Evolution of Glutamine Synthetase in Heterokonts: Evidence for Endosymbiotic Gene Transfer and the Early Evolution of Photosynthesis

Deborah L. Robertson and Aurélien Tartar

Biology Department, Clark University

Although the endosymbiotic evolution of chloroplasts through primary and secondary associations is well established, the evolutionary timing and stability of the secondary endosymbiotic events is less well resolved. Heterokonts include both photosynthetic and nonphotosynthetic members and the nonphotosynthetic lineages branch basally in phylogenetic reconstructions. Molecular and morphological data indicate that heterokont chloroplasts evolved via a secondary endosymbiosis, involving a heterotrophic host cell and a photosynthetic ancestor of the red algae and this endosymbiotic event may have preceded the divergence of heterokonts and alveolates. If photosynthesis evolved early in this lineage, nuclear genomes of the nonphotosynthetic groups may contain genes that are not essential to photosynthesis but were derived from the endosymbiont genome through gene transfer. These genes offer the potential to trace the evolutionary history of chloroplast gains and losses within these lineages.

Glutamine synthetase (GS) is essential for ammonium assimilation and glutamine biosynthesis in all organisms. Three paralogous gene families (GSI, GSII, and GSIII) have been identified and are broadly distributed among prokaryotic and eukaryotic lineages. In diatoms (Heterokonta), the nuclear-encoded chloroplast and cytosolic-localized GS isoforms are encoded by members of the GSII and GSIII family, respectively. Here, we explore the evolutionary history of GSII in both photosynthetic and nonphotosynthetic heterokonts, red algae, and other eukaryotes. GSII cDNA sequences were obtained from two species of oomycetes by polymerase chain reaction amplification. Additional GSII sequences from eukaryotes and bacteria were obtained from publicly available databases and genome projects. Bayesian inference and maximum likelihood phylogenetic analyses of GSII provided strong support for the monophyly of heterokonts, rhodophytes, chlorophytes, and plants and strong to moderate support for the Opisthokonts. Although the phylogeny is reflective of the unikont/bikont division of eukaryotes, we propose based on the robustness of the phylogenetic analyses that the heterokont GSII gene evolved via endosymbiotic gene transfer from the nucleus of the red-algal endosymbiont to the nucleus of the host. The lack of GSIII sequences in the oomycetes examined here further suggests that the GSIII gene that functions in the cytosol of photosynthetic heterokonts was replaced by the endosymbiont-derived GSII gene.

Introduction

Endosymbiotic associations have played significant roles in the evolution of photosynthetic eukaryotes (reviewed in Cavalier-Smith 2002; Bhattacharya,Yoon, and Hackett 2003; Keeling 2004). Three major lineages of photosynthetic eukaryotes (Glaucophyta, Rhodophyta [red algae], and Plantae [green algae, bryophytes, and tracheophytes]) evolved via primary endosymbiosis involving a cyanobacterium-like ancestor and a eukaryotic host. Following primary endosymbiosis and during the evolution of green and red algae, secondary endosymbiosis (involving photosynthetic eukaryotic symbionts and heterotrophic eukaryotic hosts) contributed to diverse lineages of photosynthetic eukaryotes, including euglenoids, chlorarachnids, cryptomonads, haptophytes, heterokonts, and alveolates (Cavalier-Smith 2002; Bhattacharya,Yoon, and Hackett 2003; Keeling 2004). The evolution of photosynthetic eukaryotes with green-algal–derived (e.g., euglenoids, chlorarachnids, and some alveolates) and red-algal–derived (e.g., cryptomonads, heterokonts, haptophytes, and some alveolates) plastids provides clear evidence that multiple secondary endosymbioses have occurred; however, the number and timing of the events remains unresolved.

Algal lineages derived from secondary endosymbiotic associations with an ancestor of the red algae are diverse. Three major groups, the cryptophytes, heterokonts, and alveolates, comprise both nonphotosynthetic and photosynthetic members with the nonphotosynthetic members branching basally within each lineage (Van de Peer et al. 2000; Keeling 2004; Bachvaroff, Sanchez Puerta, and Delwiche 2005). If photosynthesis evolved early within these lineages and plastids were lost secondarily, nuclear genomes of nonphotosynthetic members may contain genes that originated from the endosymbiotic genome. Given the complex molecular nature of the secondary endosymbiont, nuclear genomes of these lineages may contain genes that originated from the chloroplast (cyanobacterial in origin), the nucleus (red-algal origin), or mitochondria (alphaproteobacterial origin) genomes of the endosymbiont.

Within the heterokonts, several phylogenetic reconstructions have provided support for the early evolution of photosynthesis and gene transfer from the endosymbiont genome to the host nucleus. For example, phylogenetic analysis of genes encoding 6-phosphogluconate dehydrogenase (referred to as 6-PGD or *gnd*) demonstrated that both photosynthetic and nonphotosynthetic heterokonts grouped with the rhodophytes, chlorophytes, and vascular plants; the photosynthetic clade was the sister group of the cyanobacteria, suggesting a cyanobacterial origin (either from the red-algal chloroplast or nucleus) of the gene in heterokonts (Andersson and Roger 2002; Nozaki et al. 2004). Phylogenetic analyses of glyceradehyde-3-phosphate dehydrogenase (Liaud et al. 1997; Fagan, Hastings, and Morse 1998; Fast et al. 2001; Figge and Cerff 2001; Harper and Keeling 2003), fructose-1,6-bisphosphate aldolase (Patron, Rogers, and Keeling 2004), and several proteins involved in photosynthesis (Yoon et al. 2002; Bachvaroff, Sanchez Puerta, and Delwiche 2005) support the hypothesis that photosynthesis in heterokonts and alveolates evolved early ("chromalveolate" hypothesis;

Key words: heterokonts, diatoms, oomycetes, chloroplast, secondary endosymbiosis.

E-mail: debrobertson@clarku.edu.

Mol. Biol. Evol. 23(5):1048–1055. 2006
doi:10.1093/molbev/msj110
Advance Access publication February 22, 2006

Cavalier-Smith 1999) or serially, with chloroplast loss occurring independently in the nonphotosynthetic lineages (see Bodyl 2005 for alternative discussion).

Glutamine synthetase (GS, enzyme code 6.3.1.2) catalyzes the adenosine triphosphate–dependent synthesis of glutamine via the condensation of ammonium and glutamate and thus plays an essential role in ammonium assimilation and glutamine biosynthesis. The enzyme is found in all living organisms, and three distinct gene families (GSI, GSII, GSIII) have been identified (Southern, Parker, and Woods 1987; Bennett and Cullimore 1990; Brown et al. 1994; Reyes and Florencio 1994). Early work suggested that members of the three gene families were restricted to either prokaryotic or eukaryotic organisms. However, recent molecular studies and genome projects have shown that the gene families are broadly distributed, suggesting that the GS family members arose prior to the divergence of prokaryotes and eukaryotes (Brown et al. 1994; Mathis et al. 2000; Robertson, Smith, and Alberte 2001).

The distribution of GS gene families and the number of isoenzymes expressed varies among organisms (e.g., Beudeker and Tabita 1985; Coruzzi et al. 1989; Reyes and Florencio 1994; Crespo, Garcia-Dominguez, and Florencio 1998; Mathis et al. 2000; Turner and Young 2000; Robertson, Smith, and Alberte 2001). The majority of photosynthetic eukaryotes express multiple GS isoenzymes that are nuclear encoded yet targeted to either the cytosol or chloroplasts (Forde and Woodall 1995; Chen and Silflow 1996; Lam et al. 1996; Woodall et al. 1996). In vascular plants, the isoenzymes are encoded by members of the GSII gene family and arose through a recent gene duplication event with subsequent expansion of the cytosolic clade (Coruzzi et al. 1989).

In contrast to vascular plants and green algae, the nuclear-encoded chloroplast- and cytosol-localized GS isoenzymes in diatoms (Heterokonta) are members of the GSII and GSIII gene families, respectively (Robertson and Alberte 1996; Robertson, Smith, and Alberte 1999, 2001). Our previous phylogenetic analyses and the absence of GSII genes in extant cyanobacteria (Dufresne et al. 2003; Palenik et al. 2003) suggest that the diatom GSII gene was transferred from the nucleus of the red-algal endosymbiont that gave rise to the diatom plastid (Robertson, Smith, and Alberte 1999). Here, we expand our study of GSII evolution by including sequences from three rhodophytes, two diatoms, and four oomycetes (nonphotosynthetic heterokonts). Our phylogenetic analyses support the monophyly of GSII from plants, green algae, rhodophytes, and both photosynthetic and nonphotosynthetic heterokonts. This grouping is consistent with the hypothesis that the GSII gene of heterokonts evolved by endosymbiotic gene transfer from the nuclear genome of the red algae and provides molecular evidence that photosynthesis evolved early in the evolution of heterokonts and was secondarily lost in the oomycetes.

Materials and Methods

Amplification and Sequencing of GSII Genes from Oomycetes

Cultures of the oomycetes *Lagenidium giganteum* Couch (Agricultural Research Service Collection of Entomopathogenic Fungi [ARSEF] #373) and *Leptolegnia chapmanii* Seymour (ARSEF #2681) were obtained from the United States Department of Agriculture-Agriculture Research Service Collection of Entomopathogenic Fungal Cultures (Ithaca, NY) and grown in Sabouraud dextrose broth plus 2% yeast extract (SDY) at room temperature. Cultures (approximately 50 ml) were ground in liquid nitrogen and subjected to both DNA and RNA extraction, using a modified hexadecyltrimethylammonium bromide extraction protocol (Coyer, Robertson, and Alberte 1994) and the QIAGEN RNeasy Plant Mini Kit (Qiagene, Inc., Valencia, Calif.), respectively. The polymerase chain reaction (PCR) primers AAGAAGATGCGTGAGGACGG (forward) and CGRCGGTCCTCGWAGTAGCC (reverse) were used to amplify approximately 250 nt of the GSII gene from *L. giganteum* and *L. chapmanii* genomic DNA. Standard PCRs (25 μl; Qiagen, Inc.) had final primer concentrations of 400 nM each and the thermal conditions were: 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min performed for 30 cycles. The GSII fragments were sequenced commercially (MWG Biotech, Charlotte, N.C.) and the resulting sequences were used to design gene-specific primers for 5′ and 3′ rapid amplification of cDNA ends (RACE) PCRs. Single-stranded cDNA was synthesized from approximately 1.5 μg of total cellular RNA using an oligo-d(T) primer (GCGGCCG-CTCTAGACTAG$(T)_{18}$) for 3′ RACE or a gene-specific primer for 5′ RACE, following the manufacturer's recommendations (Invitrogen, Carlsbad, Calif.). Complete cDNA sequences, including 5′ and 3′ untranslated regions (UTRs), were obtained for each species. Predicted N-terminal signal and chloroplast transit peptides were determined using the software programs SignalP (Bendsten et al. 2004) and ChloroP (Emanuelsson, Nielsen, and von Heijne 1999), respectively.

Data Set Assembly

GSII sequences were obtained by querying GenBank with the diatom GSII sequence using Blast (Altschul et al. 1997) or by searching TaxBrowser, genome databases, and EST databases using the term GS. Retrieved amino acid sequences were aligned using ClustalX (default parameters) and manually adjusted in MacClade 4.07 (W. P. Maddison and D. R. Maddison 2000). Highly variable regions within the open reading frame (ORF) and at N- and C-termini were excluded or removed prior to phylogenetic analyses.

Phylogenetic Analysis

The GSII protein alignment used in phylogenetic analyses consisted of 335 aligned characters from 64 taxa. Models of protein evolution were evaluated using ProtTest (Drummond and Strimmer 2001; Guindon and Gascuel 2003; Abascal, Zardoya, and Posada 2005). The initial tree was obtained by Neighbor-Joining (BioNJ), and branch lengths and topology were optimized using the empirical substitution models of Jones-Taylor-Thornton, Whelan and Goldman, Blosum62, and Dayoff with all improvements. Bayesian analysis was performed using MrBayes 3.1.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). Two parallel runs, each with four chains, were run for 10^6 generations. For each run, three chains

were heated and one was "cold" with a temperature parameter of 0.20. The evolutionary models implemented in MrBayes 3.1.1 were explored using the mixed amino acid model. Rate variation across sites was approximated using a four-category gamma distribution with the proportion of invariable sites estimated from the data. Based on the results from ProtTest and MrBayes, the empirical WAG amino acid substitution model (Whelan and Goldman 2001) with a gamma distribution of rate categories plus a proportion of invariable sites was used in subsequent analyses. Trees were saved every 100th generation and, following a burn-in of 2,500 generations, 7,501 trees were generated per run. A 50% majority-rule consensus tree was generated to calculate posterior probability values.

Maximum likelihood (ML) analysis was done using PhyML (Guindon and Gascuel 2003). The initial tree was determined by Neighbor-Joining (BioNJ), using the WAG amino acid substitution model and the gamma shape parameter (1.36) and proportion of invariable sites (0.120) as determined in ProtTest. The tree topology, branch lengths, and rate parameters were optimized by the software during the run. The robustness of the data was evaluated using nonparametric boostrap analysis (500 replicates).

Results

Comparison of Heterokont GSII Sequences

Sequences of GSII transcripts were obtained from *L. giganteum* and *L. chapmanii*. PCR amplifications were designed to yield overlapping products and therefore could be assembled into single transcripts. These transcripts were 1,269 and 1,220 bp long for *L. giganteum* and *L. chapmanii*, respectively, and included the entire ORF. Both sequences have been submitted to the GenBank database and are publicly available with the accession numbers DQ173920 and DQ173921. The deduced amino acid sequences were 357 aa and 356 aa for *L. giganteum* and *L. chapmanii*, respectively. They were homologous to the 357-aa GSII sequence from *Phytophthora infestans* (GenBank accession number AAN31463) and were 84% (*L. giganteum*) and 78% (*L. chapmanii*) identical with that protein sequence. In contrast to the diatoms, which have a single-phase two intron, no introns were detected in the oomycete GSII genes when gene-specific PCR primers were used to amplify GSII from genomic DNA. In addition, the four amino acid insertion observed in diatom GSII sequences (Robertson, Smith, and Alberte 1999) was not present in the oomycete sequences.

Phylogenetic Analysis of GSII

GSII sequences were obtained for six of the major eukaryotic groups as well as bacteria. In the Bayesian analysis, 7,501 trees of each run (average scores −ln L 16624.86) were used to estimate Bayesian posterior probabilities (BPP). The topology of the tree inferred using ML (−ln L 16,594.58) was similar to that obtained by Bayesian methods. Differences in the topologies of the trees are addressed below.

In both the Bayesian and ML analyses, bacterial- and the chloroplast-localized *Chlamydomonas reinhardtii* GSII sequences formed a group that was well resolved from the other eukaryote sequences (fig. 1). There was strong (BPP = 1.00) to weak (ML bootstrap = 54%) support for the Opisthokonts (fungi and animals), and branching within the fungal and animal clades was similar to phylogenetic analyses of other genes (e.g., Baldauf et al. 2000).

GSII from the plants (green algae and vascular plants), rhodophytes (*Gelidium crinale*, *Cyanidioschyzon merolae*, *Galderia sulphuraria*), and both photosynthetic (*Thalassiosira pseudonana*, *Skeletonema costatum*) and nonphotosynthetic heterokonts (*P. infestans*, *Phytophthora ramorum*, *L. giganteum*, *L. chapmanii*) formed a single clade that received strong support in the Bayesian (BPP = 1.00) and ML (ML bootstrap = 90%) analyses. Within this larger group, there was strong support for the oomycetes, diatoms, rhodophytes, green algae, and plants. The oomycetes and diatoms were the most basally branching clades but were not monophyletic with respect to the other eukaryotes. The rhodophyte + green algae + vascular plant clade was strongly supported in the Bayesian analysis (BPP = 0.99). In contrast, while there was moderate support for the diatom + rhodophyte + green algae + vascular plant clade in the ML analysis (bootstrap value = 78%), the monophyly of rhodophyte + green algae + vascular plant GSII genes was not resolved.

Comparison of Heterokont N-Terminal Sequences

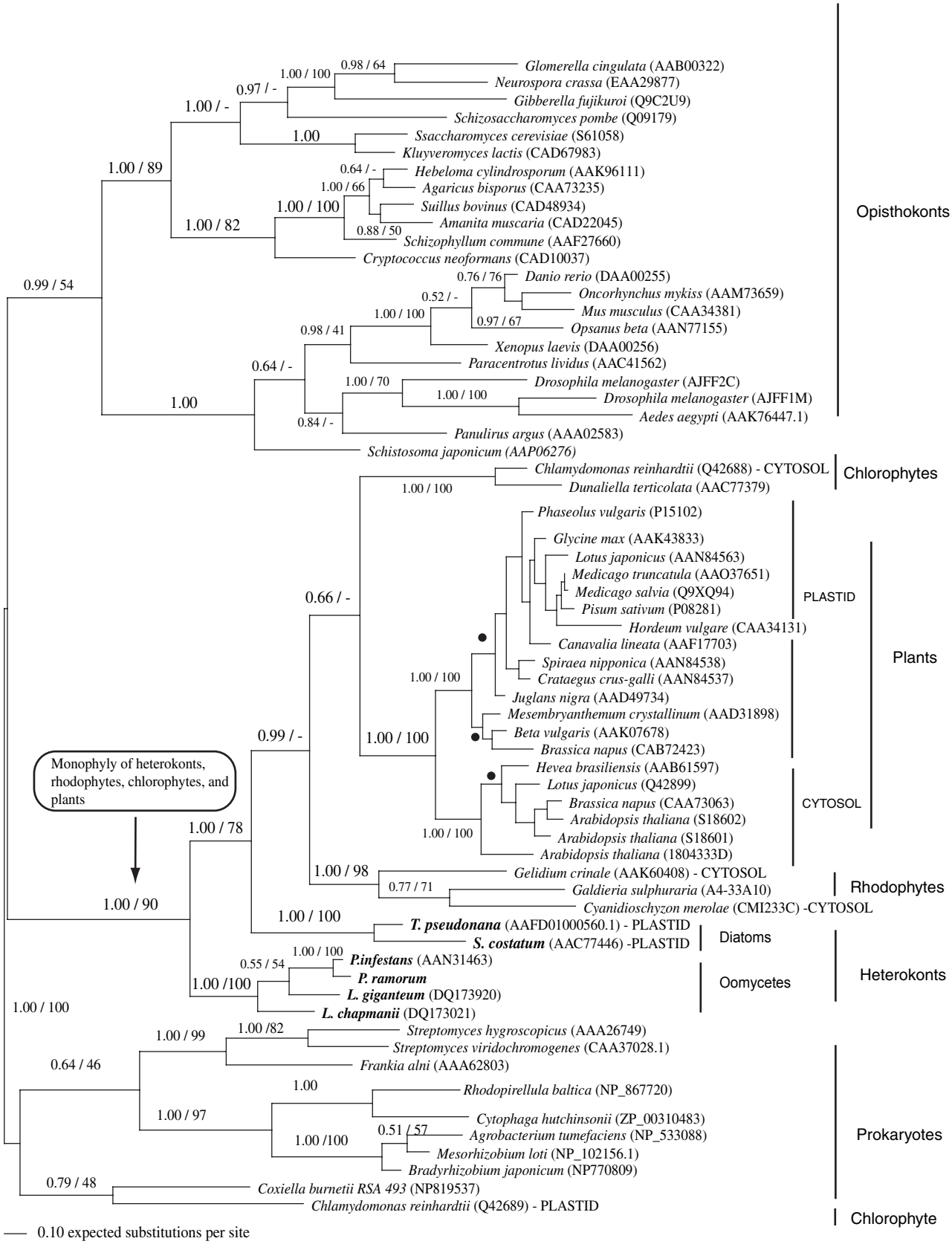
Each diatom GSII sequence contained an N-terminal, bipartite targeting sequence that directs the protein across the chloroplast-endoplasmic reticulum and chloroplast membranes. The targeting sequences of the two diatom GSII proteins were of similar length and shared several conserved residues. The predicted cleavage sites within the bipartite sequence are shown in figure 2. The oomycete GSII sequences lacked targeting sequences. However, several residues at the beginning of the ORF of the oomycete sequences were shared uniquely between diatoms and oomycetes (fig. 2).

Discussion

Phylogenetic analyses have provided strong evidence that photosynthetic heterokonts and nonphotosynthetic oomycetes (previously considered fungi), labyrinthulids, bicosoecids, and opalinids form a monophyletic group (Gunderson et al. 1987; Leipe et al. 1996; Van de Peer et al. 2000). However, the number and timing of secondary

→

FIG. 1.—Evolutionary relationships of GSII genes from six major groups of eukaryotes and bacteria. The analysis was based on 335 aligned amino acids from 64 taxa. The 50% majority-rule tree from the Bayesian analysis is shown and was inferred from 15,002 trees as described in the *Materials and Methods*. BPP are shown ranging from 0.50 to 1.00. The PhyML bootstrap values (500 replicates) are shown to the right of the BPP for each node and ranged between 46% and 100%. Support values for the branching pattern within the vascular plant cytosolic and chloroplastic clades are not shown (nodes to the right of the circles). The values are not critical to the hypotheses presented in the text. The *Phytophthora ramorum* GSII sequence is available at http://genome.jgi-psf.org/ramorum1/ramorum1.home.html.


Glomerella cingulata (AAB00322)
Neurospora crassa (EAA29877)
Gibberella fujikuroi (Q9C2U9)
Schizosaccharomyces pombe (Q09179)
Ssaccharomyces cerevisiae (S61058)
Kluyveromyces lactis (CAD67983)
Hebeloma cylindrosporum (AAK96111)
Agaricus bisporus (CAA73235)
Suillus bovinus (CAD48934)
Amanita muscaria (CAD22045)
Schizophyllum commune (AAF27660)
Cryptococcus neoformans (CAD10037)
Danio rerio (DAA00255)
Oncorhynchus mykiss (AAM73659)
Mus musculus (CAA34381)
Opsanus beta (AAN77155)
Xenopus laevis (DAA00256)
Paracentrotus lividus (AAC41562)
Drosophila melanogaster (AJFF2C)
Drosophila melanogaster (AJFF1M)
Aedes aegypti (AAK76447.1)
Panulirus argus (AAA02583)
Schistosoma japonicum (AAP06276)
Opisthokonts
Chlamydomonas reinhardtii (Q42688) - CYTOSOL
Dunaliella terticolata (AAC77379)
Chlorophytes
Phaseolus vulgaris (P15102)
Glycine max (AAK43833)
Lotus japonicus (AAN84563)
Medicago truncatula (AAO37651)
Medicago salvia (Q9XQ94)
Pisum sativum (P08281)
Hordeum vulgare (CAA34131)
PLASTID
Canavalia lineata (AAF17703)
Spiraea nipponica (AAN84538)
Crataegus crus-galli (AAN84537)
Juglans nigra (AAD49734)
Mesembryanthemum crystallinum (AAD31898)
Beta vulgaris (AAK07678)
Brassica napus (CAB72423)
Hevea brasiliensis (AAB61597)
Lotus japonicus (Q42899)
Brassica napus (CAA73063)
Arabidopsis thaliana (S18602)
Arabidopsis thaliana (S18601)
Arabidopsis thaliana (1804333D)
CYTOSOL
Plants
Monophyly of heterokonts, rhodophytes, chlorophytes, and plants
Gelidium crinale (AAK60408) - CYTOSOL
Galdieria sulphuraria (A4-33A10)
Cyanidioschyzon merolae (CMI233C) -CYTOSOL
Rhodophytes
T. pseudonana (AAFD01000560.1) - PLASTID
S. costatum (AAC77446) -PLASTID
Diatoms
P.infestans (AAN31463)
P. ramorum
L. giganteum (DQ173920)
L. chapmanii (DQ173021)
Oomycetes
Heterokonts
Streptomyces hygroscopicus (AAA26749)
Streptomyces viridochromogenes (CAA37028.1)
Frankia alni (AAK62803)
Rhodopirellula baltica (NP_867720)
Cytophaga hutchinsonii (ZP_00310483)
Agrobacterium tumefaciens (NP_533088)
Mesorhizobium loti (NP_102156.1)
Bradyrhizobium japonicum (NP770809)
Prokaryotes
Coxiella burnetii RSA 493 (NP819537)
Chlamydomonas reinhardtii (Q42689) - PLASTID
Chlorophyte
0.10 expected substitutions per site

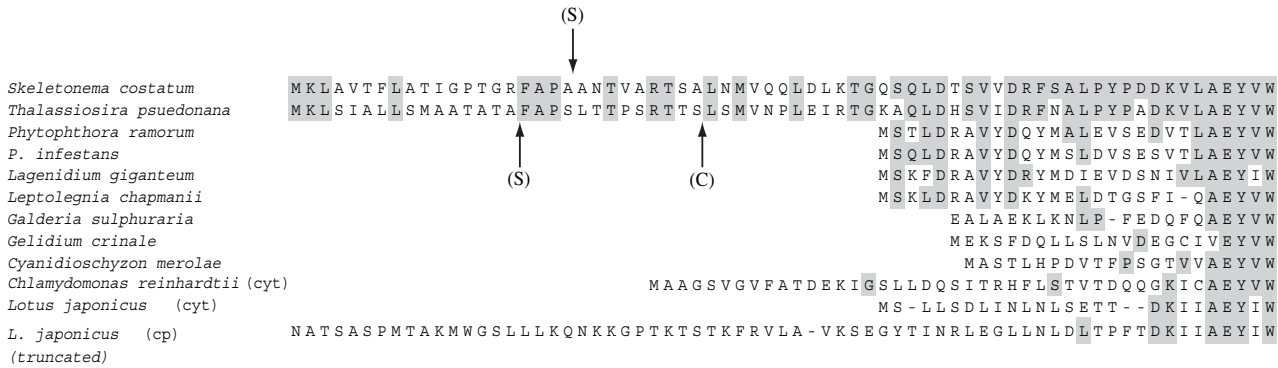


FIG. 2.—N-terminal amino acid sequences of GSII from diatoms and oomycetes. Amino acids that are identical with the *Skeletonema costatum* (diatom) GSII sequences are shaded. Arrows indicate predicted signal peptide (S) and chloroplast (C) cleavage sites in the diatom sequences, as determined by SignalP and ChloroP, respectively. A chloroplast cleavage site was not identified in the sequence of *S. costatum.*

endosymbiotic events within the heterokont lineage are unresolved. Recent studies of genes involved in photosynthesis and carbon metabolism have provided varying levels of support for the monophyly of heterokonts, haptophytes, cryptophytes, and alveolates ("chromalveolate hypothesis," Cavalier-Smith 1999). If these groups are monophyletic, a single secondary endosymbiotic association would have occurred very early in the evolution of these eukaryotes with subsequent, independent loss of plastids in the nonphotosynthetic lineages (Yoon et al. 2002, 2004; Bhattacharya,Yoon, and Hackett 2003; Harper and Keeling 2003; Harper,Waanders, and Keeling 2005). As an alternative, Bachvaroff, Sanchez Puerta, and Delwiche (2005) propose the "serial evolution hypothesis" which posits that secondary plastids arose in cryptomonads and these photosynthetic organisms were subsequently engulfed independently by heterotrophic heterokonts and alveolates following the divergence of the basally branching nonphotosynthetic groups in both lineages.

The monophyletic grouping of heterokonts, rhodophytes, chlorophytes, and plants observed in this study is consistent with two alternative hypotheses: (1) the phylogeny reflects the division the unikonts (Opisthokonts plus Amoebozoa) and bikonts (plants plus the other eukaryotes; reviewed by Horner and Hirt 2004) as rooted by the prokaryotic GSII sequences or (2) that GSII in heterokonts evolved via the endosymbiotic gene transfer from the nucleus of the endosymbiont to the nucleus of the heterotrophic host cell. In regards to the first hypothesis, the proposal of two major divisions of eukaryotes depends entirely on the proper rooting of the eukaryotic domain. Stechmann and Cavalier-Smith (2003) proposed a root for the eukaryote tree that established the bikonts and unikonts based on the shared derived fusion of dihydrofolate reductase and thymidylate synthase genes in bikonts and the shared derived fusion of carbamoyl phosphate synthase II, dihydroorotase and aspartate carbamoyltransferase in unikonts. Molecular phylogenies based on multigene mitochondrial data sets provided only moderate support for the grouping of heterokonts, jakobids, rhodophytes, green algae, and vascular plants separate from the opisthokonts (Lang et al. 2002). Within these analyses, the heterokonts were distantly related to the plants + chlorophytes and rhodophytes. Other multigene data sets and ribosomal phylogenies have provided limited resolution of the deeper branching pattern of eukaryotes with weak to no support for the close relationship of heterokonts, rhodophytes, and plants (e.g., Baldauf et al. 2000; Van de Peer et al. 2000). Although broader taxon sampling will be required to resolve the deeper branching patterns among bikont taxa (e.g., Cercozoa and Excavates), the robustness of our phylogenetic analyses suggests that GSII will be a valuable molecular marker for understanding eukaryotic evolution. However, we do not anticipate that GSII sequences alone will fully resolve the evolutionary history of eukaryotes with the support observed here. Alternatively, we suggest that the robust support for the monophyly of plants + green algae + red algae + heterokonts reflects a more recent shared history among the GSII genes.

The alternative hypothesis supported by our phylogeny is that GSII in heterokonts arose via endosymbiotic gene transfer from the nuclear genome of the endosymbiont (proto-rhodophyte) to the nuclear genome of the heterotrophic host. In diatoms, the cytosolic and chloroplastic GS isoenzymes are members of the GSIII and GSII gene families, respectively. GSIII has been identified in other heterokonts, the slime mold *Dictyostelium discoideum* and the amoeba *Entamoeba histolytica* (Robertson et al. unpublished data), suggesting that GSIII was present in the nucleus of early eukaryotes. However, we neither recover GSIII sequences from the *Phytophthora* genome data nor have we been able to amplify GSIII from *L. giganteum* and *L. chapmanii*. Thus, we propose that the GSIII gene was in the ancestral heterokont nucleus and that the GSII gene was integrated into the nuclear genome following the association with the secondary endosymbiont. This model further suggests that in the oomycetes, the GSIII gene was functionally replaced with the GSII gene following the endosymbiotic association. The nuclear genome of the endosymbiont is the predicted gene donor because extant cyanobacteria genomes do not contain GSII genes (Dufresne et al. 2003; Palenik et al. 2003). The model is consistent with the early evolution of photosynthesis in the heterokonts with subsequent loss of plastids in the oomycetes.

The endosymbiotic gene transfer hypothesis predicts that the heterokont GSII sequences will be monophyletic and branch within the red-algal clade. However, in our

analyses, the heterokonts (diatoms + oomycetes) were not monophyletic, and both groups branched basally to the red algae + green algae + vascular plant clade of GSII sequences. If our endosymbiotic transfer hypothesis is correct, the paraphyly of the heterokont GSII sequences and the lack of association between heterokont and rhodophyte GSIIs may be the result of limited taxon sampling. A broader sampling of eukaryotic GSII sequences, including members of the "chromalveolates," will help resolve the evolutionary history of the host (nonendosymbiont derived) GSII sequences and offer better resolution of gene transfer events.

Gene losses, duplications, and incomplete sampling can complicate phylogenetic reconstructions. Most photosynthetic organisms examined to date express multiple GS isoenzymes that are nuclear encoded but targeted to either the cytosol or chloroplast. The two complete red-algal GSII sequences (from *Cyanidioschzyon* and *Gelidium*) used in this study do not have chloroplast transit sequences and therefore are assumed to function in the cytosol (see also Freshwater, Thomas, and Bailey 2002). The red algae examined in this study include representatives from the most basal (Cyanidiales: *Galdieria* and *Cyanidioschyzon*) and the derived (Florideophycidae: *Gelidium*) lineages within the group. The monophyly of photosynthetic heterokonts and rhodophytes has been well supported in phylogenies of genes encoding proteins involved in photosynthesis and plastid-encoded small subunit rDNA. Within the heterokont-rhodophyte clade, members of the Cyanidiales branch basally to heterokonts and other red algae (Yoon et al. 2004) or form a sister association with the heterokonts (Müller et al. 2001). The lack of red-algal plastid GSII sequences in our analyses may have contributed to the lack of resolution of a red-algal–heterokont sister relationship. A more extensive survey of GS expression in rhodophytes is needed to determine the pattern of GS isoenzyme expression and whether there has been widespread loss of plastid-localized GSII in this lineage.

Chloroplast targeted proteins in photosynthetic heterokonts require a bipartite transit peptide that directs the protein across the endoplasmic reticulum surrounding the chloroplast and across the chloroplast membrane. If the GSII protein was targeted to the endosymbiont plastid, only the signal sequence would have been added following the transfer to the heterokont nucleus. Alternatively, if GSII was targeted to the cytosol of the endosymbiont, the gene in the heterokonts would have required the addition of the bipartite chloroplast-targeting sequence. Kilian and Kroth (2004) proposed a model for presequence evolution via recombination events mediated by introns, similar to the ideas presented by Long et al. (1996) regarding the evolution of mitochondrial targeting sequences. However, introns are not present in the N-terminal region of the heterokont GSII genes, suggesting an alternative molecular mechanism for the evolution of bipartite transit sequence. There was no evidence of a remnant transit sequence in the 5′ UTR of the oomycete sequences raising questions about the stability of the endosymbiotic association in the oomycetes. Specifically, the endosymbiont may have been lost from the oomycete lineage without being fully established as a plastid and thus prior to the evolution of the transit sequence. If plastid loss occurred independently in the nonphotosynthetic heterokont lineages, genes encoding GS may vary in these groups depending on which genes (GSIII or GSII) were fixed following plastid loss.

Genes that encode metabolic processes other than photosynthesis have the potential to be important evolutionary markers for examining the timing and stability of secondary endosymbiotic events. These genes have the potential of being retained in the nuclear genome following the loss of plastids, replacing the function of host genes (Martin and Schnarrenberger 1997; Timmis et al. 2004). For example, in vascular plants, Martin et al. (2002) estimated that approximately 18% of *Arabidopsis* nuclear genes are of cyanobacterial (endosymbiont) origin, including genes whose products are targeted to the cytosol and not involved in photosynthesis. While there has been some suggestion that oomycete genomes do not harbor a large number of plastid-derived nuclear genes (Palmer, Soltis, and Chase 2004), phylogenetic analyses of GSII (this study) and *gnd* (Andersson and Roger 2002; Nozaki et al. 2004) group the oomycetes and other heterokonts with the red algae, green algae, and vascular plants, suggesting that both of these genes were derived from the genome of the secondary endosymbiont. Analyses of additional genes involved in conserved metabolic pathways may reveal similar patterns.

Previous studies proposed that GS may be one of the oldest functioning enzymes and that it evolves in a clock-like fashion (Pesole et al. 1991; Kumada et al. 1993; Brown et al. 1994; Brown and Doolittle 1997). The phylogenetic resolution obtained here indicates that a broad analysis of GS (either as a single gene or when incorporated into multigene analyses) in chromalveolates and other under represented eukaryotes is merited and will contribute to our understanding of the timing and stability of secondary endosymbiotic events as well as the evolution of host cells.

Acknowledgments

We thank Manfred Binder, David S. Hibbett, and Sohini Ghoshroy for helpful discussions during the preparation of the manuscript. Thoughtful comments from the editor and two anonymous reviewers are appreciated. The research presented was supported by the National Science Foundation CAREER award (IBN 0238426) to D.L.R.

Literature Cited

Abascal, F., R. Zardoya, and D. Posada. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**:2104–2105.

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

Andersson, J. O., and A. J. Roger. 2002. A cyanobacterial gene in non-photosynthetic protists—an early chloroplast acquisition in eukaryotes? Curr. Biol. **12**:115–119.

Bachvaroff, T. R., M. V. Sanchez Puerta, and C. F. Delwiche. 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.

Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science **290**:972–977.

Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Evol. **340**:783–795.

Bennett, M., and J. Cullimore. 1990. Expression of three plant glutamine synthetase cDNA in *Escherichia coli*. Eur. J. Biochem. **193**:319–324.

Beudeker, R. F., and F. R. Tabita. 1985. Characterization of glutamine synthetase isoforms from *Chlorella*. Plant Physiol. **77**:791–794.

Bhattacharya, D., H. S. Yoon, and J. D. Hackett. 2003. 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.

Brown, J. R., and W. F. Doolittle. 1997. Archaea and the prokaryote-to-eukaryote transition. Microbiol. Mol. Biol. Rev. **61**: 456–502.

Brown, J. R., Y. Masuchi, F. T. Robb, and W. F. Doolittle. 1994. Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. J. Mol. Evol. **38**:566–576.

Cavalier-Smith, T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryotic family tree. J. Euk. Microbiol. **46**:347–366.

———. 2002. Chloroplast evolution: Secondary symbiogenesis and multiple losses. Curr. Biol **12**:R62–R64.

Chen, Q., and C. D. Silflow. 1996. Isolation and characterization of glutamine synthetase genes in *Chlamydomonas reinhardtii*. Plant Physiol. **112**:987–996.

Coruzzi, G. M., J. W. Edwards, S. V. Tingey, F.-Y. Tsai, and E. L. Walker. 1989. Glutamine synthetase: molecular evolution of an eclectic multi-gene family. Pp. 223–232 *in* R. Goldberg, ed. The molecular basis of plant development. Proceedings of an E.I. duPont Nermous-UCLA Symposium. Alan R. Liss, Inc., New York.

Coyer, J. A., D. L. Robertson, and R. S. Alberte. 1994. Genetic variability within a population and between diploid/haploid tissue of *Macrocystis pyrifera* (Phaeophyceae). J. Phycol. **30**:545–542.

Crespo, J. L., M. Garcia-Dominguez, and F. J. Florencio. 1998. Nitrogen control of the glnN gene that codes for GS type III, the only glutamine synthetase in the cyanobacterium *Pseudanabaena* sp. PCC 6903. Mol. Microbiol. **30**:1101–1112.

Drummond, A., and K. Strimmer. 2001. PAL: an object-oriented programming library for molecular evolution and phylogenetics. Bioinformatics **17**:662–663.

Dufresne, A., M. Salanoubat, F. Partensky, et al. (21 co-authors). 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. Proc. Natl. Acad. Sci. USA **100**:10020–10025.

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

Fagan, T., J. W. Hastings, and D. Morse. 1998. The phylogeny of glyceraldehyde-3-phosphate dehydrogenase indicates lateral gene transfer from cryptomonads to dinoflagellates. J. Mol. Evol. **47**:633–639.

Fast, N. M., J. C. Kissing, D. S. Roos, and P. J. Keeling. 2001. Nuclear-encoded plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. Mol. Biol. Evol. **18**:418–426.

Figge, R. M., and R. Cerff. 2001. GAPDH gene diversity in spirochetes: a paradigm for genetic promiscuity. Mol. Biol. Evol. **18**:2240–2249.

Forde, B. G., and J. Woodall. 1995. Glutamine synthetase in higher plants: molecular biology meets plant physiology. Pp. 1–18 *in* R. M. Wallsgrove, ed. Amino acids and their derivatives in higher plants. Cambridge University Press, Cambridge.

Freshwater, D. W., D. T. Thomas, and J. C. Bailey. 2002. Characterization of a cDNA encoding glutamine synthetase II from *Gelidium crinale* (Rhodophyta). Phycol. Res. **50**:17–21.

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

Gunderson, J. H., H. Elwood, A. Incold, K. Kindle, and M. L. Sogin. 1987. Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. Proc. Natl. Acad. Sci. USA **84**:5823–5827.

Harper, J. T., and P. J. Keeling. 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. Mol. Biol. Evol. **20**:1730–1735.

Harper, J. T., E. Waanders, and P. J. Keeling. 2005. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int. J. Syst. Evol. Microbiol. **55**:487–496.

Horner, D. S., and R. P. Hirt. 2004. An overview of eukaryote origins and evolution: the beauty of the cell and the fabulous gene phylogenies. Pp. 1–23 *in* R. P. Hirt and D. S. Horner, eds. Organelles, genomes and eukaryotic phylogeny. CRC Press, New York.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics **17**:754–755.

Keeling, P. J. 2004. Diversity and evolutionary history of plastids and their hosts. Am. J. Bot. **91**:1481–1493.

Kilian, O., and P. G. Kroth. 2004. Presequence acquisition during secondary endocytobiosis and the possible role of introns. J. Mol. Evol. **58**:712–721.

Kumada, Y., D. R. Benson, D. Hillemann, T. J. Hosted, D. A. Rochefort, C. J. Thompson, W. Wohlleben, and Y. Tateno. 1993. Evolution of the glutamine synthetase gene, one of the oldest existing and functioning genes. Proc. Natl. Acad. Sci. USA **90**:3009–3013.

Lam, H.-M., K. T. Coschigano, I. C. Oliveira, R. Melo-Oliveira, and G. M. Coruzzi. 1996. The molecular-genetics of nitrogen assimilation into amino acids in higher plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47**:569–593.

Lang, B. F., C. O'Kelly, T. Nerad, M. W. Gray, and G. Burger. 2002. The closest unicellular relatives of animals. Curr. Biol. **12**:1773–1778.

Leipe, D. D., S. M. Tong, C. L. Goggin, S. B. Slemenda, N. J. Pieniazek, and M. L. Sogin. 1996. 16S-like rDNA sequences from *Developayella elegans*, *Labyrinthuloides haliotidis*, and *Proteromonas lacertae* confirm that the stramenopiles are a primarily heterotrophic group. Eur. J. Protistol. **32**:449–458.

Liaud, M.-F., U. Brandt, M. Scherzinger, and R. Cerff. 1997. Evolutionary origin of cryptomonad microalgae: two novel chloroplast/cytosol-specific GAPDH genes as potential markers of ancestral endosymbiont and host cell components. J. Mol. Evol. **44**(Suppl.):S28–S37.

Long, M., S. J. Souza, C. Rosenberg, and W. Gilbert. 1996. Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. Proc. Natl. Acad. Sci. USA **93**:7727–7731.

Maddison, W. P., and D. R. Maddison. 2000. MacClade 4: analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Mass.

Martin, W., T. Rujan, E. Richly et al. 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.

Martin, W., and C. Schnarrenberger. 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr. Genet. **32**:1–18.

Mathis, R., P. Gamas, Y. Meyer, and J. V. Cullimore. 2000. The presence of GSI-like genes in higher plants: support for the paralogous evolution of GSI and GSII. J. Mol. Evol. **50**:116–122.

Müller, K. M., M. C. Oliveira, R. G. Sheath, and D. Bhattacharya. 2001. Ribosomal DNA phylogeny of the Bangiophycidae (Rhodophyta) and the origin of secondary plastids. Am. J. Bot. **88**:1390–1400.

Nozaki, H., M. Matsuzaki, O. Misumi, H. Kuroiwa, M. Hasegawa, T. Higashiyama, T. Shin-i, Y. Kohara, N. Ogasawara, and T. Kuroiwa. 2004. Cyanobacterial genes transmitted to the nucleus before the divergence of red algae in the chromista. J. Mol. Evol. **59**:103–113.

Palenik, B., B. Brahamsha, F. W. Larimer et al. (15 co-authors). 2003. The genome of a motile marine *Synechococcus*. Nature **424**:1037–1042.

Palmer, J. D., D. E. Soltis, and M. W. Chase. 2004. The plant tree of life: an overview and some points of view. Am. J. Bot. **91**:1437–1445.

Patron, N. J., M. B. Rogers, and P. J. Keeling. 2004. Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. Eukaryot. Cell **3**:1169–1175.

Pesole, G., M. P. Bozzetti, C. Lanave, G. Preparata, and C. Saccone. 1991. Glutamine synthetase gene evolution: a good molecular clock. Proc. Natl. Acad. Sci. USA **88**:522–526.

Reyes, J. C., and F. J. Florencio. 1994. A new type of glutamine synthetase in cyanobacteria: the protein encoded by the *gln*N gene supports nitrogen assimilation in *Synechocystis* sp. strain PCC 6803. J. Bacteriol. **176**:1260–1267.

Robertson, D. L., and R. S. Alberte. 1996. Isolation and characterization of glutamine synthetase from the marine diatom *Skeletonema costatum*. Plant Physiol. **111**:1169–1175.

Robertson, D. L., G. J. Smith, and R. S. Alberte. 1999. Characterization of a cDNA encoding glutamine synthetase from the marine diatom *Skeletonema costatum* (Bacillariophyceae). J. Phycol. **35**:786–797.

———. 2001. Glutamine synthetase in marine algae: new surprises from an old enzyme. J. Phycol. **37**:793–795.

Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572–1574.

Southern, J. A., J. R. Parker, and D. R. Woods. 1987. Novel structure, properties and inactivation of glutamine synthetase cloned from *Bacteroides fragilis*. J. Gen. Microbiol. **133**:2437–2446.

Stechmann, A., and T. Cavalier-Smith. 2003. The root of the eukaryote tree pinpointed. Curr. Biol. **13**:R665–R666.

Timmis, J. N., M. A. Ayliffe, C. Y. Huang, and W. Martin. 2004. Endosymbtiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. **5**:123–135.

Turner, S. L., and J. P. W. Young. 2000. The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. Mol. Biol. Evol. **17**:309–319.

Van de Peer, Y., S. Baldauf, W. F. Doolittle, and A. Meyer. 2000. An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. J. Mol. Evol. **51**:565–576.

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

Woodall, J., J. G. Boxall, B. G. Forde, and J. Pearson. 1996. Changing perspectives in plant nitrogen metabolism: the central role of glutamine synthetase. Sci. Prog. **79**:1–26.

Yoon, H. S., J. D. Hackett, C. Ciniglia, G. Pinto, and D. Bhattacharya. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. **21**:809–818.

Yoon, H. S., J. D. Hackett, G. Pinto, and D. Bhattacharya. 2002. The single, ancient origin of chromist plastids. Proc. Natl. Acad. Sci. USA **99**:15507–15512.

William Martin, Associate Editor

Accepted February 14, 2006