

MOLECULAR EVOLUTION OF GLUTAMINE SYNTHETASE II AND III IN THE CHROMALVEOLATES¹

Sohini Ghoshroy and Deborah L. Robertson²

Biology Department, Clark University, 950, Main Street, Worcester, MA 01610, USA

Glutamine synthetase (GS) is encoded by three distinct gene families (GSI, GSII, and GSIII) that are broadly distributed among the three domains of life. Previous studies established that GSII and GSIII isoenzymes were expressed in diatoms; however, less is known about the distribution and evolution of the gene families in other chromalveolate lineages. Thus, GSII cDNA sequences were isolated from three cryptophytes (*Guillardia theta* D. R. A. Hill et Wetherbee, *Cryptomonas phaseolus* Skuja, and *Pyrenomonas helgolandii* Santore), and GSIII was sequenced from *G. theta*. Red algal GSII sequences were obtained from *Bangia atropurpurea* (Mertens ex Roth) C. Agardh; *Compsopogon caeruleus* (Balbis ex C. Agardh) Mont.; *Flintiella sanguinaria* F. D. Ott and *Porphyridium aerugineum* Geitler; *Rhodella violacea* (Kornmann) Wehrmeyer and *Dixoniella grisea* (Geitler) J. L. Scott, S. T. Broadwater, B. D. Saunders, J. P. Thomas et P. W. Gabrielson; and *Stylonema alsidii* (Zanardini) K. M. Drew. In Bayesian inference and maximum-likelihood (ML) phylogenetic analyses, chromalveolate GSII sequences formed a weakly supported clade that nested among sequences from glaucophytes, red algae, green algae, and plants. Red algal GSII sequences formed two distinct clades. The largest clade contained representatives from the Cyanidiophytina and Rhodophytina and grouped with plants and green algae. The smaller clade (*C. caeruleus*, *Porphyra yezoensis*, and *S. alsidii*) nested within the chromalveolates, although its placement was unresolved. Chromalveolate GSIII sequences formed a well-supported clade in Bayesian and ML phylogenies, and mitochondrial transit peptides were identified in many of the sequences. There was strong support for a stramenopile-haptophyte-cryptophyte GSIII clade in which the cryptophyte sequence diverged from the deepest node. Overall, the evolutionary history of the GS gene families within the algae is complex with evidence for the presence of orthologous and paralogous sequences, ancient and recent gene duplications, gene losses and replacements, and the potential for both endosymbiotic and lateral gene transfers.

Key index words: chromalveolates; chromists; cryptophytes; dinoflagellates; glutamine synthetase II;

glutamine synthetase III; molecular evolution; Rhodophytes

Abbreviations: BPP, Bayesian posterior probability; EGT, endosymbiotic gene transfer; HGT, horizontal gene transfer; JGI/DOE, Department of Energy Joint Genome Institute; MLBS, maximum-likelihood bootstrap support; SP, signal peptide; TP, transit peptide

Marine phytoplankton plays a major role in global primary productivity and biogeochemical cycles. Several lineages within the Chromalveolata supergroup, as proposed by Cavalier-Smith (1999), are dominant members of marine phytoplankton assemblages, including diatoms, dinoflagellates, haptophytes, and cryptophytes (Falkowski et al. 2004, Falkowski and Knoll 2007, Simon et al. 2009). The evolutionary history of the chromalveolates is complex, and the composition of the supergroup and relationships among lineages are under continued revision (Parfrey et al. 2010). Currently, it is well established that these organisms gained their plastid from red-alga-like endosymbiont (Bhattacharya et al. 2004, Keeling 2004, Archibald 2005), and recent research indicates some lineages may have acquired green algal genes at some point in their evolutionary history (Frommolt et al. 2008, Moustafa et al. 2009). Moustafa et al. (2009) proposed that the acquisition of genes from these two primary photosynthetic lineages (red and green algae) may have provided the chromalveolates with the genetic composition to become ecologically successful and dominate primary productivity in the marine ecosystem.

Nitrogen availability is a key factor regulating primary productivity in marine ecosystems. Inorganic forms of nitrogen are reduced to NH_4^+ prior to the incorporation into organic compounds and glutamine synthetase (GS; EC 6.3.1.2) is an enzymatic link between carbon and nitrogen metabolism as it catalyzes the ATP-dependent condensation of NH_4^+ and glutamate forming glutamine. The nitrogen incorporated into glutamine by GS is transferred to 2-oxoglutarate by the activity of glutamine: 2-oxoglutarate amidotransferase (GOGAT, EC 1.4.7.1 or EC 1.4.1.13) yielding two molecules of glutamate. The glutamine and glutamate produced by the

¹Received 9 November 2010. Accepted 3 November 2011.

²Author for correspondence: e-mail debrobertson@clarku.edu.

GS:GOGAT cycle are used in a wide variety of essential biosynthetic pathways in the cell.

Photosynthetic eukaryotes usually express multiple GS isoenzymes that function in the cytosol and chloroplast. GS isoenzymes are also targeted to the mitochondria in some vascular plants (Taira et al. 2004), which may be true for other photosynthetic eukaryotes (Allen et al. 2006). The GS isoenzymes comprise a protein superfamily that has three distinct classes (GSI, GSII, and GSIII), which differ in molecular size and the number of subunits comprising the holoenzyme (Brown et al. 1994, Eisenberg et al. 2000). Within streptophytes, the GS isoenzymes are members of the GSII gene family and appear to have evolved via recent gene duplication, with an expansion in the number of genes encoding cytosolic-localized isoenzymes (Coruzzi et al. 1989, Ghoshroy et al. 2010). Multiple GSII isoenzymes are also expressed in green algae and early diverging plants; however, the genes encoding the chloroplast-targeted isoenzyme appear to have evolved via a horizontal gene transfer (HGT) from eubacteria to green algae early in plant evolution (Ghoshroy et al. 2010). Although, there is biochemical evidence of multiple GS isoenzymes in red algae (Casselton et al. 1986), to date, molecular studies have only identified genes encoding cytosolic-localized enzymes (Freshwater et al. 2002, Terashita et al. 2006).

Multiple GS isoenzymes are also expressed in the chromalveolates. However, in contrast to plants and green algae, the isoenzymes are members of the GSII and GSIII superfamilies. Within the chromalveolates, GSII genes have been reported from stramenopiles (Armbrust et al. 2004, Robertson and Tartar 2006, Bowler et al. 2008, Nedelcu et al. 2009), haptophytes (Nedelcu et al. 2009), dinoflagellates, and nonphotosynthetic alveolates (Slamovits and Keeling 2008). Phylogenetic analyses of GSII in diatoms and the nonphotosynthetic oomycetes suggested that GSII evolved via an endosymbiotic gene transfer (EGT) from the nuclear genome of the photosynthetic endosymbiont to the nuclear genome of the host cell (Robertson and Tartar 2006, Slamovits and Keeling 2008). If true, this possibility suggests that photosynthesis evolved early in the evolution of the stramenopiles (Robertson and Tartar 2006). Phylogenetic analyses showed that GSII from the nonphotosynthetic alveolate *Oxyrrhis marina* clustered with other photosynthetic eukaryotes and was sister to GSII from the diatom *Skeletonema costatum* (Slamovits and Keeling 2008). This phylogeny is also consistent with the hypothesis of the endosymbiotic origin of the chromalveolate GSII.

Previous studies of the molecular evolution of GSII in chromalveolates have been hampered by the limited representation of red algal taxa. Biochemical and molecular characterization of GS has been rare within red algae. Casselton et al. (1986) presented biochemical evidence that two GS isoenzymes

are expressed in *Porphyridium cruentum*. A single GSII sequence was identified in the genome of Cyanidiophyceae, *Cyanidioschyzon merolae* (Matsuzaki et al. 2004), and biochemical studies showed that the enzyme functions in the cytosol (Terashita et al. 2006). GSII sequences were also reported from *Galdieria sulphuraria* (Cyanidiophyceae) in an EST study (Weber et al. 2004), and a GSII cDNA was characterized from *Gelidium crinale* (Florideophyceae; Freshwater et al. 2002). Phylogenetic analyses suggested a sister association between the *G. crinale* GSII and that of green algae, but the association was not well supported (Freshwater et al. 2002).

Two subphyla, the Cyanidiophytina and Rhodophytina, are currently recognized within Rhodophyta (Yoon et al. 2006). Phylogenetic analyses of plastid encoded genes of red algae and chromists suggest that the secondary endosymbiotic event involving a red alga occurred after the divergence of members of Cyanidiophytina from the Rhodophytina and before the divergence of the current Florideophyceae group from the other lineages in the Rhodophytina (Yoon et al. 2002, 2004). Thus, members of the Rhodophytina, comprising the former Bangiophyceae, are predicted to be the putative donor of secondary plastids in the chromalveolates. Characterization of GSII from members of the Rhodophytina is needed to further explore the hypothesis that GSII in chromalveolates evolved by EGT.

Another key group for understanding the evolution of the chromalveolates and the molecular evolution of GSII are the cryptophytes. Cryptophytes are unique in that they retain the nuclear genome of the red algal endosymbiont as the reduced nucleomorph. Several phylogenetic studies using plastid-encoded genes suggested that cryptophytes were an early diverging clade within the chromists (Yoon et al. 2002, 2004), and single- and multigene analyses have recovered a sister association between haptophytes and cryptophytes (reviewed in Keeling 2009). More recent phylogenetic analyses using multigene data sets have grouped the stramenopiles, alveolates, and rhizaria into a clade that excluded cryptophytes and haptophytes (Yoon et al. 2008, Burki et al. 2009, Parfrey et al. 2010), and other studies using nuclear genes failed to recover the monophyly of the chromalveolates (Parfrey et al. 2010). Thus, the evolutionary relationships among the cryptophytes, haptophytes, and other members of the chromalveolate group are currently unresolved. Although, there are some uncertainties regarding chromalveolate systematics, in this study we use the term chromalveolate as defined by Cavalier-Smith (1999) to refer collectively to chromists (cryptophytes, haptophytes, and heterokonts) and alveolates (apicomplexans and dinoflagellates).

Genes encoding members of the GSIII family were first described in anaerobic members of the Bacteroidetes and later in cyanobacteria (Southern

et al. 1986, Hill et al. 1989, Reyes and Florencio 1994). Genome data now indicate that GSIII genes are more broadly distributed among the eubacteria but are not present in all eubacterial lineages and, to date, have not been described in Archaea. Eukaryotic GSIII genes were first described in diatoms (Robertson et al. 1999, 2001) and have now been identified in the genomes of haptophytes, pelagophytes, dinoflagellates, and slime molds. Inokuchi et al. (2002) examined the relationship of GSI, GSII, and GSIII in select algal lineages, but a detailed analysis of GSIII in chromalveolates is currently lacking.

In this study, we examined the distribution and evolutionary history of the GSII and GSIII enzymes in representatives from several lineages within the chromalveolates and red algae. We amplified GSII sequences from three cryptophytes (*Guillardia theta*, *Cryptomonas phaseolus*, and *Pyrenomonas helgolandii*) and GSIII sequences from *G. theta*. GSII sequences were obtained from seven members of the Rhodophytina. Our findings suggest GSII and GSIII genes in chromalveolates have different evolutionary origins, with GSII being associated with the endosymbiotic origin of plastids while GSIII genes may have been present within the genomes of early evolving eukaryotes.

MATERIALS AND METHODS

Amplification and sequencing of GSII and GSIII genes from cryptophytes and GSII sequences from red algae. Three members of the class Cryptophyceae were selected for GSII gene amplification and sequencing. Cultures of *C. phaseolus* (SAG 2013) and *P. helgolandii* (SAG 28.87) were obtained from Sammlung von Algenkulturen (SAG) at University of Göttingen (Göttingen, Germany). *G. theta* (CCMP327) was obtained from the Provasoli-Guillard National Center for Culture for Marine Phytoplankton (CCMP, West Boothbay Harbor, ME, USA). Cultures of *P. helgolandii* were grown in SWES (seawater enriched with soil extract and salts) medium, *C. phaseolus* was grown in MiEB12 medium, and *G. theta* was grown in h/2 medium, following the information posted on the respective culture center Web sites. Seven members of the subphylum Rhodophytina (Yoon et al. 2006) were selected for GSII gene amplification and sequencing. Cultures of *Bangia atropurpurea* (Bangioophyceae), *Compsopogon caeruleus* (Compsopogonophyceae), *Flintiella sanguinaria* and *Porphyridium aerugineum* (Porphyridiophyceae), *Rhodella violacea* and *Dixonella grisea* (Rhodellophyceae), and *Stylonema alsidii* (Stylonematophyceae) were obtained from Sammlung von Algenkulturen (SAG) at the University of Göttingen (Göttingen, Germany). Cultures were maintained at 17°C at 12:12 light:dark (L:D) cycle. Culture strain and media information for the red algae are provided in Table S1 (in the supplementary material). Media were prepared following information posted on the SAG Web site (<http://www.epsag.uni-goettingen.de>).

Cultures were grown to produce sufficient biomass for harvesting either by centrifugation (Eppendorf Centrifuge 5810R, Brinkmann Instruments Inc., Westbury, NY, USA) in the case of unicells or by vacuum filtration using Whatman® filter papers (circles 55 mm) (Whatman International Ltd, Maidstone England, UK). Cells and tissue were frozen in liquid nitrogen and stored at -80°C. Samples were ground in liquid nitrogen with mortar and pestle for RNA and DNA extraction.

Cells pelleted from ~50 mL of liquid culture or 100 mg of tissue material were used for RNA extraction. RNA was extracted using the RNeasy® Mini Kit from (Qiagen Inc., Valencia, CA, USA) with the modifications described by Brown et al. (2009). DNA was extracted using hexadecyltrimethylammonium bromide as described by Coyer et al. (1994). Extracted nucleic acids were quantified spectrophotometrically for downstream applications using a MWG BIOTECH Lambda Scan 200x, 96-well Microplate Reader with KCJunior Software (MWG BIOTECH, High Point, NC, USA).

cDNA was synthesized using Omniscript RT kit (Qiagen) following the manufacturer's protocols. Total RNA (~0.5–1.5 µg) was used as the template and the oligo-d (T) primer [GCGGCCGCTCTAGACTAG(T)₁₈] as the first strand primer. PCR amplifications of both cDNA and genomic DNA were done in a final volume of 25 µL with Taq PCR core kit (Qiagen) with the manufacturer's supplied Q solution to overcome problems associated with high GC content. PCR amplification of genomic DNA used ~0.1–0.5 µg of DNA as the starting template.

Initial amplification of GSII sequences from *C. phaseolus* and *P. helgolandii* was done with degenerate primers (5'-CCN III TGG AAY TWY GAY GG-3' [forward] and 5'-GGN CCN AYY TGR WAY TCC CAY TG-3' [reverse]) at a final primer concentration of 1.6 µM. Thermal conditions for amplification were: initial denaturation of 94°C for 2 min, followed by 4 cycles of 94°C for 1 min, 45°C for 1 min, 72°C for 1 min, then 30 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min with a final extension at 72°C for 15 min. GSII and GSIII genes from *G. theta* were amplified using gene specific primers based on partial sequences available at the National Center for Biotechnology Information (GSII: AJ937538.1, AJ937527.1; and GSIII: EG716812.1, EG728097.1, EG716811.1, EG718938.1). PCR reactions were done with cDNA as the template and with the final concentration of each primer at 0.4 µM. Thermal conditions for amplification were: initial denaturation of 94°C for 2 min, followed by 30 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min with a final extension at 72°C for 15 min.

Initial PCR amplification of GSII sequences from red algae was done using degenerate primers at a final primer concentration of 1.6 µM. The primer sequences are listed in Table S2 (in the supplementary material) and the combinations of forward and reverse primers used to amplify partial GSII sequences for each species are listed in Table S3 (in the supplementary material). Thermal conditions for amplification were: initial denaturation of 94°C for 2 min, followed by four cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min, then 30 cycles of 94°C for 1 min, 45°C for 1 min, 72°C for 1 min, with a final extension at 72°C for 15 min.

PCR products were either sequenced directly or cloned into TOPO vectors prior to sequencing following manufacturer's protocols (TOPO TA Cloning Kit for Sequencing; Invitrogen, Carlsbad, CA, USA). Sequences were either produced commercially (Macrogen, Seoul, South Korea) or at Clark University using an ABI 3130 genetic analyzer (Applied Biosystems, Foster City, CA, USA). Sequencing reactions were done using BigDye® Terminator v3.1 Cycle Sequencing chemistry (Applied Biosystems) and cleaned using the Agencourt CleanSEQ kit (Beckman Coulter Inc., Brea, CA, USA) following the manufacturer's protocols. Sequence data were used to design gene specific primers for 5' rapid amplification of cDNA ends (RACE) reactions using a SMART™ RACE cDNA Amplification Kit (Clontech Laboratories Inc., Mountain View, CA, USA) and following the manufacturer's recommendations. Contigs were assembled using CodonCode Aligner (Codoncode Corporation, Dedham, MA, USA), and all nucleic acid sequences were translated into amino acids in silico. Complete open reading frames were obtained for GSII sequences from the

cryptophytes, (*C. phaseolus*, *G. theta*, and *P. helgolandii*) and for two sequences from the red alga *D. grisea*.

Phylogenetic analyses of GSII and GSIII. GSII and GSIII sequences were retrieved from public databases as well as genome and EST projects. GS sequences were retrieved using diatom GSII (*Skeletonema costatum* [AAC77446]) and GSIII (*Thalassiosira pseudonana* [XP_002295274]) sequences or the keyword “glutamine synthetase” as queries. Complete information (taxa, database sources, and accession numbers) for the sequences used in this study are presented in Tables S4 and S5 in the supplementary material. Initial alignments of GS amino acid sequences were done with the web-based program MAFFT (Katoh et al. 2002) followed by manual adjustment using MacClade 4.08 (Maddison and Maddison 2000). The N- and C-terminal regions of the proteins and highly variable regions within the alignments were excluded from phylogenetic analyses. The GSII sequence from the choanoflagellate *Monosiga brevicollis* was not included in the present analyses since it had been predicted to have evolved via HGT from an unknown algal donor (Nedelcu et al. 2009). GSII sequences from ciliates and kinetoplastids were also excluded from the present study due to high substitution rates in the sequences.

Biases in amino acid composition in individual GSII and GSIII sequences relative to the multiple alignments were evaluated using the χ^2 test available in TREE-PUZZLE (Schmidt et al. 2002). The *Dictyostelium discoideum* GSIII sequence failed χ^2 test ($P = 0.05$) and was excluded from further phylogenetic analyses.

For Bayesian analysis of GSII proteins, the final alignment consisted of 170 sequences and 316 characters. Analyses were done using MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003). Two parallel runs, each with four chains (three heated and one cold) were run for 1.5×10^6 generations with a temperature of 0.15. The evolutionary models implemented in MrBayes 3.1.2 were explored using the mixed amino acid model. Rate variation across sites was approximated using a gamma distribution with the proportion of invariable sites estimated from the data. Trees were sampled every 100 generations. Likelihood tree scores of two independent runs were plotted to estimate the point of convergence to a stable likelihood, and to determine the trees to be excluded via “burn-in.” BPP of the nodes were calculated from both runs totaling 20,000 trees. Trees remaining (20,000) after a burn-in of 5,001 for each run were used to compute a 50% majority rule consensus tree.

For Bayesian analysis of GSIII proteins, the final alignment consisted of 105 sequences and 701 characters. Trees were inferred following the same procedures described above with number of generations set to 10^6 . BPP of the nodes were calculated from trees, from both the runs, totaling 10,002 trees. Trees remaining (10,002), after a burn-in of 5,000 for each run were used to compute a 50% majority rule consensus tree.

ML-based inference of the phylogenetic trees was done using the software RAxML 7.0.4 (Stamatakis 2006, Stamatakis et al. 2008). The GSII alignment for this analysis consisted of 170 taxa and 316 characters and GSIII alignment consisted of 105 taxa and 658 characters. The analyses used a random starting tree, the rapid hill-climbing algorithm (i.e., option $-f d$ in RAxML; Stamatakis et al. 2008) and the WAG model of amino acid substitution (Whelan and Goldman 2001). A random seed number was used to initiate rapid bootstrapping ($-x$) and 1000 bootstrap trees were generated by invoking $\# 1000$ and $-x$ options in RAxML. A majority rule consensus tree was created in PAUP* 4.0b (Swofford 1998).

In silico predictions of the functional localization of GSII and GSIII protein sequences. The multiple sequence alignments of GSII and GSIII were used to identify extensions in the N-terminal region of the proteins upstream of the first conserved domain that might contain putative organelle targeting sequences.

In addition, all complete chromalveolate GSII sequences were subjected to in silico analyses for the prediction of plastid-targeting presequences composed of a signal peptide (SP) and a chloroplast transit peptide (cTP) and mitochondrial transit (mTP) peptides using HECTAR (Gschloessl et al. 2008), SignalP 3.0 (Emanuelsson et al. 2007), and TargetP 1.1 (Emanuelsson et al. 2007). Analyses using SignalP and TargetP were performed in two steps. First, the complete sequence was analyzed. If a SP was predicted, it was removed from the amino acid sequence and the truncated sequence was analyzed using TargetP. The sequences from *Aphanomyces euteiches* (CU357664), *Prymnesium parvum* (DV103374.1), and *Isocryptis galbana* (ABD94151.1) were incomplete at the N-terminal ends, and thus their cellular localization could not be predicted.

All complete eukaryotic GSIII sequences were included in the in silico prediction analyses using HECTAR (Gschloessl et al. 2008), TargetP 1.1 (Emanuelsson et al. 2007), MitoProt II-v1.101 (Claros and Vincens 1996), PSORT (Nakai and Kanehisa 1991), and Predotar (Small et al. 2004). The sequences from *T. pseudonana*, *I. galbana*, and *K. micrum* were incomplete at the 5' end as determined by the multiple sequence alignment of GSIII proteins, and thus in silico predictions were not performed.

RESULTS

Characterization of new cryptophyte and rhodophyte GS sequences. Complete GSII mRNA sequences were obtained from *C. phaseolus*, *G. theta*, and *P. helgolandii* and a complete GSIII mRNA sequence was obtained from *G. theta*. GenBank accession numbers and characteristics of the cryptophyte transcripts obtained in this study are summarized in Table 1.

Two complete GSII mRNA sequences were obtained from *D. grisea*. In addition, seven partial GSII mRNA sequences were amplified from red algae. Single GSII sequences were amplified from *B. atropurpurea*, *R. violacea*, *S. alsidii*, *F. sanguinaria*, and *P. aeruginum*, and two partial GSII sequences were amplified from *C. caeruleus*. Accession numbers, percentage of amino acid sequence obtained as compared to the diatom *Skeletonema costatum* (AAC77446) GSII, and the GC content of the amplified sequences are presented in Table 2.

Phylogenetic analyses of GSII. Phylogenetic analyses of GSII amino acid sequences resulted in a well-resolved backbone structure with a clear separation of the eubacterial and eukaryotic sequences (Fig. 1 and

TABLE 1. Summary of GSII and GSIII coding sequences characterized from cryptophyte taxa in the present study. GenBank accession numbers, length of the open reading frame (ORF), and % GC content of the ORF are presented.

Taxa	Accession number	ORF length (bp)	% GC (ORF)
GSII sequences			
<i>Cryptomonas phaseolus</i>	HQ020361	1293	65
<i>Guillardia theta</i>	HQ020363	1296	59
<i>Pyrenomonas helgolandii</i>	HQ020362	1290	65
GSIII sequence			
<i>Guillardia theta</i>	HQ141405	2121	56

TABLE 2. Characteristics of GSII sequences amplified from red algae. Taxa, GenBank accession numbers (accession), number of amino acids (amino acids), percentage of protein sequence as compared to the diatom *S. costatum* GSII sequence (% protein sequence), and GC content (%GC) of the amplified sequence are provided.

Taxa	Accession	Amino acids	% protein sequence	% GC
<i>Dixonella grisea</i> cytosolic	HQ880621	350 ^a	85	50
<i>Dixonella grisea</i> (chloroplast targeted)	HQ880620	420 ^a	102	52
<i>Porphyridium aeruginum</i>	HQ880626	123	30	60
<i>Flintiella sanguinaria</i>	HQ880625	130	32	62
<i>Rhodella violacea</i>	HQ880623	301	73	66
<i>Stylonema alsidii</i>	HQ880624	141	34	54
<i>Bangia atropurpurea</i>	HQ880622	268	65	64
<i>Compsopogon caeruleus</i>	HQ880627	193	47	55
<i>Compsopogon caeruleus</i>	HQ880628	99	24	56
<i>Skeletonema costatum</i>	AAC77446	410 ^a		

^aComplete open reading frame.

Figs. S1 and S2 in the supplementary material). In accordance with previous studies, eukaryotic sequences formed six major groups identified as the metazoa, fungi, chromalveolates, red algae (with the exception of *C. caeruleus*, *P. yezoensis*, and *S. alsidii*), green algae, and vascular plants (Robertson and Tartar 2006, Nedelcu et al. 2009, Ghoshroy et al. 2010). Within the eukaryotes, there was moderate support for the separation of the opisthokonts (metazoa + fungi; maximum-likelihood bootstrap support [MLBS] = 59) from the other eukaryotes, the majority of which represent photosynthetic lineages and were well supported (BPP > 0.95, MLBS = 71). The chromalveolate GSII sequences formed a clade that included sequences from the stramenopiles (diatoms, brown alga, pelagophyte, and oomycetes), haptophytes, cryptophytes and dinoflagellates; sequences from the red algae *C. caeruleus*, *P. yezoensis*, and *S. alsidii* and the excavate *Naegleria gruberi* (Fritz-Laylin et al. 2010) were also included in this clade. However, the BPP for this clade was 0.94, which was lower than the cutoff value of 0.95 used in Figure 1.

Three well-supported groups could be distinguished within the broader chromalveolate assemblage. The three cryptophyte GSII sequences obtained in this study formed a well-supported clade (Fig. 1, Clade I; BPP > 0.95, MLBS = 100) with the two members of the Pyrenomonadales (*P. helgolandii* and *G. theta*) forming a sister association. The monophyly of oomycete GSII sequences was also well supported (Fig. 1 Clade II; BPP > 0.95, MLBS = 95). Members of the Saprolegniales (*L. chapmanii* and *A. euteiches*) and the Peronosporales (*Phytophthora* spp.) formed separate clades with the single sequence from *L. giganteum* appearing more closely related to the Peronosporales (Fig. 1). GSII sequences from the dinoflagellate *Oxyrrhis marina* and three red algae (*C. caeruleus*, *P. yezoensis*, and

S. alsidii) were associated with this clade (Clade II), but the association was not well supported.

A third group (Clade III) that contained sequences from haptophytes (*Isochrysis galbana*, *Emiliana huxleyi*, and *Prymnesium parvum*), the pelagophyte *Aureococcus anophagefferens*, and diatoms (*S. costatum*, *Phaeodactylum tricornutum*, *T. pseudonana*, and *Fragilariopsis cylindrus*) received moderate support (BPP > 0.95). Within this clade, the diatom sequences formed a well-resolved group (BPP > 0.95, MLBS = 100) as did the haptophyte sequences (BPP > 0.95, MLBS = 85). A sister association between the pelagophyte sequence (*A. anophagefferens*) and the haptophytes was also supported (BPP > 0.95, MLBS = 77). A GSII sequence from the excavate *N. gruberi* was associated with this clade (Fig. 1); however, the association was not well supported.

The diatoms, for which there were two complete genome sequences available at the time of the study (Armbrust et al. 2004, Bowler et al. 2008), have a single copy of the GSII gene. In contrast, two GSII genes were identified in the haptophyte *E. huxleyi*, and three within the genome of the brown alga, *Ectocarpus siliculosus* (Cock et al. 2010). One of the GSII sequences from *E. huxleyi* (Department of Energy Joint Genome Institute [JGI/DOE] Protein ID: 437187) grouped with other haptophyte sequences while the other sequence (JGI/DOE Protein ID: 69253) branched outside of the haptophyte + photosynthetic heterokont clade (Clade III) and formed a weak association with the three GSII sequences from *E. siliculosus*. Among the sequences from *E. siliculosus*, two appear to be more closely associated with each other than to the third sequence, suggesting a duplication event within this species or the brown algal lineage.

A red algal clade containing the majority of red algal GSII sequences showed a weak sister association with sequences from vascular plants (MLBS = 54). The red algal clade was well supported (BPP > 0.95; MLBS = 97) and within this clade, *C. merole* (Cyanidiophyceae) and *G. sulphuraria* (Cyanidiophyceae) diverged from the deepest node, similar to previously reported nuclear and plastid gene phylogenies (Yoon et al. 2002, 2006). The remaining red algal sequences within this group formed a polytomy represented by sequences from the Florideophyceae (*G. crinale*, *Gracilaria changii*, and *Chondrus crispus*; BPP > 0.95; MLBS = 70), Bangiophyceae (*B. atropurpurea* and *P. haitanensis*; BPP > 0.95; MLBS = 100), Porphyrideophyceae (*F. sanguinaria* and *P. aeruginum*; BPP > 0.95 MLBS = 65), and Rhodellophyceae (two sequences from *D. grisea*; BPP > 0.95 MLBS = 96). Sequences from Compsopogonophyceae (*C. caeruleus*) and Rhodellophyceae (*R. violacea*) were seen associated with each other, but this association was not supported.

Plastids that evolved by secondary endosymbiosis are usually surrounded by three or four membranes. In chromists, the outer most plastid membrane is

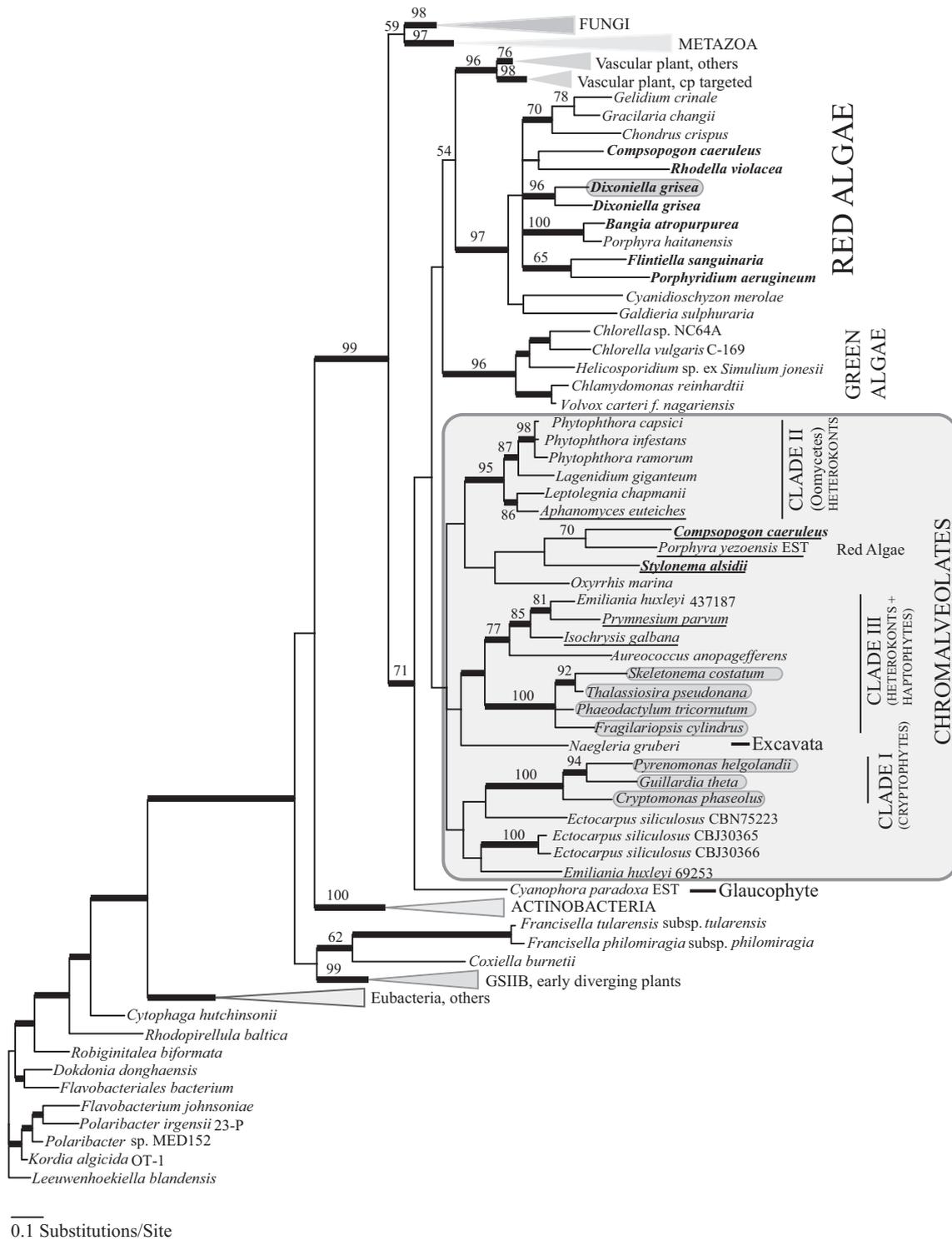


FIG. 1. Evolutionary relationships among GSII protein sequences from eubacteria and eukaryotes. Phylogenetic analyses included 170 GSII protein sequences. The unrooted 50% majority-rule consensus tree from the Bayesian analyses is shown as inferred from 20,000 trees as described in the “Materials and Methods.” Nodes with BBP support >0.95 are represented with thick lines. RAxML bootstrap values are indicated for nodes of interest and mostly among eukaryotes. GSII sequences obtained in the present study are indicated in bold typeface. Sequences identified in silico as chloroplast-targeted are identified with shaded boxes. Sequences that did not include the complete open reading frame are underlined.

continuous with the host endoplasmic reticulum (ER) and termed the chloroplast ER (cER). Proteins that are nuclear encoded and function in the plastid of these organisms have N-terminal SP and TPs that guide the protein through the cER and into the stroma of the chloroplast. In diatoms and cryptophytes, a conserved motif, ASAFAP, is observed in the SP cleavage site of chloroplast-targeted pre-proteins with the phenylalanine (F) present at the +1 position following SP cleavage (Gruber et al. 2007, Patron and Waller 2007, Kroth et al. 2008). Transit peptides that guide nuclear-encoded proteins from the cER to the stroma of plastids can resemble either mitochondrial or chloroplast TPs (e.g., Jiroutova et al. 2007). Thus, to predict the functional localization of the GSII enzymes, we first identified sequences that had an N-terminal extension upstream of the first conserved domain in the GSII amino acid alignment and then searched for conserved SP and TP elements within the sequences. Sequences that did not have complete open reading frames were excluded from this analysis.

The three cryptophyte GSII sequences (*P. helgolandii*, *G. theta*, and *C. phaseolus*) obtained in this study appear to be targeted to the plastid as SPs and TPs were identified in silico (Table 3; Fig. 1; and Fig. S3 in the supplementary material). SPs and TPs were also identified in the diatom GSII sequences (*F. cylindrus*, *P. tricornutum*, *S. costatum*, and *T. pseudonana*), indicating that they are targeted to the chloroplast. All SP sequences had the characteristic phenylalanine (F) at the +1 position of the predicted cleavage site.

An N-terminal extension was observed in the GSII sequence from *Emiliania huxleyi* (JGI/DOE Protein ID 437187) when compared with other GSII protein sequences. Although there was strong support for a predicted SP using SignalP (0.94), TargetP support for the presence of a TP following the removal of

the SP was low (0.123; Table 3). In addition, a bipartite chloroplast transit sequence was not identified using the program HECTAR. However, in the TargetP analysis of the complete open reading frame, a mitochondrial TP was predicted (Table 3). Thus the cellular location of this enzyme is currently unresolved. The *E. huxleyi* (JGI/DOE Protein ID 69253), *A. anophagefferens*, *O. marina*, and oomycete GSII protein sequences lacked N-terminal bipartite chloroplast and mitochondrial transit sequences and are assumed to be localized in the cytosol (the *O. marina* GSII sequence was predicted to be cytosolic in a previous study [Slamovits and Keeling 2008]). GSII sequences from the oomycete *A. euteiches*, the haptophytes *Prymnesium parvum* and *Isochrysis galbana*; and the red algae *C. caeruleus*, *P. yezoensis*, and *S. alsidii* lacked the N-terminal region of the protein sequence and therefore no predictions of cellular localization could be made.

One of the two complete GSII mRNA sequences obtained from *D. grisea* (GenBank: HQ880620) had an N-terminal extension when compared to other sequence (HQ880621). TargetP and ChloroP (Emanuelsson et al. 1999) analyses of the GSII sequences predicted a chloroplast transit sequence for sequence HQ880620. The chloroplast TP was predicted to be 67 amino acids in length, with one of the strongest scores (Reliability Class = 1) in the TargetP analyses (cTP = 0.916; ChloroP = 0.581). This is the first report of a chloroplast targeted GSII from rhodophytes.

Phylogenetic analyses of GSIII. Phylogenetic analysis of the GSIII protein sequences resulted in a tree with a clear resolution of eukaryotic GSIII sequences from eubacterial sequences (Fig. 2; Figs. S4 and S5 in the supplementary material), with the exception of sequences from two eukaryotes, *Perkinsus marinus* and *Trichomonas vaginalis*. These sequences were nested within the eubacterial sequences and the

TABLE 3. In silico prediction of targeting sequences in GSII proteins of chromalveolates. N-terminal sequences of GS proteins were analyzed for signal (SP) and organellar transit peptides (TP) in silico, using the web-based programs HECTAR, SignalP, and TargetP. Sequences with positive predictions are listed in the table. The scores (Score, cTP, mTP, Other), cellular location (Loc), and predicted cleavage site (CS) are provided. Signal peptide amino acids, predicted by SignalP, were removed from sequences prior to analysis in TargetP. Cellular locations include chloroplast (Cp), mitochondria (Mt) or cytosol (other). The highest value for the TargetP analyses are shown in bold.

Taxa	Protein ID	HECTAR			SignalP		TargetP		
		Loc	SP	TP	P	CS	cTP	mTP	Other
<i>Cryptomonas phaseolus</i>	HQ020361 ^a	Cp	0.798	0.951	0.97	SGA-FL	0.483	0.631	0.022
<i>Guillardia theta</i>	HQ020363 ^a	Cp	0.706	0.980	0.99	TSA-FL	0.766	0.536	0.010
<i>Pyrenomonas helgolandii</i>	HQ020362 ^a	Cp	0.702	0.953	0.99	TAA-FL	0.879	0.393	0.012
<i>Fragilariopsis cylindrus</i>	203412 ^b	Cp	0.712	0.966	0.99	ALA-FA	0.219	0.153	0.350
<i>Phaeodactylum tricornutum</i>	XP_002182209.1 ^a	Cp	0.645	0.953	0.99	ASA-FA	0.119	0.470	0.178
<i>Skeletonema costatum</i>	AAC77446.1 ^a	Cp	0.760	0.866	0.99	TAA-FA	0.173	0.273	0.369
<i>Thalassiosira pseudonana</i>	26051 ^b	Cp	0.706	0.985	0.98	ATA-FA	0.286	0.217	0.490
<i>Emiliania huxleyi</i>	437187 ^b	Cyt?	0.544	–	0.94	AMP-AT	0.078	0.123 ^c	0.130

^aGenBank accession number.

^bJGI/DOE protein ID.

^cTargetP analysis of the entire open reading frame of the *E. huxleyi* 437187 GSII sequence predicted a mitochondrial TP (mTP = 0.553, Reliability class = 5).

potential of them being derived by HGT events is discussed below. The eukaryotic clade, consisting of sequences from amoebozoa and chromalveolates, received strong support in both Bayesian and RAxML phylogenetic analyses (BPP > 0.95; MLBS = 92). Two major divisions were observed within the eukaryotic clade, one consisting of sequences from the genus *Entamoeba* (Amoebozoa) and other consisting of sequences exclusively from representatives of the supergroup Chromalveolata (*Karenia brevis*, *Karlodinium micrum*, *A. anophagefferens* [JGI/DOE Protein ID 58809], *G. theta*, *I. galbana*, *E. huxleyi*, *F. cylindrus*, *P. tricorntum*, *Chaetoceros compressum*, *T. pseudonana*, and *A. anophagefferens* [JGI/DOE

Protein ID 71821]). The chromalveolate clade received strong to moderate support in both analyses (BPP > 0.95; MLBS = 77) and the chromist clade (cryptophytes, stramenopiles, and haptophytes) received strong support (BPP > 0.95; MLBS = 100). Haptophytes and stramenopiles showed a sister association (BPP > 0.95; MLBS = 76) and the cryptophyte (*G. theta*) GSIII sequence branched from a deeper node. Two GSIII sequences were recovered in the genome of *A. anophagefferens*; one gene encoded a protein that was positioned sister to the diatoms with strong support (BPP > 0.95; MLBS = 97) while the other protein was nested within the GSIIIs from dinoflagellates (BPP > 0.95; MLBS = 100),

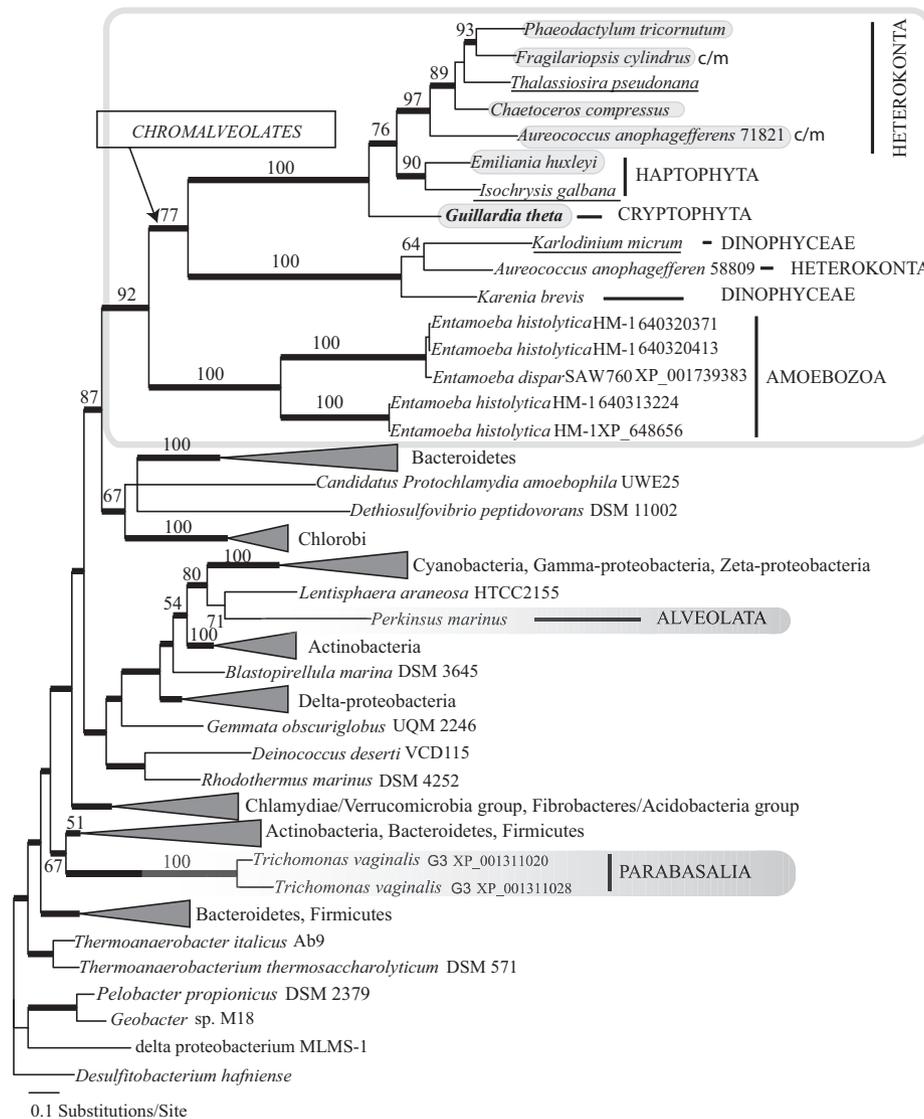


FIG. 2. Evolutionary relationships among GSIII protein sequences from prokaryotes and eukaryotes. Phylogenetic analyses included 105 GSIII protein sequences. The unrooted 50% majority-rule consensus tree from the Bayesian analyses is shown as inferred from 10,002 trees as described in the “Materials and Methods.” Nodes with BPP support >0.95 are represented with thick lines. RAxML bootstrap values are indicated for nodes of interest among the eukaryotes. The GSIII sequence obtained in the present study is indicated in bold typeface. Sequences consistently identified in silico as mitochondrial-targeted are identified with shaded boxes while those for which the predictions were ambiguous (either cytosolic or mitochondria) are shaded and labeled as c/m. Sequences that did not include the complete open reading frame are underlined.

which showed a sister association to the chromist clade.

Three eukaryotic GSIII sequences were nested among the eubacterial sequences. GSIII from the dinoflagellate *P. marinus* was nested within a clade (BPP>0.95; MLBS = 80) that included GSIII from the cyanobacteria, γ -proteobacteria, ζ -proteobacteria, and the marine bacterium *Lentisphaera araneosa* HTCC2155 (Fig. 2). Two GSIII sequences from the human protozoan parasite *T. vaginalis* grouped with sequences belonging to members of Actinobacteria, Bacteroidetes, and Firmicutes with moderate support (BPP > 0.95; MLBS = 67; Fig. 2; Fig. S4 and Table S5 in the supplementary material). The position of these eukaryotic sequences within different eubacterial clades suggests they evolved via independent HGT events.

The cellular localization of the chromalveolate GSIII enzymes was predicted by identifying N-terminal extensions upstream of the first conserved domain in the GSIII amino acid alignment and then searching for SP and TP elements within the sequences. SP sequences were not predicted for any of the chromalveolate GSIII sequences used in the present study. However, in silico analyses using HECTAR, MitoProt II-v1.101, TargetP v 1., PSORT, and Predotar identified putative mitochondrial TPs in GSIII sequences from two diatoms (*C. compressus* and *P. tricornutum*), the haptophyte *E. huxleyi*, and the cryptophyte *G. theta* (Table 4; Fig. 2; Fig. S6 in the supplementary material). Predictions for mitochondrial TPs for GSIII from *F. cylindrus* varied among algorithms. GSIII from *F. cylindrus* was predicted to be localized to the cytosol using HECTAR and Predotar whereas the other programs supported the presence of a mitochondrial TP, with prediction strengths similar to the other diatoms and with agreement as to the cleavage site position (Table 4). N-terminal targeting peptides were not identified in the GSIII sequences from the dinoflagellates (*K. brevis*, *P. marinus*), *T. vaginalis*, or *E. histolytica*, suggesting the proteins function in the cytosol.

Sequences from the haptophyte *I. galbana* and the dinoflagellate *K. micrum* were incomplete at the N-terminal end and thus not analyzed.

The *T. pseudonana* GSIII protein sequence retrieved from the JGI/DOE genome project (ver 3.0; <http://genome.jgi-psf.org/Thaps3/Thaps3.home.html>) did not contain a predicted mitochondrial TP. However, when compared to other diatom sequences in our protein alignment, the *T. pseudonana* sequence appeared incomplete at the N-terminal end. Analysis of nucleotides upstream of the 5' end of the predicted coding sequence revealed the presence of an additional start codon (AUG). When translated in silico, this upstream sequence showed a strong degree of amino acid conservation among the diatom sequences and TargetP analysis with this upstream region identified a putative mitochondrial targeting sequence. BLAST searches using GSIII with this extended region as a query against *T. pseudonana* EST sequences showed that the GSIII transcript did contain some but not all the upstream region. Further analyses of the 5' region of this transcript are required to determine the precise transcriptional and translational start sites for this gene and the cellular localization of the gene product.

One (JGI/DOE Protein ID: 71821) of the two predicted GSIII sequences identified in the genome of *A. anophagefferens* (<http://genome.jgi.doe.gov/Auran1/Auran1.home.html>) had a much longer N-terminal region than GSIII sequences from diatoms, raising the question of whether the protein was properly annotated. A BLAST (tblastn) search with the protein sequence as a query against *A. anophagefferens* EST sequences resulted in matches distributed along the 3' portion of the open reading frame corresponding to the region with highest conservation with GSIII. There was no evidence in the EST sequences that the transcripts contained the extended 5' region predicted in the genome models. Based on our multiple sequence alignment of GSIII proteins, we identified a candidate start of the open reading frame (beginning with the amino acid sequence "MMLNALKSV"). Using this open reading frame, a cytosolic location was predicted

TABLE 4. In silico prediction of targeting sequence of GSIII in chromalveolates. Sequences with positive predictions are listed in the table. In silico predictions were performed using several web-based algorithms as described in the "Materials and Methods." The scores (Score, P, or mTP), reliability (RS), predicted cellular location (Loc), likelihood of being functioning in the mitochondrial matrix, (Matrix), and the position of the predicted cleavage site (CS) are provided. Proteins predicted to be targeted to mitochondria are indicated as M, those without targeting sequences are predicted to function in the cytosol (Cyt).

Taxa	Protein ID	HECTAR		MitoProt II		Target P			PSORT			Predotar	
		Score	Loc	P	CS	Score	RS	CS	mTP	CS	Matrix	Score	Loc
<i>Chaetoceros compressus</i>	BAG39455 ^a	0.695	M	0.989	35	0.793	3	23	4.26	27	0.844	0.74	M
<i>Phaeodactylum tricornutum</i>	XP_002182898 ^a	0.614	M	0.997	33	0.868	2	18	4.52	25	0.865	0.51	M
<i>Fragilariopsis cylindrus</i>	277211 ^b	0.525	Cyt	0.944	27	0.846	2	27	4.57	27	0.869	0.20	Cyt
<i>Aureococcus anophagefferens</i>	71821 ^b	0.597	Cyt	0.685	22	0.670	3	16	2.07	22	0.601	0.60	M
<i>Emiliania huxleyi</i>	EST contig	0.834	M	0.994	21	0.941	1	30	5.61	30	0.920	0.92	M
<i>Guillardia theta</i>	HQ141405 ^a	0.824	M	0.976	29	0.923	1	28	6.28	29	0.920	0.69	M

^aGenBank accession number.

^bJGI/DOE protein ID.

by HECTAR, whereas the other programs provided evidence for a mitochondrial TP, albeit with lower support than observed for the other GSIII enzymes (Table 4). Future characterization of the 5' end of this transcript will be required to determine the start of the open reading frame.

DISCUSSION

We have illustrated the diversity and phylogenetic associations of GS isoenzymes from representatives within select chromalveolate (cryptophytes, haptophytes, stramenopiles, and dinoflagellates) and red algal lineages. The presence of GSII and GSIII enzymes has been documented previously in the diatoms (Robertson et al. 1999, 2001), and we have shown this pattern to be true for other stramenopile, cryptophyte (*G. theta*), haptophyte (*I. galbana* and *E. huxleyi*), and dinoflagellate (*K. brevis*) species.

As observed in previous phylogenetic analyses of GSII protein sequences (Robertson and Tartar 2006, Slamovits and Keeling 2008, Nedelcu et al. 2009), there was moderate support for the separation of the opisthokonts (metazoa + fungi) from the other eukaryotes, the majority of which represent photosynthetic lineages. In this study, chromalveolate GSII sequences formed a clade that included sequences from the stramenopiles (diatoms, brown alga, pelagophyte, and oomycetes), haptophytes, cryptophytes, dinoflagellates, three red algae (*C. caeruleus*, *P. yezoensis*, and *S. alsidii*), and the excavate *N. gruberi*; however, the clade was only moderately supported in the Bayesian analysis (BPP = 0.94) and unsupported by RAxML. The three red algal sequences (*C. caeruleus*, *P. yezoensis*, and *S. alsidii*) did not include complete open reading frames, which may have contributed to the low level of support for this red algal clade as well as the chromalveolates as a whole.

A second, well-supported red algal GSII clade included representatives from both the Cyanidophytina and Rhodophytina and grouped with GSIIIs from vascular plants and green algae, albeit without strong support. Within this clade, the major lineages of the Rhodophytina were resolved with moderate to strong support but the relationships among the lineages were unresolved. This clade also contained representatives for which only a portion of the open reading frame was available and therefore it is unlikely that the exclusion of *C. caeruleus*, *P. yezoensis*, and *S. alsidii* from this clade is due to the use of partial sequences.

The association of the *C. caeruleus*, *P. yezoensis*, and *S. alsidii* sequences with chromalveolates is consistent with our previous hypothesis of EGT of GSII from the nucleus of the red algal endosymbiont to the nucleus of the host (Robertson and Tartar 2006, Slamovits and Keeling 2008). The GSII sequences isolated from *C. caeruleus* were represented in the two red algal clades as were sequences from the genus *Porphyra*, suggesting a possible ancient duplication of GSII early in the evolution of the

Rhodophytina. The observed GSII phylogeny suggests that one gene may have been successfully transferred from the genome of the endosymbiont and subsequently retained by the nucleus of the chromalveolate host. Alternatively, the placement of these three red algal sequences within the chromalveolate clade could represent an HGT between the chromalveolates and the red algae. The three sequences in this clade represent three classes within the Rhodophytina (Bangiophyceae, Compsopogonophyceae, and Sytonematophyceae); if the HGT hypothesis is correct, the transfer event would have occurred early in the evolution of the Rhodophytina, with the potential loss of this gene in the other rhodophyte lineages. Further taxon sampling of members within the present classes of Bangiophyceae, Compsopogonophyceae, Porphyridiophyceae, Rhodellophyceae, and Stylonematophyceae is merited to explore the molecular evolution of GSII and distinguish between these hypotheses.

Our analyses of GSII included representatives from four major groups within the stramenopiles, which were not monophyletic and instead formed three distinct clades (oomycetes, diatoms + haptophytes, and brown algae + haptophytes). This pattern may be due to the presence of both orthologous and paralogous sequences within our analyses with differential gene loss among lineages. For example, an early expansion of the GSII family in stramenopiles with differential gene loss could explain the distribution of the heterokont GSII sequences among the three clades observed in Fig. 1.

The two GSII sequences obtained from *E. huxleyi* do not appear to have arisen by recent gene duplication event. One sequence (JGI/DOE Protein ID: 69253) was placed outside of the major haptophyte clade, without strong support with any of the major chromalveolate clades. The evolutionary relationship of this sequence can be interpreted as either an ancestral gene shared by both stramenopiles and haptophytes or the presence of this sequence in a single haptophyte taxa can be attributed to insufficient taxon sampling. The other sequence (JGI/DOE Protein ID 437187) was sister to other haptophyte sequences (*I. galbana* and *P. parvum*), and this clade was nested within the stramenopile clade that contained sequences from diatoms (*S. costatum*, *T. pseudonana*, *P. tricornutum*, and *F. cylindrus*) and the pelagophyte *A. anophagefferens*. The placement of the haptophyte clade within the stramenopiles is suggestive of a HGT event. The HGT hypothesis is favored in this case as otherwise multiple losses of GSII in several stramenopile taxa must be inferred to explain a common origin of this gene in the ancestor of both haptophytes and stramenopiles.

A GSII sequence from the Excavata, *N. gruberi* was also associated with the haptophyte + heterokont clade. Recent analysis of the *N. gruberi* genome identified seven genes with cyanobacterial ancestry (Maruyama et al. 2009). Since GSII genes have not

been identified in extant cyanobacteria to date (Dufresne et al. 2003, Palenik et al. 2003), the presence of the *N. gruberi* GSII sequence within the chromalveolate clade suggests a gene transfer event from the chromalveolates to *Naegleria*, indicating the complex nature of this organism. However, the placement of this taxon was not well supported, and alternatively, the placement of the single representative taxon from the supergroup Excavata could be attributed to phylogenetic anomaly or the topology might be reflective of the unikont (opisthokonts and Amoebozoa) and bikont (Archaeplastida, Excavate, Rhizaria, and Chromalveolates) division of eukaryotes (Cavalier-Smith 2003).

In a recent study of the cytosolic GAPDH enzyme (GapC1), expanded taxon sampling within the stramenopile and cryptophyte lineages suggested multiple HGT events between stramenopiles and dinoflagellates and stramenopiles and cryptophytes (Takishita et al. 2009). Taxon sampling of GSII within stramenopiles included species from the Oomycetes, Phaeophyceae, Pelagophyceae, and Bacillariophyceae, which represent a small portion of the current recognized diversity within the heterokont algae (Andersen 2004, Riisberg et al. 2009). Thus similar to GapC1, increased sampling of GSII among diverse eukaryotic lineages is required to further explore the relationships recovered here, including the potential for obtaining greater support of the deeper nodes within the chromalveolates and a better understanding of patterns of gene duplication, loss, and transfer within these organisms.

Nuclear-encoded, plastid-targeted proteins in cryptophytes and stramenopiles have a bipartite targeting sequence, with the combination of SP and TP used to direct nuclear-encoded proteins across the cER and into the stroma of the plastid (Kilian and Kroth 2005, Gould et al. 2006, Gruber et al. 2007, Jiroutova et al. 2007, Patron and Waller 2007). Predicted SP and TP sequences were observed in GSII sequences from cryptophytes (*P. helgolandii*, *G. theta*, and *C. phaseolus*) and stramenopiles (*T. pseudonana*, *F. cylindrus*, and *P. tricornutum*), and the majority had the characteristic F at the +1 position after the predicted SP cleavage (Kilian and Kroth 2005, Gould et al. 2006, Gruber et al. 2007, Patron and Waller 2007). The cellular localization of GSII from *E. huxleyi* (JGI/DOE Protein ID: 437187) was equivocal; there was some support for a SP and also support for the presence of a mTP. While further biochemical analyses are needed to confirm the cellular location of this protein, there was a general pattern of GSII isoenzymes being targeted to the plastids of diatoms and cryptophytes.

Genome analyses of members of Cyanidiophyceae, *C. merole* (Matsuzaki et al. 2004) and *G. sulphuraria* (Weber et al. 2004, Barbier et al. 2005), uncovered single GSII sequences in each of these species. Biochemical analyses of the singular GSII sequence identified from the genome *C. merole* confirmed the cytosolic localization of the protein

(Terashita et al. 2006). The absence of chloroplastic GS isoenzymes in this organism has been argued to be due to the low photorespiration rates in the Cyanidiophyceae (Terashita et al. 2006). However, as these thermophilic organisms were isolated from acidic environments, oxidized forms of nitrogen such as nitrate may be scarce, thus further reducing the need for multiple GS isoenzymes or the compartmentalization of N assimilation.

In contrast to the single cytoplasmic GSII in the cyanidiophytes, two GSII cDNAs were identified in *D. grisea*, one of which contained a chloroplast TP (identified by in silico analyses), suggesting the protein (HQ880620) is targeted to the chloroplast. The two sequences formed a sister association in the phylogenetic analyses, suggesting they arose through a recent gene duplication event. Although Casselton et al. (1986) presented biochemical evidence that multiple GS enzymes were expressed in the red alga *Porphyridium cruentum*, the cellular location of the isoenzymes was not determined, and thus our study presents the first evidence for a chloroplast-targeted GSII in red algae. In vascular plants, the chloroplast localized GS isoenzymes are involved in the assimilation of ammonium produced from nitrate reduction and re-assimilation of the ammonium released during photorespiration (Lam et al. 1996). Whether chloroplast-localized GS isoforms are common in mesophilic rhodophytes and serve a similar physiological function as observed in vascular plants merits further exploration.

We did extensive taxonomic sampling of the GSIII protein sequences from prokaryotes and eukaryotes. Presently, the distribution of GSIII includes the supergroup Chromalveolata, Amoebozoa (*Entamoeba* sp. and *D. discoideum*), and Excavata (*T. vaginalis* and *Euglena gracilis*); GSIII sequences have not been recovered in the genomes of members of the Archaeplastida or Opisthokonta, nor has it been reported from the Rhizaria (based on our searches of EST databases). The sequences from *D. discoideum* and *E. gracilis* were not included in our analyses as the *D. discoideum* sequences had unusual amino acid frequencies (as determined by a χ^2 analysis) and the *E. gracilis* EST data were not assembled into contigs. Although GSIII has been identified in some cyanobacteria, the cyanobacterial GSIIIs were distantly related to the eukaryotic sequences in our phylogenetic analyses. Thus, it is unlikely that GSIII in chromalveolates and Amoebozoa arose via an endosymbiotic gene transfer from the plastid. To date, GSIII sequences have not been reported from α -proteobacteria (the putative endosymbiont that gave rise to the mitochondria); however, recent analyses of mitochondrial proteome suggest that it has accumulated proteins from disparate sources including other Proteobacteria as well as nonproteobacteria, bacteriophages, Archaea, and other eukaryotes (Szkarczyk and Huynen 2010). Thus, the hypotheses that GSIII was present in the

nuclear genome of the ancestral eukaryote or introduced via EGT from the mitochondrial genome are equally likely, and both require the loss of this gene from three out of six supergroups of eukaryotes. Of note, the majority of the GSIII sequences in the chromalveolates were predicted to function in the mitochondria, based on the presence of a mitochondrial transit peptide.

GSIII sequences were obtained from photosynthetic chromalveolates but were not present in non-photosynthetic stramenopiles, the oomycetes. Within our analyses, the chromalveolate GSIII sequences were resolved into two well-supported clades. One clade contained chromist sequences while the second clade contained sequences from two dinoflagellates (both of which acquired plastids via a tertiary endosymbiosis with haptophytes) and one of the two GSIII sequences identified in the genome of the pelagophyte, *A. anophagefferens*. Thus, the stramenopiles were not monophyletic.

There are several alternative hypotheses suggested by the GSIII phylogeny. Recent phylogenetic analyses with expanded taxa and multigene data sets have supported the association of stramenopiles with dinoflagellates within the SAR (Stramenopile, Alveolates and Rhizaria) group with the exclusion of haptophytes and cryptophytes (Yoon et al. 2008, Parfrey et al. 2010). If the SAR hypothesis is correct, the GSIII phylogeny could reflect the presence of multiple gene copies in the common ancestor of stramenopiles and dinoflagellates, which were differentially lost through evolutionary time. This hypothesis would also predict the presence of GSIII sequences within members of the Rhizaria, a group that has not been extensively screened for GSIII sequences.

Alternatively, the grouping of the stramenopiles, haptophytes, and cryptophytes with the exclusion of the dinoflagellates is similar to previous multigene phylogenetic studies of plastid-encoded and plastid-targeted genes (Yoon et al. 2002, 2004, Bachvaroff et al. 2005, Khan et al. 2007). While GSIII has not been identified in extant red algae, the distribution of GSIII in the chromalveolates, amoebozoans, and excavates suggests GSIII was present in early evolving eukaryotes, including the red algal progenitor. The similarity between the GSIII phylogeny and other plastid genes could arise if GSIII was transferred following the secondary endosymbiosis that gave rise to the plastids in chromalveolates, with subsequent loss from the red algal lineage.

Finally, the grouping of the dinoflagellates and the pelagophyte is consistent with the hypothesis of an HGT event between these taxa. A recent study revealed the HGT of the methionine adenosyltransferase gene between *A. anophagefferens* and a cryptophyte (Kamikawa et al. 2009), and our analyses of GSIII genes provides evidence for an HGT between *A. anophagefferens* and haptophytes. The results from these three genes suggest that other HGT events may be uncovered in the genome of *A. anophagefferens*

and raises the question of whether HGT is common among pelagophytes or restricted to particular lineages or species. It has been hypothesized from genome studies of marine cyanobacteria that genes derived via HGT allow for continued change in cell-surface glycosylation patterns and thus reduce grazing (Palenik et al. 2003). A similar hypothesis was put forth to account for HGT derived genes in the eukaryotic marine phytoplankton *Ostreococcus* (Chlorophyceae; Palenik et al. 2007). Similarly, HGT of GS could allow for rapid changes in nitrogen assimilatory capacity, which might be advantageous for marine phytoplankton.

The number and type of GS isoenzymes expressed by stramenopiles varies, ranging from a single GSII in the oomycetes to three isoenzymes in *A. anophagefferens*, *E. huxleyi*, and *E. siliculosus*. The diatom genomes contain single copies of genes encoding GSII and GSIII while the genome of *E. siliculosus* contains genes for three GSII enzymes and *A. anophagefferens* had two GSIII and one GSII sequences. Biochemical studies provided the first evidence that the haptophyte, *E. huxleyi*, also expressed multiple GS isoenzymes (Maurin and Le Gal 1997, Robertson et al. 2001), and subsequently, two GSII and one GSIII genes were identified in the genome. The expansion of the GS gene family in the *A. anophagefferens* and *E. huxleyi* may be reflective of their physiology; both species rely more heavily on reduced and organic forms of nitrogen than the diatoms (e.g., Berg et al. 1997, Gobler et al. 2002, Bruhn et al. 2010).

The majority of the GSIII sequences from stramenopiles, haptophytes, and cryptophytes were predicted to have mitochondrial TPs. Mitochondrial presequences in these organisms are similar to other eukaryotes in that they are rich in positively charged and nonpolar residues (Liaud et al. 2000, Parker et al. 2004, Kroth et al. 2008) but can be difficult to predict accurately in silico. Mitochondrial TPs were predicted in GSIIIs from *C. compressus*, *P. tricornutum*, *E. huxleyi*, and *G. theta* by all of the algorithms used in this study; however, the position of the predicted cleavage site varied among programs. In contrast, an mTP was predicted in *F. cylindrus* by every program except HECTAR and Predotar. The support scores for the predicted mTP were similar to other diatom GSIIIs, and thus a mitochondrial location is likely for this enzyme as well, but experimental evidence is required. Although there are other enzymatic pathways in the mitochondria that can assimilate ammonium, the mitochondrial targeting of GSIII suggests this enzyme may have an important role in assimilating ammonium produced during photorespiration or in the ornithine-urea cycle (Armbrust et al. 2004, Parker et al. 2004, Parker and Armbrust 2005, Allen et al. 2011). A cytosolic function was predicted for GSIII from *A. anophagefferens* (58804), *K. brevis*, *E. histolytica*, *T. vaginalis*, and *P. marinus* while the nucleotide

sequence from the *T. pseudonana* genome was equivocal. Dual targeting of GS protein to mitochondria and chloroplast has been documented in *Arabidopsis thaliana* and has been postulated to facilitate ammonium recovery during photorespiration (Taira et al. 2004). Future studies examining the in vivo cellular location of GSIII in diatoms will be of interest to confirm the extent to which the enzyme functions in the mitochondria and if expression levels (or cellular targeting) vary in response to rates of photorespiration or primary nitrogen assimilation.

The GSIII protein sequence of *P. marinus*, a non-photosynthetic parasite of oysters and an early diverging dinoflagellate (Saldarriaga et al. 2003), grouped within the eubacterial sequences and showed a sister relationship with the sequence from the marine bacterium *Lentisphaera araneosa* HTCC2155, suggesting it evolved via HGT from Eubacteria. If GSIII were present in the nuclear genome early in eukaryotic evolution, our phylogeny suggests that *P. marinus* gained this bacteria-like sequence independently and that the host GSIII was replaced by the eubacterial GSIII sequence. Although genes of plastid origin have been reported recently in *P. marinus* (Stelter et al. 2007, Matsuzaki et al. 2008), it is unlikely that GSIII evolved via EGT from the plastid as the sequence was not sister to cyanobacteria. The GSIII protein sequence from parasitic protist *T. vaginalis* also grouped within the eubacterial GSIII sequences giving rise to the possibility of yet another HGT event. The two GSIII sequences from *T. vaginalis* diverged early within a major clade of eubacteria that also contained sequences from the genus *Clostridium*, which include human pathogens. Evidence of HGTs from eubacteria to *T. vaginalis* has been documented before and has been attributed to the parasitic nature of this organism (Markos et al. 1993, Henze et al. 1995, Hrdý and Müller 1995, de Koning et al. 2000). A recent survey of the draft genome of *T. vaginalis* identified 152 genes of HGT origin (Carlton et al. 2007), and our report of an HGT involving GSIII adds to this extensive list of horizontally derived genes. The HGT hypothesis is preferred due to the parasitic nature of the organism, although additional taxa from Excavata will provide greater understanding of the distribution and evolutionary association of the GSIII sequences.

In summary, multiple genes for GS isoenzymes are present within the chromalveolate lineages and have a diverse evolutionary history. We documented gene duplication, gene loss, and potential HGT and EGT events, but further taxon sampling is required to provide stronger support for some of the relationships observed in our phylogenetic analyses. The GS isoenzymes were predicted to function in the cytosol, chloroplast, and mitochondrion, and the number of isoenzymes varied among lineages. By identifying the different GS families present in

chromalveolates and red algae, this work provides a framework for further physiological studies of nitrogen metabolism in algae. Future studies can also be directed toward understanding the evolution of enzymes in biochemical pathways in organisms with complex genomic histories and the potential selective advantages of HGT of genes encoding essential enzymes within eukaryotes.

We thank David Hibbett and Manfred Binder for helpful discussions regarding the phylogenetic analyses and constructive review of an earlier version of the manuscript. Thoughtful comments from two anonymous reviewers are also appreciated. This research was supported by an NSF CAREER award (IBN 0238426) to D. L. R. Support for nucleic acid sequencing was provided by an award from the William Keck Foundation to Clark University. D. L. R. dedicates this work to the memory of Randall S. Alberte.

- Allen, A. E., Dupont, C. L., Obornik, M., Horak, A., Nunes-Nesi, A., McCrow, J. P., Zheng, H., Johnson, D. A., Hu, H., Fernie, A. R. & Bowler, C. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473:203–7.
- Allen, A. E., Vardi, A. & Bowler, C. 2006. An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr. Opin. Plant Biol.* 9:264–73.
- Andersen, R. A. 2004. Biology and systematics of heterokont and haptophyte algae. *Am. J. Bot.* 91:1508–22.
- Archibald, J. M. 2005. Jumping genes and shrinking genomes – probing the evolution of eukaryotic photosynthesis with genomics. *IUBMB Life* 57:539–47.
- Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306:79–86.
- Bachvaroff, T. R., Sanchez Puerta, M. V. & Delwiche, C. F. 2005. Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. *Mol. Biol. Evol.* 22:1772–82.
- Barbier, G., Oesterheld, C., Larson, M. D., Halgren, R. G., Wilkerson, C., Garavito, R. M., Benning, C. & Weber, A. P. 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.* 137:460–74.
- Berg, G. M., Glibert, P. M., Lomas, M. W. & Burford, M. A. 1997. Organic nitrogen uptake and growth by the chrysophyte *Aureococcus anophagefferens* during a brown tide event. *Mar. Biol.* 129:377–87.
- Bhattacharya, D., Yoon, H. S. & Hackett, J. D. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *BioEssays* 26:50–60.
- Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239–44.
- Brown, J. R., Masuchi, Y., Robb, F. T. & Doolittle, W. F. 1994. Evolutionary relationships of bacterial and archaeal glutamine synthetase genes. *J. Mol. Evol.* 38:566–76.
- Brown, K. L., Twing, K. I. & Robertson, D. L. 2009. Unraveling the regulation of nitrogen assimilation in the marine diatom *Thalassiosira pseudonana* (Bacillariophyceae): diurnal variations in transcript levels for five genes involved in nitrogen assimilation. *J. Phycol.* 45:413–26.
- Bruhn, A., LaRoche, J. & Richardson, K. 2010. *Emiliania huxleyi* (Prymnesiophyceae): nitrogen-metabolism genes and their

- expression in response to external nitrogen sources. *J. Phycol.* 46:266–77.
- Burki, F., Inagaki, Y., Brate, J., Archibald, J. M., Keeling, P. J., Cavalier-Smith, T., Sakaguchi, M., et al. 2009. Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol. Evol.* 1:231–8.
- Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., et al. 2007. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* 315:207–12.
- Casselton, P. J., Chandler, G., Shah, N., Stewart, G. R. & Sumar, N. 1986. Glutamine synthetase isoforms in algae. *New Phytol.* 102:261–70.
- Cavalier-Smith, T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and spirozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* 46:347–66.
- Cavalier-Smith, T. 2003. Protist phylogeny and the high-level classification of Protozoa. *Eur. J. Protistol.* 39:338–48.
- Claros, M. G. & Vincens, P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* 241:779–86.
- Cock, J. M., Sterck, L., Rouze, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard, V., et al. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* 465:617–21.
- Coruzzi, G. M., Edwards, J. W., Tingey, S. V., Tsai, F.-Y. & Walker, E. L. 1989. Glutamine synthetase: molecular evolution of an eclectic multi-gene family. In: Goldberg, R. [Ed.] *The Molecular Basis of Plant Development. Proceedings of an E.I. duPont Nermous-UCLA Symposium*. Alan R. Liss Inc., New York, pp. 223–32.
- Coyer, J. A., Robertson, D. L. & Alberte, R. S. 1994. Genetic variability within a population and between diploid/haploid tissue of *Macrocystis pyrifera* (Phaeophyceae). *J. Phycol.* 30:545–52.
- Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., Duprat, S., et al. 2003. Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl. Acad. Sci. U. S. A.* 100:10020–5.
- Eisenberg, D. H. S. G., Pfluegl, G. M. U. & Rotstein, S. H. 2000. Structure-function relationships of glutamine synthetases. *Biochim. Biophys. Acta* 1477:122–45.
- Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. *Nat. Prot.* 2:953–71.
- Emanuelsson, O., Nielsen, H. & von Heijne, G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8:978–84.
- Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O. & Taylor, F. J. R. 2004. The evolution of modern eukaryotic phytoplankton. *Science* 305:354–60.
- Falkowski, P. G. & Knoll, A. H. 2007. An introduction to primary producers in the sea: who they are, what they do, and when they evolved. In: Falkowski, P. G. & Knoll, A. H. [Eds.] *Evolution of Primary Producers in the Sea*. Academic Press, Burlington, pp. 1–441.
- Freshwater, D. W., Thomas, D. T. & Bailey, J. C. 2002. Research note: characterization of a cDNA encoding glutamine synthetase II from *Gelidium crinale* (Rhodophyta). *Phycol. Res.* 50:17–21.
- Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J. B., Carpenter, M. L., Field, M. C., Kuo, A., et al. 2010. The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* 140:631–42.
- Frommolt, R., Werner, S., Paulsen, H., Goss, R., Wilhelm, C., Zauner, S., Maier, U. G., Grossman, A. R., Bhattacharya, D. & Lohr, M. 2008. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Mol. Biol. Evol.* 25:2653–67.
- Ghoshroy, S., Binder, M., Tartar, A. & Robertson, D. L. 2010. Molecular evolution of glutamine synthetase II: phylogenetic evidence of a non-endosymbiotic gene transfer event early in plant evolution. *BMC Evol. Biol.* 10:198.
- Gobler, C. J., Renagham, M. J. & Buck, N. J. 2002. Impacts of nutrients and grazing mortality on the abundance of *Aureococcus anophagefferens* during a New York brown tide bloom. *Limnol. Oceanogr.* 47:129–41.
- Gould, S. B., Sommer, M. S., Hadfi, K., Zauner, S., Kroth, P. G. & Maier, U. G. 2006. Protein targeting into the complex plastid of cryptophytes. *J. Mol. Evol.* 62:674–81.
- Gruber, A., Vugrinec, S., Hempel, F., Gould, S., Maier, U.-G. & Kroth, P. 2007. Protein targeting into complex diatom plastids: functional characterization of a specific targeting motif. *Plant Mol. Biol.* 64:519–30.
- Gschloessl, B., Guermeur, Y. & Cock, J. M. 2008. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinf.* 9:393.
- Henze, K., Badr, A., Wettern, M., Cerff, R. & Martin, A. W. 1995. A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc. Natl. Acad. Sci. U. S. A.* 92:9122–6.
- Hill, R. T., Parker, J. R., Goodman, H. J. K., Jones, D. T. & Woods, D. R. 1989. Molecular analysis of a novel glutamine synthetase of the anaerobe *Bacteroides fragilis*. *J. Gen. Microbiol.* 135:3271–9.
- Hrdý, I. & Müller, M. 1995. Primary structure and eubacterial relationships of the pyruvate:ferredoxin oxidoreductase of the amitochondriate eukaryote *Trichomonas vaginalis*. *J. Mol. Evol.* 41:388–96.
- Huelsensbeck, J. P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–5.
- Inokuchi, R., Kuma, K. I., Miyata, T. & Okada, M. 2002. Nitrogen-assimilating enzymes in land plants and algae: phylogenetic and physiological perspectives. *Physiol. Plant.* 116:1–11.
- Jirutova, K., Horak, A., Bowler, C. & Obornik, M. 2007. Tryptophan biosynthesis in stramenopiles: eukaryotic winners in the diatom complex chloroplast. *J. Mol. Evol.* 65:496–511.
- Kamikawa, R., Sanchez-Perez, G. F., Sako, Y., Roger, A. J. & Inagaki, Y. 2009. Expanded phylogenies of canonical and non-canonical types of methionine adenosyltransferase reveal a complex history of these gene families in eukaryotes. *Mol. Phylogenet. Evol.* 53:565–70.
- Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–66.
- Keeling, P. J. 2004. Diversity and evolutionary history of plastids and their hosts. *Am. J. Bot.* 91:1481–93.
- Keeling, P. J. 2009. Chromalveolates and the evolution of plastids by secondary endosymbiosis. *J. Eukaryot. Microbiol.* 56:1–8.
- Khan, H., Parks, N., Kozera, C., Curtis, B. A., Parsons, B. J., Bowman, S. & Archibald, J. M. 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.* 24:1832–42.
- Kilian, O. & Kroth, P. G. 2005. Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J.* 41:175–83.
- de Koning, A. P., Brinkman, F. S. L., Jones, S. J. M. & Keeling, P. J. 2000. Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol. Biol. Evol.* 17:1769–73.
- Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M. S., Stanley, M. S., et al. 2008. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* 3:e1426.
- Lam, H. M., Coschigano, K. T., Oliveira, I. C., Melo-Oliveira, R. & Coruzzi, G. M. 1996. The molecular-genetics of nitrogen assimilation into amino acids in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47:569–93.
- Liaud, M.-F., Lichtl, C., Apt, K., Martin, W. & Cerff, R. 2000. Compartment-specific isoforms of TPI and GAPDH are imported into diatom mitochondria as a fusion protein: evidence in favor of a mitochondrial origin of the eukaryotic glycolytic pathway. *Mol. Biol. Evol.* 17:213–23.

- Maddison, W. & Maddison, D. 2000. *MacClade 4: Analysis of Phylogeny and Character Evolution*. Sinauer Associates Sunderland, Massachusetts.
- Markos, A., Miretsky, A. & Muller, M. 1993. A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote, *Trichomonas vaginalis*. *J. Mol. Evol.* 37:631–43.
- Maruyama, S., Matsuzaki, M., Misawa, K. & Nozaki, H. 2009. Cyanobacterial contribution to the genomes of the plastid-lacking protists. *BMC Evol. Biol.* 9:197.
- Matsuzaki, M., Kuroiwa, H., Kuroiwa, T., Kita, K. & Nozaki, H. 2008. A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*. *Mol. Biol. Evol.* 25:1167–79.
- Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S. Y., Mori, T., et al. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428:653–7.
- Maurin, C. & Le Gal, Y. 1997. Isoforms of glutamine synthetase in the marine coccolithophorid *Emiliania huxleyi*. *Comp. Biochem. Physiol.* 118B:903–12.
- Mifflin, B. J. & Habash, D. Z. 2002. The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. *J. Exp. Bot.* 53:979–87.
- Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K. & Bhattacharya, D. 2009. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324:1724–6.
- Nakai, K. & Kanehisa, M. 1991. Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins* 11: 95–110.
- Nedelcu, A. M., Blakney, A. J. & Logue, K. D. 2009. Functional replacement of a primary metabolic pathway via multiple independent eukaryote-to-eukaryote gene transfers and selective retention. *J. Evol. Biol.* 22:1882–94.
- Palenik, B., Brahamsha, B., Larimer, F. W., Land, M., Hauser, L., Chain, P., Lamerdin, J., et al. 2003. The genome of a motile marine *Synechococcus*. *Nature* 424:1037–42.
- Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., Dupont, C., et al. 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U. S. A.* 104:7705–10.
- Parfrey, L. W., Grant, J., Tekle, Y. I., Lasek-Nesselquist, E., Morrison, H. G., Sogin, M. L., Patterson, D. J. & Katz, L. A. 2010. Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst. Biol.* 59:518–33.
- Parker, M. S. & Armbrust, E. V. 2005. Synergistic effects of light, temperature, and nitrogen source on transcription of genes for carbon and nitrogen metabolism in the centric diatom *Thalassiosira pseudonana* (Bacillariophyceae). *J. Phycol.* 41:1142–53.
- Parker, M. S., Armbrust, E. V., Piovia-Scott, J. & Keil, R. G. 2004. Induction of photorespiration by light in the centric diatom *Thalassiosira weissflogii* (Bacillariophyceae): molecular characterization and physiological consequences. *J. Phycol.* 40: 557–67.
- Patron, N. J. & Waller, R. F. 2007. Transit peptide diversity and divergence: a global analysis of plastid targeting signals. *BioEssays* 29:1048–58.
- Reyes, J. C. & Florencio, F. J. 1994. A new type of glutamine synthetase in cyanobacteria: the protein encoded by the glnN gene supports nitrogen assimilation in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 176:1260–7.
- Riisberg, I., Orr, R. J., Kluge, R., Shalchian-Tabrizi, K., Bowers, H. A., Patil, V., Edvardsen, B. & Jakobsen, K. S. 2009. Seven gene phylogeny of heterokonts. *Protist* 160:191–204.
- Robertson, D. L., Smith, G. J. & Alberte, R. S. 1999. Characterization of a cDNA encoding glutamine synthetase from the marine diatom *Skeletonema costatum* (Bacillariophyceae). *J. Phycol.* 35:786–97.
- Robertson, D. L., Smith, G. J. & Alberte, R. S. 2001. Glutamine synthetase in marine algae: new surprises from an old enzyme. *J. Phycol.* 37:793–5.
- Robertson, D. L. & Tartar, A. 2006. Evolution of glutamine synthetase in heterokonts: evidence for endosymbiotic gene transfer and the early evolution of photosynthesis. *Mol. Biol. Evol.* 23:1048–55.
- Ronquist, F. & Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–4.
- Saldarriaga, J. F., McEwan, M. L., Fast, N. M., Taylor, F. J. R. & Keeling, P. J. 2003. Multiple protein phylogenies show that *Oxyrrhis marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. *Int. J. Syst. Evol. Microbiol.* 53:355–65.
- Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–4.
- Simon, N., Cras, A.-L., Foulon, E. & Lemée, R. 2009. Diversity and evolution of marine phytoplankton. *C. R. Biol.* 332:159–70.
- Slamovits, C. H. & Keeling, P. J. 2008. Plastid-derived genes in the nonphotosynthetic alveolate *Oxyrrhis marina*. *Mol. Biol. Evol.* 25:1297–306.
- Small, I., Peeters, N., Legeai, F. & Lurin, C. 2004. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4:1581–90.
- Southern, J. A., Parker, J. R. & Woods, D. R. 1986. Expression and purification of glutamine synthetase cloned from *Bacteroides fragilis*. *J. Gen. Microbiol.* 132:2827–35.
- Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–90.
- Stamatakis, A., Hoover, P. & Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* 57:758–71.
- Stelter, K., El-Sayed, N. M. & Seeber, F. 2007. The expression of a plant-type ferredoxin redox system provides molecular evidence for a plastid in the early dinoflagellate *Perkinsus marinus*. *Protist* 158:119–30.
- Swofford, D. L. 1998. *PAUP: Phylogenetic Analysis Using Parsimony, Beta Version 4.0b1*. Sinauer Associates Inc., Sunderland, Massachusetts.
- Szklarczyk, R. & Huynen, M. A. 2010. Mosaic origin of the mitochondrial proteome. *Proteomics* 10:4012–24.
- Taira, M., Valtersson, U., Burkhardt, B. & Ludwig, R. A. 2004. *Arabidopsis thaliana* GLN2-encoded glutamine synthetase is dual targeted to leaf mitochondria and chloroplasts. *Plant Cell* 16:2048–58.
- Takishita, K., Yamaguchi, H., Maruyama, T. & Inagaki, Y. 2009. A hypothesis for the evolution of nuclear-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase genes in “chromalveolate” members. *PLoS ONE* 4:e4737.
- Terashita, M., Maruyama, S. & Tanaka, K. 2006. Cytoplasmic localization of the single glutamine synthetase in a unicellular red alga, *Cyanidioschyzon merolae* 10D. *Biosci. Biotechnol. Biochem.* 70:2313–5.
- Weber, A. P., Oesterhelt, C., Gross, W., Brautigam, A., Imboden, L. A., Krassovskaya, I., Linka, N., et al. 2004. EST-analysis of the thermo-acidophilic red microalga *Galdieria sulphuraria* reveals potential for lipid A biosynthesis and unveils the pathway of carbon export from rhodoplasts. *Plant Mol. Biol.* 55:17–32.
- Whelan, S. & Goldman, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18:691–9.
- Yoon, H. S., 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, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattacharya, D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21:809–18.
- Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. 2002. The single, ancient origin of chromist plastids. *Proc. Natl. Acad. Sci. U. S. A.* 99:15507–12.
- Yoon, H. S., Müller, K. M., Sheath, R. G., Ott, F. D. & Bhattacharya, D. 2006. Defining the major lineages of red algae (Rhodophyta). *J. Phycol.* 42:482–92.

Supplementary Material

The following supplementary material is available for this article:

Fig. S1. Evolutionary relationships among GSII enzymes, partial representation of the complete tree.

Fig. S2. Evolutionary relationships among GSII enzymes, continued from Figure S1. Nodes with BBP support >0.95 are represented by thick lines. RAxML bootstrap values are indicated for nodes recovered in both analyses.

Fig. S3. Bipartite chloroplast transit sequences found chromalveolate GSII sequences.

Fig. S4. Evolutionary relationships among GSIII enzymes, partial representation of the complete tree continued on Figure S5.

Fig. S5. Evolutionary relationships among GSIII enzymes, partial representation of the complete tree continued from Figure S4.

Fig. S6. Mitochondrial target peptides were predicted *in silico* for GSIII protein sequences from select chromalveolates.

Table S1. Summary of the red algal taxa used in this study. Species names (Taxa), class, strain number (SAG#), and media used for culturing are presented. All cultures were obtained from Sammlung von Algenkulturen and the instructions for media preparation are provided at <http://www.epsag.uni-goettingen.de>.

Table S2. Degenerate primer sequences used for initial PCR amplification of GSII genes from red algae. Primer sequences are presented using IUBMB single letter codes where I represents inosine.

Table S3. Degenerate primer pairs used for initial PCR amplification of GSII sequences from red algae. Species names (taxa), the initial template used for amplification (genomic DNA or cDNA), and primers used in PCR amplification are provided. Primer sequences are provided in Table S2.

Table S4. Information for GSII protein sequences from organisms used in the present study. GenBank accession numbers and JGI DOE scaffold and protein ID information for the GSII proteins are provided.

Table S5. Information for GSIII protein sequences from organisms used in the present study. GenBank accession numbers and JGI DOE scaffold and protein ID information for the GSIII proteins are provided.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.