.S. (2008) Natural genetic variation in *lycopene epsilon cyclase* tapped for maize biofortification. *Science* **319**, 330-333.

Hiller, R. G., Wrench, P.M. and Sharples, F.P. (1995) The light-harvesting chlorophyll *a-c*-binding protein of dinoflagellates: a putative polyprotein. *FEBS Lett.* **363**, 175-178.

Houlné, G. and Schantz, R. (1988) Characterization of eDNA sequences for LHCI apoproteins in *Euglena gracilis*: The mRNA encodes a large precursor containing several consecutive divergent polypeptides. *Mol. Gen. Genet.* **213**, 479-486.

Huelsenbeck, J.P. and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Immethun, C.M., Hoynes-O'Connor, A.G., Balassy, A. and Moon, T.S. (2013) Microbial production of isoprenoids enabled by synthetic biology. *Front. Microbiol.* **4**, 75.

Jahns, P. and Holzwarth, A.R. (2012) The role of the xanthophyll cycle and of lutein in

photoprotection of photosystem II. *Biochim. Biophys. Acta* **1817**, 182-193.

Jeffrey, S.W., Wright, S.W. and Zapata, M. (2011) Micoralgal classes and their signature pigments. In *Phytoplankton Pigments. Characterization, Chemotaxonomy and Applications in Oceanography* (Roy, S., Llewellyn, C.A., Egeland, E.S. and Johnsen, G. eds). Cambridge: Cambridge University Press, pp. 3-77.

Jørgensen, K., Rasmussen, A.V., Morant, M., Nielsen, A.H., Bjarnholt, N., Zagrobelny, M., Bak, S. and Møller, B.L. (2005) Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* **8**, 280-291.

Kajiwara, S., Kakizono, T., Saito, T., Kondo, K., Ohtani, T., Nishio, N., Nagai, S. and Misawa, N. (1995) Isolation and functional identification of a novel cDNA for astaxanthin biosynthesis from *Haematococcus pluvialis*, and astaxanthin synthesis in *Escherichia coli*. *Plant Mol. Biol.* **29**, 343-352.

Karrer, P. and Eugster, C.H. (1950) Synthese von Carotinoiden V. Gleichzeitige synthetische Bildung von ε_1-Carotin, β-Carotin und *d,l*-α-Carotin. *Helv. Chim. Acta* **33**, 1952-1954.

Khachik, F. and Chang, A.N. (2009) Total synthesis of (3R,3'R,6'R)-lutein and its stereoisomers. *J. Org. Chem.* **74**, 3875-3885.

Kijlstra, A., Tian, Y., Kelly, E.R. and Berendschot, T.T.J.M. (2012) Lutein: More than just a filter for blue light. *Prog. Retin. Eye Res.* **31**, 303-315.

Koziol, A.G. and Durnford, D.G. (2008) *Euglena* light-harvesting complexes are encoded by multifarious polyprotein mRNAs that evolve in concert. *Mol. Biol. Evol.* **25**, 92-100.

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

Kummerfeld, S.K. and Teichmann, S.A. (2005) Relative rates of gene fusion and fission in multi-domain proteins. *Trends Genet.* **21**, 25-30.

Latasa, M., Scharek, R., Le Gall, F. and Guillou, L. (2004) Pigment suites and taxonomic groups in Prasinophyceae. *J. Phycol.* **40**, 1149-1155.

Laviale, M. and Neveux, J. (2011) Relationships between pigment ratios and growth irradiance in 11 marine phytoplankton species. *Mar. Ecol. Prog. Ser.* **425**, 63-U386.

Li, Z.R., Ahn, T.K., Avenson, T.J., Ballottari, M., Cruz, J.A., Kramer, D.M., Bassi, R., Fleming, G.R., Keasling, J.D. and Niyogi, K.K. (2009) Lutein accumulation in the absence of zeaxanthin restores nonphotochemical quenching in the *Arabidopsis thaliana npq1* mutant. *Plant Cell* **21**, 1798-1812.

Lohr, M. (2011) Carotenoid metabolism in phytoplankton. In *Phytoplankton Pigments. Characterization, Chemotaxonomy and Applications in Oceanography* (Roy, S., Llewellyn, C.A., Egeland, E.S. and Johnsen, G. eds). Cambridge: Cambridge University Press, pp. 113-161.

Lohr, M., Im, C.S. and Grossman, A.R. (2005) Genome-based examination of chlorophyll and carotenoid biosynthesis in *Chlamydomonas reinhardtii*. *Plant Physiol.* **138**, 490-515.

Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O. and Eisenberg, D. (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science* **285**, 751-753.

Marin, B. and Melkonian, M. (2010) Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. *Protist* **161**, 304-336.

Mayer, H. and Rüttimann, A. (1980) Synthesis of optically active natural carotenoids and structurally related compounds IV. Synthesis of (3R,3'R,6'R)-lutein. *Helv. Chim. Acta* **63**, 1451-1455.

Miles, E.W., Rhee, S. and Davies, D.R. (1999) The molecular basis of substrate channeling. *J. Biol. Chem.* **274**, 12193-12196.

Misawa, N. and Shimada, H. (1998) Metabolic engineering for the production of carotenoids in non-carotenogenic bacteria and yeasts. *J. Biotechnol.* **59**, 169-181.

Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel, M., Poulain, J., Katinka, M., Hohmann-Marriott, M.F., Piganeau, G., Rouze, P., Da Silva, C., Wincker, P., Van de Peer, Y. and Vandepoele, K. (2012) Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol.* **13**, R74.

Muchhal, U.S. and Schwartzbach, S.D. (1992) Characterization of a *Euglena* gene encoding a polyprotein precursor to the light-harvesting chlorophyll *a/b*-binding protein of photosystem II. *Plant Mol. Biol.* **18**, 287-299.

Pagano, A., Cinque, G. and Bassi, R. (1998) *In vitro* reconstitution of the recombinant photosystem II light-harvesting complex CP24 and its spectroscopic characterization. *J. Biol. Chem.* **273**, 17154-17165.

Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., Zhou, K.M., Otillar, R., Merchant, S.S., Podell, S., Gaasterland, T., Napoli, C., Gendler, K., Manuell, A., Tai, V., Vallon, O., Piganeau, G., Jancek, S., Heijde, M., Jabbari, K., Bowler, C., Lohr, M., Robbens, S., Werner, G., Dubchak, I., Pazour, G.J., Ren, Q.H., Paulsen, I., Delwiche, C., Schmutz, J., Rokhsar, D., Van de Peer, Y., Moreau, H. and Grigoriev, I.V. (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7705-7710.

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.

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., Hellsten, U., Chapman, J., Simakov, O., Rensing, S.A., Terry, A., Pangilinan, J., Kapitonov, V., Jurka, J., Salamov, A., Shapiro, H., Schmutz, J., Grimwood, J., Lindquist, E., Lucas, S., Grigoriev, I.V., Schmitt, R., Kirk, D. and Rokhsar, D.S. (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* **329**, 223-226.

Rikin, A. and Schwartzbach S.D. (1988) Extremely large and slowly processed precursors to the *Euglena* light-harvesting chlorophyll a/b binding proteins of photosystem II. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5117-5121.

Rogers, M.B., Gilson, P.R., Su, V., McFadden, G.I. and Keeling, P.J. (2007) The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: Evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol. Biol. Evol.* **24**, 54-62.

Ronen, G., Cohen, M., Zamir, D. and Hirschberg, J. (1999) Regulation of carotenoid biosynthesis during tomato fruit development: Expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. *Plant J.* **17**, 341-351.

Schmid, V.H.R. (2008) Light-harvesting complexes of vascular plants. *Cell. Mol. Life Sci.* **65**, 3619-3639.

Schoffelen, S. and van Hest, J.C.M. (2012) Multi-enzyme systems: bringing enzymes together in vitro. *Soft Matter* **8**, 1736-1746.

Siefermann-Harms, D. (1987) The light-harvesting and protective functions of carotenoids in photosynthetic membranes. *Physiol. Plantarum* **69**, 561-568.

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

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313.

Stamatakis, A., Hoover, P. and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57**, 758-771.

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

Takaichi, S. (2011) Carotenoids in algae: distributions, biosyntheses and functions. *Mar. Drugs* **9**, 1101-1118.

Telfer, A., Pascal, A. and Gall, A. (2008) Carotenoids in photosynthesis. In *Carotenoids, Vol.*

4, Natutal Functions (Britton, G., Liaaen-Jensen, S. and Pfander, H. eds). Basel, Boston, Berlin: Birkhäuser, pp. 265-308.

Turmel, M., Gagnon, M.C., O'Kelly, C.J., Otis, C. and Lemieux, C. (2009) The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol. Biol. Evol.* **26**, 631-648.

Wang, R.L., Xue, Y.M., Wu, X.X., Song, X.F. and Peng, J.J. (2010) Enhancement of engineered trifunctional enzyme by optimizing linker peptides for degradation of agricultural by-products. *Enzyme Microb. Technol.* **47**, 194-199.

Whelan, S. and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691-699.

Wilhelm, C., Kolz, S., Meyer, M., Schmitt, A., Zuber, H., Egeland, E.S. and LiaaenJensen, S. (1997) Refined carotenoid analysis of the major light-harvesting complex of *Mantoniella squamata*. *Photosynthetica* **33**, 161-171.

Worden, A.Z., Lee, J.H., Mock, T., Rouze, P., Simmons, M.P., Aerts, A.L., Allen, A.E., Cuvelier, M.L., Derelle, E., Everett, M.V., Foulon, E., Grimwood, J., Gundlach, H., Henrissat, B., Napoli, C., McDonald, S.M., Parker, M.S., Rombauts, S., Salamov, A., Von Dassow, P., Badger, J.H., Coutinho, P.M., Demir, E., Dubchak, I., Gentemann, C., Eikrem, W., Gready, J.E., John, U., Lanier, W., Lindquist, E.A., Lucas, S., Mayer, K.F.X., Moreau, H., Not, F., Otillar, R., Panaud, O., Pangilinan, J., Paulsen, I., Piegu, B., Poliakov, A., Robbens, S., Schmutz, J., Toulza, E., Wyss, T., Zelensky, A., Zhou, K., Armbrust, E.V., Bhattacharya, D., Goodenough, U.W., Van de Peer, Y. and Grigoriev, I.V. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**, 268-272.

Xiang, J.Y., Jung, J.Y. and Sampson, N.S. (2004) Entropy effects on protein hinges: The reaction catalyzed by triosephosphate isomerase. *Biochemistry* **43**, 11436-11445.

Young, A.J. (1993) Occurrence and distribution of carotenoids in photosynthetic systems. In *Carotenoids in Photosynthesis* (Young, A.J. and Britton, G. eds). London: Chapmann & Hall, pp. 16-71.

Yu, Q.J. and Beyer, P. (2012) Reaction specificities of the epsilon-ionone-forming lycopene cyclase from rice (Oryza sativa) elucidated in vitro. *FEBS Lett.* **586**, 3415-3420.

Zhang, Y.H.P. (2011) Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnol. Adv.* **29**, 715-725.

Figure Legends

Figure 1. Carotenoid biosynthetic pathway in land plants and green algae (Viridiplantae). Names of carotenoids specific to *O. lucimarinus* and other prasinophyte algae of the order Mamiellales are intalicized. LCYB = lycopene β-cyclase; LCYE = lycopene ε-cyclase.

Figure 2. Domain structure of the lycopene cyclase fusion protein from *O. lucimarinus* and phylogenetic affiliation of its three major functional domains. Domain structure of OluLCY (a), and phylogenetic affiliation of the two lycopene cyclase domains (b) and the LHC domain (c) in the LCY fusion proteins from the six Mamiellales and *B. natans*. ML (RAxML) trees were constructed from protein alignments including 383 (b) and 129 (c) amino acid positions; the lycopene cyclase sequence from *Synechococcus elongatus* PCC6301 (B) and the sequence of LHCA1 from *O. tauri* (C) were used as outgroups. The results of rapid ML bootstrap analyses with RAxML (100 replicates) are shown above the branches (only values > 50); thick branches represent ≥ 95% Bayesian posterior probability. Branch lengths are proportional to the number of substitutions per site (see scale bars). Species abbreviations in (c) are: At = *A. thaliana*; Cr = *C. reinhardtii*; Ot = *O. tauri*; Pp = *Physcomitrella patens*; Sm = *Selaginella moellendorffii*. For sequence accessions, see Table S1 and Chan et al. (2012).

Figure 3. Catalytic activity of the cyclase fusion protein from *O. lucimarinus* and of its separate LCYB and LCYE domains in lycopene-accumulating *E. coli*. Panels show HPLC chromatograms of pigment extracts from bacteria after co-transformation with plasmid pLYCO2 inducing biosynthesis of the substrate lycopene and with pBAD-constructs encoding a nearly full-length OluLCY protein (aa 50-1442; (a), lower trace), only the OluLCYB domain (aa 50-591; (b), upper trace), or only the OluLCYE domain (aa 625-1130; (c), upper trace). Plasmid pBAD-S encoding only linker S1 of OluLCY (aa 562-630; (a), upper trace) served as negative control. For comparison, the cyclases CrLCYB ((b); lower trace) and CrLCYE ((c);

lower trace) from *C. reinhardtii* were also expressed in the pLYCO2 background. Pigment abbreviations: Car = carotene; Lyco = lycopene.

Figure 4. Effect of stepwise C-terminal truncation of the OluLCY protein on the catalytic activity of its cyclase domains in lycopene-accumulating *E. coli*. The scheme on top illustrates the extent of the C-terminal truncation for each mutant, with numbers denoting amino acids. The bar chart shows the average percentage of α- and β-carotene per total carotenoids for each truncation (average and s.d. calculated from three biological replicates). The precursor carotenoids δ- and γ-carotene did not accumulate, while lycopene always accounted for less than 5% of total carotenoids.

Figure 5. Effect of partial deletion of the spacer region S1 between the LCYB- and LCYE-domain on the catalytic activity of its cyclase domains in lycopene-accumulating *E. coli*. The scheme on top illustrates the extent of the spacer deletion for each mutant, with numbers denoting amino acid positions. The bar chart shows the average percentage of α-, β- and δ-carotene per total carotenoids for each deletion (average and s.d. calculated from three biological replicates). The precursor γ-carotene did not accumulate, while lycopene accounted for 5% - 20% of total carotenoids depending on the deletion. Abbreviations: B = LCYB-domain; E = LCYE-domain; S = Linker S1; Sa = N-proximal half of Linker S1; Sb = C-proximal half of Linker S1.

Figure 6. Effect of replacing the LCYB- or/and LCYE-domain with LCYB or LCYE from *C. reinhardtii* on the catalytic activity of the chimeric protein in lycopene-accumulating *E. coli*. The bar chart shows the average percentage of α-, β- and δ-carotene per total carotenoids for each chimeric protein (average and s.d. calculated from three biological replicates). Cells expressing chimera ce-S-B accumulated 50% of total carotenoids as lycopene and minor

amounts (< 5%) as γ-carotene, while cells expressing the other chimeras accumulated between 2% (E-S-cb) and 20% (B-S-ce) of total carotenoids as lycopene, but no γ-carotene. Abbreviations: B = LCYB-domain of OluLCY; S = Linker S1 of OluLCY; E = LCYE-domain of OluLCY; cb = CrLCYB; ce = CrLCYE.

Figure 7. Accumulation of α-carotene in *E. coli* after transformation with plasmid pALPHA1 containing the constitutively expressed gene *OluLCYΔ312* or plasmid pALPHA2 containing *OluLCYΔ248* instead. Maps of pALPHA1 (a) and pALPHA2 (c) and HPLC chromatograms of pigment extracts from *E. coli* transformed either with pALPHA1 (b) or pALPHA2 (d) and grown in LB medium at 28 °C for 16 h. Gene abbreviations: cat = chloramphenicol acetyltransferase; crtB = phytoene synthase; crtE = geranylgeranyl diphosphate synthase; ctrI = phytoene desaturase; tet = tetracycline resistance; pigment abbreviations as in Figure 3.

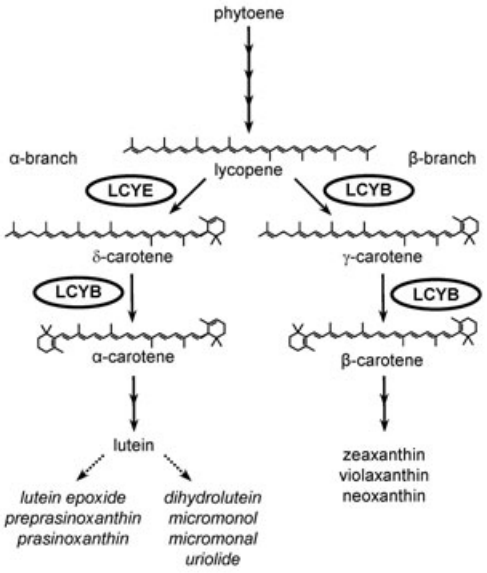

Figure 1. Carotenoid biosynthetic pathway in land plants and green algae (Viridiplantae).

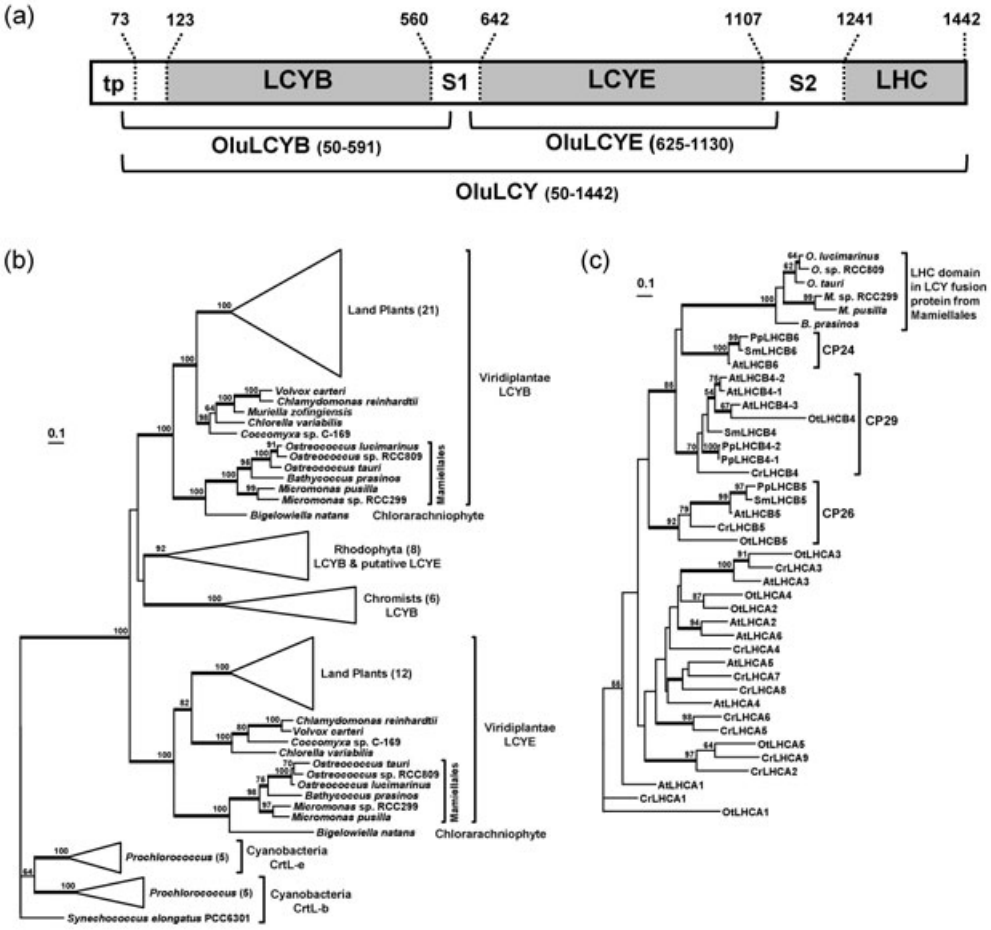

Figure 2. Domain structure of the lycopene cyclase fusion protein from *O. lucimarinus* and phylogenetic affiliation of its three major functional domains.

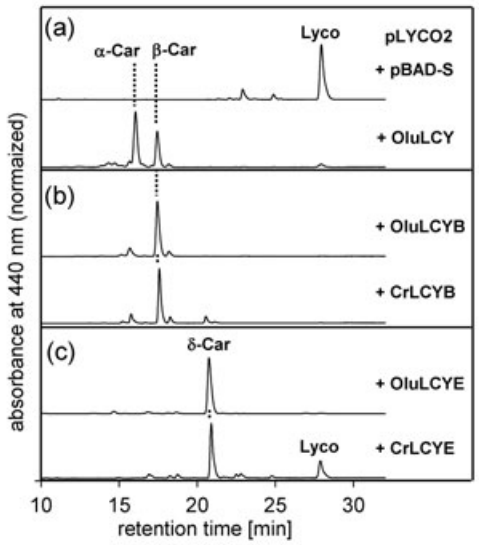

Figure 3. Catalytic activity of the cyclase fusion protein OluLCY from *O. lucimarinus* and of its separate LCYB and LCYE domains in lycopene-accumulating *E. coli*.

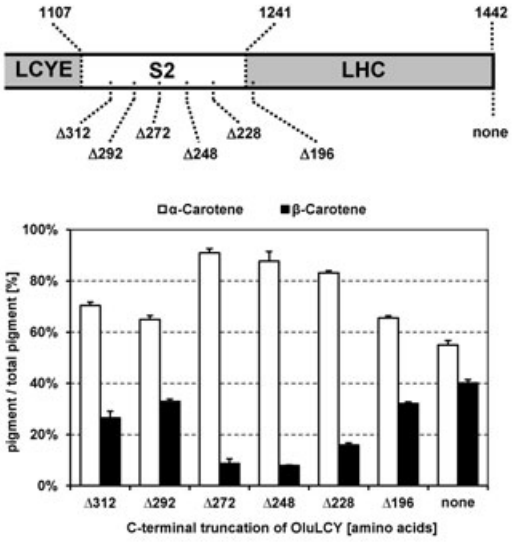

Figure 4. Effect of stepwise C-terminal truncation of the OluLCY protein on the catalytic activity of its cyclase domains in lycopene-accumulating *E. coli*.

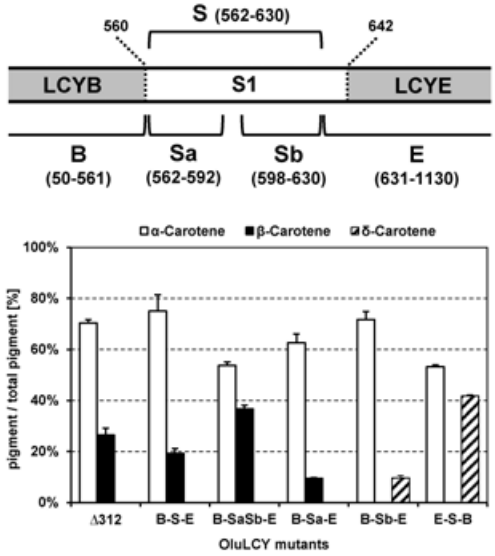

Figure 5. Effect of partial deletion of the spacer region S1 between the LCYB- and LCYE-domain on the catalytic activity of its cyclase domains in carotenogenic *E. coli*.

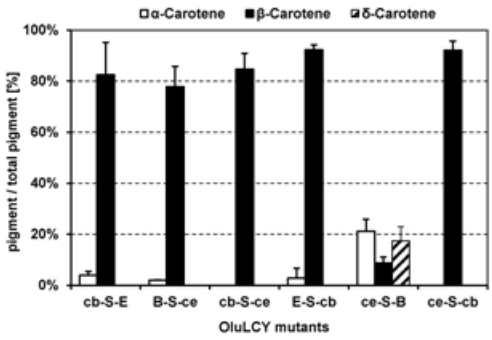

Figure 6. Effect of replacing the LCYB- or/and LCYE-domain with LCYB or LCYE from *C. reinhardtii* on the catalytic activity of the chimeric protein in lycopene-accumulating *E. coli*.

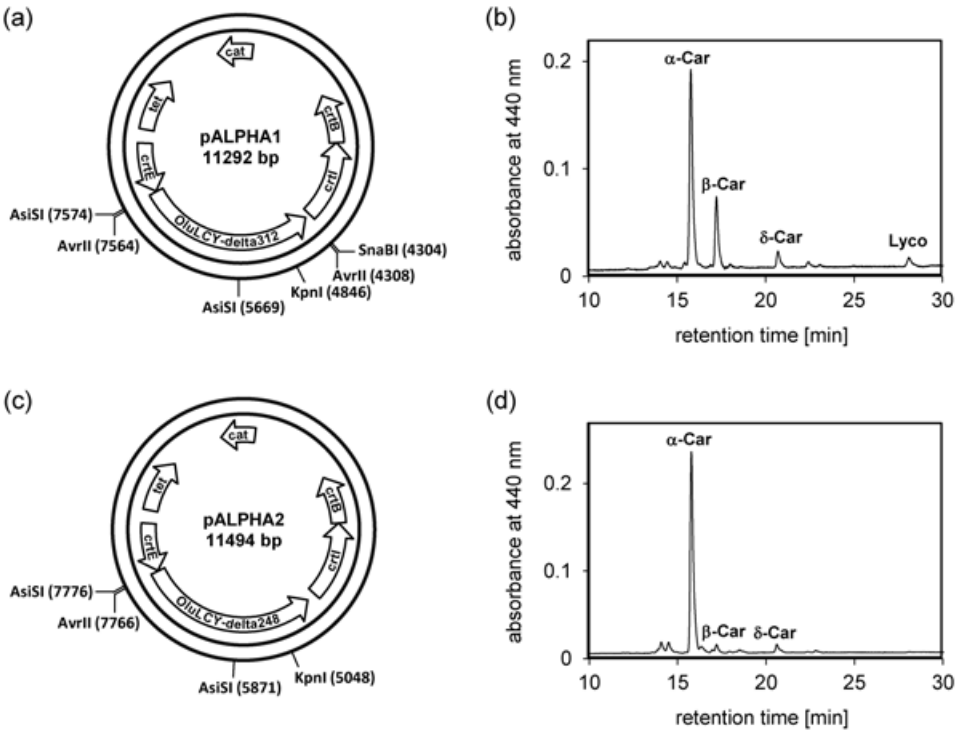

Figure 7. Accumulation of α-carotene in *E. coli* after transformation with plasmid pALPHA1 containing the constitutively expressed gene *OluLCYΔ312* or plasmid pALPHA2 containing *OluLCYΔ248* instead.