

Ancient Recruitment by Chromists of Green Algal Genes Encoding Enzymes for Carotenoid Biosynthesis

Ruth Frommolt,^{*1} Sonja Werner,[†] Harald Paulsen,[†] Reimund Goss,^{*} Christian Wilhelm,^{*} Stefan Zauner,[‡] Uwe G. Maier,[‡] Arthur R. Grossman,[§] Debashish Bhattacharya,^{||1} and Martin Lohr^{†1}

^{*}Institut für Biologie I, Pflanzenphysiologie, Universität Leipzig, Leipzig, Germany; [†]Institut für Allgemeine Botanik, Johannes Gutenberg-Universität, Mainz, Germany; [‡]Zellbiologie, Philipps-Universität Marburg, Marburg, Germany; [§]The Carnegie Institution, Department of Plant Biology, Stanford, CA; and ^{||}Department of Biological Sciences and Roy J. Carver Center for Comparative Genomics, University of Iowa

Chromist algae (stramenopiles, cryptophytes, and haptophytes) are major contributors to marine primary productivity. These eukaryotes acquired their plastid via secondary endosymbiosis, whereby an early-diverging red alga was engulfed by a protist and the plastid was retained and its associated nuclear-encoded genes were transferred to the host genome. Current data suggest, however, that chromists are paraphyletic; therefore, it remains unclear whether their plastids trace back to a single secondary endosymbiosis or, alternatively, this organelle has resulted from multiple independent events in the different chromist lineages. Both scenarios, however, predict that plastid-targeted, nucleus-encoded chromist proteins should be most closely related to their red algal homologs. Here we analyzed the biosynthetic pathway of carotenoids that are essential components of all photosynthetic eukaryotes and find a mosaic evolutionary origin of these enzymes in chromists. Surprisingly, about one-third (5/16) of the proteins are most closely related to green algal homologs with three branching within or sister to the early-diverging Prasinophyceae. This phylogenetic association is corroborated by shared diagnostic indels and the syntenic arrangement of a specific gene pair involved in the photoprotective xanthophyll cycle. The combined data suggest that the prasinophyte genes may have been acquired before the ancient split of stramenopiles, haptophytes, cryptophytes, and putatively also dinoflagellates. The latter point is supported by the observed monophyly of alveolates and stramenopiles in most molecular trees. One possible explanation for our results is that the green genes are remnants of a cryptic endosymbiosis that occurred early in chromalveolate evolution; that is, prior to the postulated split of stramenopiles, alveolates, haptophytes, and cryptophytes. The subsequent red algal capture would have led to the loss or replacement of most green genes via intracellular gene transfer from the new endosymbiont. We argue that the prasinophyte genes were retained because they enhance photosynthetic performance in chromalveolates, thus extending the niches available to these organisms. The alternate explanation of green gene origin via serial endosymbiotic or horizontal gene transfers is also plausible, but the latter would require the independent origins of the same five genes in some or all the different chromalveolate lineages.

Introduction

According to the endosymbiotic theory, plastids evolved from a cyanobacterium that was engulfed and retained by a unicellular protist. During the establishment of this primary endosymbiosis, many of the cyanobacterial genes were transferred to the nuclear genome of the host organism and their gene products retargeted to the plastid (Martin et al. 2002; Reyes-Prieto et al. 2006). The first photosynthetic eukaryotes evolved into three distinct, extant lineages, the glaucophytes, red algae, and Chloroplastida; the latter is comprised of green algae and land plants (Moreira et al. 2000; Adl et al. 2005).

The plastids of stramenopiles (e.g., diatoms and brown algae), haptophytes, and cryptophytes—groups collectively classified as “Chromista” (Cavalier-Smith 1999)—and most dinoflagellates differ from those in the Chloroplastida due to the presence of chlorophyll *c* instead of chlorophyll *b*. Whereas there is general agreement that chromist algae gained their plastids from red alga-like endosymbionts (Palmer 2003; Bhattacharya et al. 2004; Keeling 2004; Archibald 2005), current multigene phylogenetic data suggest strongly that the host cells are not monophyletic. These

trees show that stramenopiles are more closely related to alveolates (apicomplexans, ciliates, and dinoflagellates) than to a monophyletic group defined by Haptophyta and Cryptophyta (HC group; Harper et al. 2005; Burki et al. 2007, 2008; Hackett et al. 2007; Patron et al. 2007). This result is also supported by the finding of a gene replacement in the plastid genomes of the HC group via horizontal gene transfer (HGT) of a *rpl36* gene of eubacterial origin (Rice and Palmer 2006). These data open up the possibility that chromists may have gained their plastid through independent red algal secondary endosymbiotic events. Even more controversial is the notion that all chromists and alveolates are united in the supergroup “Chromalveolata” (Cavalier-Smith 1999). This hypothesis remains unsubstantiated not only because of apparent chromist paraphyly but also based on phylogenetic evidence that another supergroup, the Rhizaria, likely branches as sister to stramenopiles and alveolates, with this Stramenopiles/Alveolata/Rhizaria cluster sometimes referred to as SAR group (Burki et al. 2007, 2008; Hackett et al. 2007). Although clearly differing from and far more complex than envisioned in the original hypothesis, it is however intriguing that several sources of data continue to provide support for a red algal-derived plastid in most chromalveolates. These data include analyses of plastid-targeted proteins encoded in the nucleus (Harper and Keeling 2003), plastid-encoded proteins (Yoon et al. 2002; Bachvaroff et al. 2005; Rogers et al. 2007), and cases in which a plastid protein was replaced by a host cell-derived functional equivalent (Fast et al. 2001; Harper and Keeling 2003; Patron et al. 2004; Reyes-Prieto and Bhattacharya

¹ These authors contributed equally to this work.

Key words: secondary endosymbiosis, plastid, carotenoid biosynthesis, xanthophyll cycle, gene transfer, chromalveolates.

E-mail: lohr@uni-mainz.de.

Mol. Biol. Evol. 25(12):2653–2667. 2008

doi:10.1093/molbev/msn206

Advance Access publication September 17, 2008

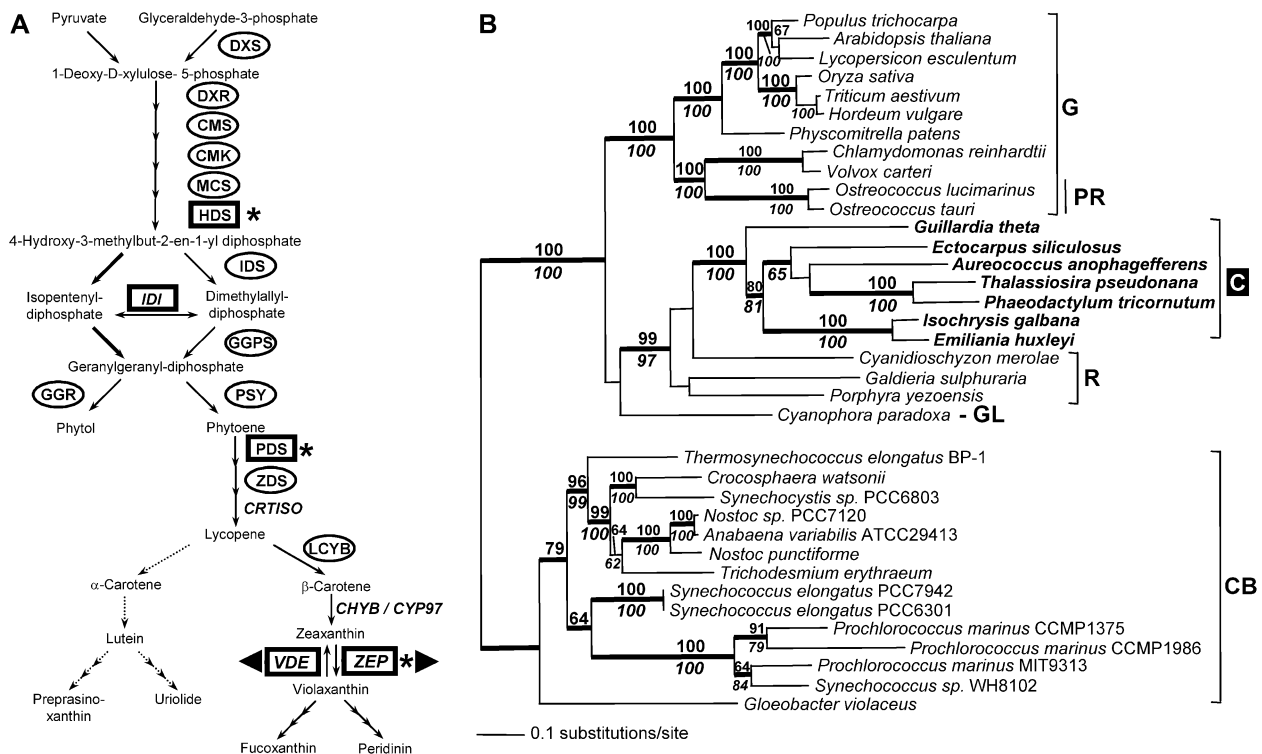


FIG. 1.—(A) Generalized scheme of carotenoid biosynthesis in plants and algae summarizing major results of our analyses. Enzymes (see table 1) encoded by homologous genes in representatives of all major taxa that we examined are in bold, others in italics. Encircled enzyme names represent chromist proteins most closely related to their red algal counterparts, whereas framed names are chromist proteins most closely related to their green algal counterparts; asterisks designate the chromist proteins that share diagnostic indels specifically with homologs from basal green algae; filled triangles label the pair of genes that are syntenic in the genomes of a diatom, a pelagophyte, and two prasinophyte algae. (B) ML (RAxML) tree of oxygenic phototrophs constructed from nine concatenated nucleus-encoded plastid proteins involved in the biosynthesis of carotenoids (DXS, DXR, 4-diphosphocytidyl-2C-methyl-D-erythritol synthase, 4-diphosphocytidyl-2C-methyl-D-erythritol kinase, MCS, IDS, geranylgeranyl diphosphate synthase, GGR, and phytoene synthase), including a total of 2,884 amino acid positions. The results of ML bootstrap analyses with RAxML (100 replicates) and PHYML (100 replicates) are shown above and below the branches, respectively; thick branches represent $\geq 95\%$ Bayesian posterior probability. Branch lengths are proportional to the number of substitutions per site (see scale bar). Species names of chromalveolate algae are in bold. Abbreviations: C, Chromista; CB, Cyanobacteria; G, green algae and land plants (Chloroplastida); GL, Glaucophyta; PR, Prasinophyceae; and R, Rhodophyta.

2007). With the possible exception of picobiliphytes (Not et al. 2007), cryptophytes are the only known chromalveolates that maintain a reduced nucleus of eukaryotic endosymbiont origin (Douglas et al. 2001). In all other cases, nuclear genes of plastid function encoded by the eukaryotic endosymbiont have been transferred to the nucleus of the new host organism resulting in loss of the captured nuclear genome.

Phylogenetic analyses are often based on data derived from random sequencing of expressed sequence tag (EST) clones and analysis of proteins inferred from these partial gene sequences. The availability of an increasing number of algal genomes enables different approaches like the phylogenetic analysis of specific pathways or ultimately of many features of whole genomes. Examples of pathway-specific phylogenetic analyses include the biosynthesis of heme (Obornik and Green 2005), chorismate (Richards et al. 2006), and tryptophan (Jiroutova et al. 2007). In eukaryotic phototrophs, these pathways are plastid localized with the corresponding enzymes being nucleus encoded. These particular analyses, however, were complicated by the fact that heme biosynthesis is a ubiquitous pathway and that the ability to synthesize chorismate and tryptophan is widespread among bacteria and lower eukaryotes. In the case of an or-

ganism resulting from a secondary endosymbiosis, the corresponding genes could have been introduced by the mitochondrial or plastid progenitors or by the two eukaryotic host cells—aside from potential origin via HGT.

In the present study, we explored the origins of nuclear-encoded, plastid-localized proteins in the paraphyletic chromists by focusing on the genes involved in carotenoid biosynthesis. Carotenoids are essential components of the photosynthetic apparatus that serve both light-harvesting and photoprotective functions (Morosinotto et al. 2003; Grossman et al. 2004). They are diterpenoids, and in plants and algae, the isoprene units utilized in their synthesis are derived from pyruvate and glyceraldehyde-3-phosphate via the plastid-localized methylerythritol phosphate (MEP) pathway (Rodríguez-Concepción and Boronat 2002; Bouvier et al. 2005). Carotenoids that contain one or more oxygen atoms are collectively termed xanthophylls. The pathway for the synthesis of carotenoids (Cunningham and Gantt 1998; Rodríguez-Concepción and Boronat 2002; Bouvier et al. 2005) is shown in figure 1A. The early stage of carotenoid biosynthesis, the MEP pathway, is not limited to plastids and cyanobacteria but is common among many other bacteria as well (Rohmer 2003); it is, however,

Table 1
Carotenogenic Enzymes That Were Examined

Enzyme (gene) ^a	Full Name	Synonyms	Chromist Affiliation ^b
<u>DXS</u>	1-deoxy-D-xylulose-5-phosphate synthase	CLA1, DEF, DXPS	Red
<u>DXR</u>	1-deoxy-D-xylulose-5-phosphate reductoisomerase	ISPC	Red
<u>CMS</u>	4-diphosphocytidyl-2C-methyl-D-erythritol synthase	ISPD	Red
<u>CMK</u>	4-diphosphocytidyl-2C-methyl-D-erythritol kinase	ISPE, CDPMEK	Red
<u>MCS</u>	2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase	ISPF, MECPS	Red
<u>HDS</u>	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	ISPG, GCPE	Green
<u>IDS</u>	Isopentenyl diphosphate:dimethylallyl diphosphate synthase	ISPH, LYTB, HDR	Red
<u>IDI</u>	Isopentenyl diphosphate:dimethylallyl diphosphate isomerase	IPI	Green
<u>GGPS</u>	Geranylgeranyl diphosphate synthase	GGPPS	Red
<u>GGR</u>	Geranylgeranyl reductase	CHLP	Red
<u>PSY</u>	Phytoene synthase		Red
<u>PDS</u>	Phytoene desaturase		Green
<u>ZDS</u>	ζ-carotene desaturase		Red
<u>LCYB</u>	Lycopene β-cyclase	CRTL-B	Red
<u>ZEP</u>	Zeaxanthin epoxidase	ABA1	Green
<u>VDE</u>	Violaxanthin de-epoxidase		Green

^a Proteins that were included in the concatenated data set are underlined.

^b Indicates whether the corresponding chromist proteins are more closely related to the homologs from red algae (red) or Chloroplastida (green).

missing from all eukaryotes that lack plastids. Yet, most of the genes encoding enzymes that are involved in the later stages of carotenogenesis in algae and plants are derived from cyanobacterial genes (Sandmann 2002) and therefore particularly suited to understanding the evolution of photosynthetic eukaryotes.

Using protein sequence similarity searches of genomic data from the haptophyte alga *Emiliania huxleyi* and four stramenopiles, that is, the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, the pelagophyte *Aureococcus anophagefferens*, and the brown alga *Ectocarpus siliculosus*, we identified candidate genes for 16 enzymes involved in carotenoid biosynthesis that have been described in vascular plants (fig. 1A and table 1). For the phylogenetic analyses, we compiled additional algal, plant, and cyanobacterial sequences available in public databases and sequenced target cDNAs from *P. tricornutum*, the cryptophyte *Guillardia theta*, the prasinophyte *Mantoniella squamata*, and the red alga *Porphyra yezoensis*.

Materials and Methods

Sequence Data Mining

The sequences of 20 different carotenogenic proteins from *Arabidopsis thaliana* (Lange and Ghassemian 2003) or *Chlamydomonas reinhardtii* (Lohr et al. 2005) were used to search the following databases by Blast (Altschul et al. 1997): 1) GenBank: protein database, dbEST, microbial genomes, and trace archives; 2) Joint Genome Institute/United States: Eukaryotic Genomics genome browsers of *Populus trichocarpa* (Sterky et al. 2004), *T. pseudonana* (Armbrust et al. 2004), *A. anophagefferens*, *Ostreococcus tauri* (Derelle et al. 2006), and *Ostreococcus lucimarinus* (Palenik et al. 2007); 3) *Cyanidioschyzon merolae* genome browser at the University of Tokyo/Japan (Matsuzaki et al. 2004); 4) *Galdieria sulphuraria* genome database at Michigan State University/United States (Barbier et al. 2005); 5) Taxonomically Broad EST Database (TBestDB) of the pan-Canadian Protist EST Program; and 6) the Plant Transcript Assemblies of The Institute for Genomic Re-

search (TIGR). URLs are supplied in supplementary table 2 (Supplementary Material online).

Assembly of genomic trace reads from *Volvox carteri*, *P. tricornutum*, *E. siliculosus*, and *E. huxleyi* was done in BioEdit 5.0.9 (Hall 1999), and the resulting contigs were analyzed by GENSCAN (<http://genes.mit.edu/GENSCAN.html>) (Burge and Karlin 1997) to predict open reading frames (ORFs) and exon/intron structures. The predicted gene models were compared with deduced protein sequences from related species (*C. reinhardtii* in case of *V. carteri* and both *C. reinhardtii* and *T. pseudonana* in case of *P. tricornutum*, *E. siliculosus*, and *E. huxleyi*) and edited manually to incorporate appropriate changes. In the assembled genes from *P. tricornutum*, we frequently found single nucleotide polymorphisms supported by two or more reads for each of two alternate sequences. Because the deduced gene products were on average 99% identical at the amino acid level, we included only one of the two alleles in the phylogenetic analyses. After the genomes of *P. tricornutum* and *V. carteri* had become publicly accessible via the JGI Eukaryotic Genomics Web site (<http://genome.jgi-psf.org/>), we checked the correctness of our gene assemblies from *P. tricornutum* and *V. carteri* against this database.

Generation of Additional cDNA Sequences

For *M. squamata*, total RNA from algal cultures was extracted with phenol/chloroform using a modified protocol as described previously (Meyer-Gauen et al. 1998). The mRNA was purified with the PolyAtract mRNA isolation system IV (Promega GmbH, Mannheim, Germany) and used to construct a cDNA library by means of the Lambda ZAP-cDNA Synthesis Kit (Stratagene Europe, Amsterdam, The Netherlands). The library was packaged and amplified using the ZAP-cDNA Gigapack III Gold Cloning Kit (Stratagene). The cDNA sequences encoding 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), phytoene desaturase (PDS), isopentenyl diphosphate synthase (IDS), and geranylgeranyl reductase (GGR) were isolated

by nested polymerase chain reaction (PCR) using *Taq* polymerase (Qiagen, Hilden, Germany) or the Expand Long Template System (Roche Diagnostics–Applied Science, Mannheim, Germany) with combinations of a heterologous primer directed against conserved motifs and a vector primer. Successful primer pairs were 1) for DXR, vector primer T3 with reverse primers CATGGACCAGTTGGGGTG (first round) and GAAGATCGCGGAGTGCTC (second round); 2) for PDS, vector primer T3 with reverse primers TGGGTCTGAACCAGATGTG (first round) and CTCCT-GIAGGAAGCGGTT (second round); 3) for IDS, vector primer T3 with reverse primers CCCTTGAGCATGGTGGTCTG (first round) and GGGGTTGTGGATGATCTC (second round); and 4) for GGR, vector primer T3 with reverse primers GACTTGGCGGCGAAGTA (first round) and AGACCCAGCCGTAGAAGTC (second round). The PCR products were purified using agarose gel electrophoresis and sequenced from both ends. The regions of the cDNAs for which no sequence information was available were amplified using nested PCR by combining target-specific primers directed against the known partial sequences with the vector primer T7, and the resulting product was purified and sequenced on both strands using commercial sequencing services (AGOWA, Berlin, Germany; GENTERprise, Mainz, Germany).

For *P. tricornutum*, partial cDNA information and genomic trace reads for genes encoding 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS) and IDS were retrieved from the GenBank database by Blast and were used to design primers for amplification of complete ORF from available cDNA libraries (Grossman et al. 1990). PCR amplification of the complete ORF-encoding HDS was achieved using PHUSION High-Fidelity DNA polymerase (Finnzymes, Espoo, Finland) with forward AAAGGCGCTATGAAGTT and reverse GGTAATTATCTGATTGGCGTA as primers, and amplification of the ORF-encoding IDS was achieved with forward CGACAC-CACCATCGAATAGA and reverse TAAGGAATTGATCGCCGAAG as primers. The PCR products were sequenced to completion on both strands by primer walking using the ABI Prism BigDye v3.1 Terminator reagents (Applied Biosystems, Foster City, CA) and a commercial sequencing service (GENTERprise).

The *P. yezoensis* EST database (Asamizu et al. 2003) was searched using Blast for translated sequences with similarity to known carotenogenic proteins from plants and algae. The following clones were identified, retrieved from the Kazusa DNA Research Institute (Chiba, Japan), and sequenced (as described for *P. tricornutum*) on both strands to completion: AU190953 (DXS), AU196956 (DXR), AU186634 (HDS), AU187853 (IDS), AU195302 (PDS), and AV435435 (GGR).

For *G. theta*, a cDNA library was generated and an EST database was set up by sequencing of 7,000 randomly chosen EST clones. Blast searches of the EST database revealed seven clones with homology to 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MCS), HDS, isopentenyl diphosphate isomerase (IDI) (2), ζ -carotene desaturase (ZDS), and zeaxanthin epoxidase (ZEP), respectively (see table 1). Identified clones were subsequently sequenced to

completion. All cDNA sequence data have been deposited with the DDBJ/EMBL/GenBank data libraries under GenBank accession numbers FJ175667–FJ175685. Database accession codes of the other sequences that were used for the phylogenetic analyses are given in supplementary table 2 (Supplementary Material online).

Phylogenetic Analyses

Protein alignments were constructed using either Muscle 3.52 (Edgar 2004) or ClustalW 1.83 (Chenna et al. 2003). Manual refinement, selection of unambiguously aligned positions, and construction of a concatenated protein alignment were made using BioEdit 5.0.9 (Hall 1999). Generally, parts of the alignments with more than one sequence having a gap extending for two or more amino acid positions were excluded from the phylogenetic analyses, although in a few cases, sequences with gaps extending over as many as four amino acid positions were included.

A maximum likelihood (ML) tree was inferred from each protein alignment using RAXML (VI-HPC, v2.2.1) (Stamatakis 2006) and the WAG substitution model with gamma rate distribution (“PROTGAMMA” implementation), four discrete rate categories, and starting from a random tree. Branch support was estimated with 100 bootstrap replicates using both RAXML (WAG substitution model and the “PROTCAT” implementation) and PhyML (Guindon and Gascuel 2003) (WAG + γ substitution model and parameters estimated during the tree search). Posterior probabilities for nodes in the RAXML trees were calculated with MrBayes 3.1 (Huelsenbeck and Ronquist 2001) running MC³ for 1 million generations using one cold and three heated chains and starting with a random tree. The pool of trees was sampled every 100th generation, and posterior probabilities were estimated using the final 5,000 trees. Chain convergence was apparent for each data set after this level of burn-in.

Dot plots were made with BioEdit 5.0.9 using a window size of 10 and the BLOSUM62 similarity matrix for analysis and 4 respectively 10, pixels as lower respectively higher, threshold for drawing.

Results and Discussion

For 16 of the 20 carotenogenic enzymes that are known from Chloroplastida, we were able to identify genes encoding homologous enzymes in chromists with sufficient sequence conservation to suggest a conserved function. For four proteins, that is, a carotenoid isomerase (CRTISO), a nonheme beta-carotene hydroxylase (CHYB), and two closely related cytochrome P450 carotene hydroxylases of the CYP97 family (see fig. 1A), we did not find genes of sufficiently high conservation to classify them as likely orthologs of the genes from Chloroplastida. Therefore, we excluded these genes/proteins from further analyses. A more detailed account of these genes will be presented elsewhere.

For 13 of the 16 chromist genes associated with carotenoid biosynthesis (table 1), homologs were detected in representatives from all major taxa that we examined; that

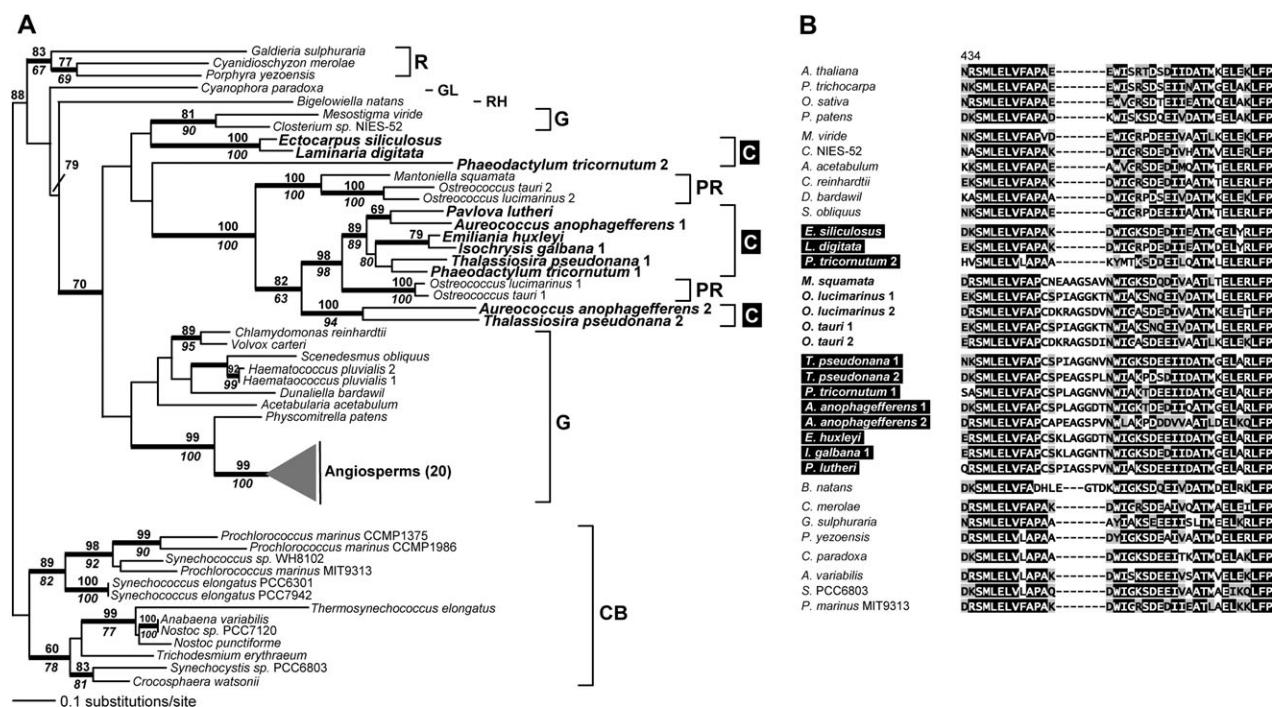


FIG. 2.—(A) Phylogeny of PDS: ML (RAXML) tree constructed from a protein alignment including 446 amino acid positions and (B) support via the presence of a diagnostic insertion specific to chromist and prasinophyte sequences. The depicted section of a full alignment is indicated by the position of the first shown amino acid in the corresponding protein sequence of *Arabidopsis thaliana*; at conserved sites, black boxes indicate identical amino acids, whereas gray boxes denote similar amino acids; species names of chromist algae are written in white letters on black background. Display features in (A) are as in figure 1B; additional abbreviation: RH, Rhizaria.

is, cyanobacteria, red algae, and Chloroplastida (green algae and land plants). We inferred a phylogeny for each of these 13 deduced proteins sequences using cyanobacterial homologs as the outgroup. The closest phylogenetic relationship for 11 of the chromist sequences (fig. 1A, circled enzyme names) was with red algal homologs. The complete set of results for this analysis is presented as supplementary figure 1 (Supplementary Material online). For a set of representative organisms, we concatenated the deduced protein sequences from nine genes with a red algal ancestry (see table 1 for genes; supplementary table 1, Supplementary Material online for data on sequence coverage) and analyzed these data using three different phylogenetic methods (ML under RAXML and PHYML and Bayesian analysis). The remaining two red genes were excluded because sequence data were missing for more than 20% of the 36 representative species. The 9-protein tree shown in figure 1B indicates that the photosynthetic eukaryotes that were sampled comprise three well-separated clades, the green line comprised of the Chloroplastida; the red line containing red algae, stramenopiles, haptophytes, and the cryptophyte *G. theta*; and the glaucophytes represented by *Cyanophora paradoxa*. Bootstrap analyses provided high statistical support for the hypothesis that the plastids in stramenopiles, haptophytes, and cryptophytes were acquired from a red alga-like organism via secondary endosymbiosis. This result is consistent with previous studies (e.g., Patron et al. 2004, 2007; Bachvaroff et al. 2005; Harper et al. 2005; Li et al. 2006; Hackett et al. 2007). The inability to recover HC monophyly using plastid genome data or as here, sequences of nuclear-encoded, plastid-targeted proteins, remains unclear but may be explain-

ed by poor taxon sampling of chromalveolates or amino acid composition bias among these data (for details, see Khan et al. 2007).

Chromist Proteins More Closely Related to Green than Red Homologs *Phytoene Desaturase*

In contrast to the 11 genes described above of red algal origin, the chromist protein sequences for PDS, HDS, and IDI were more closely related to green algal enzymes than to red algal homologs. The phylogenetic tree of PDS from cyanobacteria, algae, and land plants is shown in figure 2A. The sequences from algae and plants are grouped into five clusters: 1) red algae at the base of the eukaryotes, 2) green algae of the Ulvophyceae and Chlorophyceae, 3) Plantae (moss and angiosperms), 4) basal streptophytes of the clades Mesostigmatophyceae and Zygnematomphyceae together with PDS from two brown algae, and 5) PDS from several stramenopiles and haptophytes together with basal Chloroplastida of the class Prasinophyceae. The PDS sequences from the glaucophyte *C. paradoxa* and the rhizarian alga *Bigelowiella natans* separate the red algae from the Chloroplastida/Chromista cluster. Yet, only the nodes of the red algal, the embryophyte, and the prasinophyte/chromist cluster are resolved. Additional support for a close relationship of the sequences in the latter cluster comes from a diagnostic insertion that is exclusively associated with PDS from prasinophytes, haptophytes, diatoms, and the pelagophyte *A. anophagefferens* (see fig. 2B). The prasinophytes used for these analyses belong to the order

Mamiellales, which is regarded as one of the most basal green algal clades (Marin and Melkonian 1999). Whereas a common origin of PDS from prasinophytes and chromists with the enzyme from other green algae and land plants received only weak bootstrap support in the RAxML analyses, the prasinophyte/chromist PDS cluster is clearly separated from the red line. The combination of the phylogenetic and unambiguous insertion data suggests that the chromist PDS sequences of this cluster have been acquired from a prasinophyte-like alga before the divergence of haptophytes and stramenopiles.

Whereas PDS is encoded by a single gene in most organisms, the genomes of the prasinophytes *O. lucimarinus* and *O. tauri*, the diatoms *T. pseudonana* and *P. tricornutum*, and the pelagophyte *A. anophagefferens* each contain a pair of putative PDS genes encoding proteins with ~60% amino acid identity (75% similarity). The PDS gene that we identified in the cDNA library of another prasinophyte, *M. squamata*, is probably orthologous to PDS2 from the two *Ostreococcus* species. Although not currently identified, *M. squamata* is likely to contain another PDS gene orthologous to PDS1 from *Ostreococcus*. In the whole-genome shotgun trace reads from the haptophyte *E. huxleyi* and the brown alga *E. siliculosus*, we found only single PDS genes, whereas searches of the NCBI- and TBestDB-EST databases retrieved one PDS sequence from the haptophyte *Pavlova lutheri* but two different PDS transcripts from another haptophyte, *Isochrysis galbana*. One of these EST sequences, however, was too short for phylogenetic analyses and therefore is not included in figure 2. The predicted gene products from haptophytes that we used for the analyses are most closely related to the PDS1 paralogs from the diatoms, the pelagophyte, and *Ostreococcus* spp. (fig. 2A).

The situation becomes more complex, however, because the single PDS sequences from the two brown algae *E. siliculosus* and *Laminaria digitata* and the PDS2 protein from the diatom *P. tricornutum* do not contain the diagnostic insertion (fig. 2B) and appear to be most closely related to PDS from the basal streptophytes *Mesostigma viride* and *Closterium* sp. strain NIES-52 (fig. 2A). This result suggests that stramenopile algae may have acquired green PDS genes two times from different green algae, that is, a prasinophyte and a streptophyte alga. According to this hypothesis, *E. siliculosus* and *L. digitata* would have replaced their prasinophyte-type PDS with the streptophyte-related homolog, whereas *P. tricornutum* retained the genes from both sources.

Although the interpretation of the tree in figure 2A is complicated by the occurrence of paralogs and the putative involvement of two HGT events, the branching order within the chromist/prasinophyte cluster suggests the presence of two groups of orthologs, that is, PDS1 and PDS2. It is important to note that the gene duplication leading to PDS1 and PDS2 likely occurred before the segregation of this particular pair of paralogs into the genomes of *T. pseudonana*, *A. anophagefferens*, and the examined prasinophytes. This supports the idea that the prasinophyte-related PDS genes in chromists are an ancient acquisition and that the streptophyte-related PDS was introduced later and potentially only by stramenopiles.

4-Hydroxy-3-Methylbut-2-en-1-yl Diphosphate Synthase

For HDS, a significant evolutionary separation is found between the cyanobacterial/glaucophyte/red algal cluster and the proteins from vascular plants, green algae, chromists, and also from alveolates (fig. 3A). This observation has been made before for HDS and several other enzymes of the MEP pathway and was interpreted as the likely result of a replacement of the gene in cyanobacteria by HGT after the primary endosymbiosis (Lange et al. 2000). More recent analyses, however, suggest that a Chlamydia-like bacterium may have donated several genes (with HDS among them) to eukaryotic phototrophs at the level of the primary endosymbiosis (Huang and Gogarten 2007; Tyra et al. 2007; Moustafa et al. 2008); this is also apparent in the phylogenetic tree shown in figure 3A, with HDS from Chloroplastida and chromalveolates being most closely related to the sequences from several extant Chlamydiae. One explanation for the mixed origin of the HDS from the Chloroplastida/Chromalveolata and the Rhodophyta/Glaucophyta/Cyanobacteria cluster could be that the cyanobacterial and chlamydial HDS genes coresided in the primary host for some time and that the chlamydia-type gene was independently lost in the glaucophytes and rhodophytes which we have examined, whereas the cyanobacterial gene was lost on at least two occasions in Chloroplastida and the early rhodophytes that became the endosymbionts in chromalveolates.

The enzymes of the two distantly related clades are also distinguished by the presence of an insert of ~260 amino acids that is shared by HDS from chromalveolates, Chloroplastida, Chlamydiae, and several other related bacterial groups (Gupta 2000) but which is absent from available red algal, glaucophyte, and cyanobacterial sequences and the majority of eubacterial HDS enzymes (fig. 3B). The exceptionally large insert, which we here refer to as the “B-domain,” is about the same size as the N-terminal “A-domain” and contains a conserved motif of about 15 amino acids with similarity to the corresponding region of the A-domain. Therefore, the B-domain might have resulted from an internal duplication of the A-domain. Interestingly, the B-domain displays much higher sequence diversity than the A- and C-domains and appears to be conserved only in the enzymes from the oxygenic phototrophs. For the bacterial proteins, we found detectable sequence conservation of the insertion only among closely related species/genera (see supplementary fig. 2, Supplementary Material online). For example, the B-domain of HDS from “*Candidatus* Protochlamydia amoebophila” UWE25, an endosymbiont of amoebae of the genus *Acanthamoeba* (Collingro et al. 2005), shares no recognizable similarity with that from species of the closely related genus *Chlamydia*, whereas, for example, the B-domains in the HDS from the diatom *P. tricornutum* and members of the Chloroplastida are reasonably well conserved; in the HDS from the secondarily nonphotosynthetic Apicomplexa, the B-domain is no longer conserved (supplementary fig. 2, Supplementary Material online). This observation suggests that the B-domain in the HDS of eukaryotic phototrophs has gained a functional significance related to photosynthesis, probably a regulatory role in the formation of carotenoids

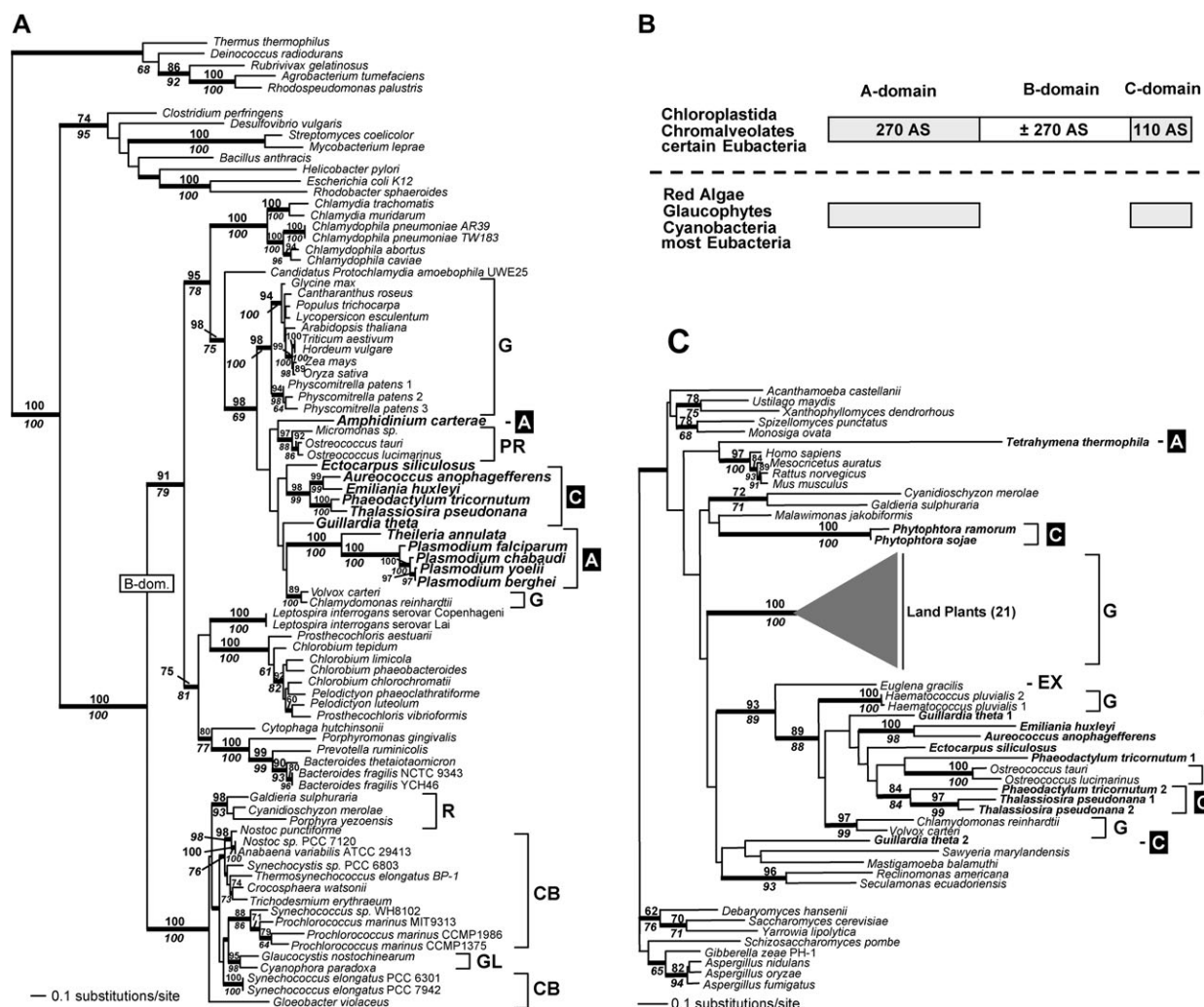


FIG. 3.—(A) Phylogeny of HDS: ML (RAxML) tree constructed from a protein alignment including 382 amino acid positions; rectangle indicates acquisition of the large insertion (B-domain) in HDS, (B) domain structure of the two types of HDS illustrating the location and taxon-specific distribution of the insertion, and (C) phylogeny of IDI: ML (RAxML) tree constructed from a protein alignment including 194 amino acid positions. Display features in (A) and (C) are as in figure 1B; additional abbreviations: A, Alveolata and EX, Excavata.

and related isoprenoids, which puts it under negative selection and thus prevents the rapid accumulation of mutations. HDS from the cryptophyte *G. theta* is an intriguing exception because it apparently has lost a large part of the insertion together with a part of the conserved N-terminal domain (supplementary fig. 2C, Supplementary Material online; see also below).

The yet unknown function of the B-domain appears to be conserved in both the green and the chromalveolate HDS and probably evolved only once. If chromalveolate HDS was transmitted by the red algal endosymbiont, then the functional role of the insertion must have been established before the split of red and green algae. Under the assumption that this function improves the performance of the HDS in some way—making it superior to the insertion-free HDS at least in the corresponding algae—it should have been preferably retained. Then it would be difficult to explain why the three red algae which we sampled contained only cyanobacterial-type HDS without the B-domain. Based on these considerations, we propose that the insertion became

functional only in the green line and that the Chlamydiae-type HDS in chromalveolates derives from a gene transfer involving a green alga which occurred early in chromalveolate evolution.

Isopentenyl Diphosphate Isomerase

In the case of IDI, two phylogenetically unrelated classes of enzymes have been described: Type-1 IDI is present in rhodophytes, chromalveolates, and the Chloroplastida, whereas cyanobacteria contain a type-2 IDI (Bouvier et al. 2005). Thus, the IDI of eukaryotic phototrophs is exceptional among the proteins in our data set because it likely does not trace back to the plastid progenitors but instead was inherited from the host cell (Cunningham and Gantt 2000). Like PDS and HDS, however, the IDI enzymes from phototrophic stramenopiles, the haptophytes, and cryptophytes form a well-supported cluster together with the IDI from green algae (fig. 3C), whereas this protein from land plants, the nonphotosynthetic stramenopiles, the ciliate

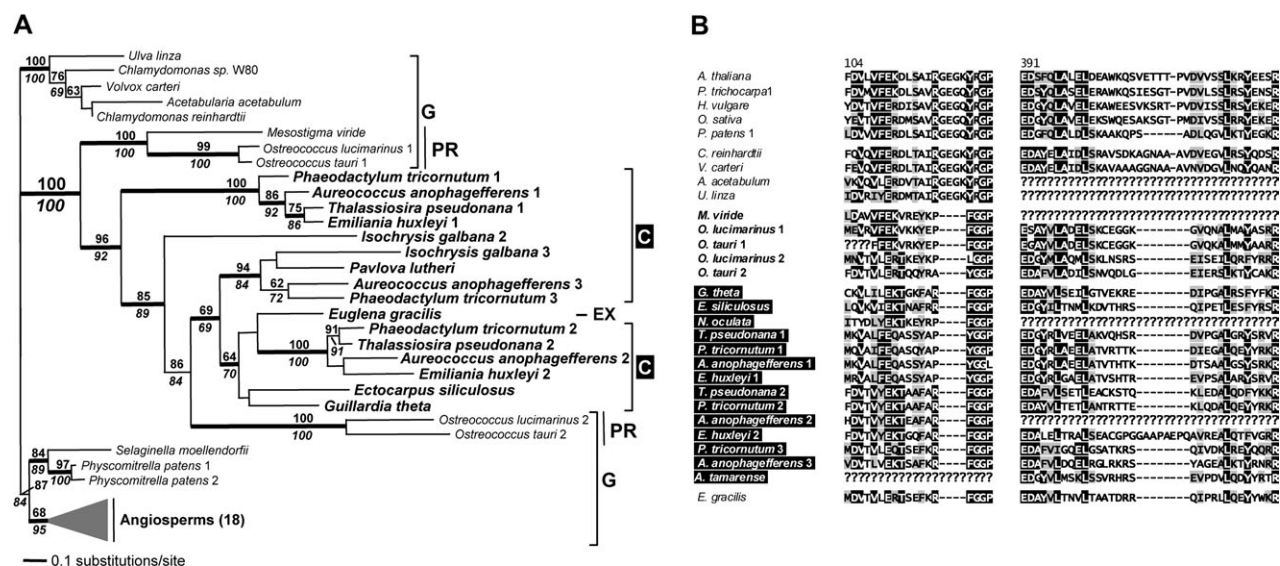


FIG. 4.—(A) Phylogeny of ZEP: ML (RAxML) tree constructed from a protein alignment including 338 amino acid positions and (B) support by the presence of two diagnostic indels specific to chromalveolate and prasinophyte sequences. The depicted sections of a full alignment are indicated by the position of the first shown amino acid in the corresponding protein sequence of *A. thaliana*; at conserved sites, black boxes indicate identical amino acids, whereas gray boxes denote similar amino acids; species names of chromalveolate algae are written in white letters on black background. Display features in (A) are as in figure 1B; additional abbreviation: EX, Excavata.

Tetrahymena thermophila, and the red algae form separate clusters of uncertain relationship. The situation becomes more complex because IDI participates not only in the plastidic MEP pathway but also in the cytosolic formation of isoprenoids for the biosynthesis of sterols and other poly-prenyls via the mevalonate pathway. This explains why genes encoding IDI are present in nonphotosynthetic protists, animals, and fungi. Moreover, there is a tendency for IDI to occur as paralogs in algae and plants, further complicating the interpretation of the phylogenetic tree. Nevertheless, the specific link between IDI from photosynthetic chromists and green algae is significant and indicates that the chromist host may have replaced its original IDI gene with a nuclear gene of green algal provenance.

Chromist Proteins of Green Algal Origin That Lack Red Algal Counterparts

The chromist genes encoding ZEP and violaxanthin de-epoxidase (VDE), the two antagonist enzymes of the photoprotective xanthophyll cycle taking place in the thylakoid membranes (Müller et al. 2001), are also most closely related to prasinophyte homologs (figs. 4A and 5A). We detected neither ZEP nor VDE genes in the genomes of cyanobacteria or the red algae *C. merolae* and *G. sulphuraria*. Both enzymes have been shown to be related to the large and diverse protein family of lipocalins (Bugos et al. 1998; Ganfornina et al. 2000). ZEP is responsible for the formation of violaxanthin by adding an epoxy group to each of the two ionone rings of zeaxanthin (fig. 1A). In the Chloroplastida, violaxanthin is both a light-harvesting pigment and the precursor of neoxanthin (Bouvier et al. 2005). Under conditions of excess light, however, VDE converts violaxanthin back to zeaxanthin which in turn promotes the radiation-less relaxation of

singlet-excited chlorophyll in the light-harvesting antennae through a mechanism that is not yet fully understood (Horton et al. 2005; Niyogi et al. 2005; Szabo et al. 2005); this regulatory dissipation of surplus excitation energy in the photosynthetic apparatus by heat is commonly referred to as nonphotochemical quenching (NPQ; for details, see Müller et al. 2001). In diatoms, haptophytes, and dinoflagellates, the major xanthophyll cycle involves diadinoxanthin and diatoxanthin (Müller et al. 2001; Horton et al. 2005), whereas only small amounts of the pigments of the violaxanthin cycle are normally present (Lohr and Wilhelm 1999). In these algae, violaxanthin is a precursor of diadinoxanthin and of the important light-harvesting pigments fucoxanthin and peridinin (fig. 1A) (Swift et al. 1980; Lohr and Wilhelm 2001).

Whereas angiosperms appear to have only single-copy genes encoding ZEP, the two diatoms, the pelagophyte, the haptophytes, and the two prasinophytes of the genus *Ostreococcus* contain multiple copies of putative ZEP genes (fig. 4A). The encoded gene products form a cluster that is separated from proteins of other green algae and vascular plants. The particularly close relationship of chromist and prasinophyte ZEP is further strengthened by the finding that they share two diagnostic indels (fig. 4B). At least one of these indels is also present in the partial sequences from another basal green alga, the streptophyte *M. viride* (Marin and Melkonian 1999) and the dinoflagellate *Alexandrium tamarensis*. Although the sequence from *A. tamarensis* was too short for reliable phylogenetic analysis, the shared indel suggests that the dinoflagellate ZEP is also closely related to the enzymes from prasinophytes.

Similarly, VDE sequences from the two *Ostreococcus* species cluster with the homologous proteins from stramenopiles, haptophytes, and the dinoflagellate *Amphidinium carterae* (fig. 5A), whereas in the two green algae *V. carteri*

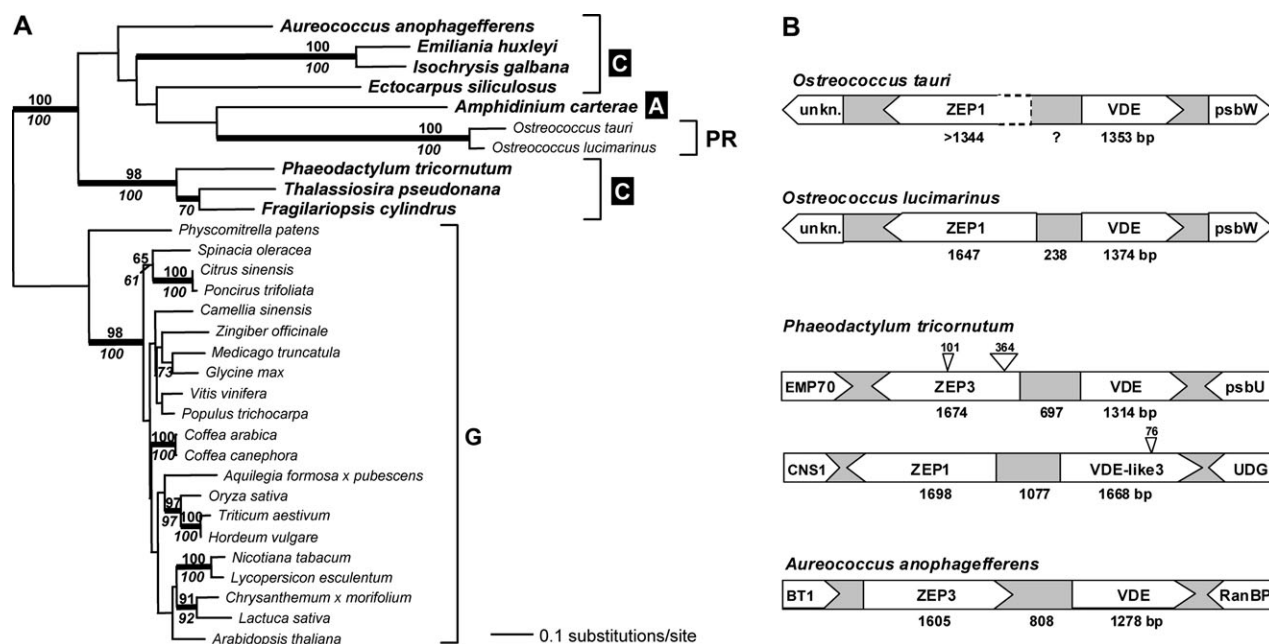


FIG. 5.—(A) Phylogeny of VDE: ML (RAXML) tree constructed from a protein alignment including 256 amino acid positions. (B) In the genomes of the prasinophytes *Ostreococcus lucimarinus* and *Ostreococcus tauri* and stramenopiles *P. tricornutum* and *A. anophagefferens*, one of the zeaxanthin (ZEP) paralogs is located next to the single gene-encoding VDE. In the case of *P. tricornutum*, a second ZEP gene (ZEP1) is adjacent to a gene encoding a VDE-like protein (see text for further details). For *O. tauri*, the question mark indicates that the exact number of base pairs which separate the ORFs of ZEP and VDE is not known due to a sequence gap in the current version of the genome assembly. Open triangles indicate position of introns. Length of ORFs not drawn to scale. Gene abbreviations: BT1, folate/biopterin transporter; CNS1, yeast-CNS1-like protein; EMP70, endomembrane transporter protein; psbU, photosystem II 12-kDa extrinsic protein; psbW, photosystem II subunit W; RanBP, Ran-binding protein; UDG, Uracil-DNA glycosylase; and unk, unknown protein. Display features in (A) are as in figure 1B; additional abbreviation: A, Alveolata.

and *C. reinhardtii*, de-epoxidation of violaxanthin appears to be catalyzed by an as-yet unknown enzyme that is unrelated to VDE from other organisms (Anwarruzzaman et al. 2003). In the case of VDE, we found only unique candidate genes in chromalveolates and Chloroplastida. We detected, however, one gene in Chloroplastida and between three and four genes in the diatoms and the pelagophyte encoding VDE-related proteins (data not shown). The conserved core domains (about 250 amino acids) of the VDE-related proteins display only 15–20% identity with VDE and therefore were not included in the phylogenetic analyses. In particular, a number of amino acid positions that have been identified to be important for catalytic function of VDE (Emanuelsson et al. 2003; Gisselsson et al. 2004) are not conserved in the VDE-like proteins (data not shown). Moreover, in preliminary analyses, we could prove the catalytic function of the recombinant VDE protein from *P. tricornutum*, and initial mutational studies of the enzyme suggest that the replacement of even short motifs in the lipocalin domain is likely to result in an inactive enzyme (Nowoisky J, Herwig S, Lohr M, unpublished data). Therefore, we suggest it unlikely that the VDE-like proteins are involved in xanthophyll cycling.

Another remarkable feature shared by the diatom *P. tricornutum*, the pelagophyte *A. anophagefferens*, and the prasinophytes *O. tauri* and *O. lucimarinus* is the syntenic arrangement of their VDE genes with one of the ZEP paralogs (fig. 5B). The synteny is not observed in *T. pseudonana*, and we found no indication of it in our partial genome assemblies from *E. huxleyi* and *E. siliculosus*.

Whereas in *Ostreococcus* spp. and *P. tricornutum*, the two genes are arranged tail to tail on opposite strands, the ZEP gene in *A. anophagefferens* appears to have changed orientation. Such small inversions involving only one or a few genes have occurred frequently in yeast genome evolution and are believed to be a general feature among eukaryotes (Seoighe et al. 2000). This local synteny is further indication of a close phylogenetic relationship of the respective genes, suggesting a gene transfer event in which this cluster was introduced into an ancestral chromist/chromalveolate from a prasinophyte-like alga. The observation that the tandem arrangement of the two genes has been conserved in several of the algal species during such a long evolutionary period indicates that they probably are coregulated and that this may offer some significant physiological advantage. Furthermore, their tandem arrangement suggests that the ZEP paralog located adjacent to the VDE gene is involved specifically in the xanthophyll cycle, whereas the other paralogs may fulfill a different role, either by catalyzing the epoxidation of carotenoids in the envelope membranes as has been proposed by Wilhelm et al. (2006) or by having acquired a different catalytic function.

It is noteworthy that in the two stramenopiles, the VDE gene is in tandem with a different ZEP paralog than in the two prasinophytes, that is, with ZEP3 instead of ZEP1 (compare figs. 4A and 5B). Interestingly, in the genome of *P. tricornutum*, ZEP1 is located next to a gene encoding one of the VDE-like proteins (fig. 5B). Together with the low but detectable sequence identity, their tandem arrangement with ZEP paralogs strongly suggests that the VDE and

VDE-like genes are paralogs as well. The presence of at least three VDE-like genes and two to three ZEP paralogs in the diatoms and the pelagophyte indicates that probably several duplications of the gene tandem took place which again complicates phylogenetic reconstruction. Looking more closely at the chromist/prasinophyte cluster in the ZEP protein tree in figure 4A, however, it appears likely that the gene duplication leading to ZEP1 and ZEP2/ZEP3 occurred before the segregation of these paralogs into the genomes of extant chromists and prasinophytes as we already have inferred for the PDS paralogs. In our view, this argues for an ancient acquisition of the ZEP1 and the ZEP2/ZEP3 paralogs in chromists from an early-diverging prasinophyte.

Their absence from cyanobacterial genomes suggests that the xanthophyll cycle enzymes originated in a photosynthetic eukaryote sometime after the primary endosymbiosis. It is unclear, however, whether this occurred before or after the split of red and green algae. Red algae are generally thought to contain zeaxanthin or lutein as their only xanthophylls. Yet, there have been several reports of the presence of the epoxy xanthophylls antheraxanthin and violaxanthin in some red algae (Marquardt and Hanelt 2004; Schubert N, Garcia-Mendoza E, and Pacheco-Ruiz I 2006; and references therein), and the operation of a xanthophyll cycle in rhodophytes of the derived genus *Gracilaria* has been suggested (Ursi et al. 2003); however, this has not been rigorously established. In other cases, no light-dependent changes in the concentration of epoxy xanthophylls in the corresponding red algae could be observed (Marquardt and Hanelt 2004; Schubert H, Andersson M, and Snoeijs P 2006). Based on current evidence, it is questionable whether a xanthophyll cycle-related NPQ would be compatible with a phycobilisome-based light-harvesting machinery. Recently, a completely different energy quenching mechanism has been found in cyanobacteria involving a water-soluble carotenoprotein that appears to interact specifically with the phycobilisomes (Wilson et al. 2006, 2008). In red algae, photoprotection against high light is not well understood, but results of fluorescence quenching studies suggested that ΔpH -dependent quenching in the PSII reaction centers may be an important component, instead (Delphin et al. 1998).

It is now also recognized that de-epoxy xanthophylls have to bind to particular light-harvesting complexes (LHCs) that mediate the increased thermal dissipation of excess excitation energy in the photosynthetic antennae. In vascular plants, the LHC protein PsbS has been implicated as an essential component in this process (Niyogi et al. 2005) but thus far PsbS has only been found in chlorophytes suggesting that it may have evolved specifically in the green lineage (Koziol et al. 2007). Recent data from the diatom *Cyclotella cryptica* (Beer et al. 2006; Gundermann and Büchel 2008) and the green alga *C. reinhardtii* (Peers et al. 2007) indicate that another type of LHC protein named LI818 or LHCSR is also involved in xanthophyll-related NPQ. LI818 is the only LHC protein known thus far that is present in the green and chromalveolate algae but missing from the available red algal genomes (Richard et al. 2000; Green 2003; Koziol et al. 2007). Taken together, these findings are compatible with the idea that

in chromalveolates the genes encoding both components necessary for xanthophyll cycle-based NPQ, that is, those of the xanthophyll cycle enzymes VDE and ZEP as well as those encoding the LI818 proteins have been acquired from a prasinophyte alga.

We currently have no information about enzymes involved in the last steps of xanthophyll formation in chromalveolates. However, there are conspicuous structural features shared between some major light-harvesting xanthophylls that are found specifically in prasinophytes of the order Mamiellales (Egeland et al. 1997) and those of chromalveolates (Young 1993), suggesting a biochemical link between the two groups (fig. 6). In light of our findings, it is possible that the formation of these xanthophylls in the chromalveolates and basal green algae is catalyzed by orthologous enzymes and that the genes encoding these enzymes may have originated in a primitive prasinophyte. Consequently, it appears that chromalveolates may have recruited genes with green algal ancestry not only for the synthesis of the core components of the xanthophyll cycle, a pathway critical for protecting the cells from excess light absorption, but also for the formation of new light-harvesting pigments with increased efficiency to absorb green light which dominates lower regions of the water column.

Evolutionary Scenarios to Explain Green Gene Acquisition and Retention by Chromists/Chromalveolates

Although none of the chromist proteins of putative green algal provenance alone provides unequivocal data, we think that the combination of our data generates sufficient circumstantial evidence to suggest that at least haptophytes and stramenopiles and putatively also cryptophytes (with the limitation of incomplete sequence data from *G. theta*) share a common set of green genes of prasinophyte derivation. Moreover, we observe the same phylogenetic mosaic of genes in the diatoms, the pelagophyte, the haptophyte, and the cryptophyte, suggesting monophyly of chromist host cells. Our results for HDS, ZEP, and VDE indicate that this hypothesis may include dinoflagellates as well, thus supporting the monophyly of these chromalveolates. This provisional result needs however to be substantiated with improved sequence coverage and taxon sampling.

It has been suggested that the red algal endosymbiont of chromalveolates was most closely affiliated to extant rhodophytes of the order Porphyridiales (Shalchian-Tabrizi et al. 2006). Given our limited taxon sampling among red algae, it could be argued that some of the putative green genes as, for example, those encoding ZEP and VDE were introduced by the red alga which became the chromalveolate endosymbiont and that descendants of this alga may exist among extant Porphyridiales, containing ZEP or VDE genes more closely related to the respective genes from chromalveolates than are those of primitive green algae. Based on the phylogenetic trees of ZEP and VDE in figures 4A and 5A, this would imply that the proteins from *Ostreococcus* are more closely related to the red algal homologs than to those of the other Chloroplastida, that is, a situation that we did not observe in any of the protein trees

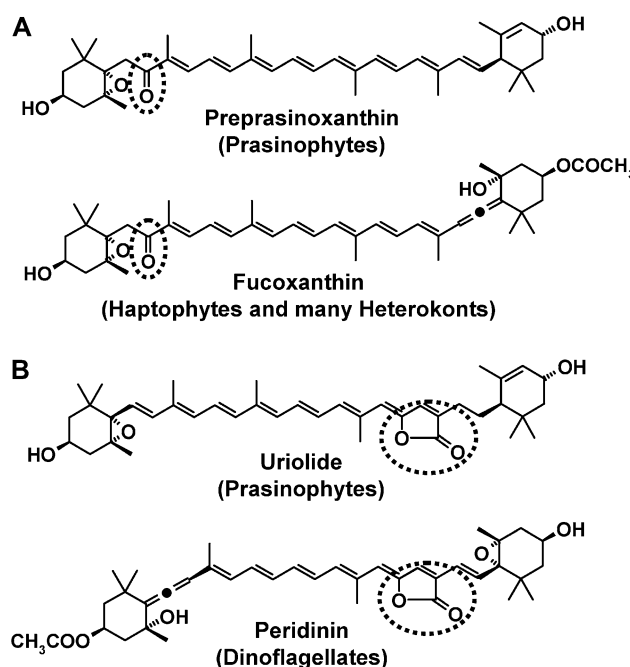


FIG. 6.—Conspicuous structural motifs are shared by xanthophylls from chromalveolates (Young 1993) and prasinophyte algae of the order Mamiellales (Egeland et al. 1997). (A) A unique keto group (indicated with dotted circle) is found at C8 in preprasinoxanthin from Mamiellales and in fucoxanthin from, for example, diatoms and haptophytes, respectively; (B) a lactone ring (indicated with dotted circle) is present in uriolide from Mamiellales and in peridinin from dinoflagellates, respectively (see also fig. 1A).

containing sequences from rhodophytes, prasinophytes, and other Chloroplastida. Therefore, we consider this explanation as unlikely.

Based on our findings and the reasoning above, we suggest the following scenarios as being most plausible for explaining the evolution of carotenoid biosynthesis in chromalveolate algae. Under the first scenario, a heterotrophic host cell engulfed a red algal endosymbiont but retained the capacity for phagocytosis and was able to feed on other algae that served as a source of foreign genes that could potentially replace host homologs. This would imply that genes in the chromalveolates that are related to those of prasinophytes were acquired by multiple, independent HGT events. Alternatively, a nonphotosynthetic protist may have engulfed and retained an early-diverging green alga, precipitating intracellular gene transfer to the host nucleus. This “cryptic” endosymbiont was superseded by a subsequent red algal endosymbiosis that triggered the loss of most green algal nuclear genes through homologous gene replacement. Under this scenario, the only relic genes of the green plastid phase of chromalveolate evolution would be those that have survived gene replacement (i.e., gene-for-gene) by red algal homologs or those that confer unique plastid properties that were not present in the new endosymbiont. Both these categories of course also apply to genes of nonplastid function.

Our data would also be compatible with the hypothesis of multiple plastid acquisitions among chromalveolates (e.g., Boudry 2005; Sanchez-Puerta et al. 2007). Even in this case, however, the scenario above would remain valid but limited to the most ancient photosynthetic chromalveolate lineage which then could have donated its plastids and nuclear genes encoding plastid proteins to other chromalveo-

late lineages via subsequent endosymbioses. Based on recent data that until now fail to provide robust support for chromalveolate monophyly but do favor the two groups SAR and HC (Burki et al. 2007, 2008; Hackett et al. 2007), one possible scenario would be that the green algal genes were introduced first by endosymbiotic gene transfer into the ancestor of the HC clade and then transferred via secondary endosymbiosis from an early-diverging alga of the HC clade to a member of the SAR group either before or shortly after the split of Rhizaria from stramenopiles and alveolates. In the former case, Rhizaria would subsequently have lost the HC endosymbiont/plastid and probably all photosynthesis-related genes, thus excluding the possibility to distinguish between the two alternatives based on analyses of plastid proteins like those involved in carotenoid biosynthesis. However, because current analyses of host-derived cytosolic proteins continue to generate conflicting results regarding the deep branching order of the eukaryote tree (Burki et al. 2007, 2008; Hackett et al. 2007; Patron et al. 2007; Rodríguez-Ezpeleta et al. 2007; Yoon et al. 2008), it appears premature to favor either single or serial acquisition of the green algal carotenogenic genes by chromalveolates.

As another potential scenario, it has been proposed that a primitive green alga may have served as the secondary host in the evolution of stramenopiles, that is, the host was already photosynthetic and would have lost its primary plastids either before or soon after the secondary endosymbiosis (Häuber et al. 1994; Nozaki et al. 2003, 2007; Nozaki 2005). In such a situation, the host cell nucleus may have already contained genes derived from a primary cyanobacterial endosymbiont, as suggested in the case of the gene for host-encoded cytosolic phosphogluconate dehydrogenase, which may have a cyanobacterial origin in plants, green,

red, and stramenopile algae as well as in a number of non-photosynthetic protists (Andersson and Roger 2002). However, recent analyses of host cell phylogeny, based on extensive protein data, suggest monophyly of all primary photosynthetic eukaryotes (i.e., the Plantae) to the exclusion of lineages that resulted from secondary endosymbiotic events (Rodríguez-Ezpeleta et al. 2005; Hackett et al. 2007; Burki et al. 2008). In light of these studies, it appears unlikely that chromophyte genes with green algal ancestry are remnants of a primary plastid.

The HGT hypothesis is used most frequently to explain the origin of genes in one lineage from multiple donors that are phylogenetically distantly related to each other (e.g., Nosenko and Bhattacharya 2007). Our data indicate however that all the green genes involved in carotenoid biosynthesis in chromalveolates originated from a specific group of basal green algae. This is consistent with the idea (not proving it) that they were acquired from a point source (i.e., a single donor). This hypothesis is well supported by our phylogenetic and diagnostic indel data for PDS and ZEP (figs. 2 and 4); however, taxon sampling among green algal species is not yet broad enough for HDS, IDI, and VDE to make the same argument. An alternative explanation for our data based on the HGT model is that prasinophytes were common prey for ancient chromalveolates and the foreign sequences arose via a gene transfer ratchet. In either case, our data suggest that the green genes encoding HDS and VDE have been acquired before the split of dinoflagellates, haptophytes, and stramenopiles, that is, early in the evolution of chromalveolates, and that the VDE gene was introduced in tandem with the prasinophyte-derived ZEP genes.

Based on the accumulating evidence, we currently favor the scenario under which the green algal genes were acquired from an ancient cryptic green algal endosymbiont. This explanation is not arbitrary but rather recognizes the fact that dinoflagellates have a documented history of plastid replacement that involves green and other types of algae. In the case of dinoflagellates such as *Karenia brevis* and *Karlodinium micrum*, the secondary plastid of red algal origin was replaced by a tertiary haptophyte-derived plastid that led to the concomitant substitution of many of the nucleus-encoded plastid proteins (for discussion, see Ishida and Green 2002; Nosenko et al. 2006; Patron et al. 2006). Nosenko et al. (2006) also identified a surviving gene from the red plastid phase of *K. brevis*, a plastid targeted form of glyceraldehyde-3-phosphate dehydrogenase (*gapCI-pd*) that is shared by peridinin-containing dinoflagellates. This is the type of evolutionary scenario we assume also happened early in chromalveolate evolution resulting in the replacement of the green algal plastid (and associated nuclear genes) with one of red algal derivation.

Given these ideas, why then would genes from a hypothetical prasinophyte plastid phase be retained by chromalveolates? In the case of the genes encoding components of the xanthophyll cycle, the answer appears to be obvious because this photoprotective mechanism ensures a high photosynthetic efficiency under fluctuating light (Müller et al. 2001; Horton et al. 2005; Lavaud 2007). The argument of functional significance may also hold true for HDS because

the insertion in the green protein is well conserved only in eukaryotic phototrophs (see above), suggesting that it has a regulatory role in the formation of carotenoids and related isoprenoids that is specific to these organisms. HDS from the cryptophyte *G. theta* is an intriguing exception because it apparently has lost a large part of the insertion together with a part of the conserved N-terminal domain (supplementary fig. 2C, Supplementary Material online). Although this may simply be coincidence, it is remarkable that similar to red algae, cryptophytes still use phycobiliproteins as light-harvesting components and lack a xanthophyll cycle. Thus, it is tempting to speculate the existence of a regulatory (feedback) mechanism involving HDS and components of the xanthophyll cycle or specific light-harvesting proteins. Based on our analyses, biochemical studies aimed at understanding the significance of the insertions in HDS and also PDS are crucial to either support or reject this hypothesis.

Finally, the high number of green proteins in the carotenoid biosynthesis pathway suggests that green genes might significantly contribute to other biosynthetic pathways within chromalveolate plastids. With the growing number of algal genome sequences, issues of evolutionary origins of individual genes and the mosaic nature of the nuclear gene content of the chromalveolates can be addressed using whole-genome comparisons. Such studies should be able to generate sufficient data to either confirm or reject the hypothesis of a cryptic green endosymbiont that we suggest has left indelible footprints in the genomes of extant chromalveolates.

Supplementary Material

Supplementary figures 1 and 2 and tables 1 and 2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>). The GenBank accession numbers for the sequences reported in this paper are FJ175667–FJ175685.

Acknowledgments

We thank the Kazusa DNA Research Institute (Chiba, Japan) for supplying *P. yezoensis* EST clones; L. Doderer, J. Nowoisky, and G. Schmidt for help with some gene assemblies; T. Burmester for discussion of the phylogenetic analyses; and two anonymous reviewers for careful reading and constructive criticism of the manuscript. Part of this work was supported by grants from the German Science Foundation (LO840/1-1 to M.L. and SFB-TR1 to U.G.M.), from the University Mainz (Forschungsfonds 2004 and 2006) awarded to M.L., from the Free State of Saxonia (Germany) awarded to R.F., and from the NSF (United States) awarded to A.R.G. (MCB-0235878) and D.B. (EF 04-31117, EF 06-25440).

Literature Cited

- Adl SM, Simpson AGB, Farmer MA, et al. (28 co-authors). 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol.* 52:399–451.

- Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Andersson JO, Roger AJ. 2002. A cyanobacterial gene in nonphotosynthetic protists—an early chloroplast acquisition in eukaryotes? *Curr Biol.* 12:115–119.
- Anwaruzzaman M, Chin BL, Li XP, Lohr M, Martinez DA, Niyogi KK. 2004. Genomic analysis of mutants affecting xanthophyll biosynthesis and regulation of photosynthetic light harvesting in *Chlamydomonas reinhardtii*. *Photosynth Res.* 82:265–276.
- Archibald JM. 2005. Jumping genes and shrinking genomes—probing the evolution of eukaryotic photosynthesis with genomics. *IUBMB Life.* 57:539–547.
- Armbrust EV, Berges JA, Bowler C, et al. (45 co-authors). 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science.* 306:79–86.
- Asamizu E, Nakajima M, Kitade Y, Saga N, Nakamura Y, Tabata S. 2003. Comparison of RNA expression profiles between the two generations of *Porphyra yezoensis* (Rhodophyta), based on expressed sequence tag frequency analysis. *J Phycol.* 39:923–930.
- Bachvaroff TR, Sanchez-Puerta MV, Delwiche CF. 2005. Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Mol Biol Evol.* 22:1772–1782.
- Barbier G, Oesterhelt C, Larson MD, Halgren RG, Wilkerson C, Garavito RM, Benning C, Weber APM. 2005. Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. *Plant Physiol.* 137:460–474.
- Beer A, Gundermann K, Beckmann J, Büchel C. 2006. Subunit composition and pigmentation of fucoxanthin-chlorophyll proteins in diatoms: evidence for a subunit involved in diadinoxanthin and diatoxanthin binding. *Biochemistry.* 45:13046–13053.
- Bhattacharya D, Yoon HS, Hackett JD. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *Bioessays.* 26:50–60.
- Bodyl A. 2005. Do plastid-related characters support the chromalveolate hypothesis? *J Phycol.* 41:712–719.
- Bouvier F, Rahier A, Camara B. 2005. Biogenesis, molecular regulation and function of plant isoprenoids. *Prog Lipid Res.* 44:357–429.
- Bugos RC, Hieber AD, Yamamoto HY. 1998. Xanthophyll cycle enzymes are members of the lipocalin family, the first identified from plants. *J Biol Chem.* 273:15321–15324.
- Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J Mol Biol.* 268:78–94.
- Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland Å, Nikolaev SI, Jakobsen KS, Pawlowski J. 2007. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE.* 2:e790.
- Burki F, Shalchian-Tabrizi K, Pawlowski J. 2008. Phylogenomics reveals a new ‘megagroup’ including most photosynthetic eukaryotes. *Biol Lett.* 4:366–369.
- Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol.* 46:347–366.
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31:3497–3500.
- Collingro A, Toenshoff ER, Taylor MW, Fritsche TR, Wagner M, Horn M. 2005. ‘*Candidatus* Protochlamydia amoebophila’, an endosymbiont of *Acanthamoeba* spp. *Int J Syst Evol Microbiol.* 55:1863–1866.
- Cunningham FX, Gantt E. 1998. Genes and enzymes of carotenoid biosynthesis in plants. *Annu Rev Plant Physiol Plant Mol Biol.* 49:557–583.
- Cunningham FX, Gantt E. 2000. Identification of multi-gene families encoding isopentenyl diphosphate isomerase in plants by heterologous complementation in *Escherichia coli*. *Plant Cell Physiol.* 41:119–123.
- Delphin E, Duval JC, Etienne AL, Kirilovsky D. 1998. Delta pH-dependent photosystem II fluorescence quenching induced by saturating, multiturnover pulses in red algae. *Plant Physiol.* 118:103–113.
- Derelle E, Ferraz C, Rombauts S, et al. (26 co-authors). 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci USA.* 103:11647–11652.
- Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu XN, Reith M, Cavalier-Smith T, Maier UG. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature.* 410:1091–1096.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Egeland ES, Guillard RRL, Liaaen-Jensen S. 1997. Algal carotenoids .63. Carotenoids from Prasinophyceae .7. Additional carotenoid prototype representatives and a general chemosystematic evaluation of carotenoids in Prasinophyceae (Chlorophyta). *Phytochemistry.* 44:1087–1097.
- Emanuelsson A, Eskling M, Åkerlund HE. 2003. Chemical and mutational modification of histidines in violaxanthin de-epoxidase from *Spinacia oleracea*. *Physiol Plant.* 119:97–104.
- Fast NM, Kissinger JC, Roos DS, Keeling PJ. 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol.* 18:418–426.
- Ganformina MD, Gutierrez G, Bastiani M, Sanchez D. 2000. A phylogenetic analysis of the lipocalin protein family. *Mol Biol Evol.* 17:114–126.
- Gisselsson A, Szilágyi A, Åkerlund HE. 2004. Role of histidines in the binding of violaxanthin de-epoxidase to the thylakoid membrane as studied by site-directed mutagenesis. *Physiol Plant.* 122:337–343.
- Green BR. 2003. The evolution of light-harvesting antennas. In: Green BR, Parson WW, editors. *Light-harvesting antennas in photosynthesis*. Dordrecht (The Netherlands): Kluwer Academic Publishers. p. 129–168.
- Grossman A, Manodori A, Snyder D. 1990. Light-harvesting proteins of diatoms: their relationship to the chlorophyll a/b binding proteins of higher plants and their mode of transport into plastids. *Mol Gen Genet.* 224:91–100.
- Grossman AR, Lohr M, Im CS. 2004. *Chlamydomonas reinhardtii* in the landscape of pigments. *Annu Rev Genet.* 38:119–173.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
- Gundermann K, Büchel C. 2008. The fluorescence yield of the trimeric fucoxanthin–chlorophyll–protein FCPa in the diatom *Cyclotella meneghiniana* is dependent on the amount of bound diatoxanthin. *Photosynth Res.* 95:229–235.
- Gupta RS. 2000. The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol Rev.* 24:367–402.
- Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D. 2007. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the

- association of Rhizaria with chromalveolates. *Mol Biol Evol.* 24:1702–1713.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.
- Harper JT, Keeling PJ. 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol Evol.* 20:1730–1735.
- Harper JT, Waanders E, Keeling PJ. 2005. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. *Int J Syst Evol Microbiol.* 55:487–496.
- Häuber MM, Müller SB, Speth V, Maier UG. 1994. How to evolve a complex plastid—a hypothesis. *Bot Acta.* 107:383–386.
- Horton P, Wentworth M, Ruban A. 2005. Control of the light harvesting function of chloroplast membranes: the LHCII-aggregation model for non-photochemical quenching. *FEBS Lett.* 579:4201–4206.
- Huang JL, Gogarten JP. 2007. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* 8:R99.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17:754–755.
- Ishida K, Green BR. 2002. Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc Natl Acad Sci USA.* 99:9294–9299.
- Jiroutova K, Horak A, Bowler C, Oborník M. 2007. Tryptophan biosynthesis in stramenopiles: eukaryotic winners in the diatom complex chloroplast. *J Mol Evol.* 65:496–511.
- Keeling PJ. 2004. Diversity and evolutionary history of plastids and their hosts. *Am J Bot.* 91:1481–1493.
- Khan H, Parks N, Kozera C, Curtis BA, Parsons BJ, Bowman S, Archibald JM. 2007. Plastid genome sequence of the cryptophyte alga *Rhodomonas salina* CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. *Mol Biol Evol.* 24:1832–1842.
- Kozioł AG, Borza T, Ishida KI, Keeling P, Lee RW, Dumford DG. 2007. Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. *Plant Physiol.* 143:1802–1816.
- Lange BM, Ghassemian M. 2003. Genome organization in *Arabidopsis thaliana*: a survey for genes involved in isoprenoid and chlorophyll metabolism. *Plant Mol Biol.* 51:925–948.
- Lange BM, Rujan T, Martin W, Croteau R. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci USA.* 97:13172–13177.
- Lavaud J. 2007. Fast regulation of photosynthesis in diatoms: mechanisms, evolution and ecophysiology. *Funct Plant Sci Biotechnol.* 1:267–287.
- Li SL, Nosenko T, Hackett JD, Bhattacharya D. 2006. Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. *Mol Biol Evol.* 23:663–674.
- Lohr M, Im CS, Grossman AR. 2005. Genome-based examination of chlorophyll and carotenoid biosynthesis in *Chlamydomonas reinhardtii*. *Plant Physiol.* 138:490–515.
- Lohr M, Wilhelm C. 1999. Algae displaying the diadinoxanthin cycle also possess the violaxanthin cycle. *Proc Natl Acad Sci USA.* 96:8784–8789.
- Lohr M, Wilhelm C. 2001. Xanthophyll synthesis in diatoms: quantification of putative intermediates and comparison of pigment conversion kinetics with rate constants derived from a model. *Planta.* 212:382–391.
- Marin B, Melkonian M. 1999. Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons. *Protist.* 150:399–417.
- Marquardt J, Hanelt D. 2004. Carotenoid composition of *Delesseria lanciafolia* and other marine red algae from polar and temperate habitats. *Eur J Phycol.* 39:285–292.
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA.* 99:12246–12251.
- Matsuzaki M, Misumi O, Shin-I T, et al. (42 co-authors). 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature.* 428:653–657.
- Meyer-Gauen G, Herbrand H, Pahnke J, Cerff R, Martin W. 1998. Gene structure, expression in *Escherichia coli* and biochemical properties of the NAD(+)–dependent glyceraldehyde-3-phosphate dehydrogenase from *Pinus sylvestris* chloroplasts. *Gene.* 209:167–174.
- Moreira D, Le Guyader H, Philippe H. 2000. The origin of red algae and the evolution of chloroplasts. *Nature.* 405:69–72.
- Morosinotto T, Caffarri S, Dall'Osto L, Bassi R. 2003. Mechanistic aspects of the xanthophyll dynamics in higher plant thylakoids. *Physiol Plant.* 119:347–354.
- Moustafa A, Reyes-Prieto A, Bhattacharya D. 2008. Chlamydiae have contributed over sixty genes to plantae with a variety of different functions. *PLoS ONE.* 3:e2205.
- Müller P, Li XP, Niyogi KK. 2001. Non-photochemical quenching. A response to excess light energy. *Plant Physiol.* 125:1558–1566.
- Niyogi KK, Li XP, Rosenberg V, Jung HS. 2005. Is PsbS the site of non-photochemical quenching in photosynthesis? *J Exp Bot.* 56:375–382.
- Nosenko T, Bhattacharya D. 2007. Horizontal gene transfer in chromalveolates. *BMC Evol Biol.* 7:173.
- Nosenko T, Lidie KL, Van Dolah FM, Lindquist E, Cheng JF, Bhattacharya D. 2006. US Department of Energy-Joint Genome Institute, Chimeric plastid proteome in the Florida “red tide” dinoflagellate *Karenia brevis*. *Mol Biol Evol.* 23:2026–2038.
- Not F, Valentin K, Romari K, Lovejoy C, Massana R, Tobe K, Vaulot D, Medlin LK. 2007. Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science.* 315:253–255.
- Nozaki H. 2005. A new scenario of plastid evolution: plastid primary endosymbiosis before the divergence of the “Plantae,” emended. *J Plant Res.* 118:247–255.
- Nozaki H, Iseki M, Hasegawa M, Misawa K, Nakada T, Sasaki N, Watanabe M. 2007. Phylogeny of primary photosynthetic eukaryotes as deduced from slowly evolving nuclear genes. *Mol Biol Evol.* 24:1592–1595.
- Nozaki H, Matsuzaki M, Takahara M, Misumi O, Kuroiwa H, Hasegawa M, Shin-I T, Kohara Y, Ogasawara N, Kuroiwa T. 2003. The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *J Mol Evol.* 56:485–497.
- Oborník M, Green BR. 2005. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol Biol Evol.* 22:2343–2353.
- Palenik B, Grimwood J, Aerts A, et al. (38 co-authors). 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci USA.* 104:7705–7710.
- Palmer JD. 2003. The symbiotic birth and spread of plastids: how many times and whodunit? *J Phycol.* 39:4–11.

- Patron NJ, Inagaki Y, Keeling PJ. 2007. Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. *Curr Biol.* 17:887–891.
- Patron NJ, Rogers MB, Keeling PJ. 2004. Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. *Eukaryot Cell.* 3:1169–1175.
- Patron NJ, Waller RF, Keeling PJ. 2006. A tertiary plastid uses genes from two endosymbionts. *J Mol Biol.* 357:1373–1382.
- Peers G, Truong T, Elrad D, Grossman A, Niyogi K. 2007. A non-photochemical quenching mutant of *Chlamydomonas* reveals a role for L1818/LHCSR proteins. *Photosynth Res.* 91:250.
- Reyes-Prieto A, Bhattacharya D. 2007. Phylogeny of Calvin cycle enzymes supports Plantae monophyly. *Mol Phylogenet Evol.* 45:384–391.
- Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D. 2006. Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. *Curr Biol.* 16:2320–2325.
- Rice DW, Palmer JD. 2006. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biol.* 4:31.
- Richard C, Ouellet H, Guertin M. 2000. Characterization of the LI1818 polypeptide from the green unicellular alga *Chlamydomonas reinhardtii*. *Plant Mol Biol.* 42:303–316.
- Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, McLeod R, Roberts CW. 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot Cell.* 5:1517–1531.
- Rodríguez-Concepción M, Boronat A. 2002. Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. *Plant Physiol.* 130:1079–1089.
- Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol.* 15:1325–1330.
- Rodríguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H, Lang BF. 2007. Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr Biol.* 17:1420–1425.
- Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 2007. The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol Biol Evol.* 24:54–62.
- Rohmer M. 2003. Mevalonate-independent methylerythritol phosphate pathway for isoprenoid biosynthesis. Elucidation and distribution. *Pure Appl Chem.* 75:375–387.
- Sanchez-Puerta MV, Bachvaroff TR, Delwiche CF. 2007. Sorting wheat from chaff in multi-gene analyses of chlorophyll c-containing plastids. *Mol Phylogenet Evol.* 44:885–897.
- Sandmann G. 2002. Molecular evolution of carotenoid biosynthesis from bacteria to plants. *Physiol Plant.* 116:431–440.
- Schubert H, Andersson M, Snoeijis P. 2006. Relationship between photosynthesis and non-photochemical quenching of chlorophyll fluorescence in two red algae with different carotenoid compositions. *Mar Biol.* 149:1003–1013.
- Schubert N, García-Mendoza E, Pacheco-Ruiz I. 2006. Carotenoid composition of marine red algae. *J Phycol.* 42:1208–1216.
- Seoighe C, Federspiel N, Jones T, et al. (20 co-authors). 2000. Prevalence of small inversions in yeast gene order evolution. *Proc Natl Acad Sci USA.* 97:14433–14437.
- Shalchian-Tabrizi K, Skånseng M, Ronquist F, Klaveness D, Bachvaroff TR, Delwiche CF, Botnen A, Tengs T, Jakobsen KS. 2006. Heterotachy processes in rhodophyte-derived secondhand plastid genes: implications for addressing the origin and evolution of dinoflagellate plastids. *Mol Biol Evol.* 23:1504–1515.
- Stamatakis A. 2006. RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 22:2688–2690.
- Sterky F, Bhalerao RR, Unneberg P, et al. (19 co-authors). 2004. A *Populus* EST resource for plant functional genomics. *Proc Natl Acad Sci USA.* 101:13951–13956.
- Swift IE, Milborrow BV, Jeffrey SW. 1980. Formation of neoxanthin, diadinoxanthin and peridinin from [¹⁴C]zeaxanthin by a cell-free system from *Amphidinium carterae*. *Phytochemistry.* 21:2859–2864.
- Szabo I, Bergantino E, Giacometti GM. 2005. Light and oxygenic photosynthesis: energy dissipation as a protection mechanism against photo-oxidation. *EMBO Rep.* 6:629–634.
- Tyra HM, Linka M, Weber AP, Bhattacharya D. 2007. Host origin of plastid solute transporters in the first photosynthetic eukaryotes. *Genome Biol.* 8:R212.
- Ursi S, Pedersen M, Plastino E, Snoeijis P. 2003. Intraspecific variation of photosynthesis, respiration and photoprotective carotenoids in *Gracilaria birdiae* (Gracilariales: Rhodophyta). *Mar Biol.* 142:997–1007.
- Wilhelm C, Büchel C, Fisahn J, et al. (12 co-authors). 2006. The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. *Protist.* 157:91–124.
- Wilson A, Ajlani G, Verbavatz JM, Vass I, Kerfeld CA, Kirilovskya D. 2006. A soluble carotenoid protein involved in phycobilisome-related energy dissipation in cyanobacteria. *Plant Cell.* 18:992–1007.
- Wilson A, Punginelli C, Gall A, et al. (11 co-authors). 2008. A photoactive carotenoid protein acting as light intensity sensor. *Proc Natl Acad Sci USA.* 105:12075–12080.
- Yoon HS, Grant J, Tekle Y, Wu M, Chaon B, Cole J, Logsdon J, Patterson D, Bhattacharya D, Katz L. 2008. Broadly sampled multigene trees of eukaryotes. *BMC Evol Biol.* 8:14.
- Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. *Proc Natl Acad Sci USA.* 99:15507–15512.
- Young AJ. 1993. Occurrence and distribution of carotenoids in photosynthetic systems. In: Young AJ, Britton G, editors. *Carotenoids in photosynthesis*. London: Chapman & Hall. p. 16–71.

Martin Embley, Associate Editor

Accepted September 11, 2008