Nucleotide sequence and phylogeny of a plastocyanin gene in the marine diatom, *Thalassiosira oceanica*.

Edward Samuel Woo

Department of Biology
McGill University
Montreal, Quebec, Canada

August 2008

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science.

© Edward Woo 2008
Table of Contents

Table of Contents ................................................................. ii
Index of Tables ................................................................. iii
Index of Figures.................................................................. iii
Index of Supplemental Figures........................................... iii
Preface.................................................................................... iv
Acknowledgements ............................................................. v

Chapter 1. Background ..................................................... 6
Introduction............................................................................. 6
Trace Metal Requirements ...................................................... 6
Responses to Trace Metal Limitations.................................... 9
Horizontal and Vertical Gene Transfers.............................. 10
Determination of Inheritance ................................................ 13
Thesis Objectives ............................................................... 17
References............................................................................ 18

Chapter 2. Nucleotide sequence and phylogeny of a plastocyanin gene in the marine diatom, Thalassiosira oceanica. ................................. 27
Abstract.................................................................................. 28
Résumé..................................................................................... 29
Introduction.............................................................................. 30
Materials and Methods........................................................ 34
Results................................................................................... 40
Discussion.............................................................................. 45
References............................................................................ 55
Index of Tables

Table 1. Primers and probes used to identify and isolate the plastocyanin gene in *Thalassiosira oceanica* (CCMP 1005) using 5’ and 3’ RACE and degenerate PCR. ................................................................................................................................... 66

Table 2. List of species and accession numbers of plastocyanin protein sequences used for phylogenetic analysis. ......................................................................................................................... 68

Table 3. Species, clones, collection sites and accession numbers of 5.8s rRNA gene sequences of diatoms screened for plastocyanin by degenerate PCR........................................ 70

Index of Figures

Figure 1. Genomic structure of plastocyanin and the inferred protein sequence for *Thalassiosira oceanica* (CCMP 1005). ........................................................................................................... 72

Figure 2. Modeled globular surface images of the tertiary plastocyanin protein structure using the inferred protein sequence of *Thalassiosira oceanica* (CCMP 1005) (A, B) and the plastocyanin protein sequence of *Chlamydomonas reinhardtii* (C, D) ....... 74

Figure 3. Phylogenetic analysis of plastocyanin from *Thalassiosira oceanica* (CCMP 1005). ........................................................................................................................ 76

Figure 4. Phylogenetic tree of *Thalassiosira* spp. inferred by Bayesian analysis using 5.8s rRNA sequences ........................................................................................................ 78

Figure 5. Aligned sequences of plastocyanin from *Thalassiosira oceanica* (CCMP 1005) and *T. oceanica* (CCMP 1006) ......................................................................................... 80

Index of Supplemental Figures

Supplemental Figure 1. Southern blot analysis of coastal and oceanic species of phytoplankton using probes designed to plastocyanin from *Thalassiosira oceanica* (CCMP 1005)........................................................................................................ 82

Supplemental Figure 2. Maximum likelihood analysis of 18s rRNA gene sequences of *Thalassiosira* species ........................................................................................................... 85

Supplemental Figure 3. Distribution of cytochrome c₆ and plastocyanin in photosynthetic organisms containing chlorophyll c. ........................................................................... 88
Preface

All experiments and analyses were designed and carried out by the author. Holly Nikiforuk performed a large portion of the work associated with culturing of the specimen used in the experiments. Sarah-Ann Quesnel translated the English abstract into French. All ideas, experiments and written communication were completed in collaboration with the author’s supervisor, Dr. Neil Price.
Acknowledgements

My first thank you is to my supervisor, Dr. Neil Price, without his guidance, support and patience none of this would have been possible. His trust and encouragement allowed for true freedom to test personal limitations and, sometimes expensive, research methodology. He challenged me to see both the small details and the bigger picture. I am grateful to him for not only being a research advisor but also a mentor.

My supervisory committee, comprised of Dr. Thomas Bureau and Dr. Daniel Schoen, was a tremendous asset during my thesis and invaluable as a source of validation and as a sounding board for ambitious research goals. The Bureau and Schoen lab were all incredibly generous with their time and the “borrowing” of their equipment. All research carried out was truly a group effort, in one-way or another. Rod Docking and Dr. J. Busch were especially helpful as colleagues and floor mates. Thank you to all my undergraduate friends who listened to me drone on about my research and Chris Young for helping me with protein modeling.

Lastly, and in my Ed-esque way, I’d like to thank my lab mates and friends (really?). Thank you for your patience and encouragement. Thank you to Graham Peers and Carol Adly for bringing a naïve undergraduate into the fantastic world of diatom research. Thank you to Holly for being such a bubbly spring of enthusiasm and labor. Thank you to Sarah and Ada for the numerous cups of coffee and listening to me talk about ridiculous and sometimes, nonsensical things, and thank you to Kyle for all your “manly” advice, hanging out and for driving me to Quebec City and back to Alberta.

...And of course, a hearty thank you to my parents, without whom, I would not be where I am.
Chapter 1. Background

Introduction

The world’s oceans are a major sink for carbon dioxide and are increasingly important to offset rising atmospheric CO$_2$ levels as terrestrial sinks become limited (Hurtt et al., 2002, Canadell et al., 2007). Phytoplankton, which contribute to the oceanic uptake, represent only 1-2% of global carbon but account for roughly 40% of total global carbon (C) fixation (Falkowski et al., 2000). Their photosynthetic production drives the biological pump that delivers carbon to the deep ocean through deposition of particulate organic matter (Raven & Falkowski, 2002). A variety of chemical and environmental factors such as nutrient resources limit the rate of oceanic photosynthetic C fixation.

Trace metal nutrients are especially important because they are present at exceedingly low concentrations in surface water and are essential co-factors for phytoplankton growth. Current estimates are that iron, for example, limits phytoplankton growth in ~30% of the world’s oceans (Morel et al., 1991, Tsunda et al., 2003). To better understand the constraints on oceanic C production we need to know how trace metals affect phytoplankton growth and specifically how phytoplankton have adapted to lower trace metal concentrations.

Trace Metal Requirements

Trace metals are important constituents of enzymes and proteins (Da Silva & Williams, 1991) and in some secreted proteins such as ferretin (Castruita et al., 2006). In
virtually all organisms, they function in such enzymes as Cu, Fe, Zn or Ni superoxide dismutases, Fe-S proteins and cytochromes. In oxygenic photosynthetic organisms, trace metals are largely involved in the photosynthetic machinery (Raven et al., 1999). Iron, for example, is required for core photosynthetic proteins in photosynthetic center (PS) I and II and in the electron transport carriers ferredoxin and cytochrome b₆f (Raven et al., 1999). Other trace metals such as Cu are required for such proteins as plastocyanin (Sigfridsson, 1998), a congeneric protein to Fe-containing cytochrome c₆ (Kerfeld & Krogmann, 1998). The amounts and types of trace metals required by photoautotrophs may be largely determined by the presence and use of specific photosynthetic proteins. For example, most of the Cu requirement in some green algae and diatoms can be attributed to the use of plastocyanin (Li et al., 1996, Peers & Price, 2006).

Differences exist in the trace metal requirements of the two evolutionary distinct lineages of photosynthetic organisms; the red and green supergroups. These two supergroups of phytoplankton diverged early in the evolution of eukaryotic phytoplankton and gave rise to 2 plastid lineages (Falkowski et al., 2004). The green group, which utilizes chlorophyll b, has increased requirements for Cu, Zn and Fe while the red group, which utilizes chlorophyll c, has increased requirements for Cd, Co and Mn (Quigg et al., 2003). Variations in trace metal availability are thus predicted to affect the relative abundance of the red and green phytoplankton by affecting differentially their photosynthetic ability and viability.

Trace metal concentrations, in the ocean, varies in time and space and thus may affect phytoplankton ecology and evolution. In both the proterozoic and archean oceans, Fe was much more bioavailable than present day. An abundance of Fe is thought to have
contributed to the early ecological dominance of the green algae and their relatives (Falkowski et al., 2004); groups of organisms unified by a common plastid origin. Oxygenation of the environment led to Fe precipitation and decreased availability and eventual succession by members of the red plastid line (Falkowski et al., 2004). Other trace metals, such as Cu, increased as Fe decreased. The low Fe and Mn requirements and high Cd, Co and Mo requirements of the red supergroup are thought to be an adaptation to changes in availability of trace metals that allowed for its ecological success (Quigg et al., 2003). The disparity in metal concentrations between the paleo and modern day oceans also exists today between the near shore and offshore waters. Coastal waters contain far more nutrients, including trace metals, than oceanic waters because they receive inputs from terrestrial sources. Although the red supergroup has come to dominate modern oceans, certain members have a greater ability and propensity to proliferate in low nutrient waters than their counterparts found under high nutrient conditions (Ryther & Kramer, 1961, Sunda & Huntsman, 1995, Peers et al., 2005, Price, 2005).

Regardless of evolutionary heritage, trace metal requirements for growth of phytoplankton exceeds availability in the ocean (Boyd et al., 2000, Coale et al., 1996, Morel et al., 1991). This imbalance is a common theme in biology and is found in many systems. In the human body, for example, Fe is sequestered to limit bacterial growth (Moeck & Coulton, 1998) and provides a barrier to potential pathogens. Likewise, some regions of the oceans contain such low concentrations of some metals that primary production and hence photosynthesis is inhibited (Boyd et al., 2000, Coale et al., 1996, Morel et al., 1991, Marchetti et al., 2006). For organisms in the oceanic environment, the
imbalance is exacerbated by complex chemical reactions that affect the speciation of the metals and greatly reduce their bioavailability (Hudson, 2005).

**Responses to Trace Metal Limitations**

Reconciling the imbalance between the concentrations of metal required for growth and the concentrations present in the environment is a significant obstacle for photosynthetic organisms to overcome. Some species of phytoplankton have adapted and developed specific mechanisms to respond to trace nutrient limitation. Under Fe limiting conditions, for example, phytoplankton from Fe-limited environments increase Fe transport capacity (Maldonado & Price, 1999), produce siderophores (Trick *et al.*, 1983), reduce Fe (III) chelates (Maldonado & Price, 2001) and ingest insoluble particulate Fe (Nodwell & Price, 2001). Other physiological and biochemical responses act to decrease the cellular requirements for metals.

One such adaptation involves changes in the stoichiometry of components of the light reaction of photosynthesis to decrease the amount of Fe needed. Comparison of *Thalassiosira oceanica* and *Thalassiosira weissflogii* reveals that the former species has reduced the amount of photosystem I (PSI) and cytochrome b₆f, which are disproportionately iron rich (Strzepek & Harrison, 2004). Other adaptations in other species include the biochemical pathways that completely replace metal demanding biochemical pathways (Posey & Gherardini, 2000) or the replacement of metal-containing enzymes with non-metal containing enzymes. Indeed, the replacement of the Fe-based photosynthetic electron carrier, ferrodoxin, with a functionally equivalent flavin-based flavodoxin under Fe-limiting conditions has been observed in the laboratory.
and in the field (LaRoche et al., 1996). This adaptive response appears to be common to many species of plankton with the exception of some species isolated from Fe-rich environments (Erdner et al., 1999). The expression and presence of flavodoxin follows a habitat-dependent pattern that corresponds to the availability of iron in the environment, occurring in habitats where trace-nutrients are limiting (Erdner & Anderson, 1999, Erdner et al., 1999). Adaptations such as these have helped oceanic phytoplankton survive in metal poor regions of the sea. What other adaptations do oceanic phytoplankton posses?

**Horizontal and Vertical Gene Transfers**

The adaptations described above depend on the genetic make up of the organism in question. After all, the sum of the genetic composition of an organism is the product of its historical or recent adaptations. The genetic composition of an organism is determined by the inheritance of genes. These genes may be inherited vertically from progenitors or, less frequently, horizontally from exogenous sources. In bacteria, horizontal transfer of genes from exogenous sources may be a significant mechanism of gene acquisition and play a role in survival. Indeed, it is been proposed that the majority of operational genes are continuously exchanged among some species (Jain et al., 1999). The observation is not restricted to terrestrial bacterial communities. Within the marine environment, multiple species of the cyanobacteria, *Prochlorococcus* and *Synechococcus*, have acquired the eukaryotic Calvin cycle genes and fructose bisphosphate aldolase, resulting in partial or complete replacement of the endogenous homologues (Rogers et al., 2007). Other cases of transfer between domains have been reported, for example, as much as
24% of the genome of the bacterium, *Thermotoga maritime*, originated from archaea (Nelson *et al.*, 1999).

Despite a pervasive number of examples of horizontal gene transfer in bacteria much less is known about gene transfer among eukaryotes. Three cases of interchange of mitochondrial genes have been documented in species of higher plants (Andersson, 2005). In two of these, the mechanisms of transfer remain unknown but viruses have been implicated (Bergthorsson *et al.*, 2003, Won & Renner, 2003). In the third, an intermediary vector is thought not to be required because of prolonged contact between the donor and host (Davis & Wurdack, 2004). The mixotroph, *Bigelosiella natans*, contains a chloroplast derived from secondary endosymbiosis of a green alga, but roughly 20% of its 78 plastid targeted proteins have been acquired from foreign sources other than the green alga symbiont (Archibald *et al.*, 2003). For the most part, examples of horizontal gene transfer in protists can be traced to a phagocytic event (Doolittle, 1998). It should be noted that although some modern day photosynthetic organisms lack phagocytic ability, at one time their ancestors must have been phagocytic to acquire the chloroplast. Lack of phagocytic ability does not exclude the possibility of recent lateral gene transfer by non-phagocytic protists by other mechanisms such as viral vectors that have been observed to carry photosynthetic genes (Zeidner *et al.*, 2005). Neither do domain boundaries seem to hinder lateral transfer of genes into eukaryotes. Several cases of horizontal gene transfer from prokaryote to eukaryote have been postulated, including the transfer of 3-hydroxy-3-methylgutaryl-coenzyme A class 2 reductase found in *Giardia* (Loftus *et al.*, 2005), iron hydrogenase in *Nyctotherus* and the fungal catalases (Hall *et al.*, 2005).
Endosymbiotic gene transfer and subsequent vertical inheritance has largely determined an organism’s trace metal requirement (Quigg et al., 2003). Contemporary photosynthetic eukaryotes are composed of two separate evolutionary lineages that diverged following the primary photosynthetic endosymbiosis. Member of these lineages use unique combinations of chlorophyll molecules and extrinsic proteins associated with the oxygen evolving center (De Las Rivas et al., 2004). The “green” lineage uses chlorophyll a and chlorophyll b while the “red” lineage uses chlorophyll a and chlorophyll c. The endosymbiont (plastid) within these photosynthetic eukaryotes underwent gene reduction and transfer of genes to the host nucleus. Each lineage has inherited a unique composition of cellular proteins which has largely determined its trace metal requirements. The red supergroup preferentially using Cd, Co and Mn metalloproteins, and the green supergroup Cu, Zn and Fe metalloproteins (Quigg et al., 2003). These trace metal requirements have in turn affected the distribution and activity of those photoautotrophs in the sea.

The acquisition of genes encoding proteins that function in the mitochondria or chloroplast requires a signal peptide to target the protein to the location of function. In the chloroplast of chlorophyll a/b-containing organisms this necessitates an encoded, translated and attached peptide to the protein of function (Morse & Nassoury, 2005). In chlorophyll a/c-containing organisms, greater complexities exist because of the presence of one or two additional membranes surrounding the chloroplast; a result of secondary or tertiary endosymbiosis. The presence of a signal peptide allows for the proper translocation of the protein through these additional membranes.
Inheritance of specific genes through vertical or horizontal gene transmission can be determined using an established set of criteria. The evidence of horizontal transfer lies in the signal or marker of the event verses the event itself. Codon composition difference in organisms also allows us to detect a horizontal gene transfer event (Palenik et al., 2006). Codon composition of different strains and of different species may vary (Rocap et al., 2003). Provided that the sequence between donor and recipient organisms differ in their codon bias, transferred genes may be delineated by the difference. For example, the oceanic cyanobacteria, *Synechococcus* (WH 8501) possess genomic regions encoding metal-related genes that have atypical GC content compared to the rest of it’s genome but are absent from its sequenced coastal cognate, *Synechococcus* (CC 9331) (Palenik et al., 2006). Although straightforward, this criterion is limited and more difficult to apply or evaluate due to the high quantity of sequence data that is needed.

Another criterion to identify transferred genes can be the unusual distribution of the gene. In a particular species the presence of a gene that is otherwise found in distant relatives but not close relatives may be an indicator (Brown, 2003). Clearly, this criterion alone is insufficient to determine lateral gene transfer because gene loss can also create such a pattern. The unusual distribution of glutamate dehydrogenase genes has been used as evidence of lateral gene transfer within and between prokaryotes and eukaryotes (Andersson & Rogers, 2003).

A third common criterion for lateral gene transfer is the appearance of unexpected gene homologies (Campbell, 2000). Structural and sequence homology has been used to
suggest lateral gene transfer of dermonecrotic toxin between spiders and bacteria (Cordes & Binford, 2006). Programs such as Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1997) provide a quick way to screen for putative genes that may have been transferred through other sources. Finally, the strongest indicator for putative horizontal gene transfer event may be the discordant relationship between the gene sequence or encoded protein (Avise, 2004, Syvanen, 1994) and other phylogenetic markers between species. Incongruence of the phylogenetic tree for two Archeal genes in *Thermotoga maritima*, together with gene structure and distribution data, has been used to identify multiple lateral transfers between Achaeae and Bacteria (Nesbø et al., 2001).

In general, both the gene tree and the species tree reach greater accuracy as the number of characters (genes and species, respectively) and size of the character (nucleotide sequence length) increases (Harrison & Langdale, 2006). Datasets may be analyzed using several different methods such as maximum likelihood, maximum parsimony, neighbour-joining and Bayesian analysis. These methods vary in the algorithms they employ (Holder & Lewis, 2003), and each has its own strengths and weaknesses.

No single phylogenetic method performs well under all conditions (Li, 1997). Some of the methods make explicit assumptions about the pattern and rate of character substitution while others make some and others none at all. A method that makes no assumptions is not necessarily better than ones that do. The performance and usefulness of a particular method is dependent on the assumptions and also the computational procedure and optimality criterion used. These issues are considered below.
The maximum likelihood (ML) searches for the maximum likelihood value for the observed character states for each possible tree and chooses the tree that results with the largest ML value (Felsenstein, 2005). It makes explicit assumptions about the rate of evolution and pattern of substitution of characters and hence allows for the choice of different probabilistic models. In general, ML methods are considered to be consistent, but inconsistency can occur if the substitution model that is chosen is unrealistic or if the rate of evolution is assumed to be uniform although it is not.

On the other hand, maximum parsimony (MP) inference makes no explicit assumptions at all and searches for the tree that requires the smallest number of evolutionary changes to explain the observed character states of the outer taxonomic units (Salemi, 2003). Minimizing the number of substitutions also conversely minimizes the number of homeoplastic events and hence this approach does not work well for sequences that have had a large number of parallel, convergent or back reversal substitutions. When the degree of divergence becomes so large that homeoplasys are common, this method may result in misleading inferences.

Neighbour-joining methods seek to find neighbouring outer taxonomic units (OTU) sequentially such that the total length of all the branches forming the tree is minimized. This method, like MP, does not assume a rate constancy but assumes that the effects of unequal rates among branches can be corrected from the distances established in the distance matrix used to construct it (Salemi, 2003). Hence, the accuracy of the distance matrix will affect the accuracy of the NJ inference. If the lengths of the sequences are short then large statistical errors will occur in the distance matrix and the accuracy of this inference will be affected.
A concern with all three of these methods is that when a sufficiently large number of OTUs are used, these methods will use heuristic algorithms, finding the best solution to a subset of sequences, and then sequentially adding branches to that subset to find the best inferred tree. However, this approach does not fully appraise the sample space that incorporates all trees and may result in the inference of a tree that is the best local solution but not global. The use of Bayesian analysis minimizes the probability of this occurring (Holder & Lewis, 2003).

Bayesian analysis (Ronquist & Huelsenbeck, 2005) uses a method based on the Bayes theorem. This type of analysis starts with a random tree typology with random branch lengths and random likelihood parameters and slight changes to the parameter that are either accepted or rejected depending on the posterior probability. After a significant number of cycles the trees generated by Markov sampling start to converge around a posterior probability maxima. A posterior probability can be best described as a probability that the deduced tree fits the observations, in our case, sequence data. The deduced posterior probability may be either the best local posterior probability or the best global posterior probability. To ensure that the tree converged upon is the tree with the highest posterior probability the program uses Metropolis couple Markov chain Monte Carlo to create heated chains to ensure that the resultant tree is indeed the correct one.

Depending on the results of these types of analysis we can infer the relationship of a specific sequence to homologous sequences present in other organisms, thus allowing us to deduce the possible source of our sequence. To explain why a specific gene has been acquired we can examine current ecological data, such as the distribution of the organisms that posses it, as well as, make inferences from geo-historical information.
Thesis Objectives

Recently, an oceanic diatom, *T. oceanica* (CCMP 1005), isolated from a Fe-limited region of the sea was found to have drastically reduced Fe requirements. Differential spectroscopy identified a small soluble blue copper protein that had a high degree of sequence similarity to plastocyanin (Peers and Price, 2006), a Cu-containing photosynthetic electron transport protein.

In the following body of work I have isolated the gene for plastocyanin from chlorophyll a/c-containing diatom, *T. oceanica* (CCMP 1005), an organism that has been found to be especially adapted to low Fe environments. Until now, plastocyanin has been observed to be restricted to the chlorophyll a/b-containing lineage of photosynthetic organisms (Sandmann, 1986, Sandmann *et al.*, 1983, Sigfridsson, 1998). Using the sequence of the plastocyanin gene I have evaluated its phylogenetic relationship to other plastocyanins and have speculated on the possible reason for such an acquisition. The occurrence of the plastocyanin gene in other chlorophyll a/c-containing diatoms was investigated.
References


Morse, D. & Nassoury, N. 2005. Protein targeting to the chloroplast of photosynthetic eukaryotes: getting there is half the fun. *Biochimica et Biophysica Acta* **1743**:5-19.
Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L.,
Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton,
G. G., Fleischmann, R. D., Eisen, J. A., White, O., Salzberg, S. L., Smith, H. O.,
Venter, J. C. & Fraser, C. M. 1999. Evidence for lateral gene transfer between
Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature*
399:323-29.

‘archaeal’ genes in *Thermotoga maritima* reveal multiple transfers between

mixotrophic phytoplankton. *Limnology and Oceanography* 46.

Palenik, B., Ren, Q., Dupont, C. L., Myers, G. S., Heidelberg, J. F., Badger, J. H.,
Madupu, R., Nelson, W. C., Brinkac, L. M., Dodson, R. J., Durkin, A. S.,
Daugherty, S. C., Sullivan, S. A., Khouri, H., Mohamoud, Y., Halpin, R. &
Paulsen, I. T. 2006. Genome sequence of *Synechococcus* CC9311: Insights into
adaptation to a coastal environment. *Proceedings of the National Academy of
Sciences* 103:13555-59.


Examination of hydroxamate-siderophore production by neritic eukaryotic marine

Tsunda, A., Takeda, S., Saito, H., Nishioka, J., Nojiri, Y., Kudo, I., Kiyosawa, H.,
Shiomoto, A., Imai, K., Tsuneo, O., Simamoto, A., Tsumune, D., Yoshimura, T.,
Aono, T., Hiuma, A., Kinugasa, M., Suzuki, K., Sohrun, Y., Nori, Y., Tani, H.,
2003. A mesoscale iron enrichment in the western Subarctic Pacific induces a
large centric diatom bloom. Science 300.

Won, H. and Renner, S. S. 2003. Horizontal gene transfer from flowering plants to


Potential photosynthesis gene recombination between Prochlorococcus and


(Submitted to the Journal of Phycology)

Edward S. Woo¹, Neil M. Price¹

¹Department of Biology, McGill University, 1205 Ave. Dr. Penfield, Montréal, Québec, Canada, H3A 1B1.
Abstract

Diatoms are thought to have acquired an Fe-containing cytochrome (cyt) c$_6$ to transfer electrons between cyt b$_6$f and photosystem (PS) I of the photosynthetic apparatus like other chlorophyll a/c-containing phytoplankton. Here we report the isolation and cloning of a plastocyanin gene from *Thalassiosira oceanica* (CCMP 1005). The gene encoded a Cu-containing protein that is known in other organisms to functionally replace the Fe-containing cyt c$_6$. The inferred protein sequence had the highest identity with the green haptophyte, *Emiliania huxleyi*, and possessed many of the globular properties necessary for function and interaction with upstream and downstream partners. Eleven strains of oceanic and coastal diatoms were screened for the presence of the plastocyanin gene using degenerate primers: one other species was observed to contain the gene; *T. oceanica* (CCMP 1006). Phylogenetic analysis of the 5.8 rRNA gene of these species showed that both *T. oceanica* strains with plastocyanin were closely related to each other. The cloned sequence of T. oceanica (CCMP 1006) contained greater than 80% of the protein-coding region and shared 99% nucleotide identity and 100% conserved unique intronic region. The inferred protein sequence of this species had 100% identity with the inferred protein sequence of *T. oceanica* (CCMP 1005).
Résumé

Durant l'évolution de la photosynthèse, il est pensé que les diatomées ont acquis cytochrome (cyt) c6, une protéine contenant un atome de Fe, pour le transfert des électrons entre le complexe cyt b6f et le photosystème (PS) I de l'appareil photosynthétique, comme d'autres phytoplanctons ayant les chlorophyll a/c. Ici, nous rapportons l'isolement et clonage du gène de plastocyanin du diatomée Thalassiosira oceanica (CCMP 1005). Le gène codait une protéine contenant un atome de Cu qui est connue de remplacer fonctionnellement cyt c6 (contenant un atome de Fe). La sequence inférée de la protéine montrait la plus grande identité avec Emiliania huxleyi, un haptophyte vert (green haptophyte), et possédait de nombreuses propriétés globulaires nécessaires aux fonctions et intéractions avec les protéines partenaires en amont et en aval. Onze espèces de diatomées océanique et côtières ont été examinées pour la présence du gène codant pour plastocyanin en utilisant les amorces dégénérées (degenerate primers): une autre espèce a révélé la présence du gène de plastocyanin; T. oceanica (CCMP 1006). L'analyse phylogénétique des gènes 5.8 rRNA de cet espèce a démontré que la souche de T. oceanica possédaient le plastocyanin sont étroitement reliées entre elles. Les séquences clonées de T. oceanica (CCMP 1006) recelaient plus de 80% de la région cryptant la protéine et partageaient 99% des nucléotides et 100% des régions introniques uniques. La séquence de protéine inférée de ces espèces de phytoplanctons ont démontré 100% d'identité avec celle inférée pour T. oceanica (CCMP 1005).
Introduction

Oxygenic photosynthesis evolved within an ancestral prokaryote (cyanobacterium) and subsequently spread to eukaryotes through primary, secondary and tertiary endosymbiosis. Its essential features include a manganese-containing oxygen evolving center, two photosystems with light-harvesting chlorophyll pigments, and electron carriers that participate in cyclic and non-cyclic electron flow (Nelson & Ben-Shem, 2005). Although differences are found in the types of accessory pigments used to capture light energy and in the extrinsic proteins of the oxygen evolving center (De Las Rivas et al., 2004), most components of the light reaction of photosynthesis have remained conserved among eukaryotes with two notable exceptions. First, in algae, flavodoxin and ferrodoxin are functionally interchangeable redox proteins that transfer electrons to NADP+. When Fe is in short supply flavodoxin can replace ferrodoxin (a heme-containing protein) (Erdner et al., 1999), but some higher plants can only use ferrodoxin (Arabidopsis Genome Initiative, 2000). Second, in some cyanobacteria and unicellular green algae cytochrome c₆ or plastocyanin can be used in electron transport depending on environmental conditions (Inda & Pelato, 2002, Merchant & Bogorad, 1986). Cytochrome c₆ in absent from higher plants and plastocyanin is used constitutively (De la Rosa et al., 2002). Unicellular algae with red algal-derived plastids are thought to use only cytochrome c₆ and to lack plastocyanin (Sandmann, 1986, Sandmann et al., 1983).

Cytochrome c₆ and plastocyanin are soluble metalloproteins in the thylakoid lumen that transfer electrons between cytochrome b₆f complex and photosystem I
Plastocyanin is a 10,000-12,000 kDa Cu-containing protein that has been extensively studied and well characterized (Sigfridsson, 1998). All plastocyanins exhibit a characteristic oxidized-minus-reduced spectrum and contain several conserved amino acids crucial to Cu binding (Nersissian & Shipp, 2002). Cytochrome c$_6$ is a low molecular weight soluble redox carrier typically comprised of 83-90 amino acids and a single haem group (Howe et al., 2006). Both proteins have similar isoelectric points and midpoint redox potentials (De la Rosa et al., 2002). As far as it is known, there is little reason to use one protein over the other (Raven et al., 1999), except if the necessary metal co-factor is unavailable.

In cyanobacteria and certain unicellular green algae, plastocyanin expression is regulated by the availability of copper. Chlamydomonas reinhardtii, for example, synthesizes plastocyanin preferentially in the presence of Cu in adequate amounts required for photosynthesis (Merchant & Bogorad, 1986). When Cu is limiting, plastocyanin production ceases and the heme-containing cytochrome c$_6$ is induced as a substitute (Li et al., 1996). In the green alga Scenedesmus vacuolatus, cytochrome c$_6$ concentrations increases when Fe concentration increases (Peleato et al., 2003). When one of these proteins is absent and insufficient amounts of the metal co-factor is available for synthesis of the other, growth is inhibited.

The distribution of plastocyanin and cytochrome c$_6$ among photosynthetic eukaryotes is believed to be determined by evolutionary history (Sandmann, 1986, Sandmann et al., 1983). Both proteins existed in oxygenic photosynthetic bacteria that eventually became plastids of green algae and higher plants. Depending on gene loss, living cyanobacteria may posses both plastocyanin and cytochrome c$_6$. Unicellular
chlorophyll a/b-containing organisms, such as the unicellular green alga, *Chlamydomonas reinhardtii*, may use either protein to transport electrons from cytochrome b₆f to PSI (Merchant & Bogorad, 1986). The lineage that gave rise to higher plants appears to have lost cytochrome c₆ (De la Rosa et al., 2002) which only use plastocyanin. A cytochrome c₆-like protein has been identified in *Arabidopsis* (Gupta et al., 2002), but it contains an additional 12 amino acid loop that alters the surface properties of the protein and is not believed to be a functional cytochrome c₆ (Howe et al., 2006). Chlorophyll a/c-containing organisms, derived from secondary endosymbiosis of a red alga, are thought to have only cytochrome c₆ (Sandmann, 1986, Sandmann et al., 1983).

The presence of cytochrome c₆ and absence of plastocyanin in algae with chlorophyll a/c was originally established by spectroscopic analysis (Sandmann, 1986, Sandmann et al., 1983). In eight species of chromophyte algae that were examined the oxidized-minus-reduced spectra lacked the diagnostic absorption maxima at 597 nm for a type-1 copper protein (Sandmann et al., 1983). Since this work, many more classes and species of algae have been obtained in culture and are available for study. Very recent results challenge the pattern of distribution of plastocyanin and cytochrome c₆ originally established. Plastocyanin sequences have now been identified from expressed sequence tags (EST) of *Emiliania huxleyi* (haptophyte) (Nosenko et al., 2006) and *Karenia brevis* (dinoflagellate) (Nosenko et al., 2006) and the nucleotide database of *Karlodinium micrum* (dinoflagellate) (Patron et al., 2006). Peers and Price (2006) purified a plastocyanin from *Thalassiosira oceanica* (CCMP1005) and showed that it had a high similarity to plastocyanin from cyanobacteria. Reduced-minus-oxidized spectra indicate
that *T. oceanica* (CCMP1005) has reduced amounts of cytochrome b$_6$f, as compared to its coastal cognate, *Thalassiosira weissflogii* (CCMP 1335), and non-detectable amounts of cytochrome c$_6$ even in Fe-replete media (Strzepek & Harrison, 2004).

The distribution of plastocyanin in *Karenia brevis* is thought to have occurred through tertiary endosymbiosis of a haptophyte ancestor that itself acquired the plastocyanin from either a green alga or cyanobacterium (Nosenko et al., 2006). *Karlodinium micrum*, like *Karenia brevis*, is also capable of phagocytosis and may have acquired plastocyanin in a similar manner. Phylogenetic inference indicates that plastocyanin in these organisms diverged from plastocyanin from the green lineage (Nosenko et al., 2006). Acquisition of plastocyanin through phagocytosis is unlikely in a diatom because of its rigid outer Si frustule.

In this paper we used published protein sequence data for primer design to isolate a fragment of the gene encoding plastocyanin in *T. oceanica* (CCMP 1005). We subsequently cloned and sequenced the complete mature protein coding sequence and used phylogenetic analysis to infer the origin of this gene in *T. oceanica* (CCMP 1005). Using our deduced protein sequence we screened 10 other strains of chlorophyll a/c-containing diatoms for the presence of the plastocyanin gene.
Materials and Methods

Algal Material used. Different species and strains of coastal and oceanic Thalassiosira were chosen to represent the branches of the maximum-likelihood phylogenetic tree inferred from 5.8 rRNA sequences. Minutocellus spp. was selected as an outgroup. Phytoplankton were obtained from the Culture Collection of Marine Phytoplankton (CCMP), Bigelow Laboratories for Ocean Sciences (West Boothbay Harbor, ME, USA) and maintained in f/2 media. Filter sterilized nutrients were added to 0.22 μm filtered natural seawater according to protocol (Guillard, 1975), autoclaved and filtered again. Thalassiosira oceanica (CCMP 1005, CCMP 1006, CCMP 999); T. rotula (CCMP 1647); T. pseudonana (CCMP 1015, CCMP1335); T. weissflogii (CCMP 1336); Minutocellus polymorphous (CCMP 497) and Minutocellus sp. (CCMP 1701) were grown under continuous light with an irradiance of 180 μmol quanta m⁻² s⁻¹ at 22°C. Thalassiosira guillardii (CCMP 988) and T. pseudonana (CCMP 1014) were grown under a 13:11 light: dark cycle with an irradiance of 100 μmol quanta m⁻² s⁻¹. Cultures used for degenerate PCR were grown in 10L volumes of f/2 media in clear polycarbonate carboys (Nalgene, Rochester, NY, USA), harvested by filtration onto 2 μm polycarbonate filters (GE Osmotics, Minnetonka, MN, USA), snap frozen in liquid nitrogen and stored at –80°C.

Extraction of nucleic acids. Genomic DNA was extracted using DNeasy Mini-Prep Kit (Qiagen, Mississauga, ON, CAN) according the manufacture’s protocol and further purified by ethanol precipitation (Sambrook & Russell, 2004). The quality of the DNA
was determined by performing a full spectrum (220-750 nm) scan using a NanoDrop ND-1000 spectrometer (NanoDrop Technologies Inc., Wilmington, DE, USA). Only DNA samples with a 260 nm/280 nm ratio greater than 1.8 and a 260/230 ratio greater than 2.2 were used. Purified DNA was stored at -20°C until use. Total RNA was extracted using RNeasy (Qiagen) according to the manufacture’s instructions. RNA was stored at -80°C. RNA integrity was verified by comparing the 28s and 18s rRNA Swedberg peaks on an Agilent 2100 Bioanalyzer (Genome Quebec, Montreal, QC, CAN).

**Plastocyanin isolation, cloning and annotation.** Two degenerate primers, NTERM forward and IPTERP reverse, were manually designed to the plastocyanin’s N-terminal sequence AQTVEVKM and a conserved amino acid sequence, PHNVVFDEDNIP, of plastocyanin (Table 1). PCR amplification was performed using 0.5 μM of primers with 2.5 U TAQ DNA polymerase (Qiagen) and 1.5 mM MgCl₂, 10 mM dNTP and 250 ng of *T. oceanica* (CCMP 1005) DNA in a 50 μL reaction volume. The PCR cycling condition used was: 94°C for 10 min followed by 35 cycles of 94°C for 1 min, 42°C for 1 min and 72°C for 1 min with a final elongation step at 72°C for 10 min. PCR products were visualized on a 1% high purity agarose gel (Invitrogen, Burlington, ON, CAN), excised and extracted with the QIAquick PCR Purification kit (Qiagen). Fragments were ligated into the linearized pCR2.1 vector with TA overhangs and covalently bound Topoisomerase (Invitrogen). The resultant plasmids were subsequently used to transform chemically competent *Escherichia coli* DH5α cells (Invitrogen). Transformed cells were grown on LB agar containing 50 μg/ml kanamycin. Ten clones were subsequently chosen randomly and grown in liquid LB with 50 μg/ml kanamycin. Plasmids were isolated from
the bacterial cells using a QIAprep Spin Miniprep kit (Qiagen). Plasmids were sequenced by Genome Quebec using a 3730xl DNA Analyzer system (Applied Biosystems, Foster City, CA, USA) using M13 primers. Identities of the sequences were checked by BLASTX searches of the National Center for Biotechnology Information’s (NCBI) non-redundant database (Altschul et al. 1990) using default parameters. A sequence of high homology to plastocyanin was identified and used to design a 3’ internal forward primer for 3’ Restriction Ligase Mediated Rapid Amplification of DNA ends (Ambion, Palo Alta, CA, USA) (Table 1). A 5’ internal reverse primer (Table 1) was designed to the original internal genomic fragment and 5’RACE was completed using a 5’SMART RACE cDNA Amplification Kit (BD Biosciences, Mississauga, ON Canada). Ten products were cloned using TOPO-pCR2.1 vector as described above. The complete coding sequence was determined by overlapping fragments and blasted against the NCBI non-redundant database using the BLASTX program with default parameters.

**Comparative Modeling.** Using the previously determined N-terminal sequence (Peers & Price, 2006) to identify the proper reading frame, the mature protein sequence was deduced and blasted against the protein database. The primary amino acid sequence was then threaded to a number of best templates using the first approach option of the SwissModel program (Schwede et al., 2003). PyMol (DeLano, 2004) was then used to generate a 3-D representation and identify hydrophobic, acidic and basic patches on the protein’s surface.

**Sequence analysis of plastocyanin and rRNA sequences from photosynthetic organisms and phylogenetic inference.** The protein sequence was used to search the NCBI non-
redundant proteins sequence database (Altschul et al., 1990). Protein sequences with an expected score greater than 6.0e-4 were retrieved. The accession numbers of the sequences and the species used for phylogenetic analysis are listed (Table 2).

Conserved blocks of sequences were constructed using the Block Maker program (http://blocks.fhcrc.org/blockmkr/make_blocks.html) (Henikoff et al., 1995) and concatenated. A multiple alignment was created with ClustalW using the default settings and manually edited with BioEdit. This procedure resulted in a 26-taxon dataset with 86 amino acids. A substitution matrix was chosen using the best BIC and AIC values as calculated by PROTTEST (http://darwin.uvigo.es/software/prottest_server.html). Protein maximum likelihood phylogenies were inferred using PHYML with a Whelan and Goldman (WAG) substitution matrix and with the assumption of both invariant residuals and variable rates of evolution. The alpha values were calculated using 8 rate categories. Using TREE-PUZZLE (http://www.tree-puzzle.de) a second protein maximum likelihood tree was constructed using branch swamping and a WAG substitution model with an eight-model discrete-gamma model. Both trees resulted in identical typologies. Each maximum likelihood inference was bootstrapped 200 times. For all other methods, the assumption was made that rate of evolution was constant between amino acid residuals. The same multiple alignment was used in subsequent analysis. Neighbour-joining tree was produced with the NJ program from the PHYLIP package (http://evolution.genetics.washington.edu/phylip) using 200 replicates and the Jones-Taylor-Thornton matrix assuming invariant sites and 8-category rate of substitution. Consensus tree was created using CONSENSUS from the PHYLIP package and
visualized with TREEVIEW (http://taxonomy.zoology.gla.ac.uk/rod/treeview) for each analysis.

Bayesian analysis was performed using Mr. Bayes (Ronquist & Huelsenbeck, 2005) with a WAG substitution matrix. Every 100 generations were sampled and the analysis was terminated when convergence was reached (2,500,000 generations). Deviation between posterior probabilities was 0.005837. Four thousand generations were burnt at the end of the run and the potential scale factor equalled one. TREEVIEW was used to create a graphic representation.

Phylogenetic analysis of species used for degenerate PCR was performed with Mr. Bayes, as described above. The accession numbers of 5.8S rRNA sequences used for this analysis are listed (Table 3).

*Degenerate PCR.* Degenerate PCR was used to amplify a portion of the sequence encoding plastocyanin using 2.5U TAQ DNA polymerase (Qiagen) and 1.5 mM MgCl₂, 10 mM dNTP and 250 ng of genomic DNA in a 50 μl reaction volume. BLOCKB forward and BLOCKC reverse, designed to the conserved regions of plastocyanin, CEPHQGAG and PHNVVF, and degenerate NTERM forward and CEPH reverse primers were used (Table 1). For each species a variety of PCR conditions were tested. Primer concentration varied from 0.1 μM to 1 μM, melting temperatures varied from 32°C to 47°C and cycle durations varied from 1 to 3 min, genomic DNA quantity ranged from 100 ng to 1μg. In some samples, DMSO was used to reduce the possibility of inhibition by secondary structures. The following PCR reaction conditions were found to
amplify plastocyanin: 94°C for 3 min followed by 35 cycles of 94°C for 1 min, 45°C for 1 min, 72°C for 1 min and with final 10 min extension step at 72°C. Products were visualized on a 1% gel and were excised, cloned and sequenced as described above.
Results

Gene Cloning. A 395 nucleotide gene fragment was isolated from *Thalassiosira oceanica* (CCMP 1005) using degenerate primers to the N-terminus of plastocyanin (NTERM forward) (Peers & Price, 2006) and the conserved motif, CEPHQGAG Block C reverse (Table 1). The conserved region was identified by a multiple alignment of 25 plastocyanin sequences from the NCBI database. The isolated gene fragment encoded for an amino acid sequence, PHNVVDEDIP, common to most plastocyanins, and included one of the two histidines (H) required for Cu-binding by the mature protein. It also encoded a portion of the N-terminal sequence determined from protein sequencing (Peers & Price, 2006).

The gene fragment possessed a 95 nucleotide intron that had no appreciable similarity to any sequence in the NCBI database. Nucleotides surrounding the 5’ splice site, nucleotide 95 to 96, were similar to the canonical AGGT(A/G)AGT sequence (Shapiro & Senapathy, 1987, Mount, 1982) except thymidine replaced adenine at nucleotide position 93. Such substitution occurs in 17% of eukaryotic splice sites (Shapiro & Senapathy, 1987). Translation of the gene sequence and alignment with known plastocyanin protein sequences also identified these residuals as a splice site. An automated exon intron detection program, GENESCAN (Burge & Karline, 1998), predicted with 98.5% probability that this sequence was a splice site. The 3’ splice site, nucleotides 192 to 193, followed canonical rules and corresponded exactly to conserved 3’ splice site signal (Shapiro & Senapathy, 1987). The intron contained a polypyrimidinetrack of less than 11 nucleotides (nt 175-85) and a branch point similar to that highly
conserved in yeast and *Giardia lamblia* (Nixon et al., 2002a). Results from the 3’ RACE showed that the putative intron was missing in amplified cDNA.

Rapid amplification of 3’ and 5’ cDNA ends successfully amplified the remaining nucleotides upstream and downstream of the original primers (and as well those nucleotides bound by the original degenerate primers). The inferred protein sequence was found to possess Cys-X (2)-His-X (4)-Met and an upstream His residue necessary for the binding of the Cu$^{2+}$ co-factor (Sigfridsson, 1998, Nersissian & Shipp, 2002). This motif is a hallmark of the family of type I blue copper proteins, of which plastocyanin is a member. The presence of such a motif further confirms the gene sequence identification as plastocyanin. The beginning of the coding region was determined by the presence of a start codon and the amino acid sequence information from the N-terminus (Peers & Price, 2006). The presence of a stop codon and the poly-A tail of our transcript determined the end of the codon regions. The complete gene sequence encoded for a protein of 100 amino acids in length with a molecular mass of 11.1 kDa (Fig. 1). Sequence comparison of the inferred protein showed the highest similarity with plastocyanin of *Scenedesmus obliquus* (71%) and highest identity with plastocyanin of *Emiliania huxleyi* (57%).

**Comparative Modeling.** Comparative modeling of the plastocyanin sequence demonstrated the inferred product possessed the structural elements and correct globular properties needed for interaction with upstream and downstream partners (Fig. 2). The protein model identified a single hydrophobic patch containing His-82, a residual involved in coordinated covalent bonding of the Cu co-factor. This hydrophobic patch is a common property of all blue copper proteins (Sigfridsson, 1998). Tyrosine-78 was
located in an acidic patch known to play a role in electron transfer. The presence of these patches and their orientation with respect to each other, the hydrophobic patch on the northern side and the acidic patch on the eastern side, are consistent with other plastocyanins and needed for proper functioning and interaction (De Rienzo et al., 2000).

The structure and surface typology of plastocyanin from *Chlamydomonas reinhardtii* has been well studied (Redinbo et al., 1993) and is included for comparison (Fig. 2).

**Sequence analysis of plastocyanin and rRNA sequences from photosynthetic organisms and phylogenetic inference.** Phylogenetic analysis of the inferred protein sequence was performed using Bayesian analysis (Ronquist & Huelsenbeck, 2005), neighbour-joining (Felsenstein, 2005) and maximum likelihood (Guindon & Gascuel, 2003) methods. Shown in Figure 3 is a Bayesian inferred tree with posterior probability and the bootstrap support values from maximum likelihood and neighbour joining at each node. All phylogenetic methods produced similar superclade typologies. The twenty-six sequences in the analysis were divided into a monophyletic and a polyphyletic group. A monophyletic group contained the eukaryote plastocyanins; including the green algae, chromophytes and higher plants. The cyanobacterial sequences of plastocyanin formed a polyphyletic group. The relationship between gene sequences differed depending on the phylogenetic method of analysis, but the overall structure of the tree was maintained. Within the green algal clade, low bootstrap value for maximum likelihood analysis occurred at the node separating red algal sequences from green algal and higher plant sequences. This made the relationship between the red green algal sequences uncertain. However, all phylogenetic inference methods placed *T. oceanica*’s sequence into a clade closely related to green algal ancestral plastocyanins, suggesting that the group of red
algal sequences is indeed closely related to green algal sequences. Maximum likelihood, neighbour-joining and Bayesian analyses inferred *T. oceanica*'s plastocyanin to be monophyletic with green algae and those plastocyanin sequences found in dinoflagellates, separate from cyanobacterial sequences. Sequence analysis of the complete inferred protein sequence for plastocyanin indicated that it was most similar to *Scenedesmus obliquus*, a green (chlorophyll a/b) chloroplast-containing alga.

Using published sequences of 5.8s rRNA and 18s rRNA, we inferred the relationship of *T. oceanica* (CCMP 1005) to other species of *Thalassiosira* to ensure that its classification within this genus was justified by genetic relatedness in addition to observed morphological similarities (Round *et al.*, 1990). Phylogenetic analysis using either 18s rRNA (supplemental) or 5.8s rRNA (Fig. 4) showed that *T. oceanica* (CCMP 1005) was closely related to *T. weissflogii* and distantly related to the sequenced diatom, *T. pseudonana* (CCMP 1335), but firmly within the *Thalassiosira* clade, a red chloroplast-containing algal group.

**Degenerate PCR.** We tested other diatoms for the presence of a plastocyanin gene using degenerate primers designed to conserve regions. Two combinations of primers were used: BLOCKB forward and BLOCKC reverse, as well as, NTERM forward and BLOCKC reverse. These primers successfully amplified products from *Arabidopsis thaliana*, a higher plant containing plastocyanin (results not shown). *Thalassiosira oceanica* (CCMP 1005) was used as a positive control in the experiments. Degenerate PCR products were cloned as described above. A sequence with homology to plastocyanin was found in *T. oceanica* (CCMP 1006) (Figure 4). The sequenced product,
including intronic and exonic regions, had 99% nucleotide identity to plastocyanin from *T. oceanica* (CCMP 1005). The inferred protein sequence had 100% protein identity.
Discussion

The pathway of electron flow in the light reactions of oxygenic photosynthesis has been largely delineated in model organisms (e.g. *Chlorella*) and extrapolated to other photoautotrophs. Significant differences in the types of accessory and primary photosynthetic pigments were recognized early on as one of the distinguishing features of different alga taxa (Fott, 1974). Those algae possessing chlorophyll c or chlorophyll b, recently referred to as the red and green plastid lines (Falkowski & Raven, 1997), not only possessed different pigments but a number of other biochemical and cytological features that were used to classify algae into their respective groups.

Plastocyanin and cytochrome c₆ are interchangeable redox proteins that transfer electrons from cytochrome b₆f to PSI. Although many photosynthetic organisms contain both proteins, many do not. Indeed, plastocyanin is the only carrier in higher plants, and some green algae and cyanobacteria (Weigel *et al.*, 2003, Merchant & Bogorad, 1986, Ho & Krogmann, 1984). One of the unique characteristics of chlorophyll c-containing phytoplankton was thought to be their exclusive use of cytochrome c₆ (Raven *et al.*, 1999). This conclusion (paradigm) was established in the mid 1980’s by studies that examined only a few representative species in the red plastid lineage (Sandmann, 1986, Sandmann *et al.*, 1983). The method of analysis at that time was difference spectroscopy which was used to identify the characteristic absorption maximum of plastocyanin at 597 nm in partially-purified protein extracts. It remains one of the key methods to quantify plastocyanin. Recent efforts to sequence whole genomes of a wide diversity of protists and the creation of expressed sequence tag (EST) databases have provided an alternative
means to look at the taxonomic distribution of plastocyanin and cytochrome c\(_6\). Sequence search of the genomes of chlorophyll c-containing organisms, including the hot spring alga *Cyanidioschyzon merolae*, the thermophile *Galdieria sulphuraria* and the EST and genome databases of the diatoms, *T. pseudonana* (CCMP 1335) and *Phaeodactylum tricornutum* (CAPP1051/1, CCMP 632) are in agreement with the original results (Sandmann et al., 1983). These organisms lack plastocyanin and instead contain cytochrome c\(_6\). The molecular results are in agreement with the original survey of the *Chrysophyceae*, *Xanthophyceae*, *Phaeophyceae* and *Rhodophyceae* that were unable to detect plastocyanin in these chlorophyll c-containing organisms (Sandmann et al., 1983).

The results presented here confirm that the red algal derived diatom, *Thalassiosira oceanica* (CCMP 1005), contains a plastocyanin gene. The inferred mature protein sequence had the highest identity with *Emiliania huxleyi* and similarity with *Scenedesmus obliquus*. The necessary motifs for function were present and comparative modeling illustrated the presence of structural elements and globular properties needed for interaction with upstream and downstream partners, as well as for function. These molecular data supports biochemical and physiological data from our lab that identified this protein and its role in photosynthetic electron flow in *T. oceanica* (CCMP 1005) (Peers & Price, 2006).

Using published sequences of 5.8s and 18s rRNA gene, we inferred the relationship of *T. oceanica* (CCMP 1005) to other *Thalassiosira* species. Phylogenetic analysis using either gene showed that *T. oceanica* (CCMP 1005) was closely related to
other *Thalassiosira* species (Supplemental Fig. 2; Fig. 5). Thus, this oceanic strain appears to be correctly assigned to the *Thalassiosira* genus, belonging to the red (chlorophyll a/c) plastid lineage. Previous work supports the relatedness of the Thalassiosiraceae (Hoppenrath *et al.*, 2007, Kaczmarska *et al.*, 2006). *Thalassiosira oceanica* (CCMP 1005) was not analyzed in these published data, however, our results show that it is closely related to *T. weissflogii* (CCMP 1336) and distantly related to *T. pseudonana* (CCMP 1335), the sequenced diatom (Armbrust *et al.*, 2004). Our 18s rRNA data show that two of the *T. oceanica* species studied here (CCMP 1001 and 1005) are more closely related to coastal species than to each other (Supplemental Fig. 2). The inclusion of CCMP 1001 in a clade with *T. pseudonana* was strongly supported by Kaczmarska *et al.* (2007).

Species within the same genus may possess vastly different physiologies (Marchetti *et al.*, 2006, Peers *et al.*, 2005, Dupont *et al.*, 2006, Rocap *et al.*, 2003, Palenik *et al.*, 2006). A good example of this is the tolerance of closely related diatoms to Fe limitation. *Thalassiosira oceanica* (CCMP 1005) and *T. weissflogii* (CCMP 1336) are genetically similar yet the former species, isolated from metal poor waters of the open ocean, is able to grow at near maximum rates with 1/10 of the amount to Fe as its coastal congener. This is possible because of specific adaptations that reduce the cellular requirements for the limiting metal, including the use of alternative biochemical pathways (Archibald, 1983, Posey & Gherardini, 2000, Erdner *et al.*, 1999), up-regulation of resource scavengers (Trick *et al.*, 1983), and changes in the stoichiometry of nutrient requiring cellular components (Strzepek & Harrison, 2004).
Under Fe limitation, Fe-containing proteins are replaced with alternative proteins with non-Fe metal cofactors (Kunert et al., 1976, Wood, 1978, Merchant & Bogorad, 1986, Sandmann, 1986, Erdner et al., 1999, Peers & Price, 2004). Such replacements include increased use of a Mn-superoxide dismutase (SOD) for a Fe-SOD (Peers & Price, 2004), loss of Fe requiring secretary products (Posey & Gherardini, 2000) or the total replacement of all pathways requiring Fe (Archibald, 1983). Most trace metals in *Thalassiosira* are used in the light reactions of photosynthesis (Raven et al., 1999). Under Fe-limitation, a decrease in PSI (which contains 12 Fe atoms) and a reduction Fe-containing cytochrome b₆f occurs in *T. oceanica* (Strzepek & Harrison, 2004). The Fe-containing electron transport protein ferrodoxin is also replaced by an non-Fe-containing flavodoxin (LaRoche et al., 1996), further economizing on Fe. Collectively, these adaptations account for a 3-fold reduction in the cellular Fe requirements of the oceanic compared to the coastal isolate. Similarly, the substitution of plastocyanin for it’s Fe containing homologue would further decrease Fe requirements by 10% (Peers & Price, 2006), although it would increase by 10-fold the need for Cu. Given the greater relative abundance of Cu than of Fe in the sea, this trade off may be advantageous. Importantly, such a substitution would possibly maintain electron transport from PSII to PSI and production of ATP in situations where Fe is limiting (Peers & Price, 2006).

The loss of the electron carrier between PSII and PSI in autotrophs contributes to excessive superoxide radical formation (Tognetti et al., 2006) and decreased growth (Tognetti et al., 2007). In the absence of Cu or Fe or impaired electron transport on the donor side of PSI, a decrease in autotrophic growth is observed (Peers & Price, 2006). Organisms that possess both cytochrome c₆ and plastocyanin may use either protein to
transfer electrons between PSII and PSI, depending on the availability of the metal co-
factors, Cu and Fe. For example, in copper deficient cells of the cyanobacteria,
Synechocystis, (Zhang et al., 1992) and the green algae, Chlamydomonas reinhardtii (Li
et al., 1996, Eriksson et al., 2004), plastocyanin is decreased and cytochrome c₆
production is induced. However, in many marine phytoplankton, cytochrome c₆ has been
the only observed primary electron carrier to PSI. The presence of the Fe-containing
cytochrome c₆ may have been beneficial during the evolution of the red plastid lineage
since the availability of Fe in the early ocean was greater than it is now (Anbar & Knoll,
2002). As the ocean became increasingly Fe-poor, growth of species that were able to
use other redox proteins would have been selected for with strong selection on organisms
using a Cu-containing plastocyanin in regions where Fe was limited.

Our PCR method amplified a plastocyanin gene in two strains of diatoms. These
strains (T. oceanica (CCMP 1006 and 1005)) were of oceanic provenance and had low Fe
requirements for growth (Peers et al., 2005). Plastocyanin was not detected in cultures
isolated from coastal regions, even in closely related species (e.g. T. oceanica (CCMP
999) and T. weissflogii (CCMP 1336)). Physiological experiments in our lab have
demonstrated different physiological reactions under low trace metal concentrations.
Both diatoms, T. oceanica (CCMP 1005) and T. oceanica (CCMP 1006) are able to
persist at iron concentrations that are non permissive to coastal diatoms such as CCMP
1335 and 1336 (Peers & Price, 2004). So as argued above, it is not surprising that both
CCMP 1005 and 1006 were found to have plastocyanin while the coastal diatoms, T.
pseudonana (CCMP 1335 and 1336), did not. We note that we are unable at this time to
evaluate the phylogenic relationship between the two oceanic strains to determine if they
are truly different species. Nor can we explain why plastocyanin was not detected in some of the other oceanic strains (CCMP 1014, CCMP 497). For one of these species, *Thalassiosira pseudonana* (CCMP 1014), however, the lack of plastocyanin was consistent with its unusually high Fe requirements for growth compared to other open ocean isolates. This species possesses a more coastal-like physiology than other oceanic diatoms and may be an ephemeral resident of the Pacific gyre that is only intermittently supplied to offshore waters by coastal jets and eddies that transport water away from the coast.

One possible explanation for the occurrence of plastocyanin in these diatom species may be horizontal gene transfer (HGT). Other than that of plastid acquisition and symbiosis, lateral gene transfer into oceanic plankton from an exogenous source is not surprising. Although most examples of horizontal gene transfer occur between prokaryotes themselves or between the chloroplast and the nucleus of photosynthetic organism, intra and inter domain transfer between other kingdoms have occurred (Pollack *et al.*, 2005, Waller *et al.*, 2006, Archibald *et al.*, 2003) and may be a major driving force in eukaryotic algae evolution (Archibald *et al.*, 2003). Many individual genes such as thymidine kinase (White *et al.*, 2005), NADH oxidase (Nixon *et al.*, 2002b), and flavohemoglobin (Anderson *et al*. 2003) have been acquired in some protists by HGT. However, sequence analysis suggest that since the initial acquisition, plastocyanin has been vertically inherited – an observation that doesn’t not exclude the possibility of gene transfer through the acquisition of a new symbiont as previously suggested for the green dinoflagellate (Nosenko *et al.*, 2006).
Sequence analysis of the complete inferred protein sequence for plastocyanin indicated that it shared the highest identity with the haptophyte *E. huxleyi* and highest similarity with the unicellular green alga *Scenedesmus obliquus* and not to cyanobacteria, as suggested by N-terminal sequence analysis (Peers & Price, 2006). Analysis of the plastocyanin sequence of *T. oceanica* (CCMP 1005) using NJ, ML, MP (supplemental) and Bayesian analyses (Figure 3), placed it in a monophyletic clade with the dinoflagellates and the haptophyte. This branch was closely related to the sequences of chlorophyll b-containing algae, relative to those sequences from cyanobacteria. Early branching of the plastocyanin sequences in the chlorophyll c-containing species is consistent with the early separation of the red algal lineage from cyanobacteria and green algae. The grouping of the dinoflagellate, haptophyte, and now, diatom sequences with each other suggests a common source and vertical inheritance since the acquisition of the gene.

The trace metal requirements of photosynthetic protists are thought to reflect the ancestral condition of the original plastid prior to endosymbiosis and to be transferred by vertical inheritance to subsequent generations (Quigg et al., 2003). A schism gave rise to different clades of photosynthetic eukaryotes to create the “green” lineage (chlorophyll a/b) and the “red” lineage (chlorophyll a/c) of photosynthetic organisms. Two competing hypotheses exist to explain the origin of the wide diversity of organisms with a red algal plastid. The chromalveolate hypothesis, supported by morphological, biogeochemical and shared-derived characteristics suggests that a single secondary endosymbiotic event occurred to give rise to this lineage containing dinoflagellates, heterokonts (including diatoms), and haptophytes (including coccolithophores) (Cavalier-Smith, 2002a). The
competing hypothesis proposes that the diverse number of extant organisms serving as hosts to secondary red plastids acquired a red plastid in separate independent endosymbiotic events and do not share a recent common ancestor (Grzebyk et al., 2003). If plastocyanin were acquired by the endosymbiont before the initial secondary endosymbiotic event, then we would expect to see a more frequent distribution of the gene among extant red algal derived species than currently observed. This distribution would be determined by the extent of gene loss by the plastid before and after endosymbiosis (portable hypothesis) or after (chromalveolate hypothesis). Plastocyanin has been found within the red plastid-containing protist, *Karenia brevis*, and blast searches for the plastocyanin sequence have identified it in *Karlodinium micrum* and the haptophyte, *E. huxleyi*. Both dinoflagellates, *Karlodinium micrum* and *Karenia brevis*, have been examined and it has been suggested that the secondary plastid in these organisms has been replaced by a tertiary plastid of haptophyte origin (Yoon et al., 2002, Tengs et al., 2000, Nosenko et al., 2006). The plastid-encoded plastocyanin of *Karenia brevis* has been examined in detail and distribution and phylogenetic evidence (Patron et al., 2006) indicates that a haptophyte ancestor that became the tertiary endosymbiont most likely acquired the plastocyanin gene from a green alga. Because all eukaryotic photosynthetic organisms were phagotrophic, at one point in their evolution, it is quite possible that the gene may have been horizontally transferred from an organism of green algal lineage. Our analysis supports this scenario since all our phylogenetic inferences resulted in the grouping of the red plastocyanin sequences with those of the green lineage.

The acquisition and utilization of plastocyanin by the common ancestor of the coccolithophores, diatoms and dinoflagellates would have been beneficial during
changing habitat conditions. Estimates based on the evolution of ribosomal genes and the molecular clock theory indicate that diatoms originated as much as 251 million years ago (Falkowski et al., 2004). Iron concentrations were decreasing at this time while Cu was becoming progressively more available as the ocean changed from anoxic to oxic (Saito et al., 2003). Responding, and thus adapting, to such a changing environment would have been necessary for ecological success. Selection of specific chemical elements would have depended on abundance, availability and utility in biochemical reactions (Da Silva & Williams, 1991). Utilizing elements that would be less costly in terms of energy would therefore be beneficial. Hence acquiring the use of a Cu based protein, an element that was increasingly becoming more abundant, and replacing the use of a functional equivalent costlier Fe based protein was advantageous. This advantage would be increasingly important as red algae came to dominate the ocean over those of green origin during the next 250 million years (Falkowski et al., 2004). Indeed, *E. huxleyi*, a member of the coccolithophore, a group of phytoplankton possessing chlorophyll c, but in steady decline over the last 55 million years, is present and abundant in the world's ocean (Falkowski et al., 2004). Diatoms, such as the *T. oceanica*, radiated about 125 million years ago and potentially proliferated as late as 33 million years ago were undoubtedly under similar selection pressures. Those processing this adaptation would have an advantage over others and may have given rise to extant species. Those species possessing plastocyanin, but inhabiting relatively Fe rich environments, would have eventually lost the gene.

The maintenance and use of plastocyanin in diatoms today may provide a distinct ecological advantage over other species that reply on Fe-containing cytochrome. We
hypothesize that the ancestral plastid (the endosymbiont) contained plastocyanin and that selective pressure maintained it in some *Thalassiosira* species but that others lost it or no longer express it. The acquisition of this gene coupled with the need for diatoms to reduce their iron requirements may have provided a necessary condition for diatoms to inhabit Fe-limiting environments. The presence of plastocyanin in oceanic diatoms may thus be an example of an ecologically relevant gene acquisition.
References


Redinbo, M., Cascio, D., Choukair, M., Rice, D., Merchant, S. & Yeates, T. 1993. The 1.5-A crystal structure of plastocyanin from the green alga *Chlamydomonas reinhardti*. *Biochemistry* 32.


Table 1. Primers and probes used to identify and isolate the plastocyanin gene in Thalassiosira oceanica (CCMP 1005) using 5’ and 3’ RACE and degenerate PCR. The NTERM forward primer was designed to the amino acids, LVFEPAK, of the N-terminal region of the protein (Peers & Price, 2006) and the IPTERP reverse primer was designed to the conserved amino acid residuals, PHNVVFDEPNIP. Both of these primers were used for the initial isolation and identification of plastocyanin and for subsequent probing and blotting. The 5’ internal reverse and the 3’ internal forward primers were designed to a genomic fragment from T. oceanica (CCMP 1005) and were used in 5’ and 3’ RACE procedures. The BLOCK C reverse primer was designed to the amino acids, CEPHQGAG. The BLOCK B forward primer was designed to the motif PHNVVF and used for degenerate PCR. Primer sequences are listed using IUPAC nomenclature where Y = C or T, R = A or G and N = A, G, T or C. Inosine (I) was used instead of a nucleotide base at some positions.
<table>
<thead>
<tr>
<th>Primers / Probe</th>
<th>Target</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTERM forward</td>
<td>QTVEVKM</td>
<td>5’-CYC ARA CYG TYG ARG TYA ARA TGG G-3’</td>
</tr>
<tr>
<td>IPTERP reverse</td>
<td>PHNVVFDEDNIP</td>
<td>5’-GGI ATR TTR TCY TCR TCR TCC ACC ACR TTR TGN GG-3’</td>
</tr>
<tr>
<td>Block B forward</td>
<td>PHNVVF</td>
<td>5’-CCI CAY AAY GTN GTN TTY-3’</td>
</tr>
<tr>
<td>Block C reverse</td>
<td>CEPHQGAG</td>
<td>5’-CCI GCI CCY TGR TGI GGY TCR CA-3’</td>
</tr>
<tr>
<td>3’ internal forward</td>
<td></td>
<td>5’-CTC CGG ACT TCT TGT CTT CG-3’</td>
</tr>
<tr>
<td>5’ internal reverse</td>
<td></td>
<td>5’-AAG GAG CGT TGA GCG ATC CCC TCC A-3’</td>
</tr>
</tbody>
</table>
Table 2. List of species and accession numbers of plastocyanin protein sequences used for phylogenetic analysis.
<table>
<thead>
<tr>
<th>Species</th>
<th>Supergroup</th>
<th>Division*</th>
<th>Class*, φ</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crocosphaera watsonii WH 8501</td>
<td>Cyanobacteria</td>
<td>Chroobacteria</td>
<td>ZP_00517105</td>
<td></td>
</tr>
<tr>
<td>Phormidium laminosum</td>
<td>Cyanobacteria</td>
<td>Chroobacteria</td>
<td>Q51883</td>
<td></td>
</tr>
<tr>
<td>Prochlorococcus marinus CCMP 1375</td>
<td>Cyanobacteria</td>
<td>Chroobacteria</td>
<td>NP_875473</td>
<td></td>
</tr>
<tr>
<td>Synechococcus sp. WH 8102</td>
<td>Cyanobacteria</td>
<td>Chroobacteria</td>
<td>NP_897590</td>
<td></td>
</tr>
<tr>
<td>Synechocystis sp. PCC 6803</td>
<td>Cyanobacteria</td>
<td>Chroobacteria</td>
<td>NP_442157</td>
<td></td>
</tr>
<tr>
<td>Trichodesmium erythraeum IMS101</td>
<td>Cyanobacteria</td>
<td>Chroobacteria</td>
<td>Q111R2</td>
<td></td>
</tr>
<tr>
<td>Gloeobacter violaceus PCC 7421</td>
<td>Cyanobacteria</td>
<td>Chroobacteria</td>
<td>Q7NIA7</td>
<td></td>
</tr>
<tr>
<td>Nostoc punctiforme PCC 73102</td>
<td>Cyanobacteria</td>
<td>Chroobacteria</td>
<td>B2IVD7</td>
<td></td>
</tr>
<tr>
<td>Karenia brevis</td>
<td>Protozoa</td>
<td>Dinophyceae</td>
<td>DQ531587</td>
<td></td>
</tr>
<tr>
<td>Karlodinium micrum</td>
<td>Protozoa</td>
<td>Dinophyceae</td>
<td>Q2IA53</td>
<td></td>
</tr>
<tr>
<td>Emiliania huxleyi</td>
<td>Protozoa</td>
<td>Prymnesiophyceae</td>
<td>GM1.2200045</td>
<td></td>
</tr>
<tr>
<td>Physcomitrella patens</td>
<td>Plantae</td>
<td>Bryopsida</td>
<td>Q9SXW9</td>
<td></td>
</tr>
<tr>
<td>Chlamydomonas reinhardtii</td>
<td>Plantae</td>
<td>Chlorophyceae</td>
<td>XP_001702952</td>
<td></td>
</tr>
<tr>
<td>Chlorella vulgaris</td>
<td>Plantae</td>
<td>Chlorophyceae</td>
<td>P00300</td>
<td></td>
</tr>
<tr>
<td>Pediastrum boryanum</td>
<td>Plantae</td>
<td>Chlorophyceae</td>
<td>AB017810</td>
<td></td>
</tr>
<tr>
<td>Pediastrum duplex</td>
<td>Plantae</td>
<td>Chlorophyceae</td>
<td>BAE00064</td>
<td></td>
</tr>
<tr>
<td>Scenedesmus obliquus</td>
<td>Plantae</td>
<td>Chlorophyceae</td>
<td>P26956</td>
<td></td>
</tr>
<tr>
<td>Hordeum vulgare</td>
<td>Plantae</td>
<td>Lilopsida</td>
<td>P08248</td>
<td></td>
</tr>
<tr>
<td>Oryza sativa</td>
<td>Plantae</td>
<td>Lilopsida</td>
<td>A2Y886</td>
<td></td>
</tr>
<tr>
<td>Zea mays</td>
<td>Plantae</td>
<td>Lilopsida</td>
<td>P16002</td>
<td></td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>Plantae</td>
<td>Magnoliopsida</td>
<td>Q8LD06</td>
<td></td>
</tr>
<tr>
<td>Nicotiana tabacum</td>
<td>Plantae</td>
<td>Magnoliopsida</td>
<td>P35476</td>
<td></td>
</tr>
<tr>
<td>Pism sativum</td>
<td>Plantae</td>
<td>Magnoliopsida</td>
<td>P16002</td>
<td></td>
</tr>
<tr>
<td>Solanum crispum</td>
<td>Plantae</td>
<td>Magnoliopsida</td>
<td>P00297</td>
<td></td>
</tr>
<tr>
<td>Scherffelia dubia</td>
<td>Plantae</td>
<td>Prasinophytae</td>
<td>AJ919618</td>
<td></td>
</tr>
</tbody>
</table>

(Cavalier-Smith, 2002a)

(Cavalier-Smith, 2002b)

(Adl et al., 2005)
Table 3. Species, clones, collection sites and accession numbers of 5.8s rRNA gene sequences of diatoms screened for plastocyanin by degenerate PCR. The provenance of each clone was determined according to its collection site reported in the CCMP database.
<table>
<thead>
<tr>
<th>Species</th>
<th>Strain Number</th>
<th>Collection Site</th>
<th>5.8s rRNA accession number</th>
<th>Provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thalassiosira guillardii</td>
<td>CCMP 988</td>
<td>North Atlantic, Continental Shelf 39.6000N 70.1000W</td>
<td>EF208788</td>
<td>Uncertain*</td>
</tr>
<tr>
<td>T. oceanica</td>
<td>CCMP 999</td>
<td>North Atlantic, Continental Slope 39.0833N 71.9333W</td>
<td>EF134953</td>
<td>Uncertain</td>
</tr>
<tr>
<td>T. oceanica</td>
<td>CCMP 1005</td>
<td>Sargasso Sea 33.1833N 65.2500W</td>
<td>EF362630</td>
<td>Oceanic</td>
</tr>
<tr>
<td>T. oceanica</td>
<td>CCMP 1006</td>
<td>Sargasso Sea 32.3000N 64.8400W</td>
<td>EF362631</td>
<td>Oceanic</td>
</tr>
<tr>
<td>T. pseudonana</td>
<td>CCMP 1014</td>
<td>North Pacific Gyre 28.000N 155.0000W</td>
<td>EF208791</td>
<td>Oceanic</td>
</tr>
<tr>
<td>T. pseudonana</td>
<td>CCMP 1015</td>
<td>San Juan Island, Washington State, USA 48.5440N 123.0100W</td>
<td>EF208792</td>
<td>Coastal</td>
</tr>
<tr>
<td>T. pseudonana</td>
<td>CCMP1335</td>
<td>Moriches Bay, New York, USA 40.7560N 72.8200W</td>
<td>EF208793</td>
<td>Coastal</td>
</tr>
<tr>
<td>T. rotula</td>
<td>CCMP 1647</td>
<td>Gulf of Naples, Italy 40.7500N 14.3300E</td>
<td>EF208798</td>
<td>Coastal</td>
</tr>
<tr>
<td>T. weissflogii</td>
<td>CCMP 1336</td>
<td>Gardiners Island New York, USA 41.1100N 72.1000W</td>
<td>DQ469927</td>
<td>Coastal</td>
</tr>
<tr>
<td>Minutocellus polymorphus</td>
<td>CCMP 497</td>
<td>Sargasso Sea 32.0000N 64.0000W</td>
<td>N/A</td>
<td>Oceanic</td>
</tr>
<tr>
<td>Minutocellus sp.</td>
<td>CCMP 1701</td>
<td>Gulf of Oman, Arabian Sea 23.5643N 58.8530E</td>
<td>N/A</td>
<td>Coastal</td>
</tr>
</tbody>
</table>

*These species were isolated from warm core eddies over the continental shelf and slope and are most likely of oceanic origin, but they could also be coastal species that were transported into the eddy.
Figure 1. Genomic structure of plastocyanin and the inferred protein sequence for *Thalassiosira oceanica* (CCMP 1005). The top and bottom rows of each line represent the genomic coding sequence and the deduced amino acid sequence, respectively. The gap between amino acid residuals, nucleotides 95 to 191, indicates an intronic region identified by classical intronic excision signals.
1 - GCCCAAACGTGAGGTTAAAGATGGGAGCCGACTCCGGACTT - 42
1 - A Q T V E V K M G A D S G L - 14

43 - CTTGTCTTGCAGCCTGCCAAGGTCACTGCTGCAAGGGAGAC - 84
15 - L V F E P A K V T V C K G D - 28

85 - ACTGTCAGGTGAAGTCTTGGATCTCTGATTTTTTCTGCA - 126
29 - T V K W - 42

127 - CATGGACCACGAATGGACGAGCCTAAACTCACCAAAACCGG - 168
43 - - 56

169 - ATGACGAATCGATTCGTCTCAGGATCAACAACAGCTGAGC - 210
57 - I N N K A G - 70

211 - CCCACCAACGTGATGGGACGAGGATAACATCCCTGATGGCG - 252
71 - P H N V V D D E D N I P D G - 84

253 - TCGATCAGGAGAAGATTTCATGGACAAGCCAGCTCGCGACGC - 294
85 - V D Q E K I S M D D Q L G E - 98

295 - CGGGAGACACCTTCGAAATGAAATTCGATACCGCCGGAACCT - 336
99 - P G D T F E M K F D T A G T - 112

337 - ATGGTTAATCGGAACCTACCGCGGTCGCCGATGCAGG - 378
113 - Y G Y Y C E P H R G A G M Q - 126

379 - CCACCCTTGTTGTTCA - 395
127 - A T L V V Q - 132
Figure 2. Modeled globular surface images of the tertiary plastocyanin protein structure using the inferred protein sequence of *Thalassiosira oceanica* (CCMP 1005) (A, B) and the plastocyanin protein sequence of *Chlamydomonas reinhardtii* (C, D).

*Chlamydomonas reinhardtii* protein sequence was deduced from the gene sequence (Merchant *et al.*, 1990) and the structure determined by crystallography (Redinbo *et al.*, 1993). The right hand image (eastern side) is rotated 90 degrees to the left relative to the left hand image (northern side). Shades of red coloration indicate the degree of acidic residuals in a region; shades of blue coloration indicate the degree of basic residuals in a region and white area represent hydrophobic regions of the protein. PyMOL was used to generate this figure.
Figure 3. Phylogenetic analysis of plastocyanin from *Thalassiosira oceanica* (CCMP 1005). Values at nodes represent maximum likelihood bootstrap values, Bayesian analysis posterior probabilities and neighbour-joining bootstrap values, respectively. Dashes indicate situations where a bootstrap value was not possible. The tree was constructed using the inferred mature protein sequence of plastocyanin for each organism represented. Green lettering denotes chlorophyll a/b-containing organisms, blue lettering indicates cyanobacteria and red lettering indicates chlorophyll a/c-containing organisms.
Figure 4. Phylogenetic tree of *Thalassiosira* spp. inferred by Bayesian analysis using 5.8s rRNA sequences. The green algae, *Chlamydomonas reinhardtii*, was used as the outgroup. Oceanic species are marked with an asterisk. Values at the nodes indicate posterior probabilities.
Figure 5. Aligned sequences of plastocyanin from *Thalassiosira oceanica* (CCMP 1005) and *Thalassiosira oceanica* (CCMP 1006). Italicized nucleotides represent primer sequences; asterisks indicate variable nucleotides.
Supplemental Figure 1. Southern blot analysis of coastal and oceanic species of phytoplankton using probes designed to plastocyanin from *Thalassiosira oceanica* (CCMP 1005).

Numbers above blot represent clone numbers of *Thalassiosira oceanica* (CCMP 1005, CCMP 1006, CCMP 999); *T. guillardii* (CCMP 988); *T. weissflogii* (CCMP 1336); *T. rotula* (CCMP 1647); *T. pseudonana* (CCMP 1014, CCMP 1015, CCMP1335); *Minutocellus polymorphus* (CCMP 497), and *Minutocellus* sp. (CCMP 1701). A fully degenerate probe was designed to the conserved plastocyanin sequence CEPHQGAG (Table 1) using the CODEHOP program (Rose et al., 1998) with all plastocyanin protein sequences from the protein database, InterPro (http://www.ebi.ac.uk/interpro/). Fifteen micrograms of genomic DNA was digested to completion as indicted by complete co-digestion of 0.25μg of pBR355 plasmid DNA (Invitrogen) using 35U/μg ALU1 (New England Biolabs, Ipswich, MA, USA), and loaded onto a 1% agarose, 0.5X TBE (Tris Borate EDTA) gel. Transfer of nucleic acids from the gel to a solid support for southern blotting was done using upward capillary action using alkaline transfer (Sambrook & Russell, 2004) onto Hybond XL nylon (Amersham, Piscataway, NJ, USA) overnight. Probes were labeled using the 5’-[$\gamma ^{-32}P$] ATP T4 terminal labeling kit (Amersham) and purified using a G25 Sephadex column (GE Illustra, Piscataway, NJ, USA). To confirm 5’-[$\gamma ^{-32}P$] ATP labelling of the probe, the probe was run on a 6% urea PAGE gel and assessed for the presence of proper labelled product and exposed overnight. Hybridization was performed at 42°C and washes were performed as follows: 5X SSC (sodium chloride sodium citrate), 0.1% SDS (sodium dodecyl sulfate), 2 X 5 minutes at 21°C; 1X SSC, 0.1% SDS, 15 minutes at 42°C, 0.5X SSC, 0.1% SDS 20 minutes at 42°C.
An alkaline phosphate screen was exposed for 24, 48 and 64 hours and visualized using a Phosphoimager. Film and conventional autoradiography was used.
Supplemental Figure 2. Maximum likelihood analysis of 18s rRNA gene sequences of *Thalassiosira* species. Sequences were aligned using the Clustal W program and the alignment was manually corrected to exclude gaps and ambiguously coded regions. The PHYLIP suite of programs, including SEQBOOT, PROML, CONSENSUS was used to analyze and compare the sequences. PROML was used to construct ML tree with the Dayhoff substitution matrix. The analysis was bootstrapped 100 times using SEQBOOT. A consensus tree was created with CONSENSUS using the majority rule. The tree was visualized using TREEVIEW (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). The sequences used in this analysis can be found in the GenBank database under the following accession numbers: AM235383.1, *Thalassiosira profunda* (POMX); DQ093369.1, *Thalassiosira aestivalis* (CCMP 975); DQ093368.1, *Thalassiosira tumida* (CCMP 1465); AF374481.2, *Thalassiosira pseudonana* (CCMP 1335)*, DQ093367.1, (CCMP 1007), AY485452.1 (CCMP 1335)*, AJ535169.1 (p11); DQ093366.1, *Thalassiosira minima* (CCMP 991); DQ093365.1, *Thalassiosira nordenskioeldii* (CCMP 997); AM050629.1, *Thalassiosira hendeyi* (MHth1T); AJ810854.1, *Thalassiosira anguste-lineata* (Mhta1); AJ810856.1, *Thalassiosira punctigera* (MHtp1), AY485526.1 (AWI); AJ810855.1, *Thalassiosira delicatula* (MHtd1); AF374477.2, *Thalassiosira weissflogii*, AY485445.1 (CCMP1049); AJ535171.1, *Thalassiosira sp.* (CCMP 1281); AJ535170.1, *Thalassiosira fluviatilis* (clone p928); AF374482.2, *Thalassiosira antarctica*; AF374479.2, *Thalassiosira oceanica* (CCMP 1005), DQ093364.1 (CCMP 1001); AF374478.2, *Thalassiosira guillardii*; AF374480.2, *Thalassiosira rotula*, AF462059.1 (CCMP1018), AF462058.1
(CCMP1647), X85397.1 (CCAP 1085/4); X85396.1, Thalassiosira.eccentrica;
AY665727.1, Chlamydomonas reinhardtii.
Supplemental Figure 3. Distribution of cytochrome c₆ and plastocyanin in photosynthetic organisms containing chlorophyll c.
<table>
<thead>
<tr>
<th>Class</th>
<th>Species</th>
<th>Cytochrome c&lt;sub&gt;i&lt;/sub&gt; identified (Yes/No)</th>
<th>Plastocyanin identified (Yes/No)</th>
<th>Point of Isolation</th>
<th>Type of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillariophyceae</td>
<td>Phaeodactylum tricornutum (1090-1b)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Skeletonema costatum (LB 1077/1, C)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira oceanica (CCMP 999)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-</td>
<td>No</td>
<td>Uncertain</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira guillardii (CCMP 988)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-</td>
<td>No</td>
<td>Uncertain</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira weissflogii (CCMP 1336)&lt;sup&gt;b,d&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira rotula (CCMP 1647)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-</td>
<td>No</td>
<td>Coastal</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira pseudonana (CCMP 1335)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>T. minuscula (CCMP 1036)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-</td>
<td>No</td>
<td>Coastal</td>
<td>Physiological</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira oceanica (CCMP 1003)&lt;sup&gt;c,d&lt;/sup&gt;</td>
<td>-</td>
<td>Yes</td>
<td>Oceanic</td>
<td>Physiological</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira oceanica (CCMP 1006)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>No</td>
<td>Yes</td>
<td>Oceanic</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira oceanica (CCMP 1005)&lt;sup&gt;b,c,e&lt;/sup&gt;</td>
<td>No</td>
<td>Yes</td>
<td>Oceanic</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira pseudonana (CCMP 1015)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-</td>
<td>Yes</td>
<td>Coastal</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>Minutocellus sp. (CCMP 1701)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-</td>
<td>No</td>
<td>Coastal</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>Minutocellus polymorphus (CCMP 497)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-</td>
<td>No</td>
<td>Oceanic</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td></td>
<td>Thalassiosira pseudonana (CCMP 1014)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-</td>
<td>No</td>
<td>Oceanic</td>
<td>Degenerate PCR / Southern blot</td>
</tr>
<tr>
<td>Xanthophyceae</td>
<td>Bumilleropsis filiformis (Konstanz stock)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Vischeria stellata (887-2, G)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Tribonema aequale (880-1, G)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Bumilleria sicula</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td>Xanthophyceae</td>
<td>Ectocarpus siliculosus (Konstanz)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Cutleria multiforma&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td>Phaeophyceae</td>
<td>Polysiphonia sp. (888-1, G)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Phaeocystis pouchetii (CCMP 1072)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Rhodophyceae</td>
<td>Cystocarpus spongiosus (CCMP 756)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
</tr>
<tr>
<td></td>
<td>Cyanidium caldarium (RK-1)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Phorphyridium cruentum (1380-1a, G)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Phorphyridium aerugineum (1380-2, G)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Polysiphonia sp. (888-1, G)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Yes</td>
<td>No</td>
<td>Coastal</td>
<td>Differential Spectroscopy</td>
</tr>
<tr>
<td></td>
<td>Prymnesiophyceae</td>
<td>Karenia brevis (CCMP 718)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>-</td>
<td>Yes</td>
<td>Coastal</td>
</tr>
<tr>
<td></td>
<td>Dinophyceae</td>
<td>Karlodinium micrum (CCMP 415)&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Yes</td>
<td>Yes</td>
<td>Coastal</td>
</tr>
<tr>
<td></td>
<td>Prymnesiophyceae</td>
<td>Emiliana huxleyi (CCMP 371)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>-</td>
<td>Yes</td>
<td>Oceanic</td>
</tr>
</tbody>
</table>

---

*Sandmann et al. 1983; * own result; * Peers, Quesnel and Price 2005; * Strzepek and Harrison 2004; * Peers and Price 2006; * Nosenko et al 2006; * Patron et al. 2006*